MULTIVARIATE BEHAVIORAL RESEARCH
2018, VOL. 53, NO. 1, 36–56
https://doi.org/10.1080/00273171.2017.1383224

SEM Based CARMA Time Series Modeling for Arbitrary N

Johan H.L. Oud[a], Manuel C. Voelkle[b], and Charles C. Driver[c]

[a]Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands; [b]Department of Psychology, Humboldt University, Berlin, Germany, Max Planck Institute for Human Development, Berlin, Germany; [c]Max Planck Institute for Human Development, Berlin, Germany

ABSTRACT
This article explains in detail the state space specification and estimation of first and higher-order autoregressive moving-average models in continuous time (CARMA) in an extended structural equation modeling (SEM) context for $N = 1$ as well as $N > 1$. To illustrate the approach, simulations will be presented in which a single panel model ($T = 41$ time points) is estimated for a sample of $N = 1{,}000$ individuals as well as for samples of $N = 100$ and $N = 50$ individuals, followed by estimating 100 separate models for each of the one-hundred $N = 1$ cases in the $N = 100$ sample. Furthermore, we will demonstrate how to test the difference between the full panel model and each $N = 1$ model by means of a subject-group-reproducibility test. Finally, the proposed analyses will be applied in an empirical example, in which the relationships between mood at work and mood at home are studied in a sample of $N = 55$ women. All analyses are carried out by ctsem, an R-package for continuous time modeling, interfacing to OpenMx.

KEYWORDS
Continuous time; time series; state space modeling; CARMA models; structural equation modeling; $N = 1$-research; state space modeling; interindividual and intraindividual research

Introduction

Time series analysis, made popular by Box and Jenkins (1970), has greatly benefited from the introduction of the state space approach. The state space approach stems from control engineering (Kalman, 1960; Zadeh & Desoer, 1963) and distinguishes the state of a system, which is a vector of latent variables driven by the system dynamics in the state transition equation, from the observations. It turns out that any Box-Jenkins autoregressive and moving-average (ARMA) model as well as any extended ARMAX model, in which exogenous variables are added to the model, can be represented as a state space model (Aoki, 1987; Caines, 1988; Deistler, 1985; Harvey, 1981; Ljung, 1985). However, the state space representation is much more flexible, allowing one to formulate many time series models that cannot easily be handled by the Box-Jenkins ARMA approach, and makes important state space modeling techniques such as the Kalman filter and smoother accessible for time series analysis (Durbin & Koopman, 2001). These advantages are well-known in control theory but less so in the behavioral and related sciences. Therefore, the ARMA model in continuous time presented in this article, called a CARMA model (Brockwell, 2004; Singer, 1991; Tómasson, 2011; Tsai & Chan, 2000), will be formulated as a state space model.

The same argument that in discrete time leads from a simple autoregressive model to a more complicated ARMA representation, namely that the past is influencing the present in this more complicated fashion, applies for the choice of a basic or more complicated CARMA model. The realization that most real-life processes evolve in continuous time is one reason to model from the start in continuous time and to consider discrete-time ARMA processes to be "embedded" in true underlying CARMA processes. Chow, Lu, Zhu, and Sherwood (2016), Oud and Delsing (2010), Oud and Jansen (2000), Voelkle, Oud, Davidov, and Schmidt (2012) discussed a series of problems connected with discrete time models. Two main problems are the difficulty of fitting discrete time models to irregularly spaced longitudinal data and the dependence of discrete time results on the time interval selected. Continuous time analysis is needed to make the different and, possibly, contradictory effects in discrete time independent of the interval. The definition of the time scale and thus of the interval may of course depend on the research context. In educational research, a year may be restricted to the days that education is given and

CONTACT Johan H. L. Oud j.oud@pwo.ru.nl Behavioural Science Institute, Radboud University, P.O. Box 9104, 6500 HE, Nijmegen, The Netherlands

in medical research bone age may sometimes be more appropriate than chronological age. Situations in which the system follows a truly discrete-time pattern are rare. A truly discrete time system functions at and only at the time points it is observed and so the length of the time interval between observations becomes irrelevant. Most psychological processes, however, evolve continuously over time, even if unobserved. The discrete time systems considered in this article are so-called sampled continuous time systems, sampled at the discrete observation time points but functioning in continuous time (Zadeh & Desoer, 1963).

Structural equation modeling (SEM) was introduced by Jöreskog with two seminal papers (Jöreskog, 1973, 1977) along with LISREL (Jöreskog & Sörbom, 1976), the first published SEM program. The strong relationships between the state space model (SSM) and SEM have been highlighted by MacCallum and Ashby (1986), Oud (1978), Oud, van den Bercken, and Essers (1990), and more recently by Chow, Ho, Hamaker, and Dolan (2010). Both consist of a measurement part and an explanatory part and in both the explanatory part specifies the relationships between latent variables. While in the state space approach the latent explanatory part is a recursive dynamic model, in SEM it is a latent structural equation model. The debate about which of the two, the state space model (SSM) or SEM, is more general (Chow et al., 2010), should be considered in connection with the associated estimation procedures: filtering for SSM, which is a recursive stepwise procedure from time point to time point, and an overall procedure applied on the entire time series at once for SEM. While for long time series, filtering techniques are computationally faster, the possibility to allow arbitrary measurement error structures, spanning the entire time range of the model is an important advantage of SEM (Voelkle, Oud, von Oertzen, & Lindenberger, 2012). Especially in behavioral science latent variables are typically measured by multiple measures which in addition to a common part may be composed of a specific part. SEM's overall procedure allows the specific parts to correlate across time independently from and thus not confounding the relationships between the latent variables in the structural equation.

The CARMA time series analysis procedure to be presented in this article is based on SEM continuous-time state-space modeling, developed for panel data by Oud and Jansen (2000) and Voelkle and Oud (2013) and implemented in the R-package ctsem (Driver, Oud, & Voelkle, 2017). The package ctsem interfaces to OpenMx (Boker et al., 2011; Neale et al., 2016), an open source extended structural equation modeling framework. The kernel of the procedure is a multivariate stochastic differential equation coupled with a measurement model, for which maximum likelihood estimation is performed via the so-called exact discrete model EDM (Bergstrom, 1984). The EDM uses the exact solution of the stochastic differential equation to link the underlying continuous time parameters exactly to the parameters of the discrete-time model describing the data. Despite the nonlinear connection between discrete and continuous time, the stochastic differential equation on which the procedure and ctsem are based, is linear. This differs from nonlinear SSMs that several authors in the behavioral and related sciences have turned attention to recently (e.g., Chow et al., 2016; King, Nguyen, & Ionides, 2016; Molenaar & Newell, 2003; Singer, 2011).

As pointed out above and emphasized by Chow et al. (2010), different estimation procedures have traditionally been associated with SSM and SEM. In addition, SSM is typically applied in a $N = 1$ context and SEM on large N data. Chow et al. (2010) also note that different traditions in handling the initial conditions (means and variances-covariances) might lead to different analysis results from SSM and SEM. In a series of simulations, Oud (2007) and Oud and Singer (2008) compared the results of SEM and filtering in maximum likelihood estimation of various continuous time models. It turned out that in case of identical models, being appropriate for both procedures, the parameter estimates as well as the associated standard errors are equal. Making use of the extended SEM framework underlying OpenMx, which offers Kalman filtering (Neale et al., 2016), ctsem now combines and fully integrates the SEM and Kalman filter procedures. By the keywords “mxRAM” and “Kalman” one can easily switch between the SEM and Kalman filter procedure (for $N = 1$ “Kalman” is default but this may be overridden by “mxRam”). Under both procedures N may range from large to 1 and initial conditions may be handled by extra parameters as in the SEM tradition or by the stationarity assumption (with regard to means and/or variances-covariances) using the keyword “stationary” as in the SSM tradition. Computing time depends on these choices but most importantly on the number of observation time points T. For small T and small or no variation in observation intervals, SEM (“mxRAM”) takes less computing time than the Kalman filter. SEM will take increasingly more computing time for larger T. For the analysis of the classic example of the Wolfer sunspot data ($N = 1$, $T = 167$) by a CARMA(2,1) model (Phadke & Wu, 1974) ctsem took 1.44 seconds with the Kalman filter and 224.8 seconds with SEM. The longer computing time for larger T is related to the fact that the size of most of the matrices increases by a multiple of the increase in T.

So far there is no comprehensive treatment of how to specify CARMA models by means of SEM and ctsem. Thus, the aim of the present article is, first, to discuss the state space specification of CARMA models in an SEM context and to show how the analysis of CARMA models of increasing order in simulated data sets is performed by ctsem. Second, we will examine the correctness and the quality of the proposed estimation procedure for the specified models and for samples of $N = 1000$, $N = 100$, $N = 50$, and $N = 1$. A large sample of $N = 1000$ is used, because examining the correctness of derivations and procedure with regard to CARMA requires parameter estimates to hardly differ from the simulated values; $N = 100$ and $N = 50$ are used to evaluate the decrease in quality under more common conditions; $N = 1$ is used, because this is a novel situation. Contrary to some suggestions in the literature, there is no doubt that SEM may be applied in situations of $N < T$, including $N = 1$ (Hamaker, Dolan, & Molenaar, 2003; Singer, 2010; Voelkle et al., 2012). Except for a small simulation study in Singer (2012), the SEM continuous time procedure has not yet been applied to $N = 1$, in particular not on empirical data and by ctsem. The analysis is repeated over the 100 separate models for each of the 100 $N = 1$ cases in the sample of $N = 100$. While the different traditions of $N = 1$ and large N in, respectively, SSM and SEM, made it difficult to compare a group model to a model estimated in an individual subject, this is easily done by the SEM procedure implemented in ctsem. We will, third, demonstrate how to test the difference between the full panel model and the $N = 1$ model by means of a subject-group-reproducibility test. Fourth, the proposed analyses will be applied in an empirical example in which the relationships between mood at work and mood at home are analyzed in a sample of $N = 55$ women. The example illustrates many features of the continuous-time procedure, its applicability on $N = 1$, the subject-group-reproducibility, and the extreme diversity of the observation intervals entertained by the subjects.[1]

Continuous time model

Basic model

In discrete time, the multivariate autoregressive moving-average model ARMA(p,q) with p the maximum lag of the dependent variables vector $\mathbf{y}_t$ and q the maximum lag of the error components vector $\mathbf{e}_t$ reads for a model with, for example, $p = 2$ and $q = 1$

$$\mathbf{y}_t = \mathbf{F}_{t,t-1}\mathbf{y}_{t-1} + \mathbf{F}_{t,t-2}\mathbf{y}_{t-2} + \mathbf{G}_t\mathbf{e}_t + \mathbf{G}_{t,t-1}\mathbf{e}_{t-1}. \quad (1)$$

The autoregressive part with **F**-matrices specifies the lagged effects of the dependent variables, while the moving-average part with **G**-matrices handles the incoming errors and lagged errors. So, an ARMA model applies two different mechanisms to let past values influence present values in a time series. By the **F**-matrices observed past values, in Equation (1) $\mathbf{y}_{t-1}, \mathbf{y}_{t-2}$ with successive lags 1, 2, directly influence the present value $\mathbf{y}_t$, while by the **G**-matrices the successive unobserved error components, $\mathbf{e}_t$, $\mathbf{e}_{t-1}$ in Equation (1), are used to predict $\mathbf{y}_t$. The autoregressive **F**-matrices need not be diagonal nor even symmetric, allowing reciprocal effects with different values in both directions to be specified. The errors in vectors $\mathbf{e}_t$, $\mathbf{e}_{t-1}$ are assumed to be independently standard-normally distributed (having covariance matrices **I**) with the matrices $\mathbf{G}_t$, $\mathbf{G}_{t,t-1}$ being lower-triangular. Because $\mathbf{e}_t$, $\mathbf{e}_{t-1}$ have covariance matrices **I**, the moving-average effects, $\mathbf{G}_t\mathbf{e}_t$, $\mathbf{G}_{t,t-1}\mathbf{e}_{t-1}$, get covariance matrices $\mathbf{Q}_t = \mathbf{G}_t\mathbf{G}'_t$[2], $\mathbf{Q}_{t,t-1} = \mathbf{G}_{t,t-1}\mathbf{G}'_{t,t-1}$, which may be nondiagonal and with arbitrary variances on the diagonal. Specifying moving-average effects is no less general than the specification of covariance matrices. Any covariance matrix **Q** can be written as $\mathbf{Q} = \mathbf{G}\mathbf{G}'$ in terms of a lower-triangular matrix (Cholesky factor) **G**. In addition, estimating **G** instead of directly **Q** has the advantage of avoiding possible negative variance estimates showing up in the direct estimate of **Q**.

The moving-average part $\mathbf{G}_t\mathbf{e}_t + \mathbf{G}_{t,t-1}\mathbf{e}_{t-1}$ in Equation (1) may equivalently be written as $\mathbf{G}_{t,t-1}\mathbf{e}_{t-1} + \mathbf{G}_{t,t-2}\mathbf{e}_{t-2}$ with the time indices shifted backward in time from t and $t-1$ to $t-1$ and $t-2$. Replacing the instantaneous error component $\mathbf{G}_t\mathbf{e}_t$ by the lagged one $\mathbf{G}_{t,t-1}\mathbf{e}_{t-1}$(and $\mathbf{G}_{t,t-1}\mathbf{e}_{t-1}$ by $\mathbf{G}_{t,t-2}\mathbf{e}_{t-2}$) could be considered more appropriate, if the errors are taken to stand for the unknown causal influences on the system, which need some time to operate and to affect the system. The fact that the two unobserved consecutive error components get other names but retain their previous values will result in an observationally equivalent (equally fitting) system. Although equivalent, the existence of different representations in discrete time (forward or instantaneous representation in terms of t and $t-1$ and backward or lagging representation in terms of $t-1$ and $t-2$) is nevertheless unsatisfactory. The forward representation puts everything that happens in between t and $t-1$ forward in time at t, the backward representation puts the same information backward in time at $t-1$. From a causal standpoint, though, the backward representation is no less problematic than the forward representation, since it is anticipating effects that in true time will happen only later.

The ambiguous representation in discrete time of the behavior between t and $t-1$ in an ARMA(p, q) model disappears in the analogous continuous-time CARMA

[1] ctsem input and output code for the simulations and the empirical example can be found on the first author's website http://www.socsci.ru.nl/~hano/

[2] In this article the prime sign ′ indicates transpose of a matrix.

(p, q) model, which reads for $p = 2$ and $q = 1$

$$\frac{d^2\mathbf{y}(t)}{dt^2} = \mathbf{F}_0\mathbf{y}(t) + \mathbf{F}_1\frac{d\mathbf{y}(t)}{dt} + \mathbf{G}_0\frac{d\mathbf{W}(t)}{dt} + \mathbf{G}_1\frac{d^2\mathbf{W}(t)}{dt^2}. \tag{2}$$

Writing $\mathbf{y}(t)$ instead of $\mathbf{y}_t$ emphasizes the development of $\mathbf{y}$ across continuous time. The role of successive lags in discrete time is taken over by successive derivatives in continuous time. The causally unsatisfactory instantaneous and lagging representations meet, so to speak, in the derivatives, which instead of using a discrete time interval $\Delta t = t - (t-1) = 1$, let the time interval go to zero: $\Delta t \to 0$. Derivative $d\mathbf{y}(t)/dt$, for example, informs about the value of the difference $\mathbf{y}_t - \mathbf{y}_{t-1}$ per time interval unit, as the time interval Δt goes to zero (gradient at t).

The error process in continuous time is the famous Wiener process[3] $\mathbf{W}(t)$ or random walk through continuous time. Its main defining properties are the conditions of independently and normally distributed increments, $\Delta\mathbf{W}(t) = \mathbf{W}(t) - \mathbf{W}(t - \Delta t)$, having mean $\mathbf{0}$ and covariance matrix $\Delta t\mathbf{I}$. This means that the increments with arbitrary Δt are for $\Delta t = 1$ standard-normally distributed as assumed for $\mathbf{e}_t, \mathbf{e}_{t-1}$ in discrete time. Likewise, the role of lower-triangular $\mathbf{G}_0, \mathbf{G}_1$ in (2) is analogous to the role of $\mathbf{G}_t, \mathbf{G}_{t,t-1}$ in discrete time. Derivative $d\mathbf{W}(t)/dt$ (white noise) does not exist in the classical sense but can be defined in the generalized function sense and also integral $\int_{t_0}^{t} \mathbf{G}d\mathbf{W}(t)$ can be defined rigorously (Kuo, 2006, pp. 104–105 and pp. 260–261). Integrals are needed to go back again from the continuous time specification in Equation (2) to the observed values in discrete time, which will be the topic of the next subsection. The next subsection will show in more detail how discrete time and continuous time equations such as (1) and (2) become connected as $\Delta t \to 0$.

Equation (2) is sometimes written as

$$\mathbf{F}_0\mathbf{y}(t) + \mathbf{F}_1\frac{d\mathbf{y}(t)}{dt} + \mathbf{F}_2\frac{d^2\mathbf{y}(t)}{dt^2} = \mathbf{G}_0\frac{d\mathbf{W}(t)}{dt} + \mathbf{G}_1\frac{d^2\mathbf{W}(t)}{dt^2}. \tag{3}$$

with $\mathbf{F}_2 = -\mathbf{I}$ and opposite signs for $\mathbf{G}_0, \mathbf{G}_1$, making it clear that the CARMA(2,1) model has $\mathbf{F}_2$ as the highest degree matrix in the autoregressive part and $\mathbf{G}_1$ as the highest degree in the moving-average part.

We will now show how the CARMA(2,1) model in Equation (2) and the general CARMA(p, q) model can be formulated as special cases of the continuous-time state space model. The continuous-time state space model consists of two equations: A latent dynamic equation (4) with so-called drift matrix $\mathbf{A}$ and diffusion matrix $\mathbf{G}$, and a measurement equation (5) with loading matrix $\mathbf{C}$ and measurement error vector $\mathbf{v}(t)$:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{G}\frac{d\mathbf{W}(t)}{dt}, \tag{4}$$
$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{v}(t). \tag{5}$$

Many dynamic phenomena in continuous time physics are described by drift and diffusion. In meteorology, for example, the movement of clouds from one place to another is called drift, shrinkage or expansion on the same place is called diffusion. In general, the variables in state vector $\mathbf{x}(t)$ are assumed to be latent and only indirectly measured by the observed variables in $\mathbf{y}(t)$ with measurement errors in $\mathbf{v}(t)$. The measurement error vector $\mathbf{v}(t)$ is assumed independent of $\mathbf{x}(t)$ and normally distributed: $\mathbf{v}(t) \sim N(\mathbf{0}, \mathbf{R})$. For the initial state $\mathbf{x}(t_0)$ we assume $\mathbf{x}(t_0) \sim N(\boldsymbol{\mu}_{\mathbf{x}(t_0)}, \boldsymbol{\Phi}_{\mathbf{x}(t_0)})$. Often, but not necessarily, it is assumed $E[\mathbf{x}(t_0)] = \boldsymbol{\mu}_{\mathbf{x}(t_0)} = \mathbf{0}$. The latter would imply that $E[\mathbf{x}(t)] = E[\mathbf{x}(t_0)] = \mathbf{0}$ and, in case all eigenvalues of $\mathbf{A}$ have negative real part, the mean trajectory of the model would then have $\mathbf{0}$ as stable equilibrium state.

In state space form, the observed second-order model CARMA(2,1) in Equation (2) gets a state vector $\mathbf{x}(t) = [\mathbf{x}_1(t)'\mathbf{x}_2(t)']'$, which is two times the size of the observed vector $\mathbf{y}(t)$. The first part $\mathbf{x}_1(t)$ is not directly related to the observed variables and thus belongs to the latent part of the state space model. This special case of the state space model equates the second part to the observed vector: $\mathbf{y}(t) = \mathbf{x}_2(t)$. Equation (2) then follows from state space model (4)-(5) by specification

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{F}_0 \\ \mathbf{I} & \mathbf{F}_1 \end{bmatrix}, \mathbf{G} = \begin{bmatrix} \mathbf{G}_0 & \mathbf{0} \\ \mathbf{G}_1 & \mathbf{0} \end{bmatrix}, \mathbf{C} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \end{bmatrix}, \mathbf{v}(t) = \mathbf{0}. \tag{6}$$

Applying (6) we find first

$$\frac{d\mathbf{x}_1(t)}{dt} = \mathbf{F}_0\mathbf{x}_2(t) + \mathbf{G}_0\frac{d\mathbf{W}(t)}{dt}, \tag{7}$$

$$\begin{aligned} \frac{d\mathbf{x}_2(t)}{dt} &= \mathbf{x}_1(t) + \mathbf{F}_1\mathbf{x}_2(t) + \mathbf{G}_1\frac{d\mathbf{W}(t)}{dt} \Rightarrow \\ \frac{d^2\mathbf{x}_2(t)}{dt^2} &= \frac{d\mathbf{x}_1(t)}{dt} + \mathbf{F}_1\frac{d\mathbf{x}_2(t)}{dt} + \mathbf{G}_1\frac{d^2\mathbf{W}(t)}{dt^2}. \end{aligned} \tag{8}$$

Substituting (7) into the implication in (8) gives

$$\frac{d^2\mathbf{x}_2(t)}{dt^2} = \mathbf{F}_0\mathbf{x}_2(t) + \mathbf{F}_1\frac{d\mathbf{x}_2(t)}{dt} + \mathbf{G}_0\frac{d\mathbf{W}(t)}{dt} + \mathbf{G}_1\frac{d^2\mathbf{W}(t)}{dt^2}, \tag{9}$$

which for $\mathbf{y}(t) = \mathbf{x}_2(t)$ leads to the CARMA(2,1) model in (2).

The specification in the previous paragraph can be generalized to find the CARMA(p,q) model as follows in state

[3] In this article, we follow common practice to write the Wiener process by capital letter $\mathbf{W}$, although it is a vector whose size should be inferred from the context.

space form, where $r = \max(p, q+1)$.

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{F}_0 \\ \mathbf{I} & \mathbf{0} & \ddots & \mathbf{0} & \mathbf{F}_1 \\ \mathbf{0} & \mathbf{I} & \ddots & \mathbf{0} & \mathbf{F}_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I} & \mathbf{F}_{r-1} \end{bmatrix},$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_0 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{G}_1 & \mathbf{0} & \ddots & \mathbf{0} & \mathbf{0} \\ \mathbf{G}_2 & \mathbf{0} & \ddots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{G}_{r-1} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} \end{bmatrix}, \mathbf{v}(t) = \mathbf{0}. \tag{10}$$

$\mathbf{G}$ leads to diffusion covariance matrix $\mathbf{Q} = \mathbf{G}\mathbf{G}'$, which has Cholesky factor based covariance matrices $\mathbf{G}_i\mathbf{G}_i{}'$ on the diagonal and in case $q > 0$ off-diagonal matrices $\mathbf{G}_i\mathbf{G}_j{}'$ $(i, j = 0, 1, \ldots, r-1)$. Instead of (10) which is based on Singer (1992b), one often finds the alternative state space form (11) (see e.g. Singer, 1992b; Tómasson, 2011; Tsai & Chan, 2000). Here the moving average matrices $\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_{r-1}$ are rewritten as $\mathbf{G}_i = \mathbf{H}_i\mathbf{G}_0$ in terms of corresponding matrices $\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_{r-1}$, specified in the measurement part of the state space model, and $\mathbf{G}_0$. For two reasons we prefer (10) in the case of CARMA(p,q) models with $q > 0$. Application of (11) requires the matrices $\mathbf{H}_i$ and $\mathbf{F}_j$ to commute, which in practice restricts the applicability to the univariate case. In addition, using the measurement part for the moving average specification would make it difficult to specify at the same time measurement parameters. For CARMA(p,0) models (all $\mathbf{H}_i = \mathbf{0}$) we prefer (11), because each state variable in state vector $\mathbf{x}(t)$ is easily interpretable as the derivative of the previous one. The interpretation of the state variables in (7)-(8) is less simple.

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \ddots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{I} \\ \mathbf{F}_0 & \mathbf{F}_1 & \mathbf{F}_2 & \cdots & \mathbf{F}_{r-1} \end{bmatrix},$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{G}_0 \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{I} & \mathbf{H}_1 & \mathbf{H}_2 & \cdots & \mathbf{H}_{r-1} \end{bmatrix}, \mathbf{v}(t) = \mathbf{0}. \tag{11}$$

The fact that the general CARMA(p,q) model fits seamlessly into the state space model, means that all continuous-time time series problems in modeling and estimation can be handled by state space form (4)-(5). The state space approach reformulates higher-order models ($p > 1$ and/or $q > 0$) as a first-order model (with r components) and this will be applied in the sequel.

Connecting discrete and continuous time model in the EDM

The EDM combines the discrete time and continuous time model and does so in an exact way. It is by the EDM that the exact procedure in this article differentiates from many approximate procedures found in the literature (e.g., Gasimova et al., 2014; Steele & Ferrer, 2011a, 2011b), which try to avoid the nonlinearity in the connection the EDM makes between discrete and continuous time. In an extensive simulation study (Oud, 2007), the exact procedure was compared to a well-known example of those approximate procedures and found to give considerably lower biases and root mean squared error values for the continuous time parameter estimates.

Our aim is to show what the exact connection looks like between the general continuous-time state space model and its discrete-time counterpart, derived from it. The first-order models ARMA(1,0) and CARMA(1,0) in state space form differ from the general discrete and continuous time state-space model only in a simpler measurement equation. So, having made the exact connections between the general state space models and thus between ARMA(1,0) and CARMA(1,0) and knowing that each CARMA(p,q) model can be written as a special case of the general state space model, the exact connections between CARMA(p,q) and ARMA(p,q), where the latter is derived from the former, follow. Next we consider the question of making an exact connection between an arbitrary ARMA(p^*,q^*) model and a CARMA(p,q) model, where the degrees p^* and p as well as q^* and q need not be equal.

Comparing discrete time equation (1) to the general discrete-time state space model in (12)-(13) one observes that the latter becomes immediately the ARMA(1,0) model for $\mathbf{y}_t = \mathbf{x}_t$.

$$\mathbf{x}_t = \mathbf{A}_{\Delta t}\mathbf{x}_{t-\Delta t} + \mathbf{G}_{\Delta t}\mathbf{e}_{t-\Delta t}, \tag{12}$$
$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{v}_t. \tag{13}$$

Inserting arbitrary lag Δt instead of fixed lag $\Delta t = 1$ enables us to put discrete time models with different intervals (e.g., years and months) on the same time scale and to connect them to the common underlying continuous time model for $\Delta t \to 0$.

State equation (12) can be put in the equivalent difference quotient form

$$\frac{\Delta \mathbf{x}_t}{\Delta t} = \mathbf{A}_{*\Delta t}\mathbf{x}_{t-\Delta t} + \mathbf{G}_{\Delta t}\frac{\mathbf{e}_{t-\Delta t}}{\Delta t} \text{ for}$$
$$\mathbf{A}_{*\Delta t} = (\mathbf{A}_{\Delta t} - \mathbf{I})/\Delta t \text{ implying } \mathbf{A}_{\Delta t} = \mathbf{I} + \mathbf{A}_{*\Delta t}\Delta t. \tag{14}$$

So we have the discrete-time state space model in two forms: difference quotient form (14) and solution form (12). Equation (12) is called the solution of (14), because it describes the actual state transition across time in accordance with (14) and is so said to satisfy the difference quotient equation. Note that analogously the general continuous-time state space model (4)-(5) immediately accommodates the special CARMA(1,0) model for $\mathbf{y}(t) = \mathbf{x}(t)$ and can be put in two forms: stochastic differential equation (4) and its solution (15) (Arnold, 1974; Singer, 1990):

$$\mathbf{x}(t) = \mathrm{e}^{\mathbf{A}\Delta t}\mathbf{x}(t-\Delta t) + \int_{t-\Delta t}^{t} \mathrm{e}^{\mathbf{A}(t-s)}\mathbf{G}\mathrm{d}\mathbf{W}(s). \tag{15}$$

In the exact discrete model EDM the connection between discrete and continuous time is made by means of the solutions, which in both cases describe the actual transition from the previous state at $t - \Delta t$ to the next state at t. The EDM thus combines both models and connects them exactly by the equalities:

$$\mathbf{A}_{\Delta t} = \mathrm{e}^{\mathbf{A}\Delta t} \text{ and } \mathbf{Q}_{\Delta t} = \int_{t-\Delta t}^{t} \mathrm{e}^{\mathbf{A}(t-s)}\mathbf{Q}\mathrm{e}^{\mathbf{A}'(t-s)}\mathrm{d}s. \tag{16}$$

While discrete time autoregression matrix $\mathbf{A}_{\Delta t}$ and continuous time drift matrix $\mathbf{A}$ are connected via the highly nonlinear matrix exponential (Moler & Van Loan, 2003), the errors are indirectly connected by their covariance matrices $\mathbf{Q}_{\Delta t} = \mathbf{G}_{\Delta t}\mathbf{G}'_{\Delta t}$ and $\mathbf{Q} = \mathbf{G}\mathbf{G}'$. In estimating, after finding the drift matrix $\mathbf{A}$ on the basis of $\mathbf{A}_{\Delta t}$, next on the basis of $\mathbf{G}_{\Delta t}$ the diffusion matrix $\mathbf{G}$ is found.

Let us illustrate the connection between CARMA(1,0) and ARMA(1,0) by an example that will be used later on in the simulations. If $\mathbf{A} = \begin{bmatrix} -1.0 & 0.2 \\ 0.3 & -1.5 \end{bmatrix}$, the exact connection in the EDM is for $\Delta t = 1$ made by $\mathbf{A}_{\Delta t=1} = \mathrm{e}^{\mathbf{A}\Delta t} = \mathrm{e}^{\mathbf{A}} = \begin{bmatrix} 0.377 & 0.058 \\ 0.088 & 0.231 \end{bmatrix}$. $\mathbf{A}$ in the differential equation may be compared to $\mathbf{A}_{*\Delta t=1} = (\mathrm{e}^{\mathbf{A}\Delta t} - \mathbf{I})/\Delta t = \mathrm{e}^{\mathbf{A}} - \mathbf{I} = \begin{bmatrix} -0.623 & 0.058 \\ 0.088 & -0.769 \end{bmatrix}$ in the difference equation for $\Delta t = 1$. For $\Delta t = 0.1$ we get $\mathbf{A}_{*\Delta t=0.1} = (\mathrm{e}^{\mathbf{A}\times 0.1} - \mathbf{I})/0.1 = \begin{bmatrix} -0.949 & 0.177 \\ 0.264 & -1.390 \end{bmatrix}$ which is much closer to $\mathbf{A}$ and for $\Delta t = 0.001$ $\mathbf{A}_{*\Delta t=0.001} = (\mathrm{e}^{\mathbf{A}\times 0.001} - \mathbf{I})/0.001 = \begin{bmatrix} -0.999 & 0.200 \\ 0.300 & -1.499 \end{bmatrix}$ becomes virtually equal to $\mathbf{A}$.

Making an exact connection between an ARMA(p^*,q^*) model and a model CARMA(p,q) such that the ARMA process $\{\mathbf{y}_t; t = 0, t = \Delta t, t = 2\Delta t, \ldots\}$ generated by ARMA(p^*,q^*) is a subset of the CARMA process $\{\mathbf{y}(t); t \geq 0\}$ generated by CARMA(p,q) is called "embedding" in the literature. The degrees p^* and q^* of the embedded model and p and q of the embedding model need not be equal. Embeddability is a much-debated issue. Embedding is not always possible and need not be unique. Embedding is clearly possible for the case of ARMA(1,0) model $\mathbf{y}_t = \mathbf{A}_{\Delta t}\mathbf{y}_{t-\Delta t} + \mathbf{G}_{\Delta t}\mathbf{e}_{t-\Delta t}$ derived from CARMA(1,0) model $\frac{\mathrm{d}\mathbf{y}(t)}{\mathrm{d}t} = \mathbf{A}\mathbf{y}(t) + \mathbf{G}\frac{\mathrm{d}\mathbf{W}(t)}{\mathrm{d}t}$ with $\mathbf{A}_{\Delta t} = \mathrm{e}^{\mathbf{A}\Delta t}$, $\mathbf{Q}_{\Delta t} = \int_{t-\Delta t}^{t} \mathrm{e}^{\mathbf{A}(t-s)}\mathbf{Q}\mathrm{e}^{\mathbf{A}'(t-s)}\mathrm{d}s$, $\mathbf{Q}_{\Delta t} = \mathbf{G}_{\Delta t}\mathbf{G}'_{\Delta t}$, $\mathbf{Q} = \mathbf{G}\mathbf{G}'$ as shown above. The same is true for the higher-order ARMA(p,q) model, derived from CARMA(p,q). However, in general it is nontrivial to prove embeddability and to find the parameters of the CARMA(p,q) model embedding an ARMA(p^*,q^*) process. For example, not all ARMA(1,0) processes have a CARMA(1,0) process in which it can be embedded. A well-known example is the simple univariate process $y_t = a_{\Delta t}y_{t-\Delta t} + g_{\Delta t}e_{t-\Delta t}$ with $-1 < a_{\Delta t} < 0$, because there does not exist any a for which $a_{\Delta t} = \mathrm{e}^{a\Delta t}$ can be negative. However, Chan and Tong (1987) showed that for this ARMA(1,0) process with $-1 < a_{\Delta t} < 0$ a higher order CARMA(2,1) process can be found, in which it can be embedded.

Also embeddability need not be unique. Different CARMA models may embed one and the same ARMA model. A classic example is "aliasing" in the case of matrices $\mathbf{A}$ with complex conjugate eigenvalue pairs $\lambda_{1,2} = \alpha \pm \beta i$ with i the imaginary unit (Hamerle, Nagl, & Singer, 1991; Phillips, 1973). Such complex eigenvalue pairs imply processes with oscillatory movements. Adding $\pm k2\pi/\Delta t$ to β leads for arbitrary integer k to a different $\mathbf{A}$ with a different oscillation frequency but does not change $\mathbf{A}_{\Delta t} = \mathrm{e}^{\mathbf{A}\Delta t}$ and so may lead to the same ARMA model. The consequence is that the CARMA model cannot uniquely be determined (identified) by the ARMA model and the process generated by it. Fortunately, the number of aliases in general is limited in the sense that there exists only a finite number of aliases that lead for the same ARMA model to a real $\mathbf{G}$ and so to a positive definite $\mathbf{Q}$ in the CARMA model (Hansen & Sargent, 1983). The size of the finite set additionally depends on the observation interval Δt, a smaller Δt leading to less aliases. The number of aliases may also be limited by sampling the observations in the discrete time process at unequal intervals (Oud & Jansen, 2000; Tómasson, 2015; Voelkle & Oud, 2013).

An important point with regard to the state space modeling technique of time series is the latent character of the state. Even in the case of an observed ARMA(p,q) or CARMA(p,q) model of such low dimension as $p = 2$, we have seen that part of the state is not directly connected to the data. This especially has consequences for the initial time point. Suppose for the ARMA(2,0) model in state space form (17)-(18)

$$\begin{array}{ccccccc} \begin{bmatrix} \mathbf{x}_{1,t-\Delta t} \\ \mathbf{x}_{1,t} \end{bmatrix} & = & \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{F}_{t,t-2\Delta t} & \mathbf{F}_{t,t-\Delta t} \end{bmatrix} & \begin{bmatrix} \mathbf{x}_{1,t-2\Delta t} \\ \mathbf{x}_{1,t-\Delta t} \end{bmatrix} & + & \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{t,t-\Delta t} \end{bmatrix} & \mathbf{e}_{t-\Delta t}, \\ \mathbf{x}_t & = & \mathbf{A}_{\Delta t} & \mathbf{x}_{t-\Delta t} & + & \mathbf{G}_{\Delta t} & \mathbf{e}_{t-\Delta t} \end{array} \tag{17}$$

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{0} & \mathbf{I} \end{bmatrix} \mathbf{x}_t, \tag{18}$$

the initial time point, where the initial data are located, is $t_0 = t - \Delta t$. It means that there are no data directly or indirectly connected to (the lagged) part of $\mathbf{x}_{t-\Delta t}$. The initial parameters related to this part can nevertheless be estimated but become highly dependent on the model structure and the uncertainty will be reflected in high standard errors. It does not help to start the model at later time point $t_0 + \Delta t$, as this would result in data loss, since the $\mathbf{0}$ in (18) simply eliminates the lagged part of $\mathbf{x}_{t-\Delta t}$ without any connection to the data. Similar remarks apply to the initial derivative $\mathrm{d}\mathbf{x}_1(t)/\mathrm{d}t$

$$\begin{array}{ccccccc} \begin{bmatrix} \dfrac{\mathrm{d}\mathbf{x}_1(t)}{\mathrm{d}t} \\ \dfrac{\mathrm{d}^2\mathbf{x}_1(t)}{\mathrm{d}t^2} \end{bmatrix} & = & \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{F}_0 & \mathbf{F}_1 \end{bmatrix} & \begin{bmatrix} \mathbf{x}_1(t) \\ \dfrac{\mathrm{d}\mathbf{x}_1(t)}{\mathrm{d}t} \end{bmatrix} & + & \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_0 \end{bmatrix} & \dfrac{\mathrm{d}\mathbf{W}(t)}{\mathrm{d}t}, \\ \dfrac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} & = & \mathbf{A} & \mathbf{x}(t) & + & \mathbf{G} & \dfrac{\mathrm{d}\mathbf{W}(t)}{\mathrm{d}t}, \end{array} \tag{19}$$

$$\mathbf{y}(t) = \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \mathbf{x}(t), \tag{20}$$

in (19)-(20), which is not directly connected to the data and cannot be computed at the initial time point. Again, the related initial parameters can be estimated in principle. It is important to realize, however, that dependent on the model structure, the number of time points analyzed and the length of the observation intervals, these initial parameter estimates can become extremely unreliable. In a simulation study of a CARMA(2,0) model with oscillating movements, Oud and Singer (2008) found in the case of long interval lengths extremely large standard errors for the estimates related to the badly measured initial $\mathrm{d}\mathbf{x}_1(t)/\mathrm{d}t$. This lack of data and relative unreliability of estimates is the price one has to pay for choosing higher order ARMA(p, q) and CARMA(p, q) models.

Extended continuous time model

The extended continuous-time state space model reads

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \boldsymbol{\gamma} + \mathbf{G}\frac{\mathrm{d}\mathbf{W}(t)}{\mathrm{d}t}, \tag{21}$$

$$\mathbf{y}_{t_i} = \mathbf{C}\mathbf{x}(t_i) + \mathbf{D}\mathbf{u}(t_i) + \boldsymbol{\kappa} + \mathbf{v}_{t_i}. \tag{22}$$

Technical details of the extended model are discussed by Hamerle, Singer, and Nagl (1993). In comparison to the basic model in (4)-(5), the extended model exhibits one minor notational change and two major additions. The minor change is in the measurement equation and is only meant to emphasize the discrete time character of the data at the discrete time points $t_i (i = 0, \ldots, T-1)$ with $\mathbf{x}(t_i)$ and $\mathbf{u}(t_i)$ sampling the continuous time vectors $\mathbf{x}(t)$ and $\mathbf{u}(t)$ at the observation time points. One major addition are the effects $\mathbf{B}$ and $\mathbf{D}$ of fixed exogenous variables in the vector $\mathbf{u}(t)$. The other is the addition of random subject effect vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\kappa}$ to the equations. While the (statistically) fixed variables in $\mathbf{u}(t)$ may change across time (time-varying exogenous variables), the subject specific effects $\boldsymbol{\gamma}$ and $\boldsymbol{\kappa}$ with possibly a different value for each subject in the sample are assumed to be constant across time, but normally distributed random variables: $\boldsymbol{\gamma} \sim N(\mathbf{0}, \boldsymbol{\Phi}_{\boldsymbol{\gamma}})$, $\boldsymbol{\kappa} \sim N(\mathbf{0}, \boldsymbol{\Phi}_{\boldsymbol{\kappa}})$. To distinguish them from the changing states the constant random effects in $\boldsymbol{\gamma}$ are called traits. Because trait vector $\boldsymbol{\gamma}$ is modeled to influence $\mathbf{x}(t)$ continuously, before as well as after t_0, $\boldsymbol{\Phi}_{\mathbf{x}(t_0),\boldsymbol{\gamma}}$, the covariance matrix between initial state and traits, cannot in general be assumed zero. The additions in state equation (21) lead to the following extended solution:

$$\begin{aligned} \mathbf{x}(t) = {} & \mathrm{e}^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0) + \int_{t_0}^{t} \mathrm{e}^{\mathbf{A}(t-s)}\mathbf{B}\mathbf{u}(s)\mathrm{d}s \\ & + \mathbf{A}^{-1}[\mathrm{e}^{\mathbf{A}(t-t_0)} - \mathbf{I}]\boldsymbol{\gamma} + \int_{t_0}^{t} \mathrm{e}^{\mathbf{A}(t-s)}\mathbf{G}\mathrm{d}\mathbf{W}(s). \end{aligned} \tag{23}$$

Exogenous variables

We have seen in the case of stability, defined as all eigenvalues of $\mathbf{A}$ having negative real part, that basic model (4)-(5) has in its mean trajectory $\mathbf{0}$ as stable equilibrium state. Exogenous $\mathbf{B}\mathbf{u}(t)$ and $\mathbf{D}\mathbf{u}(t)$ accommodate nonzero constant as well as nonconstant mean trajectories $E[\mathbf{x}(t)]$ and $E[\mathbf{y}(t)]$ even in the case of stability. By far the most popular exogenous input function is the unit function, $e(s) = 1$ for all s over the interval, with the effect $\mathbf{b}_e$ called intercept and integrating over interval $[t_0, t)$ into $\int_{t_0}^{t} \mathrm{e}^{\mathbf{A}(t-s)}\mathbf{b}_e e(s)\mathrm{d}s = \mathbf{A}^{-1}[\mathrm{e}^{\mathbf{A}(t-t_0)} - \mathbf{I}]\mathbf{b}_e$. In the measurement equation the effect of the unit variable in $\mathbf{D}$ is called measurement intercept or origin and allows measurement instruments to have scales with different starting points in addition to the different units specified in $\mathbf{C}$.

Useful in describing sudden changes in the environment is the intervention function, a step function that takes on a certain value a until a specific time point t' and changes to value b at that time point until the end of the interval: $i(s) = a$ for all $s < t'$, $i(s) = b$ for all $s \geq t'$. An effective way of handling the step or piecewise constant function is a two-step procedure, in which Equation (24) is applied twice. First with $\mathbf{u}(t_0)$ containing the step function value of the first step before t' and next with $\mathbf{u}(t_0)$

containing the step function value of the second step.

$$\mathbf{x}(t) = \mathrm{e}^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0) + \mathbf{A}^{-1}[\mathrm{e}^{\mathbf{A}(t-t_0)} - \mathbf{I}]\mathbf{Bu}(t_0) + \mathbf{A}^{-1}[\mathrm{e}^{\mathbf{A}(t-t_0)} - \mathbf{I}]\boldsymbol{\gamma} + \int_{t_0}^{t} \mathrm{e}^{\mathbf{A}(t-s)}\mathbf{G}\mathrm{d}\mathbf{W}(s). \tag{24}$$

In the second step the result $\mathbf{x}(t)$ of the first step is inserted as $\mathbf{x}(t_0)$. The relatively simple solution equation (24) that will be used in the sequel, has much more general applicability, though, than just for step functions. It can be used to approximate any exogenous behavior function in steps and approximate its effect arbitrarily closely by oversampling (dividing the observation interval in smaller intervals) and choosing the oversampling intervals sufficiently small.[4]

The handling of exogenous variables takes another twist, when it is decided to endogenize them. The problem with oversampling is that it is often not known, what the exogenous behavior function looks like in between observations. By endogenizing exogenous variables they are taken out of the exogenous vector $\mathbf{u}(t)$ in (21), added as extra variables to the state vector $\mathbf{x}(t)$ in $\mathrm{d}\mathbf{x}(t)/\mathrm{d}t$ and in the same way as the other state variables related to their past values by means of drift matrix $\mathbf{A}$ and made subject to continuous-time error by means of diffusion matrix $\mathbf{G}$. Advantages of endogenizing are its non-approximate nature and the fact that the new state variables may not only be modeled to influence the other state variables but also to be reciprocally influenced by them.

In addition to analyzing differences between time points within subjects, exogenous variables also allow for analyzing differences between subjects in case of an $N > 1$ sample. Suppose the first element of $\mathbf{u}(t)$ is the unit variable, corresponding in $\mathbf{B}$ with first column $\mathbf{b}_e$, and the second element is a dummy variable differentiating boys and girls (boys 0 at all time points and girls 1 at all time points) and corresponding to second column $\mathbf{b}_d$. Supposing all remaining variables have equal values, the mean or expectation $E[\mathbf{x}(t)]$ of girls over the interval $[t_0, t)$ will then differ by the amount of $\mathbf{A}^{-1}[\mathrm{e}^{\mathbf{A}(t-t_0)} - \mathbf{I}]\mathbf{b}_d$ from the one of boys. This amount will be zero for $t - t_0 = 0$, but relating the dummy variable at initial time point t_0 in a simple regression also to the state variables at initial time point t_0 allows to distinguish different initial means $E[\mathbf{x}(t_0)]$ for boys and girls. Thus, the same dummy variable at t_0 impacts both the state variables at t_0 and according to the state space model over the interval $t - t_0$ the state variables at the next observation time point.

Traits

Although, as we have just seen, there is some flexibility in the mean or expected trajectory, because subjects in different groups can have different mean trajectories, it would nevertheless be a strange implication of the model, if a subject's expected current and future behavior is totally dependent on the group of which he or she is modeled to be a member. It should be noted that the expected trajectories are not only interesting per se, but they also play a crucial role in the estimated latent sample trajectory of a subject, defined as the conditional mean $E[\mathbf{x}(t)|\mathbf{y}]$, where $\mathbf{y}$ is the total data vector of the subject (Kalman smoother), or $E[\mathbf{x}(t)|\mathbf{y}[t_0, t]]$, where $\mathbf{y}[t_0, t]$ is all data up to and including t (Kalman filter). In a model without traits, the subject regresses towards (in the case of a stable model) or egresses from (in an unstable model) the mean trajectory of its group. The consequences are particularly dramatic for predictions, because after enough time is elapsed, the subject's trajectory in a stable model will be coinciding with its group trajectory.

From solution equation (23) it becomes clear, however, that in the state-trait model each subject gets its own mean trajectory that differs from the group's mean. After moving the initial time point of a stable model sufficiently far into the past, $t_0 \to -\infty$, the subject's expected trajectory is

$$E[\mathbf{x}(t)|\boldsymbol{\gamma}] = \int_{-\infty}^{t} \mathrm{e}^{\mathbf{A}(t-s)}\mathbf{Bu}(s)\mathrm{d}s - \mathbf{A}^{-1}\boldsymbol{\gamma}, \tag{25}$$

which keeps a subject specific distance $-\mathbf{A}^{-1}\boldsymbol{\gamma}$ from the subject's group mean trajectory $E[\mathbf{x}(t)] = \int_{-\infty}^{t} \mathrm{e}^{\mathbf{A}(t-s)}\mathbf{Bu}(s)\mathrm{d}s$. As a result the subject's sample trajectory regresses towards its own mean instead of its group mean. A related advantage of the state-trait model is that it clearly distinguishes trait variance (diagonals of $\boldsymbol{\Phi}_{\gamma}$), also called unobserved heterogeneity between subjects, from stability. Because in a pure state model ($\boldsymbol{\gamma} = \mathbf{0}$) all subject-specific mean trajectories coincide with the group mean trajectory, trait variance and stability are confounded in the sense that an actually nonzero trait variance leads to a less stable model (eigenvalues of $\mathbf{A}$ having less negative real part) as a surrogate for keeping the subject-specific mean trajectories apart. In a state-trait model, however, stability is not hampered by hidden heterogeneity. A similar distinction between state and trait is made by Hamaker, Dolan, and Molenaar (2005), although in their approach traits tend to replace states (some individuals are more "traited" than others; p. 228), while in a state-trait-model both are specified.

[4] A better approximation than a step function is given by a piecewise linear or polygonal approximation (Oud & Jansen, 2000; Singer, 1992a). Then we write $\mathbf{u}(t)$ in (23) as $\mathbf{u}(t) = \mathbf{u}(t_0) + (t - t_0)\mathbf{b}_{(t_0,t]}$ and (24) becomes: $\mathbf{x}(t) = \mathrm{e}^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0) + \mathbf{A}^{-1}[\mathrm{e}^{\mathbf{A}(t-t_0)} - \mathbf{I}]\mathbf{Bu}(t_0) + \{\mathbf{A}^{-2}[\mathrm{e}^{\mathbf{A}(t-t_0)} - \mathbf{I}] - \mathbf{A}^{-1}(t - t_0)\}\mathbf{Bb}_{(t_0,t]} + \mathbf{A}^{-1}[\mathrm{e}^{\mathbf{A}(t-t_0)} - \mathbf{I}]\boldsymbol{\gamma} + \int_{t_0}^{t} \mathrm{e}^{\mathbf{A}(t-s)}\mathbf{G}\mathrm{d}\mathbf{W}(s)$. (24A)

It should be noted that the impact of the fixed and random effects $\mathbf{Bu}(t)$ and $\boldsymbol{\gamma}$ in the state equation (21) is quite different from that of $\mathbf{Du}(t)$ and $\boldsymbol{\kappa}$ in the measurement equation (22). The latter is a one-time snapshot event with no consequences for the future. It just reads out in a specific way the current contents of the system's state. However, the state equation is a dynamic equation where influences may have long lasting and cumulative future effects that are spelled out by Equation (23) or (24). In particular, the traits $\boldsymbol{\gamma}$ differ fundamentally from the nondynamic or "random measurement bias" $\boldsymbol{\kappa}$, earlier proposed for panel data by Goodrich and Caines (1979), Jones (1993), and Shumway and Stoffer (2000).

Model estimation by SEM

As emphasized above, if the data are collected in discrete time, we need the EDM to connect the continuous time parameter matrices to the discrete time parameter matrices describing the data. The continuous time model in state space form contains eight parameter matrices that are connected to the corresponding discrete time matrices as shown in Table 1. In Table 1 $\otimes$ is the Kronecker product and the row operator puts the elements of the $\mathbf{Q}$ matrix row-wise in a column vector, whereas irow stands for the inverse operation.

While there are only eight continuous time parameter matrices, there may be many more discrete time parameter matrices. This is typically the case for the dynamic matrices $\mathbf{A}_{\Delta t_{i,j}}, \mathbf{B}_{\Delta t_{i,j}}, \mathbf{Q}_{\Delta t_{i,j}}$. The observation time points $t_i (i = 0, \ldots, T-1)$ may differ for different subjects $j (j = 1, \ldots, N)$ but also the observation intervals $\Delta t_{i,j} = t_{i,j} - t_{i-1,j}$ $(i = 1, \ldots, T-1)$ between the observation time points. Different observation intervals can lead to many different discrete time matrices $\mathbf{A}_{\Delta t_{i,j}}, \mathbf{B}_{\Delta t_{i,j}}, \mathbf{Q}_{\Delta t_{i,j}}$ but all based on the same underlying continuous time matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{Q}$. The most extreme case is that none of the intervals is equal to any other interval, a situation a traditional discrete time analysis would be unable to cope with but is unproblematic in continuous time analysis (Oud & Voelkle, 2014).

Table 1. Relationships between the single set of eight continuous-time parameter matrices/vector (left) and possibly many discrete-time matrices (right) as specified in the EDM; initial parameter vector and matrix $\boldsymbol{\mu}_{\mathbf{x}(t_0)}$ and $\boldsymbol{\Phi}_{\mathbf{x}(t_0)}$ are equal between continuous and discrete time; measurement parameter matrices **C**,**D**, and **R** are chosen equal to realize measurement invariance.

Continuous time	Discrete time
$\mathbf{A}$	$\mathbf{A}_{\Delta t_{i,j}} = e^{\mathbf{A}\Delta t_{i,j}}$
$\mathbf{B}$	$\mathbf{B}_{\Delta t_{i,j}} = \mathbf{A}^{\circ}_{\Delta t_{i,j}}\mathbf{B}$ for $\mathbf{A}^{\circ}_{\Delta t_{i,j}} = \mathbf{A}^{-1}(e^{\mathbf{A}\Delta t_{i,j}} - \mathbf{I})$
$\mathbf{Q} = \mathbf{GG}'$	$\mathbf{Q}_{\Delta t_{i,j}} = \int_{t-\Delta t_{i,j}}^{t} e^{\mathbf{A}(t-s)}\mathbf{Q}e^{\mathbf{A}'(t-s)}ds = \text{irow}[\mathbf{A}_{\#}^{-1}(e^{\mathbf{A}_{\#}\Delta t_{i,j}} - \mathbf{I})\text{row}\mathbf{Q}]$ for $\mathbf{A}_{\#} = \mathbf{A} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{A}$
$\mathbf{C}$	$\mathbf{C}_{t_{i,j}} = \mathbf{C}$
$\mathbf{D}$	$\mathbf{D}_{t_{i,j}} = \mathbf{D}$
$\mathbf{R}$	$\mathbf{R}_{t_{i,j}} = \mathbf{R}$
$\boldsymbol{\mu}_{\mathbf{x}(t_0)} = E[\mathbf{x}(t_0)] = \mathbf{B}_{t_0}\mathbf{u}(t_0)$	
$\boldsymbol{\Phi}_{\mathbf{x}(t_0)} = E[(\mathbf{x}(t_0) - \mu_{\mathbf{x}(t_0)})(\mathbf{x}(t_0) - \mu_{\mathbf{x}(t_0)})']$	

The initial parameter matrices $\boldsymbol{\mu}_{\mathbf{x}(t_0)}$ and $\boldsymbol{\Phi}_{\mathbf{x}(t_0)}$ deserve special attention. In a model with exogenous variables the initial state mean may take different values in different groups defined by the exogenous variables. Since the mean trajectories $E[\mathbf{x}(t)]$ may be deviating from each other because of exogenous influences after t_0, it is natural to let them already differ at t_0 as a result of past influences. These differences are defined regression-wise by $E[\mathbf{x}(t_0)] = \mathbf{B}_{t_0}\mathbf{u}(t_0)$ with $\mathbf{B}_{t_0}\mathbf{u}(t_0)$ absorbing all unknown past influences. For example, if $\mathbf{u}(t_0)$ consists of two variables, the unit variable (1 for all subjects) and a dummy variable defining gender (0 for boys and 1 for girls), there will be two means $E[\mathbf{x}(t_0)]$, one for the boys and one for the girls. If $\mathbf{u}(t_0)$ contains only the unit-variable, the single remaining vector $\mathbf{b}_{t_0}$ in $\mathbf{B}_{t_0}$ will become equal to the initial mean: $E[\mathbf{x}(t_0)] = \mathbf{b}_{t_0}$. Because the instantaneous regression matrix $\mathbf{B}_{t_0}$ just describes means and differences as a result of unknown effects from before t_0, it should not be confused with the dynamic $\mathbf{B}$ and as a so-called predetermined quantity in estimation not undergo any constraint from $\mathbf{B}$. Similarly, $\boldsymbol{\Phi}_{\mathbf{x}(t_0)}$ should not undergo any constraint from the continuous-time diffusion covariance matrix $\mathbf{Q}$.

A totally new situation for the initial parameters arises, however, if we assume the system to be in equilibrium. Equilibrium means first $\boldsymbol{\mu}_{\mathbf{x}(t)} = \boldsymbol{\mu}_{\mathbf{x}(t_0)}$ as well as all exogenous variables $\mathbf{u}(t) = \mathbf{u}(t_0)$ being constant. Evidently, the latter is the case, if the only exogenous variable is the unit variable, reducing $\mathbf{B}$ to a vector of intercepts, but also if it contains additionally gender or any other additional exogenous variables, differentiating subjects from each other but constant in time. The assumption of equilibrium, $\boldsymbol{\mu}_{\mathbf{x}(t)} = \boldsymbol{\mu}_{\mathbf{x}(t_0)}$, –possibly but not necessarily a stable equilibrium– leads to equilibrium value

$$\boldsymbol{\mu}_{\mathbf{x}(t)} = \boldsymbol{\mu}_{\mathbf{x}(t_0)} = -\mathbf{A}^{-1}\mathbf{Bu}_c, \tag{26}$$

with $\mathbf{u}_c$ the value of the constant exogenous variables $\mathbf{u}(t) = \mathbf{u}(t_0) = \mathbf{u}_c$. If we assume the system to be stationary, additionally $\boldsymbol{\Phi}_{\mathbf{x}(t)} = \boldsymbol{\Phi}_{\mathbf{x}(t_0)}$ is assumed to be in equilibrium, leading to equilibrium value

$$\boldsymbol{\Phi}_{\mathbf{x}(t)} = \boldsymbol{\Phi}_{\mathbf{x}(t_0)} = \text{irow}[-\mathbf{A}_{\#}^{-1}\text{row}\mathbf{Q}]\,. \tag{27}$$

The novelty of the stationarity assumption is that the initial parameters are totally defined in terms of the dynamic parameters as is clearly seen from (26) and (27). It means, in fact, that the initial parameters disappear and the total number of parameters to be estimated is considerably reduced. Although attractive and present as an option in ctsem (Driver et al., 2017), the stationarity assumption is quite restrictive and can be unrealistic in practice.

To estimate the EDM as specified in Table 1 by SEM we put all variables and matrices of the EDM into SEM model

$$\eta = \boldsymbol{B}\eta + \zeta \text{ with } \boldsymbol{\Psi} = E(\zeta\zeta'), \tag{28}$$

$$\mathbf{y} = \boldsymbol{\Lambda}\eta + \varepsilon \text{ with } \boldsymbol{\Theta} = E(\varepsilon\varepsilon'). \tag{29}$$

The SEM model consists of two equations, structural equation (28) and measurement equation (29), in terms of four vectors, η, $\mathbf{y}$, ζ, ε, and four matrices, $\boldsymbol{B}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Theta}$. From Equations (28)-(29) one easily derives the model implied mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ and next the raw maximum likelihood equation (30) (see e.g., Bollen, 1989)

$$\begin{aligned}\text{RML} = \sum_{j=1}^{N} [m_j \log(2\pi) \\ + \log(|\boldsymbol{\Sigma}_j|) + (\boldsymbol{y}_j - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}_j^{-1} (\boldsymbol{y}_j - \boldsymbol{\mu}_j)].\end{aligned} \tag{30}$$

The subscript j makes the SEM procedure extremely flexible by allowing any number of subjects, including $N = 1$, and any missing value pattern for each of the subjects j, as the number of variables m_j ($m = pT$), the data vector $\boldsymbol{y}_j$, the mean vector $\boldsymbol{\mu}_j$ and the covariance matrix $\boldsymbol{\Sigma}_j$ may all be subject specific. In case of missing values, the corresponding rows and columns of the missing elements for that subject j are simply deleted.

For obtaining the maximum likelihood estimates of the EDM, it suffices to show how the SEM vectors η, ζ, $\mathbf{y}$, ε and matrices $\boldsymbol{B}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Theta}$ include the variables and matrices of the EDM (see Appendix A). In the vector of exogenous variables $\mathbf{u}(t) = [\mathbf{u}_c{}' \;\; \mathbf{u}_v{}'(t)]'$ we distinguish two parts: the part $\mathbf{u}_c$, consisting of the unit variable and, for example, gender and other variables that differ between subjects but are constant across time, and the part $\mathbf{u}_v(t)$, that at least for one subject in the sample is varying across time. Exogenous variables like weight and income, for example, are to be put into $\mathbf{u}_v(t)$.

For the crucial property of measurement invariance we need to specify

$$\begin{aligned}\mathbf{C}_{t_{0,j}} &= \mathbf{C}_{t_{1,j}} = \cdots = \mathbf{C}_{t_{T-1,j}} = \mathbf{C},\\ \mathbf{D}_{c,t_{0,j}} &= \mathbf{D}_{c,t_{1,j}} = \cdots = \mathbf{D}_{c,t_{T-1,j}} = \mathbf{D}_c,\\ \mathbf{D}_{v,t_{0,j}} &= \mathbf{D}_{v,t_{1,j}} = \cdots = \mathbf{D}_{v,t_{T-1,j}} = \mathbf{D}_v,\\ \mathbf{R}_{t_{0,j}} &= \mathbf{R}_{t_{1,j}} = \cdots = \mathbf{R}_{t_{T-1,j}} = \mathbf{R}.\end{aligned} \tag{31}$$

Although measurement invariance is important for substantive reasons, statistically speaking, the assumption of strict measurement invariance may be relaxed if necessary. An additional advantage of the SEM approach is the possibility of specifying measurement error covariances across time in $\boldsymbol{\Theta}_{\text{cov}}$. This can be done in ctsem by the MANIFESTTRAIT option. Measurement instruments often measure specific aspects, which they do not have in common with other instruments and can be taken care of by freeing corresponding elements in $\boldsymbol{\Theta}_{\text{cov}}$. In view of identification, however, one should be cautious in choosing elements of $\boldsymbol{\Theta}_{\text{cov}}$ to be freed.

Simulations

The aim of this section is to evaluate the quality of continuous-time model estimation for increasing complexity (first-order autoregressive, second-order autoregressive, second-order autoregressive with moving-average parameters) and decreasing sample size ($N = 1000$, $N = 100$, $N = 50$ and $N = 1$). The $N = 1$ analysis is repeated for each of the subjects in the $N = 100$ sample. In addition a test is described to test the difference between subject and group results, which is applied on each subject and group of remaining 99 subjects in the $N = 100$ sample.

CARMA(1,0) model

The first model simulated is a simple bivariate CARMA(1,0) model, whose true parameter values are displayed in Table 2. The model consists of a 2 × 2 drift matrix with auto-effects a_{11}, a_{22} and reciprocal cross-effects a_{12}, a_{21}, intercepts b_1, b_2, uncorrelated error variances $q_{11} = g_{11}^2$, $q_{22} = g_{22}^2$ and initial parameters $\mu_{y_1(t_0)}$, $\mu_{y_2(t_0)}$ and $\sigma^2_{y_1(t_0)}$, $\sigma^2_{y_2(t_0)}$, $\sigma_{y_1(t_0),y_2(t_0)}$. In total, the model has 13 parameters, for which data were randomly generated for $N = 1{,}000$ subjects over $T = 41$ time points, separated by 40 observation intervals of equal length $\Delta t = 1$. The analyses were carried out using ctsem. To check the validity of the data generation procedure and to serve as a reference point for the subsequent analyses, we started by analyzing the total sample of $N = 1{,}000$. As expected under these favorable circumstances, the parameter estimates hardly differ from the true values used in the data generation (see Table 2). The saturated model contains 3485 parameters [41 × 2 = 82 means and (82 × 83)/2 = 3403 (co)variances], the degrees of freedom of the restrictive model with as few as 13 parameters is $df = 3472$. Despite the highly restrictive character of the model and the huge sample, $\chi^2 = 3603.7$ with $df = 3472$ does not lead to significance ($p = 0.06$). We may conclude that the data generation and estimation procedure has been successful. Please note that while under these favorable circumstances the computation of a saturated model is possible and useful for reasons of comparison, for other models (e.g., time series with

Table 2. True parameter values and estimation results of the simulation study of a CARMA(1,0) model.

	True Parameter Value	$N = 1{,}000$	$N = 100$ (se)	$N = 50$ (se)	$N = 1$ (x 99)[a]		
					$\overline{\text{est}}$	$\overline{\text{se}}$	$\sqrt{\text{var(est)}}$
a_{11}	− 1.0	− 1.00	− 1.00(0.03)	− 1.06 (0.05)	− 1.08	0.46	0.37
a_{22}	− 1.5	− 1.50	− 1.39 (0.05)	− 1.44 (0 .07)	− 1.67	1.17	1.29
a_{12}	0.2	0.19	0.24 (0.03)	0.27 (0.06)	0.28	0.64	0.57
a_{21}	0.3	0.31	0.24 (0.04)	0.29(0.05)	0.38	0.69	0.73
b_1	10	10.1	9.6(0.28)	10.0(0.41)	10.2	3.8	3.5
b_2	12	12.0	11.5 (0.31)	11.5 (0.45)	12.8	6.1	5.9
q_{11}	4	4.0	3.9 (0.12)	4.0(0.17)	3.8	1.3	1.2
q_{22}	4	4.0	3.9(0.14)	3.8 (0.20)	4.2	2.5	2.4
$\mu_{y_1(t_0)}$	0	− 0.05	0.11 (0.10)	0.09 (0.12)	0.11	1.00	0.98
$\mu_{y_2(t_0)}$	0	0.02	− 0.08(0.10)	− 0.07 (0.12)	− 0.08	1.00	0.98
$\sigma^2_{y_1(t_0)}$	1	1.14	0.95(0.13)	0.68 (0.14)			
$\sigma^2_{y_2(t_0)}$	1	1.01	0.95 (0.13)	0.76 (0.15)			
$\sigma_{y_1(t_0),y_2(t_0)}$	0	0.00	− 0.13(0.10)	− 0.03 (0.10)			
χ^2		3603.7					
df		3472					
p		0.058					

[a]One improper $N = 1$ solution because of out-of-bound standard errors was deleted.

$N = 1$) the computation of a saturated model may be impossible or even undesirable.

Because a sample size of $N = 1{,}000$ is quite large for many practical applications, we next concentrated on samples of $N = 100$ and $N = 50$. As apparent from Table 2, in these conditions the parameter estimates are not deviating much more from the true parameter values than for the $N = 1{,}000$ sample. For one parameter ($\sigma^2_{y_1(t_0)}$) the $N = 100$ estimate is even closer than the $N = 1{,}000$ one. Also, in several cases the $N = 50$ sample is closer to the true values than the $N = 100$ sample. Whereas in the $N = 1{,}000$ case the restrictive continuous time model (H_0) could be tested by means of the likelihood ratio test against the saturated model (H_1), this was not possible in the $N = 100$ case, because the determinant of the huge 82×82 data covariance matrix turned out to be virtually zero. By definition, computing the likelihood ratio test against the saturated model is impossible for samples $N \leq 82$ and the more so for $N = 1$.

Next we took the same 100 subjects of the $N = 100$ sample but performed on each of them an $N = 1$ analysis. A classic problem in $N = 1$ analysis is the estimate of initial parameter values. Because for a single subject there is only one datum available for each variable at the initial time point, one could estimate the mean $\mu_{y_i(t_0)}$ but not simultaneously also the variance $\sigma^2_{y_i(t_0)}$, let alone additionally the covariance $\sigma_{y_i(t_0),y_{i'}(t_0)}$ with any other variable $y_{i'}(t_0)$. Usually, however, the initial covariance matrix is not of great concern in state space modeling and rather arbitrarily set, for example, to $\mathbf{\Phi}_{\mathbf{y}(t_0)} = \mathbf{I}$. This is justified by the fact that for a stable model and many observation time points T, the effect of the initial value chosen will be "forgotten" after a while and have minimal effect on the other estimates (Jazwinski, 1970, p. 243). We followed standard practice and fixed $\mathbf{\Phi}_{\mathbf{y}(t_0)}$ at $\mathbf{I}$ in the $N = 1$ analyses of the simulation, while estimating only the mean $\boldsymbol{\mu}_{\mathbf{y}(t_0)}$, which for $N = 1$ is equal to the observed data.

As apparent from Table 2, the average parameter estimates $\overline{\text{est}}$ over the 100 $N = 1$ analyses are only a bit more deviating from the true values than the estimates in the $N = 100$ analysis. In one case (b_1) $\overline{\text{est}}$ was closer to the true value. Observe that, as expected, the mean $\boldsymbol{\mu}_{\mathbf{y}(t_0)}$ in the $N = 100$ analysis is equal to the average over the 100 $N = 1$ analyses. Not only the parameter estimates, but also the average standard error estimates $\overline{\text{se}}$ show little or no bias, when compared to the actual standard deviations of the parameter estimates $\sqrt{\text{var(est)}}$ in Table 2.

Subject-group-reproducibility test

Whereas it is not possible in $N = 1$ research to perform a likelihood ratio test for the comparison of the estimated model with the saturated model, it is possible to compare and test the results on the individual level with the results on the group level, in particular, to test whether the models on the individual and group level coincide. The total sample of $N = 100$ is divided into two subsamples, one sample of $N = 99$ and one of $N = 1$, and a combined multiple group analysis is carried out. By leaving out each time one different subject from the total group the test has some similarity with the jackknife. In one analysis, the restricted model is estimated, in which all parameters are constrained to be equal in both samples. In a second analysis, parameters are estimated without any restrictions across the two submodels, except on the initial variances/covariances (H_1). Because the combined

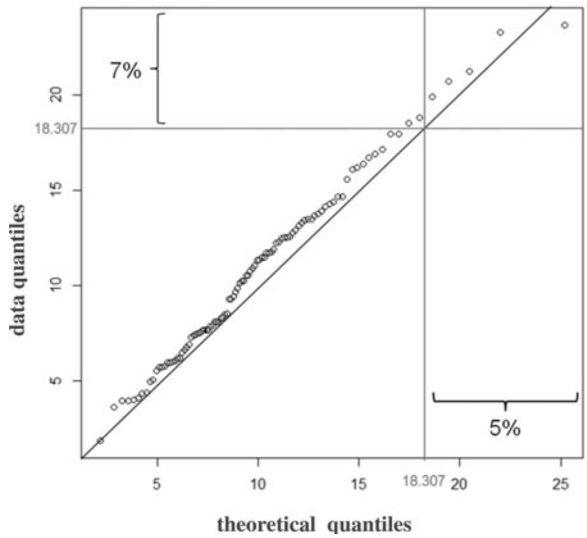

Figure 1. Q-Q plot for the 100 χ^2-values resulting from the subject-group-reproducibility tests in which 100 subjects, generated from the same model, were individually compared to each sample of remaining subjects.

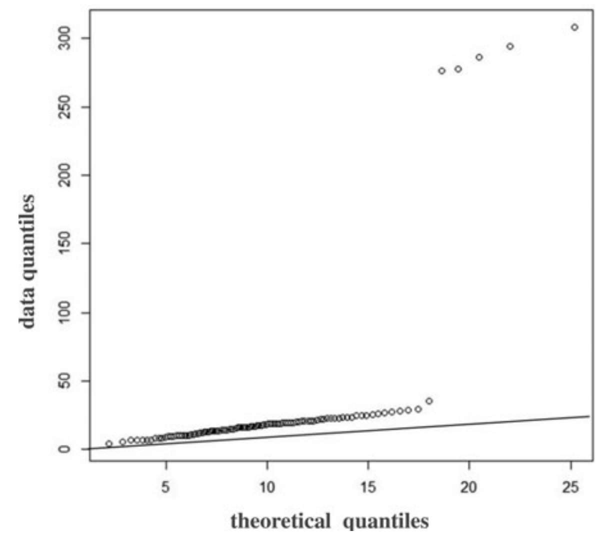

Figure 2. Q-Q plot for the 100 χ^2-values resulting from the subject-group-reproducibility tests on 100 subjects, of whom 95 were generated by one model and 5 others by a different model.

analysis has also the group sample available, we have the opportunity to use equality constraints to take over the variances/covariances estimates of the group sample to the $N = 1$ sample. In both analyses the total $-2ll$ (-2 times the log likelihood) is computed as the sum of the two $-2ll's$ in the two samples. The difference between $-2ll$ under H_0 and $-2ll$ under H_1, that is $(-2ll\_H_0) - (-2ll\_H_1)$, is χ^2-distributed with degrees of freedom df equal to the number of additional equality constraints under H_0 (Voelkle et al., 2012). In case of the simulation model in Table 2 there are 10 additional equality constraints under H_0 and so $df = 10$. Rejecting H_0 means that the model of the tested individual cannot be equated with the model of the remaining group of 99 individuals.

For the 100 individuals 100 subject-group-reproducibility tests were performed. The 100 resulting χ^2-values are displayed in the Q-Q plot of Figure 1. Because all 100 individual models were generated from the same model, H_0 is true for all individuals and the empirically found χ^2-distribution should coincide with the theoretical χ^2-distribution, that is, the values should be located on the diagonal in Figure 1. As apparent from Figure 1, the empirical χ^2-values follow rather precisely the diagonal and are only a little higher than theoretically expected. In agreement with that, the 5% rejection region becomes empirically 7% because of two subjects with a χ^2-value just above the critical value of 18.307.

The question may arise what happens and what it means, if the subject-group-reproducibility test is not passed. We distinguish two ideal-typical situations. In the first situation, every subject differs differently from every other subject in the $N = 100$ sample. In this situation, the χ^2-values will be far above the diagonal in the Q-Q plot and the implication is that a general model for all studied subjects does not hold. In the second situation the vast majority of the group shares a model, while only a small group of individuals share a different model. We simulated the second situation by replacing 5 subjects in the $N = 100$ sample by 5 subjects that were generated by a model, in which 4 of the 13 parameters had true values that differed substantially from the ones in Table 2: $b_1 = 10$, $b_2 = 12$, $q_{11} = 4$, $q_{22} = 4$ were changed for those subjects into $b_1 = 5$, $b_2 = 6$, $q_{11} = 1$, $q_{22} = 1$. The resulting Q-Q plot is shown in Figure 2. The figure first shows that the χ^2-values of the 5 deviating subjects are clearly differentiated from those of the majority group. The χ^2-values of the majority group are not far above the diagonal but more than in Figure 1. This is explained by the fact that in the tests of the non-deviating subjects the deviating subjects are part of the sample of 99 remaining subjects.

CARMA(2,0) and CARMA(2,1) models

The second and third simulated models are a CARMA(2,0) and CARMA(2,1) model as shown in Table 3 and 4. The CARMA(2,0) model extends the CARMA(1,0) model in Table 2 with the following second-order effect matrix $\mathbf{F}_1$ (see Equation (2))

$$\mathbf{F}_1 = \begin{bmatrix} f_{1.11} & 0 \\ 0 & f_{1.22} \end{bmatrix} = \begin{bmatrix} -2.4 & 0 \\ 0 & -2.6 \end{bmatrix}. \tag{32}$$

In addition the following second-order MA effect matrix $\mathbf{G}_1$ was added to the CARMA(2,1) model (see

Table 3. True parameter values and estimation results of the simulation study of a CARMA(2,0) model.

					$N=1$ (x 92)[a]		
	True Parameter Value	$N=1,000$	$N=100$ (se)	$N=50$ (se)	$\overline{\text{est}}$	$\overline{\text{se}}$	$\sqrt{\text{var(est)}}$
$f_{0,11}$	−1.0	−1.00	−1.02 (0.04)	−1.03 (0.05)	−1.37	0.71	0.65
$f_{0,22}$	−1.5	−1.50	−1.42 (0.05)	−1.35 (0.07)	−1.71	0.74	0.65
$f_{0,12}$	0.2	0.20	0.21 (0.04)	0.24 (0.05)	0.44	0.57	0.63
$f_{0,21}$	0.3	0.29	0.22 (0.03)	0.16 (0.04)	0.33	0.38	0.44
$f_{1,11}$	−2.4	−2.41	−2.38(0.09)	−2.36(0.12)	−2.90	1.62	1.61
$f_{1,22}$	−2.6	−2.60	−2.62 (0.11)	−2.62(0.15)	−2.93	1.42	1.62
b_1	10	10.0	10.1 (0.37)	9.8 (0.52)	12.0	6.1	4.8
b_2	12	12.1	12.0 (0.44)	12.0 (0.69)	13.8	6.9	6.7
q_{11}	4	4.0	4.1 (0.24)	4.1 (0.33)	5.8	6.7	5.3
q_{22}	4	4.1	4.0 (0.25)	4.0 (0.37)	5.1	5.4	5.9
$\mu_{y_1(t_0)}$	0	−0.07	0.19 (0.10)	0.13 (0.13)	0.15	1.00	1.04
$\mu_{y_2(t_0)}$	0	−0.02	0.04(0.11)	0.09 (0.16)	0.04	1.00	1.13
$\sigma^2_{y_1(t_0)}$	1	1.02	1.04 (0.15)	0.82 (0.16)			
$\sigma^2_{y_2(t_0)}$	1	0.99	1.23(0.17)	1.25 (0.25)			
$\sigma_{y_1(t_0),y_2(t_0)}$	0	−0.09	0.10 (0.11)	0.13 (0.14)			
$\sigma_{x_1(t_0)}$	1	1.17	0.93(0.48)	1.09(0.70)			
$\sigma_{x_2(t_0)}$	1	1.20	1.28(0.56)	2.10(0.95)			
χ^2		3582.3					
df		3468					
p		0.086					

[a]Eight improper $N=1$ solutions because of out-of-bound standard errors were deleted.

again Equation (2))

$$\mathbf{G}_1=\begin{bmatrix} g_{1,11} & 0 \\ 0 & g_{1,22}\end{bmatrix}=\begin{bmatrix} 0.5 & 0 \\ 0 & 0.5\end{bmatrix}\Rightarrow$$

$$\mathbf{Q}_1=\mathbf{G}_1\mathbf{G}'_1=\begin{bmatrix} q_{1,11} & 0 \\ 0 & q_{1,22}\end{bmatrix}=\begin{bmatrix} g^2_{1,11} & 0 \\ 0 & g^2_{1,22}\end{bmatrix}=\begin{bmatrix} 0.25 & 0 \\ 0 & 0.25\end{bmatrix}. \quad (33)$$

As apparent from Table 3, the quality of the CARMA(2,0) estimates in the $N>1$ samples is hardly worse than found for CARMA(1,0) in Table 2. An exception are the initial variances $\sigma_{x_1(t_0)}$ and $\sigma_{x_2(t_0)}$ of the latent variables in the model. However, these initial variables actually are not part of the model itself (see Equation (2)) but function as a kind of byproduct in its state space formulation. Oud and Singer (2008) found for these latent variables also bias and large standard errors but little influence on the quality of the other parameter estimates. We did not fix these variances at arbitrary values, because this is even less satisfactory than estimating, which had in fact more positive than negative effects on the other parameter estimates. As mentioned above, a possible solution would be to replace the initial parameters by their stationary values, which, however, may not be realistic in practice and was not simulated in the present case. Eight of the 100 $N=1$ analyses were useless because of clearly out-of-bound standard errors (>100). Some small sample bias is undoubtedly present in the parameter values of the remaining $N=1$ analyses. Also in these analyses with only $T=41$ data points per subject the standard errors are pretty large and would in practice prevent drawing clear conclusions. It is nevertheless comforting that the estimated standard errors are in most cases close to the true standard deviations of the parameter estimates.

Although the number of parameters to be estimated increases from 13 in CARMA(1,0) to 17 in CARMA(2,0) and 19 in CARMA(2,1), the $N=1,000$ samples hardly show deterioration in the quality of the parameter estimates. With the exception of the initial latent variances, the estimates of the CARMA(2,1) parameters are quite accurate with small standard errors (see Table 4). Also the results of the $N=100$ and even $N=50$ sample are not much worse for the CARMA(2,1) model than for the simpler models, although the estimates are somewhat less accurate and the standard errors clearly larger than for the simpler models. The results of the $N=1$ analyses (not reported in Table 4) were not acceptable, however. Forty of the 100 analyses gave a nonconvergent solution (38 had OpenMx status code 6 and 2 had code 5). We repeated the analysis for sample sizes ranging between $N=50$ and $N=1$, and found that especially the estimates of the variances $\sigma_{x_1(t_0)}$ and $\sigma_{x_2(t_0)}$ of the moving average components deteriorate rapidly with decreasing N, becoming even negative for $N=24$. We conclude that with $T=41$ time points (40 observation intervals) estimation of the CARMA(2,1) model can be done by samples of at least $N=25$ or larger but that the use of smaller samples may then be questionable. In $N=1$ ARMA and CARMA modeling, one sometimes finds shorter time series but mostly univariate cases without effects estimated between variables.

Table 4. True parameter values and estimation results of the simulation study of a CARMA(2,1) model.

	True Parameter Value	$N = 1{,}000$ (se)	$N = 100$ (se)	$N = 50$ (se)
$f_{0,11}$	− 1.0	− 0.99(0.02)	− 0.92(0.07)	− 0.87(0.09)
$f_{0,22}$	− 1.5	− 1.48(0.03)	− 1.57(0.10)	− 1.68 (0.16)
$f_{0,12}$	0.2	0.19(0.01)	0.16(0.05)	0.16(0.06)
$f_{0,21}$	0.3	0.29(0.01)	0.32(0.04)	0.39(0.06)
$f_{1,11}$	− 2.4	− 2.38(0.05)	− 2.29(0.15)	− 2.20(0.20)
$f_{1,22}$	− 2.6	− 2.60 (0.05)	− 2.70(0.16)	− 2.69(0.20)
b_1	10	10.0(0.19)	9.3 (0.57)	8.8 (0.77)
b_2	12	12.0(0.21)	12.5 (0.73)	12.8 (1.14)
$q_{0,11}$	4	4.0(0.18)	3.5 (0.53)	3.2 (0.68)
$q_{0,22}$	4	4.0(0.16)	4.0 (0.56)	4.6 (0.95)
$q_{1,11}$	0.25	0.25(0.01)	0.29(0.04)	0.34(0.05)
$q_{1,22}$	0.25	0.26(0.01)	0.28(0.04)	0.25(0.06)
$\mu_{y_1(t_0)}$	0	0.00(0.03)	− 0.11(0.10)	− 0.07(0.16)
$\mu_{y_2(t_0)}$	0	− 0.05(0.03)	0.19(0.11)	0.27(0.15)
$\sigma^2_{y_1(t_0)}$	1	1.07(0.05)	1.08 (0.15)	1.35 (0.27)
$\sigma^2_{y_2(t_0)}$	1	0.93 (0.04)	1.13(0.16)	1.12 (0.22)
$\sigma_{y_1(t_0),y_2(t_0)}$	0	− 0.09(0.03)	− 0.03 (0.11)	0.02 (0.17)
$\sigma_{x_1(t_0)}$	1	0.89(0.22)	0.87(0.65)	2.02(0.96)
$\sigma_{x_2(t_0)}$	1	0.87(0.24)	1.20(0.85)	3.03(1.53)
χ^2		3550.8		
df		3466		
p		0.154		

Empirical example: Mood at home and mood at work, CARMA(1,0) on $N = 55$, $T = 48$

The data were collected as part of a more comprehensive project on work and family (Klumb, Hoppmann, & Staats, 2006; Klumb, Voelkle, & Siegler, 2017). In this study, dual-career couples, with at least one child under five years, were investigated during a seven-day ambulatory assessment period. During this period, participants completed short questionnaires on momentary affect, situational variables including their current location (i.e., work vs. home), and social interactions using a handheld computer (Psion 3a, Psion PLC, London, Great Britain). During 7 workdays, people had to complete a questionnaire after waking up and before going to bed. In between, they received 5 signals with an interval length of about 3 hours between them, resulting in a total of $T = 7 \times 7 = 49$ measurement occasions. For the purpose of the present work we will focus on 55 women and their reported mood across the measurement occasions. Mood was assessed on a 9-point-scale from 1 (very bad) to 5 (very good): 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5. The aim of this exemplary analysis is to find out, whether mood at work (MW) influences mood at home (MH), MH influences MW, both effects take place, or none of both. This will be evaluated by the cross-effects in the drift matrix, by the implied cross-lagged effects across time, and by trajectories described by subjects through the two-dimensional state space in the phase portrait (Butner, Gagnon, Geuss, Lessard, & Story, 2015). The models CARMA(1,0) and CARMA(2,0), used in the simulation in the previous section, were applied to the data.

Depending on the place where the subject was at the moment of the measurement, each of a maximum of 48 measurements per subject was assigned to the variable MW or the variable MH. Because all values were missing in one of the variables at the start, we skipped the first of the 49 measurement occasions. The pattern of measurement intervals and home-work shifts was extremely irregular between subjects. In fact, each measurement came from a unique point in time. A correct discrete time analysis would therefore not be possible as it would require at least as many variables as there are values in the data set and many more extra phantom variables to take care of the irregular time intervals between measurement time points (Oud & Voelkle, 2014). In contrast, the continuous time analysis performed here requires "only" $2 \times 48 = 96$ variables. Even if analyzed in continuous time, the data set contains a huge amount of missing values. This is the case, because at the person level it is impossible to have simultaneous observations at work and at home. So, if MW is present, MH is missing, and vice versa. These missing values were taken care of by OpenMx's FIML function.

As in the simulations we started with analyses on the total sample of $N = 55$ women. Based on the experience in the simulations we tried the CARMA(2,0) and CARMA(1,0) models. As in ARMA modeling, extending the order of the model and using a model selection criterion such as AIC and BIC to select the best fitting model, is customary in CARMA modeling. The BIC measure, $-2ll - \ln(\mathit{number\ of\ observations}) \times df$, indicates that the simpler model CARMA(1,0) has to be preferred to CARMA(2,0), CARMA(1,0)'s BIC-value of

Table 5. Parameter estimates and standard errors in the total sample of $N = 55$ for CARMA(1,0) model of the empirical example.

Parameter	Estimate	Standard error
a_{11}	− 0.465	0.060
a_{22}	− 0.165	0.023
a_{12}	0.324	0.051
a_{21}	0.064	0.022
b_1	0.499	0.095
b_2	0.415	0.048
q_{11}	0.458	0.044
q_{22}	0.160	0.011
$\mu_{y_1(t_0)}$	3.923	0.171
$\mu_{y_2(t_0)}$	4.118	0.123
$\sigma^2_{y_1(t_0)}$	0.343	0.137
$\sigma^2_{y_2(t_0)}$	0.708	0.155
$\sigma_{y_1(t_0)y_2(t_0)}$	0.306	0.177
$-2ll$	5004.4	
observations	2403	
df	2390	

5105.6 being 11.1 lower than the one of CARMA(2,0). The CARMA(2,0) solution has to be rejected on several other counts too: Two parameters had no standard errors, four standard errors were out-of-bound (>100) and most of the estimates were unacceptably high or low. In contrast, the CARMA(1,0) gave a solution that is acceptable in all respects with parameter estimates that are all highly significant.

As apparent from Table 5, in CARMA(1,0) MH (variable 1) has an auto-effect of −0.465 in continuous time which is substantively lower than the auto-effect of −0.165 of MW (variable 2), meaning that MH is a less persistent property than MW. Figure 3 displays what this means in terms of the autoregressive effects over different intervals across time. Both MW and MH are rather short-lived. While after 24 hours about 10% of MW is left, MH is almost totally gone after the same interval.

The main research question, however, concerns the cross-effects between MW and MH. It turns out that both the effect of MW on MH ($a_{12} = 0.324$) and MH on MW ($a_{21} = 0.064$) are significant but the effect in the first direction (from work to home) is much stronger. These effects can also be visualized across time by means of the cross-lagged effects (see Figure 4). As apparent from Figure 4 the maximum cross-lagged effect for MW → MH is a little higher than 0.4 and for MH → MW a little below 0.1. In both cases the maximum is already reached after just a few hours.

Figure 5 shows the mean development of MH and MW over time. Starting from an initial value of 3.923 at t_0,

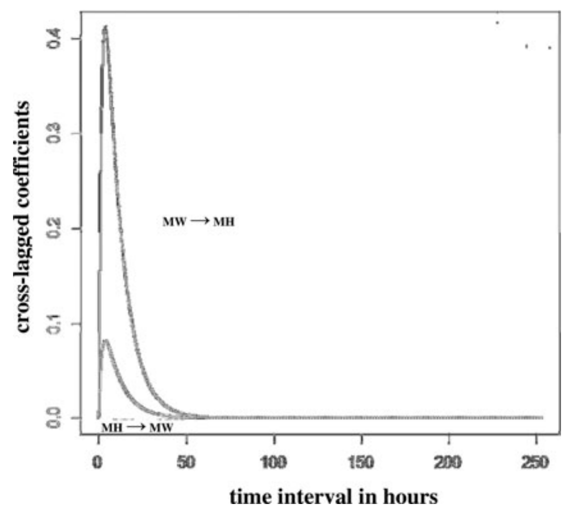

Figure 4. Cross-lagged effects of MW on MH (MW → MH) and MH on MW (MH → MW) over intervals in terms of hours, computed as the off-diagonal elements of matrix exponential $e^{\mathbf{A}\Delta t}$.

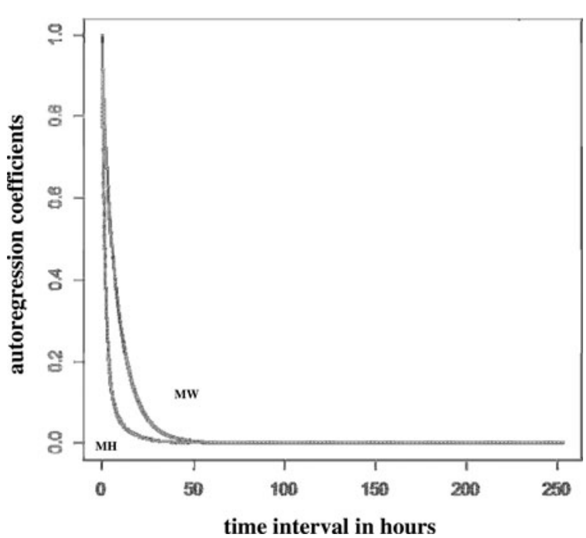

Figure 3. Autoregressive effects of mood at work (MW) and mood at home (MH) over intervals in terms of hours, computed as the diagonal elements of matrix exponential $e^{\mathbf{A}\Delta t}$.

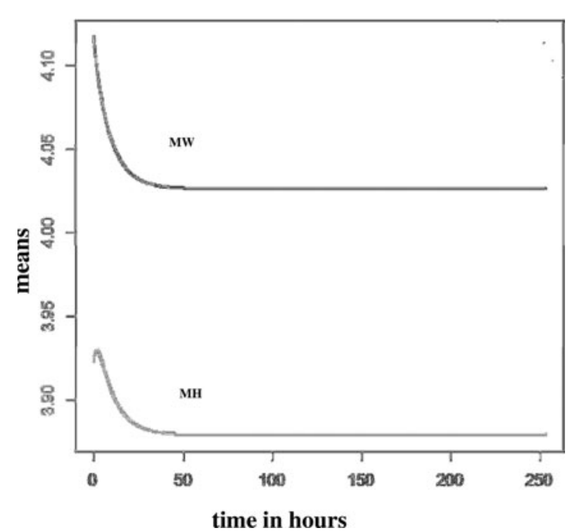

Figure 5. Mean development of MH and MW across time in terms of hours from measurement start, computed as $e^{\mathbf{A}(t-t_0)}E[\mathbf{x}(t_0)] + \mathbf{A}^{-1}[e^{\mathbf{A}(t-t_0)} - \mathbf{I}]\mathbf{b}$.

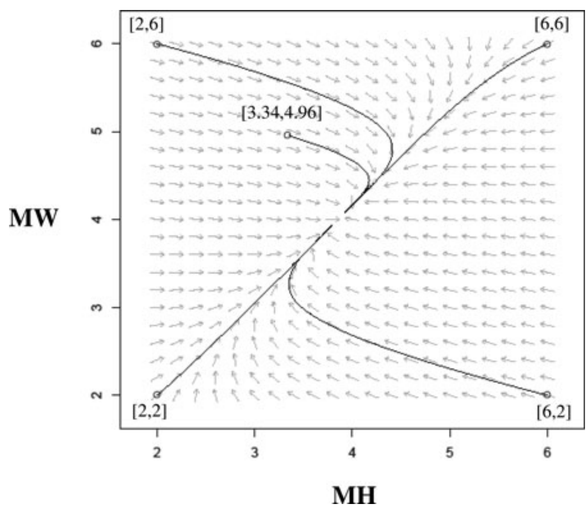

Figure 6. Phase portrait of the estimated model with arrows indicating change directions in the state space of MH and MW, the stable equilibrium state (attractor) located at [3.88, 4.03], and five trajectories starting from initial states [2,6], [6,6], [2,2], [6,2] and [3.34,4.03].

the mean of MH goes up a tiny fraction before converging rapidly to the stable equilibrium state. The mean of MW goes from 4.118 immediately in the direction of its stable equilibrium state. Note that the difference between the two initial means as well as the difference between initial means and stable equilibrium states are in fact small. Nevertheless, the phase portrait below will make visible big differences in the movement from initial state to final equilibrium state. Considerable and significant difference can be observed between the variances of MH and MW. The differences in mood between women are much bigger at work than at home (Table 5 gives at t_0 variance values of 0.708 and 0.343, respectively).

Important information about the dynamics of the state variables MH and MW in their interaction can be obtained from the phase portrait in Figure 6. Use of phase portraits in psychological research to generate and test theories of change is advocated, for example, by Butner et al. (2015). Phase portraits are easily produced by the R-package phaseR (Grayling, 2014). The portrait displays the two-dimensional state space with arrows indicating the direction of changes in the space, implied by the model. Because drift matrix $\mathbf{A}$ has both eigenvalues real and negative, the model has a stable equilibrium state, located at the point [MH = 3.88, MW = 4.03] defined by $-\mathbf{A}^{-1}\mathbf{b}$, and all arrows point more or less strongly in that direction. A stable equilibrium point in a phase portrait is called attractor and located at such a point of no change.

How this works out for the trajectories which subjects describe through the state space, is shown by means of five cases with different initial values [MH, MW]: Four exemplary subjects start from the rather extreme initial value pairs [2,6], [6,6], [2,2] and [6,2] and one from a more realistic initial value pair [mean − SD for MH, mean + SD for MW]. It turns out that subjects starting with low or high scores on both dimensions MH and MW go almost linearly to the stable equilibrium position. However, subjects low on MH but high on MW have a tendency to maintain this high mood on MW relatively long and to improve simultaneously their low level on MH considerably, temporarily even beyond the final equilibrium position of MH, before joining the [6,6] trajectory toward the equilibrium position. A similar trajectory pattern of change is followed by subjects starting at or close to [3.34, 4.96]. The trajectory from [6,2] mirrors the one from [2,6]. Before joining the [2,2] trajectory to the equilibrium point, the low level on MW in the [6,2] trajectory is maintained relatively long and the high level on MH is deteriorating relatively rapidly, temporarily even to a lower HM level than in the final equilibrium point. So, the portrait nicely shows how for these women a bad mood at work is relatively persistent and tends at least temporarily to ruin the mood at home. It should be noted, that in a discrete-time approach with only a limited number of snap shots given, it would be impossible to reconstruct the detailed picture of the continuous-time trajectory.

Finally, the analysis was repeated for each of the 55 women separately in combination with each remaining group of 54 women for comparison as explained before. The resulting χ^2-values of the comparison between each individual woman and the remaining group (subject-group-reproducibility test value) are displayed in Figure 7. Because of lacking standard errors in the combined solution, 11 individuals were omitted, so that Figure 7 contains only 44 values. As can be observed in Figure 7,

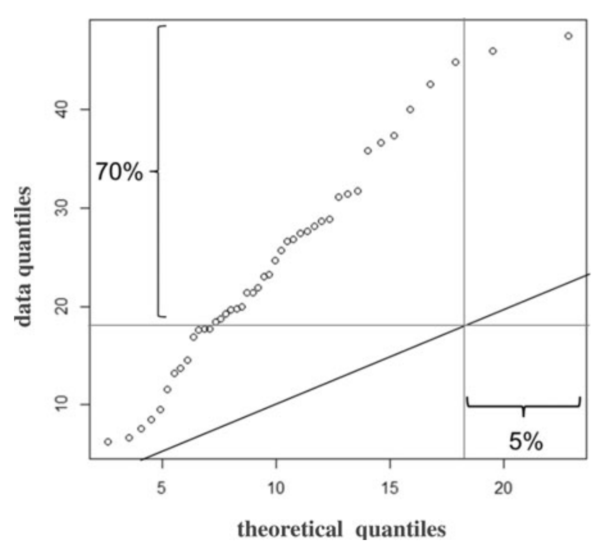

Figure 7. Q-Q plot for 44 χ^2-values resulting from the subject-group-reproducibility tests in which the 55 women in the real life example were individually compared to each sample of 54 remaining subjects; 11 χ^2-values for analyses with lacking standard errors were omitted.

the empirical rejection rate of 70% is much higher than the nominal level of 5% and so the hypothesis that all 55 individuals follow the same CARMA(1,0) has to be rejected. Unfortunately, the Q-Q plot does not reveal a clear subgroup of similarly deviating individuals as shown in the simulation study (see Figure 2), so that a straightforward solution by simply partitioning the sample, seems unlikely. Thus, caution should be exercised when attempting to generalize from the model estimated for the entire group to any specific woman. One could decide to refrain from interpreting the model on the group level and to concentrate exclusively on the individual subjects. However, similar patterns in individual subject analyses may also point to the value of such higher-level analysis and in the case of short series of individual subjects no other possibility might be left.

Conclusion

In their comprehensive introduction to modern time series analysis Prado and West (2010) observed that in many statistical models "the assumption that the observations are realizations of independent random variables is key. In contrast, time series analysis is concerned with describing the dependence among the elements of a sequence of random variables" (p. 1). Without doubt, SEM for a long period took position in the first group of models, which hampered the development of $N = 1$ modeling and time series analysis in an SEM context. The present article attempts to reconcile both perspectives by putting time series of independently drawn subjects in one and the same overall SEM model, while using continuous time state space modeling to simultaneously account for the dependence between observations in each time series over time. The present article explained in detail how this may be achieved for first and higher order CARMA(p,q) models in an extended SEM framework.

The analyses in this article are done by the R-package ctsem, which interfaces to the flexible SEM package OpenMx. Both the more recent SEM procedure and the Kalman filter procedure, common in state space modeling, are options in OpenMx and ctsem. If both procedures are applicable, the results are equal. However, the SEM procedure allows estimation of models with arbitrary measurement error structures across time, whereas the stepwise recursive Kalman filter procedure essentially assumes the measurement errors to be uncorrelated between time points, but is less time consuming for large T. In both procedures one easily performs $N =$ 1 and $N > 1$ analyses and both enable to combine the N $N = 1$ analyses with the N group analyses of remaining subjects in one overall analysis and to test in subject-group-reproducibility tests whether the subject models coincide with the model for the entire group.

Attempts to combine time series of different subjects in a common model are rare in traditional time series analysis and state space modeling. A rather isolated proposal was done by Goodrich and Caines (1979). They call a data set consisting of $N > 1$ time series "cross-sectional", thereby using this term in a somewhat different meaning from what is customary in behavioral science. They give a consistency proof for state space model parameter estimates in this kind of data in which "the number T of observations on the transient behavior is fixed but the number N of independent cross-sectional samples tends to infinity" (p. 403).

Special in the proposal of Goodrich and Caines (1979) is that they made the measurement intercepts random over subjects ($\boldsymbol{\kappa}$). The present publication allows not only the measurements intercepts but also the dynamic intercepts in the state equation ($\boldsymbol{\gamma}$) to be random (as 'traits" differentiated from the changing "states"). We could have gone one step further by randomizing also other parameters and so making the common overall model maximally subject specific. We refrained from doing that, because this would have led to complex interaction terms, which are difficult to handle in a frequentist approach (Boulton, 2014). As remarked by Boulton, the specification of such random effects much more naturally arises in the Bayesian paradigm. Some Bayesian work in this direction has been done by Oravecz, Tuerlinckx, and Vandekerckhove (2009, 2011, 2016) and is implemented in the latest version of ctsem.

As a matter of fact, an important advantage of $N >$ 1 models is to not be forced to $T \to \infty$ asymptotics, which at least in the social and behavioral sciences is often unrealistic. Arguably, there are not many processes with, for example, exactly the same parameter values over the whole time range until infinity. As argued by Yu (2014, p. 738), an extra advantage offered by continuous time modeling in this respect is, that asymptotics can be applied on the time dimension, even if T is taken as fixed. Supposing the discretely observed data to be recorded at $0, \Delta t, 2\Delta t, n\Delta t (= T)$, this so-called "in-fill" asymptotics takes T as fixed but lets $n \to \infty$ in continuous time. By letting N as well as n go to infinity a kind of double asymptotics results, which may be particularly useful for typical applications in the social sciences, where it is often hard to argue that T will approach infinity.

Although continuous time modeling is still rare in behavioral science, compelling reasons exist to switch from discrete time to continuous time and thus from ARMA to CARMA. One main point is the analysis of irregularly spaced data. An extreme case of such irregularly spaced data was found in the empirical example.

Forcing irregularly spaced data into the equal interval framework of a discrete time analysis leads to missing data and is possible only to a limited degree (Oud & Voelkle, 2014). In fact, because the time intervals are arbitrary in a continuous time analysis, missing data of this type totally disappear in continuous time. In discrete time research usually a lot of effort is put into avoiding missing data. However, by distributing different observation time points over the sample units instead of giving them all the same observation time points, usually a better representation of the underlying continuous time process is obtained but one deliberately enhances the missingness problem from a discrete time perspective (Voelkle & Oud, 2013).

Article Information

Conflict of Interest Disclosures: Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical Principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: This work was not supported by a Grant.

Role of the Funders/Sponsors: None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Acknowledgments: The authors would like to thank Petra Klumb for providing the data for the empirical example and four anonymous reviewers for their extensive and insightful comments on prior versions of this manuscript. The ideas and opinions expressed herein are those of the authors alone, and endorsement by the authors' institutions is not intended and should not be inferred.

References

Aoki, M. (1987). *State space modeling of time series*. Berlin: Springer.

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

Bergstrom, A. R. (1984). Continuous time stochastic models and issues of aggregation over time. In Z. Griliches & M. D. Intriligator (Eds.), *Handbook of econometrics: Vol.2* (pp. 1145–1212). Amsterdam: North-Holland. doi: 10.1016/s1573-4412(84)02012-2

Boker, S. M., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., Spies, J., Estabrook, R., Kenny, S., Bates, T., Mehta, P., & Fox, J. (2011). OpenMx: An open source extended structural equation modeling framework. *Psychometrika*, *76*(2), 306–317. doi: 10.1007/s11336-010-9200-6

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley. doi: 10.1002/9781118619179

Boulton, A. J. (2014). Bayesian estimation of a continuous-time model for discretely observed panel data (Doctoral dissertation, University of Kansas, 2014).

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. Oakland, CA: Holden-Day.

Brockwell, P. J. (2004). Representations of continuous-time ARMA processes. *Journal of Applied Probability*, *41a*, 375–382. doi:10.1017/s0021900200112422

Butner, J. E., Gagnon, K. T., Geuss, M. N., Lessard, D. A., & Story, D. N. (2015). Utilizing topology to generate and test theories of change. *Psychological Methods*, *20*, 1–25. doi: 10.1037/a0037802

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

Chan, K., & Tong, H. (1987). A note on embedding a discrete time ARMA model in a continuous parameter ARMA model. *Journal of Time Series Analysis*, *8*(3), 277–281. doi: 10.1111/j.1467-9892.1987.tb00439.x

Chow, S-M., Lu, Z., Zhu, H., & Sherwood, A. (2016). Fitting nonlinear ordinary differential equation models with random effects and unknown initial conditions using the Stochastic Approximation Expectation Maximization (SAEM) algorithm. *Psychometrika*, *81*, 102–143. doi: 10.1007/s11336-014-9431-z

Chow, S-M., Ho, M. R., Hamaker, E. J., & Dolan, C. V. (2010). Equivalences and differences between structural equation modeling and state-space modeling techniques. *Structural Equation Modeling*, *17*, 303–332. doi: 10.1080/10705511003661553

Deistler, M. (1985). General structure and parametrization of ARMA and state-space systems and its relation to statistical problems. In E. J. Hannan, P. R. Krishnaiah, & M. M. Rao (Eds.), *Handbook of statistics: Vol. 5. Time series in the time domain* (pp. 257–277). Amsterdam: North-Holland. doi: 10.1016/s0169-7161(85)05011-8

Driver, C. C., Oud, J. H. L., & Voelkle, M. C. (2017). Continuous time structural equation modelling with R package ctsem. *Journal of Statistical Software*, *7*, 1–35. doi:10.18637/jss.v077.i05

Durbin, J., & Koopman, S. J. (2001). *Time series analysis by state space methods*. Oxford: Oxford University Press. doi: 10.1093/acprof:oSo/9780199641178.001.0001

Gasimova, F., Robitzsch, A., Wilhelm, O., Boker, S., Hu, Y., & Hülür, G. (2014). Dynamical systems analysis applied to working memory data. *Frontiers of Psychology*, *5*, 687. doi:10.3389/fpsyg.2014.00687

Goodrich, R. L., & Caines, P. E. (1979). Linear system identification from nonstationary cross-sectional data. *IEEE Transactions on Automatic Control*, *24*(3), 403–411. doi: 10.1109/TAC.1979.1102037

Grayling, M. (2014). *phaseR, version 1.3*. Cambridge: MRC Biostatistics Unit.

Hamaker, E. L., Dolan, C. V., & Molenaar, C. M. (2003). ARMA-based SEM when the number of time points T exceeds the number of cases N: Raw data maximum likelihood. *Structural Equation Modeling*, *10*(3), 352–379. doi: 10.1207/s15328007sem1003_2

Hamaker, E. L., Dolan, C. V., & Molenaar, C. M. (2005). Statistical modeling of the individual: Rationale and application of multivariate stationary time series analysis. *Multivariate Behavioral Research*, *40*(2), 207–233. doi: 10.1207/s15327906mbr4002_3

Hamerle, A., Nagl, W., & Singer, H. (1991). Problems with the estimation of stochastic differential equations using structural equations models. *Journal of Mathematical Sociology*, *16*(3), 201–220. doi: 10.1080/0022250X.1991.9990088

Hamerle, A., Singer, H., & Nagl, W. (1993). Identification and estimation of continuous time dynamic systems with exogenous variables using panel data. *Economic Theory*, *9*, 283–295. doi: 10.1017/S0266466600007544

Harvey, A. C. (1981). *Time series models*. Oxford: Philip Allen.

Hansen, L., & Sargent, T. J. (1983). The dimensionality of the aliasing problem. *Econometrica*, *51*(2), 377–388. doi: 10.2307/1911996

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

Jones, R. H. (1993). *Longitudinal data with serial correlation: A state space approach*. London: Chapman & Hall. doi: 10.1007/978-1-4899-4489-4

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

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

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

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering (Trans. ASME, ser. D)*, *82*(1), 35–45. doi:10.1115/1.3662552

King, A. A., Nguyen, D., & Ionides, E. L. (2016). Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software*, *69*, 1–43. doi: 10.18637/jss.v069.i12

Klumb, P., Hoppmann, C., & Staats, M. (2006). Division of labor in German dual-earner families: Testing equity theoretical hypotheses. *Journal of Marriage and Family*, *68*(4), 870–882. doi:10.1115/1.3662552

Klumb, P. L., Voelkle, M. C., & Siegler, S. (2017). How negative social interactions at work seep into the home: A prosocial and an antisocial pathway. *Journal of Organizational Behavior*, *38*, 629–649. doi:10.1002/job.2154

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

Ljung, L. (1985). Estimation of parameters in dynamical systems. In E. J. Hannan, P. R. Krishnaiah, & M. M. Rao (Eds.), *Handbook of statistics: Vol. 5. Time series in the time domain* (pp. 189–211). Amsterdam: North-Holland. doi: 10.1016/s0169-7161(85)05009-x

MacCallum, R., & Ashby, F. G. (1986). Relationships between linear systems theory and covariance structure modeling. *Journal of Mathematical Psychology*, *30*, 1–27. doi: 10.1016/0022-2496(86)90039-8

Moler, C., & Van Loan, C. (2003). Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review*, *45*(1), 3–49. doi: 10.1137/S00361445024180

Molenaar, P. C. M., & Newell, K. M. (2003). Direct fit of a theoretical model of phase transition in oscillatory finger motions. *British Journal of Mathematical and Statistical Psychology*, *56*, 199–214. doi: 10.1348/000711003770480002

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., … Boker, S. M. (2016). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika*, *81*, 535–549. doi:10.1007/s11336-014-9435-8

Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2009). A hierarchical Ornstein-Uhlenbeck model for continuous repeated measurement data. *Psychometrika*, *74*, 395–418. doi: 10.1007/s11336-008-9106-8

Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2011). A hierarchical latent stochastic differential equation model for affective dynamics. *Psychological Methods*, *16*, 468–490. doi: 10.1037/a0024375

Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2016). Bayesian data analysis with the bivariate hierarchical Ornstein-Uhlenbeck process model. *Multivariate Behavioral Research*, *51*(1), 106–119. doi: 10.1080/00273171.2015.1110512

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

Oud, J. H. L., van den Bercken, J. H., & Essers, R. J. (1990). Longitudinal factor score estimation using the Kalman filter. *Applied Psychological Measurement*, *14*, 395–418. doi: 10.1177/014662169001400406

Oud, J. H. L., & Delsing, M. J. M.H. (2010). Continuous time modeling of panel data by means of SEM. In K. van Montfort, J. Oud & A. Satorra, (Eds.), *Longitudinal research with latent variables* (pp. 201–244). New York: Springer. doi: 10.1007/978-3-642-11760-2_7

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. doi: 10.1007/BF02294374

Oud, J. H. L., & Singer, H. (2008). Continuous time modeling of panel data: SEM versus filter techniques. *Statistica Neerlandica*, *62*, 4–28. doi: 10.1111/j.1467-9574.2007.00376.x

Oud, J. H. L., & Voelkle, M. C. (2014). Do missing values exist? Incomplete data handling in cross-national longitudinal studies by means of continuous time modeling. *Quality & Quantity*, *48*, 3271–3288. doi 10.1007/s11135-013-9955-9

Phadke, S. M., & Wu, M. S. (1974). Modeling of continuous stochastic processes from discrete observations with application to sunspots data. *Journal of the American Statistical Association*, *69*, 325–329. doi: 10.1111/j.1467-9574.2007.00376.x

Phillips, P. C. B. (1973). The problem of identification in finite parameter continuous time models. In A. R. Bergstrom (Ed.), *Statistical inference in continuous time models* (pp. 135–173). Amsterdam: North-Holland.

Prado, R., & West, M. (2010). *Time series: Modeling, computation, and inference*. Boca Raton: Chapman & Hall.

Shumway, R. H., & Stoffer, D. S. (2000). *Time series analysis and its applications*. New York: Springer. doi: 10.1007/978-1-4419-7865-3

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. (1992a). The aliasing-phenomenon in visual terms. *Journal of Mathematical Sociology*, *17*, 39–49. doi: 10.1080/0022250X.1992.9990097

Singer, H. (1992b). *Zeitkontinuierliche dynamische Systeme [Continuous time dynamic systems]*. Frankfurt:Campus.

Singer, H. (2010). SEM modeling with singular moment matrices part I: ML estimation of time series. *Journal of Mathematical Sociology*, *34*, 301–320. doi: 10.1080/0022250X.2010.509524

Singer, H. (2011). Continuous-discrete state-space modeling of panel data with nonlinear filter algorithms. *AStA Advanced Statistical Analysis*, *95*, 375–415. doi: 10.1007/s10182-011-0172-3

Singer, H. (2012). SEM modeling with singular moment matrices part II: ML estimation of sampled stochastic differential equations. *Journal of Mathematical Sociology*, *36*, 22–43. doi: 10.1080/0022250X.2010.532259 https://doi.org/10.1080/0022250X.2010.532259

Steele, J. S., & Ferrer, E. (2011a). Latent differential equation modeling of self-regulatory and coregulatory affective processes. *Multivariate Behavioral Research*, *46*, 956–984. doi:10.1080/00273171.2011.625305

Steele, J. S., & Ferrer, E. (2011b). Response to Oud & Folmer: Randomness and residuals. *Multivariate Behavioral Research*, *46*, 994–1003. doi: 10.1080/00273171.625308

Tómasson, H. (2011). *Some computational aspects of Gaussian CARMA modelling*. Vienna: Institute for Advanced Studies.

Tómasson, H. (2015). Some computational aspects of Gaussian CARMA modelling. *Statistical Computation*, *25*, 375–387. doi 10.1007/s11222-013-9438-9

Tsai, H., & Chan, K. (2000). A note on the covariance structure of a continuous-time ARMA process. *Statistica Sinica*, *10*, 989–998.

Voelkle, M. C., Oud, J. H.L., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological Methods*, *17*, 176–192. doi: 10.1037/a0027543

Voelkle, M. C., Oud, J. H.L., von Oertzen, T., & Lindenberger, U. (2012). Maximum likelihood dynamic factor modeling for arbitrary N and T using SEM. *Structural Equation Modeling*, *19*, 329–350. doi:10.1080/10705511.2012.687656

Voelkle, M. C., & Oud, J. H. L. (2013). Continuous time modelling with individually varying time intervals for oscillating and nonoscillating processes. *British Journal of Mathematical and Statistical Psychology*, *66*, 103–126. doi:10.1111/j.2044-8317.2012.02043.x

Yu, J. (2014). Econometric analysis of continuous time models: A survey of Peter Philips's work and some new results. *Econometric Theory*, *30*(4), 737–774. doi: 10.1017/S0266466613000467

Zadeh, L. A., & Desoer, C. A. (1963). *Linear system theory: The state space approach*. New York: McGraw-Hill. doi: 10.1137/1006079

Appendix

Including the variables and matrices of the EDM into the SEM vectors η, ζ, $\mathbf{y}$, ε and SEM matrices $\boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}, \boldsymbol{\Theta}$

$$
\begin{aligned}
\eta &= [\mathbf{x}'_j \mathbf{u}'_j]' \text{ with } \mathbf{x}_j = [\mathbf{x}'(t_{0,j})\mathbf{x}'(t_{1,j}) \cdots \mathbf{x}'(t_{T-1,j})]' \\
&\qquad \text{and } \mathbf{u}_j = [\mathbf{u}'_{c,j}\mathbf{u}'_{\upsilon}(t_{0,j})\mathbf{u}'_{\upsilon}(t_{1,j}) \cdots \mathbf{u}'_{\upsilon}(t_{T-1,j})]', \\
\zeta &= [\mathbf{w}'_j \mathbf{u}'_j]' \text{ with } \mathbf{w}_j = [\mathbf{x}'(t_{0,j}) - \boldsymbol{\mu}'_{\mathbf{x}(t_{0,j})}\ \mathbf{w}'(t_{1,j} - \Delta_{1,j}) \cdots \\
&\qquad \mathbf{w}'(t_{T-1,j} - \Delta_{T-1,j})]', \\
\mathbf{y} &= [\mathbf{y}'_j \mathbf{u}'_j]' \text{ with } \mathbf{y}_j = [\mathbf{y}'(t_{0,j})\mathbf{y}'(t_{1,j}) \cdots \mathbf{y}'(t_{T-1,j})]', \\
\varepsilon &= [\mathbf{v}'_j \mathbf{0}']' \text{ with } \mathbf{v}_j = [\mathbf{v}'(t_{0,j})\mathbf{v}'(t_{1,j}) \cdots \mathbf{v}'(t_{T-1,j})]',
\end{aligned}
$$

$$
\boldsymbol{B} = \begin{bmatrix}
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{B}_{c,t_0} & \mathbf{B}_{\upsilon,t_0} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{A}_{\Delta t_{1,j}} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{B}_{c,\Delta t_{1,j}} & \mathbf{B}_{\upsilon,\Delta t_{1,j}} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{A}_{\Delta t_{2,j}} & \cdots & \mathbf{0} & \mathbf{B}_{c,\Delta t_{2,j}} & \mathbf{0} & \mathbf{B}_{\upsilon,\Delta t_{2,j}} & \cdots & \mathbf{0} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_{\Delta t_{T-1,j}} & \mathbf{B}_{c,\Delta t_{T-1,j}} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{B}_{\upsilon,\Delta t_{T-1,j}} \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0}
\end{bmatrix},
$$

$$
\boldsymbol{\Psi} = \begin{bmatrix}
\boldsymbol{\Phi}_{\mathbf{x}(t_0)} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\Phi}_{\mathbf{x}(t_0),\mathbf{u}} \\
\mathbf{0} & \mathbf{Q}_{\Delta t_{1,j}} & \cdots & \mathbf{0} & \mathbf{0} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{Q}_{\Delta t_{T-1,j}} & \mathbf{0} \\
\boldsymbol{\Phi}_{\mathbf{u},\mathbf{x}(t_0)} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\Phi}_{\mathbf{u}}
\end{bmatrix},
$$

$$
\boldsymbol{\Lambda} = \begin{bmatrix}
\mathbf{C}_{t_{0,j}} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{D}_{c,t_{0,j}} & \mathbf{D}_{\upsilon,t_{0,j}} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{C}_{t_{1,j}} & \cdots & \mathbf{0} & \mathbf{D}_{c,t_{1,j}} & \mathbf{0} & \mathbf{D}_{\upsilon,t_{1,j}} & \cdots & \mathbf{0} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_{t_{T-1,j}} & \mathbf{D}_{c,t_{T-1,j}} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D}_{\upsilon,t_{T-1,j}} \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}
\end{bmatrix},
$$

$$
\boldsymbol{\Theta} = \begin{bmatrix}
\mathbf{R}_{t_{0,j}} & & \boldsymbol{\Theta}_{\text{cov}} & & \mathbf{0} \\
 & \mathbf{R}_{t_{1,j}} & & & \mathbf{0} \\
\boldsymbol{\Theta}'_{\text{cov}} & & \ddots & & \vdots \\
 & & & \mathbf{R}_{t_{T-1,j}} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0}
\end{bmatrix}.
$$

We abbreviate the dynamic errors $\int_{t_{i,j}-\Delta t_{i,j}}^{t_{i,j}} \mathbf{e}^{\mathbf{A}(t_{i,j}-s)}\mathbf{G}d\mathbf{W}(s)$ to $\mathbf{w}(t_{i,j} - \Delta_{i,j})$. The mean sum of squares and cross-products matrix of $\mathbf{u}_j$ over sample units is called $\boldsymbol{\Phi}_{\mathbf{u}}$ and the mean sum of cross-products between $\mathbf{x}(t_{0,j}) - \boldsymbol{\mu}_{\mathbf{x}(t_{0,j})}$ and $\mathbf{u}_j$ is called $\boldsymbol{\Phi}_{\mathbf{x}(t_0),\mathbf{u}}$. The latter must be estimated, though, if the state is latent.

The traits $\boldsymbol{\gamma}$ and $\boldsymbol{\kappa}$ are not explicitly displayed but can be viewed as a special kind of constant zero-mean exogenous variables $\mathbf{u}_{c,j}$ in $\mathbf{u}_j$, whose covariance matrices $\boldsymbol{\Phi}_\gamma$ and $\boldsymbol{\Phi}_\kappa$ in $\boldsymbol{\Phi}_\mathbf{u}$ as well as $\boldsymbol{\Phi}_{\mathbf{x}(t_0),\gamma}$ and $\boldsymbol{\Phi}_{\mathbf{x}(t_0),\kappa}$ in $\boldsymbol{\Phi}_{\mathbf{x}(t_0),\mathbf{u}}$ are not fixed quantities but have to be estimated. These latent variables have no loadings in $\boldsymbol{\Lambda}$ and have $\mathbf{B}_{c,t_0} = \mathbf{0}$ in $\boldsymbol{B}$. For $\boldsymbol{\gamma}$ the $\mathbf{B}_{c,\Delta t_{i,j}}$ in $\boldsymbol{B}$ are replaced by $\mathbf{A}^{\circ}_{\Delta t_{i,j}}$(see Table 1) and for $\boldsymbol{\kappa}$ the $\mathbf{D}_{c,t_{i,j}}$ in $\boldsymbol{\Lambda}$ by $\mathbf{I}$.