Kees van Montfort · Johan H.L. Oud
Manuel C. Voelkle   *Editors*

Continuous Time Modeling in the Behavioral and Related Sciences

Springer

Continuous Time Modeling in the Behavioral and Related Sciences

Kees van Montfort • Johan H. L. Oud •
Manuel C. Voelkle

Editors

Continuous Time Modeling in the Behavioral and Related Sciences

Springer

Editors
Kees van Montfort
Marketing and Supply Chain Management
Nyenrode Business University
Breukelen, The Netherlands

Johan H. L. Oud
Behavioural Science Institute
University of Nijmegen
Nijmegen, The Netherlands

Manuel C. Voelkle
Department of Psychology
Humboldt-Universität zu Berlin
Berlin, Germany

Preface

Over the past decades behavioral scientists have increasingly realized the potential of longitudinal data to address specific research questions, in particular those of a cause-effect nature. As a consequence, the methodology of longitudinal research and analysis has made much progress and an increasing number of large-scale longitudinal (time series and panel) data sets have become available for analysis. However, in accordance with the way longitudinal data are collected, at a restricted number of discrete time points, the statistical analysis is typically based on discrete time models. As argued by the authors in the present book, a series of problems is connected to this type of models, which make their results highly questionable. One main issue is the dependence of discrete time parameter estimates on the chosen time interval in dynamic modeling, which leads to incomparability of results across different observation intervals and, if unaccounted for, may even lead to contradictory conclusions.

Continuous time modeling, in particular by means of differential equations, offers a powerful solution to these problems, yet the use of continuous time models in the behavioral and related sciences such as psychology, sociology, economics, and medicine is still rare. Fortunately, recent initiatives to introduce and adapt continuous time models in a behavioral science context are gaining momentum. The purpose of the book is to assess the state of the art and to bring together the different initiatives. Furthermore, we emphasize the applicability of continuous time methods in applied research and practice.

Bergstrom, one of the pioneers of continuous time modeling in econometrics, credits Bartlett for the first significant contribution to the problem of estimating continuous time stochastic models from discrete data. As Bartlett succinctly stated in 1946: "The discrete nature of our observations in many economic and other time series does not reflect any lack of continuity in the underlying series. ...An unemployment index does not cease to exist between readings, nor does Yule's pendulum cease to swing" (p. 31).

At the time of Bartlett's contribution, intensive discussions took place in econometrics on the interpretation, identification, and estimation of nonrecursive simultaneous equations models. Simultaneity was considered by Bergstrom and

others as an artifact from short-term causal effects between variables within relatively long discrete time intervals (often years in economics), being in the case of nonrecursive simultaneity reciprocal effects within such intervals. In his 1966 publication "Nonrecursive models as discrete approximations to systems of stochastic differential equations," Bergstrom further concretized this interpretation of nonrecursiveness by describing the underlying causal effects by a system of stochastic differential equations. In fact, he connected this system to the discrete time data in two ways: the exact discrete model (EDM) and the approximate discrete model (ADM). Whereas the EDM connects the system to the discrete time data by highly nonlinear restrictions, the ADM uses the approximate linear restrictions of a (recursive or nonrecursive) simultaneous equations model. This book is greatly indebted to Bergstrom, as most of the chapters more or less explicitly continue and elaborate on the kind of exact or approximate methods, started by Bergstrom.

To implement the exact approach, traditionally maximum likelihood by means of filtering techniques, in particular Kalman filtering, is used and several of the chapters apply these techniques. In behavioral science, the application of maximum likelihood by means of structural equation modeling (SEM) became very popular and several other chapters make use of this alternative procedure.

In total, the book contains 16 chapters which cover a vast range of continuous time modeling approaches, going from one closely mimicking traditional linear discrete time modeling to highly nonlinear state space modeling techniques. Each chapter describes the type of research questions and data that the approach is most suitable for, provides detailed statistical explanations of the models, and includes one or more applied examples. To allow readers to implement the various techniques directly, accompanying computer code is made available online.

The book addresses the great majority of researchers in the behavioral and related sciences, who are interested in longitudinal data analysis. This includes readers who are involved in research in psychology, sociology, economics, education, management, and medical sciences. It is meant as a reference work for scientists and students working with longitudinal data and wanting to apply continuous time methods. The book also provides an overview of various recent developments for methodologists and statisticians. Especially for PhD students it offers the means to carry out analyses of longitudinal data in continuous time. Readers are supposed to have knowledge of statistics as taught at Master's or early PhD level.

We thank the authors for their willingness to contribute to the book, Eva Hiripi of Springer for her help in realizing the project, and Yoram Clapper, Leonie Richter, and Vincent Schmeits for their excellent typesetting work.

Breukelen, The Netherlands	Kees van Montfort
Nijmegen, The Netherlands	Johan H. L. Oud
Berlin, Germany	Manuel C. Voelkle

Contents

Contributors

Frits D. Bijleveld Vrije Universiteit Amsterdam School of Business and Economics, Department of Econometrics, Amsterdam, The Netherlands

SWOV Institute for Road Safety Research, The Hague, The Netherlands

Steven M. Boker University of Virginia, Department of Psychology, Charlottesville, VA, USA

Andreas M. Brandmaier Max Planck Institute for Human Development, Center for Lifespan Psychology, Berlin, Germany

Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Berlin, Germany

Marcus J. Chambers Department of Economics, University of Essex, Colchester, UK

Meng Chen The Pennsylvania State University, Department of Human Development and Family Studies, State College, PA, USA

Sy-Miin Chow The Pennsylvania State University, Department of Human Development and Family Studies, State College, PA, USA

David A. Cole Vanderbilt University, Department of Psychology & Human Development, Peabody College of Education and Human Development, Nashville, TN, USA

Jacques J. F. Commandeur Vrije Universiteit Amsterdam School of Business and Economics, Department of Econometrics, Amsterdam, The Netherlands

SWOV Institute for Road Safety Research, The Hague, The Netherlands

Pascal R. Deboeck University of Utah, Department of Psychology, College of Social and Behavioral Sciences, Salt Lake City, UT, USA

Charles C. Driver Max Planck Institute for Human Development, Center for Lifespan Psychology, Berlin, Germany

Emilio Ferrer University of California, Davis, Department of Psychology, Davis, CA, USA

Joseph E. Gonzales University of Massachusetts, Lowell, Department of Psychology, Lowell, MA, USA

Ellen L. Hamaker Department of Methodology and Statistics, Faculty of Social Sciences, Utrecht University, Utrecht, The Netherlands

Michael D. Hunter School of Psychology, Georgia Institute of Technology, Atlanta, GA, USA

Siem Jan Koopman Vrije Universiteit Amsterdam School of Business and Economics, Department of Econometrics, Amsterdam, The Netherlands

CREATES, Aarhus University, Aarhus, Denmark

Tinbergen Institute, Amsterdam, The Netherlands

Rebecca M. Kuiper Department of Methodology and Statistics, Faculty of Social Sciences, Utrecht University, Utrecht, The Netherlands

Omar Licandro University of Nottingham, School of Economics, Nottingham, UK

IAE-CSIC and Barcelona GSE, Barcelona, Spain

J. Roderick McCrorie School of Economics and Finance, University of St Andrews, St Andrews, UK

Heiner Meulemann University of Cologne, Institute for Sociology and Social Psychology, Cologne, Germany

Robert G. Moulder University of Virginia, Department of Psychology, Charlottesville, VA, USA

Nynke M. D. Niezink Carnegie Mellon University, Department of Statistics & Data Science, Pittsburgh, PA, USA

Zita Oravecz Pennsylvania State University, Department of Human Development and Family Studies, State College, PA, USA

Johan H. L. Oud Behavioural Science Institute, University of Nijmegen, Nijmegen, The Netherlands

Kristopher J. Preacher Vanderbilt University, Department of Psychology & Human Development, Peabody College of Education and Human Development, Nashville, TN, USA

Luis A. Puch Universidad Complutense de Madrid, Department of Economic Analysis and ICAE, Madrid, Spain

Nilam Ram Pennsylvania State University, Department of Human Development and Family Studies, State College, PA, USA

Jesús Ruiz Department of Economics and ICAE, Universidad Complutense de Madrid, Madrid, Spain

Oisín Ryan Department of Methodology and Statistics, Faculty of Social Sciences, Utrecht University, Utrecht, The Netherlands

Hermann Singer FernUniversität in Hagen, Department of Economics, Hagen, Germany

Tom A. B. Snijders University of Groningen, Department of Sociology, Groningen, The Netherlands

University of Oxford, Nuffield College, Oxford, UK

Joel S. Steele Portland State University, Department of Psychology, Portland, OR, USA

Michael A. Thornton Department of Economics and Related Studies, University of York, York, UK

Stacey S. Tiberio Oregon Social Learning Center, Eugene, OR, USA

Helgi Tómasson University of Iceland, Faculty of Economics, Reykjavík, Iceland

Kees van Montfort Nyenrode Business Universiteit, Department of Marketing and Supply Chain Management, Breukelen, The Netherlands

Erasmus University Rotterdam, Department of Biostatistics, Rotterdam, The Netherlands

Manuel C. Voelkle Humboldt-Universität zu Berlin, Department of Psychology, Berlin, Germany

Max Planck Institute for Human Development, Center for Lifespan Psychology, Berlin, Germany

Sunčica Vujić University of Antwerp, Department of Economics, Antwerp, Belgium

University of Bath, Department of Economics, Bath, UK

Julie Wood Pennsylvania State University, Department of Human Development and Family Studies, State College, PA, USA

Chapter 1
First- and Higher-Order Continuous Time Models for Arbitrary *N* Using SEM

Johan H. L. Oud, Manuel C. Voelkle, and Charles C. Driver

1.1 Introduction

This chapter is closely based on an earlier article in *Multivariate Behavioral Research* (Oud et al. 2018). The chapter extends the previous work by the analysis of the well-known Wolfer sunspot data along with more detailed information on the relation between the discrete time and continuous time model in the so-called exact discrete model (EDM). The previously published MBR article provides additional information on the performance of CARMA(p, q) models by means of several simulations, an empirical example about the relationships between mood at work and mood at home along with a subject-group-reproducibility test. However, to improve readability, we repeat the basic introduction to CARMA(p, q) modeling in the present chapter.

J. H. L. Oud (✉)
Behavioural Science Institute, University of Nijmegen, Nijmegen, The Netherlands
e-mail: j.oud@pwo.ru.nl

M. C. Voelkle
Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany
e-mail: manuel.voelkle@hu-berlin.de

C. C. Driver
Max Planck Institute for Human Development, Berlin, Germany
e-mail: driver@mpib-berlin.mpg.de

Time series analysis has a long history. Although the roots can be traced back to the earlier work of statisticians such as Yule, Kendall, and Durbin (see the historical overview in Mills 2012), it was the landmark publication "Time Series Analysis, Forecasting and Control" by Box and Jenkins (1970) that made time series analysis popular in many fields of science. Examples are Harvey (1989) and Lütkepohl (1991) in economics or Gottman (1981) and McCleary and Hay (1980) in the social sciences.

Time series analysis has greatly benefited from the introduction of the state space approach. The state space approach stems from control engineering (Kalman 1960; Zadeh and Desoer 1963) and sharply 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. The measurement part of the state space model specifies the relation between the observed and the underlying latent variables. 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 (Caines 1988; Deistler 1985; Harvey 1981; Ljung 1985). However, the state space representation is much more flexible, allows 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 and Koopman 2001). Therefore, the ARMA model in continuous time presented in this chapter, called a CARMA model (Brockwell 2004; Tómasson 2011; Tsai and Chan 2000), will in this chapter be formulated as a state space model.

Structural equation modeling (SEM) was introduced by Jöreskog (1973, 1977) along with the first SEM software: LISREL (Jöreskog and Sörbom 1976). The strong relationships between the state space approach and SEM were highlighted by Oud (1978) and Oud et al. (1990). Both consist of a measurement part and an explanatory part, and in both the explanatory part specifies the relationships between latent variables. Whereas in the state space approach the latent explanatory part is a recursive dynamic model, in SEM it is a latent structural equation model. As explained in detail by Oud et al. (1990), SEM is sufficiently general to allow specification of the state space model as a special case and estimation of its parameters by maximum likelihood and other estimation procedures offered by SEM. By allowing arbitrary measurement error structures, spanning the entire time range of the model, SEM further enhances the flexibility of the state space approach (Voelkle et al. 2012b).

An important drawback of almost all longitudinal models in SEM and time series analysis, however, is their specification in discrete time. Oud and Jansen (2000), Oud and Delsing (2010), and Voelkle et al. (2012a) discussed a series of problems connected with discrete time models, which make their application in practice highly questionable. One main problem is the dependence of discrete time results on the chosen time interval. This leads, first, to incomparability of results over different observation intervals within and between studies. If unaccounted for, it can easily lead to contradictory conclusions. In a multivariate model, one researcher could

find a positive effect between two variables *x* and *y*, while another researcher finds a negative effect between the same variables, just because of a different observation interval length in discrete time research. Second, because results depend on the specific length of the chosen observation interval, the use of equal intervals in discrete time studies does not solve the problem (Oud and Delsing 2010). Another interval might have given different results to both researchers. Continuous time analysis is needed to make the different and possibly contradictory effects in discrete time independent of the interval for equal as well as unequal intervals.

The CARMA time series analysis procedure to be presented in this chapter is based on SEM continuous-time state-space modeling, developed for panel data by Oud and Jansen (2000) and Voelkle and Oud (2013). The first of these publications used the nonlinear SEM software package Mx (Neale 1997) to estimate the continuous time parameters, the second the ctsem program (Driver et al. 2017). ctsem interfaces to OpenMx (Boker et al. 2011; Neale et al. 2016), which is a significantly improved and extended version of Mx. In both publications the kernel of the model is a multivariate stochastic differential equation, and in both 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. The exact solution turns out to impose nonlinear constraints on the discrete time parameters. Several authors tried to avoid the implementation of nonlinear constraints by using approximate procedures, although acknowledging the advantages of the exact procedure (e.g., Gasimova et al. 2014; Steele and Ferrer 2011a,b). In an extensive simulation study, Oud (2007) compared, as an example, the multivariate latent differential equation procedure (MLDE; Boker et al. 2004), which is an extension of the local linear approximation (LLA; Boker 2001), to the exact procedure. The exact procedure was found to give considerably lower biases and root-mean-square error values for the continuous time parameter estimates.

Instead of SEM, most time series procedures, particularly those based on the state space approach (Durbin and Koopman 2001; Hannan and Deistler 1988; Harvey 1989), use filtering techniques for maximum likelihood estimation of the model. Singer (1990, 1991, 1998) adapted these techniques for continuous time modeling of panel data. 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. Because both procedures maximize the likelihood, this ought to be the case. Considerable technical differences between the two procedures, however, make it nevertheless worthwhile to find that results indeed coincide. In view of its greater generality and its popularity outside of control theory, especially in social science, SEM can thus be considered a useful alternative to filtering. Making use of the extended SEM framework underlying OpenMx (Neale et al. 2016), both SEM and the Kalman filter procedure, originally developed in control theory, are implemented in ctsem.

An argument against SEM as an estimation procedure for time series analysis could be that it is known and implemented as a large N procedure, while time series analysis is defined for $N = 1$ or at least includes $N = 1$ as the prototypical case. In the past, different minimum numbers of sample units have been proposed for SEM, such as $N \geq 200$, at least 5 units per estimated parameter, at least 10 per variable, and so on, with all recommendations being considerably larger than $N = 1$. Recently, Wolf et al. (2013), on the basis of simulations, settled on a range of minimum values from 30 to 460, depending on key model properties. For small N, in particular $N = 1$, leading to a nonpositive definite sample covariance matrix, some SEM programs simply refuse to analyze the data or they change the input data before the analysis. For example, LISREL gives a warning ("Matrix to be analyzed is not positive definite") and then increases the values on the diagonal of the data covariance matrix S to get this matrix positive definite ("Ridge option taken with ridge constant = . . . "), before starting the analysis.

As made clear by Hamaker et al. (2003), Singer (2010), and Voelkle et al. (2012b), however, the long-standing suggestion that a SEM analysis cannot be performed because of a nonpositive definite covariance matrix S is wrong. Nothing in the likelihood function requires the sample covariance matrix S to be positive definite. It is the model implied covariance matrix Σ that should be positive definite, but this depends only on the chosen model and in no way on the sample size N. For the same reason, minimum requirements with regard to N do not make sense, and $N = 1$ is an option as well as any other N. The quality of the estimates as measured, for example, by their standard errors, depends indeed on the amount of data. The amount of data, however, can be increased both by the number of columns in the data matrix (i.e., the number of time points T in longitudinal research) and by the number of rows (i.e., the number of subjects N in the study). The incorrect requirement of a positive definite S for maximum likelihood estimation (Jöreskog and Sörbom 1996) may be explained by the modified likelihood function employed in LISREL and other SEM programs. This contains the quantity $\log |S|$ (see Equation (3) in Voelkle et al. 2012b), and thus these programs cannot handle $N = 1$ and N smaller than the number of columns in S. By employing the basic "raw data" or "full-information" maximum likelihood function (RML or FIML), OpenMx avoids this problem and (a) allows any N, including $N = 1$, (b) permits the nonlinear constraints of the EDM, (c) allows any arbitrary missing data pattern under the missing at random (MAR) assumption, and (d) allows individually varying observation intervals in continuous time modeling by means of so-called definition variables (Voelkle and Oud 2013).

Until now, however, the SEM continuous time procedure has not yet been applied on empirical $N = 1$ data, although the possibility is discussed and proven to be statistically sound by Singer (2010), and the software to do so is now readily available (Driver et al. 2017). Thus, the aim of the present chapter is, first, to discuss the state space specification of CARMA models in a SEM context. Second, to show that the SEM continuous time procedure is appropriate for $N = 1$ as well as for $N > 1$. Third, the proposed analysis procedure using ctsem will be applied on the well-known Wolfer sunspot data. This $N = 1$ data set has been analyzed by several

continuous time analysts before. The results from ctsem will be compared to those found previously.

1.2 Continuous Time Model

1.2.1 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 $q = 1$

$$\mathbf{y}_t = \mathbf{F}_{t,t-1}\mathbf{y}_{t-1} + \mathbf{F}_{t,t-2}\mathbf{y}_{t-2} + \mathbf{G}_t\mathbf{e}_t + \mathbf{G}_{t,t-1}\mathbf{e}_{t-1}. \tag{1.1}$$

The autoregressive part with $\mathbf{F}$-matrices specifies the lagged effects of the dependent variables, while the moving-average part with $\mathbf{G}$-matrices handles the incoming errors and lagged errors. Assuming the errors in the vectors $\mathbf{e}_t$, $\mathbf{e}_{t-1}$ to be independently standard normally distributed (having covariance matrices $\mathbf{I}$) and $\mathbf{G}_t$, $\mathbf{G}_{t,t-1}$ lower-triangular, the moving-average effects $\mathbf{G}_t\mathbf{e}_t$, $\mathbf{G}_{t,t-1}\mathbf{e}_{t-1}$ get covariance matrices $\mathbf{Q}_t = \mathbf{G}_t\mathbf{G}_t'$, $\mathbf{Q}_{t,t-1} = \mathbf{G}_{t,t-1}\mathbf{G}_{t,t-1}'$, which may be nondiagonal and with arbitrary variances on the diagonal. Specifying moving-average effects is no less general than the covariance matrices. Any covariance matrix $\mathbf{Q}$ can be written as $\mathbf{Q} = \mathbf{G}\mathbf{G}'$ in terms of a lower-triangular matrix (Cholesky factor) $\mathbf{G}$. In addition, estimating $\mathbf{G}$ instead of directly $\mathbf{Q}$ has the advantage of avoiding possible negative variance estimates showing up in the direct estimate of $\mathbf{Q}$.

The moving-average part $\mathbf{G}_t\mathbf{e}_t + \mathbf{G}_{t,t-1}\mathbf{e}_{t-1}$ in Eq. (1.1) may also 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(p, q) model, which reads for $p = 2$ and $q = 1$

$$\frac{d^2\mathbf{y}(t)}{dt^2} = \mathbf{F}_0\mathbf{y}(t) + \mathbf{F}_1\frac{d\mathbf{y}(t)}{dt} + \mathbf{G}_0\frac{d\mathbf{W}(t)}{dt} + \mathbf{G}_1\frac{d^2\mathbf{W}(t)}{dt^2}. \tag{1.2}$$

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: $\Delta t \rightarrow 0$. Equation (1.2) is sometimes written as

$$\mathbf{F}_0\mathbf{y}(t) + \mathbf{F}_1\frac{d\mathbf{y}(t)}{dt} + \mathbf{F}_2\frac{d^2\mathbf{y}(t)}{dt^2} = \mathbf{G}_0\frac{d\mathbf{W}(t)}{dt} + \mathbf{G}_1\frac{d^2\mathbf{W}(t)}{dt^2}, \tag{1.3}$$

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.

The next subsection will show in more detail how discrete time and continuous time equations such as (1.1) and (1.2) become connected as $\Delta t \rightarrow 0$. The error process in continuous time is the famous Wiener process[1] $\mathbf{W}(t)$ 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 standard normally distributed for $\Delta t = 1$ 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 (1.2) 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 Eq. (1.2) to the observed values in discrete time.

We will now show how the CARMA(2,1) model in Eq. (1.2) 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 (1.4) with so-called drift matrix $\mathbf{A}$ and diffusion matrix $\mathbf{G}$ and a measurement equation (1.5) with loading matrix $\mathbf{C}$ and measurement error

---

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

vector $\mathbf{v}(t)$:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{G}\frac{d\mathbf{W}(t)}{dt}, \tag{1.4}$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{v}(t). \tag{1.5}$$

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 the model has an equilibrium state: $E[\mathbf{x}(t)] = E[\mathbf{x}(t_0)] = \mathbf{0}$, and in case all eigenvalues of $\mathbf{A}$ have negative real part, $\mathbf{0}$ is the stable equilibrium state in the model.

In state space form, the observed second-order model CARMA(2,1) in Eq. (1.2) 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 (1.2) then follows from state space model (1.4)–(1.5) by specification

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{F}_0 \\ \mathbf{I} & \mathbf{F}_1 \end{bmatrix}, \; \mathbf{G} = \begin{bmatrix} \mathbf{G}_0 & \mathbf{0} \\ \mathbf{G}_1 & \mathbf{0} \end{bmatrix}, \; \mathbf{C} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \end{bmatrix}, \; \mathbf{v}(t) = \mathbf{0}. \tag{1.6}$$

Applying (1.6) we find first

$$\frac{d\mathbf{x}_1(t)}{dt} = \mathbf{F}_0\mathbf{x}_2(t) + \mathbf{G}_0\frac{d\mathbf{W}(t)}{dt}, \tag{1.7}$$

$$\frac{d\mathbf{x}_2(t)}{dt} = \mathbf{x}_1(t) + \mathbf{F}_1\mathbf{x}_2(t) + \mathbf{G}_1\frac{d\mathbf{W}(t)}{dt} \; \Rightarrow$$

$$\frac{d^2\mathbf{x}_2(t)}{dt^2} = \frac{d\mathbf{x}_1(t)}{dt} + \mathbf{F}_1\frac{d\mathbf{x}_2(t)}{dt} + \mathbf{G}_1\frac{d^2\mathbf{W}(t)}{dt^2}. \tag{1.8}$$

Substituting (1.7) into the implication in (1.8) gives

$$\frac{d^2\mathbf{x}_2(t)}{dt^2} = \mathbf{F}_0\mathbf{x}_2(t) + \mathbf{F}_1\frac{d\mathbf{x}_2(t)}{dt} + \mathbf{G}_0\frac{d\mathbf{W}(t)}{dt} + \mathbf{G}_1\frac{d^2\mathbf{W}(t)}{dt^2}, \tag{1.9}$$

which for $\mathbf{y}(t) = \mathbf{x}_2(t)$ leads to the CARMA(2,1) model in (1.2).

The specification in the previous paragraph can be generalized to find the CARMA(p, q) model in state space form in (1.10), where $r = \max(p, q + 1)$. $\mathbf{G}$ 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

$\mathbf{G}_i \mathbf{G}_j'$ ($i, j = 0, 1, \ldots, r - 1$). In the literature one often finds the alternative state space form (1.11) (see e.g., Tómasson 2011; Tsai and 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

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

$$
\mathbf{C} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} \end{bmatrix}, \ \mathbf{v}(t) = \mathbf{0}. \tag{1.10}
$$

model and $\mathbf{G}_0$. For two reasons we prefer (1.10) in the case of CARMA(p, q) models with $q > 0$. The derivation of (1.11) 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 (1.11), 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 (1.7)–(1.8) is less simple.

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

$$
\mathbf{C} = \begin{bmatrix} \mathbf{I} & \mathbf{H}_1 & \mathbf{H}_2 & \cdots & \mathbf{H}_{r-1} \end{bmatrix}, \ \mathbf{v}(t) = \mathbf{0}. \tag{1.11}
$$

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 (1.4)–(1.5). The state space approach in fact reformulates higher-order models as a first-order model, and this will be applied in the sequel.

1.2.2 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 chapter differentiates from many approximate procedures found in the literature. We show how 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 (1.1) to the general discrete time state space model in (1.12)–(1.13), one observes that the latter becomes immediately the ARMA(1,0) model for $\mathbf{y}_t = \mathbf{x}_t$ but is more flexible in time handling.

$$\mathbf{x}_t = \mathbf{A}_{\Delta t}\mathbf{x}_{t-\Delta t} + \mathbf{G}_{\Delta t}\mathbf{e}_{t-\Delta t} \tag{1.12}$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{v}_t. \tag{1.13}$$

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 (1.12) can be put in the equivalent difference quotient form

$$\frac{\Delta\mathbf{x}_t}{\Delta t} = \mathbf{A}_{*\Delta t}\mathbf{x}_{t-\Delta t} + \mathbf{G}_{\Delta t}\frac{\mathbf{e}_{t-\Delta t}}{\Delta t}$$
$$\text{for } \mathbf{A}_{*\Delta t} = (\mathbf{A}_{\Delta t} - \mathbf{I})/\Delta t \text{ implying } \mathbf{A}_{\Delta t} = \mathbf{I} + \mathbf{A}_{*\Delta t}\Delta t. \tag{1.14}$$

So we have the discrete time state space model in two forms: difference quotient form (1.14) and solution form (1.12). Equation (1.12) is called the solution of (1.14), because it describes the actual state transition across time in accordance with (1.14) and is so said to satisfy the difference quotient equation. Note that analogously the general continuous time state space model (1.4)–(1.5) 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 (1.4) and its solution (1.15) (Arnold 1974; Singer 1990):

$$\mathbf{x}(t) = e^{\mathbf{A}\Delta t}\mathbf{x}(t - \Delta t) + \int_{t-\Delta t}^{t} e^{\mathbf{A}(t-s)}\mathbf{G}d\mathbf{W}(s). \tag{1.15}$$

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:

$$\mathbf{A}_{\Delta t} = \mathrm{e}^{\mathbf{A}\Delta t} \quad \text{and} \quad \mathbf{Q}_{\Delta t} = \int_{t-\Delta t}^{t} \mathrm{e}^{\mathbf{A}(t-s)} \mathbf{Q} \mathrm{e}^{\mathbf{A}'(t-s)} \mathrm{d}(s) . \qquad (1.16)$$

While discrete time autoregression matrix $\mathbf{A}_{\Delta t}$ and continuous time drift matrix $\mathbf{A}$ are connected via the highly nonlinear matrix exponential, 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.

The connection between discrete and continuous time becomes further clarified by two definitions of the matrix exponential $\mathrm{e}^{\mathbf{A}\Delta t}$. The standard definition (1.17)

$$\begin{aligned}
\mathrm{e}^{\mathbf{A}\Delta t} &= \sum_{k=0}^{\infty} (\mathbf{A}\Delta t)^k / k! \\
&= \mathbf{I} + \mathbf{A}\Delta t + \left(\frac{1}{2}\mathbf{A}^2 \Delta t^2 + \frac{1}{6}\mathbf{A}^3 \Delta t^3 + \frac{1}{24}\mathbf{A}^4 \Delta t^4 + \cdots \right)
\end{aligned} \qquad (1.17)$$

shows the rather complicated presence of $\mathbf{A}$ in $\mathbf{A}_{\Delta t}$ and the formidable task to extract continuous time $\mathbf{A}$ from discrete time $\mathbf{A}_{\Delta t}$ in an exact fashion. However, it also shows that for $\Delta t \to 0$, the quantity between parentheses becomes arbitrarily small, and the linear part $\mathbf{I} + \mathbf{A}\Delta t$ could be taken as an approximation of $\mathrm{e}^{\mathbf{A}\Delta t}$, leading precisely to $\mathbf{A}_{*\Delta t}$ in the difference quotient equation (1.14) as approximation of $\mathbf{A}$ in the differential equation (1.4). Depending on the length of the interval Δt, however, the quality of this approximation can be unduly bad in practice. An alternative definition is based on oversampling (Singer 2012), meaning that the total time interval between measurements is divided into arbitrary small subintervals: $\delta = \frac{\Delta t}{D}$ for $D \to \infty$. The definition relies on the multiplication property of autoregression: $\mathbf{x}_t = \mathbf{A}_{0.5\Delta t} \mathbf{x}_{t-0.5\Delta t}$ and $\mathbf{x}_{t-0.5\Delta t} = \mathbf{A}_{0.5\Delta t} \mathbf{x}_{t-\Delta t} \Rightarrow \mathbf{x}_t = \mathbf{A}_{0.5\Delta t} \mathbf{A}_{0.5\Delta t} \mathbf{x}_{t-\Delta t} = \mathbf{A}_{\Delta t} \mathbf{x}_{t-\Delta t}$ which is equally valid in the continuous time case: $\mathbf{x}_t = \mathrm{e}^{\mathbf{A}(0.5\Delta t)} \mathrm{e}^{\mathbf{A}(0.5\Delta t)} \mathbf{x}_{t-\Delta t} = \mathrm{e}^{\mathbf{A}\Delta t} \mathbf{x}_{t-\Delta t}$, leading to multiplicative definition:

$$\mathrm{e}^{\mathbf{A}\Delta t} = \lim_{D \to \infty} \prod_{d=0}^{D-1} (\mathbf{I} + \mathbf{A}\delta_d) \text{ for } \delta_d = \frac{\Delta t}{D}. \qquad (1.18)$$

$\mathbf{I} + \mathbf{A}\Delta t$ in (1.17), becoming $\mathbf{I} + \mathbf{A}\delta$ in (1.18), is no longer an approximation of $\mathrm{e}^{\mathbf{A}\delta}$ but becomes equal to $\mathrm{e}^{\mathbf{A}\delta}$ for $D \to \infty$ and $\delta \to 0$, while then $\mathbf{A}_{*\delta} = (\mathrm{e}^{\mathbf{A}\delta} - \mathbf{I})/\delta$ in difference equation (1.14) becomes equal to drift matrix $\mathbf{A}$ in differential equation (1.4). So, the formidable task to extract $\mathbf{A}$ from $\mathbf{A}_{\Delta t}$ is performed here in the same simple way as done by just taking approximate $\mathbf{A}_{*\Delta t}$ from $\mathbf{A}_{\Delta t}$ in (1.14), but instead of once over the whole interval Δt, it is repeated

over many small subintervals, becoming for sufficiently large D equal to the exact procedure in terms of (1.17). We write $\mathbf{A}_{*\Delta t} \approx \mathbf{A}$, but one should keep in mind that by taking a sufficiently small interval in terms of the right-hand side of (1.18), one can get $\mathbf{A}_{*\Delta t}$ as close to $\mathbf{A}$ as wanted. An adapted version[2] of the oversampling procedure is described in Voelkle and Oud (2013).

Let us illustrate the connection between CARMA(1,0) and ARMA(1,0) by an example. 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 $\{\mathbf{y}(t); t \geqslant 0\}$ 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

---

[2]The adapted version uses (1.18A)

$$\mathrm{e}^{\mathbf{A}\Delta t} = \lim_{D \to \infty} \prod_{d=0}^{D-1} \left[\left(\mathbf{I} - \tfrac{1}{2}\mathbf{A}\delta_d \right)^{-1} \left(\mathbf{I} + \tfrac{1}{2}\mathbf{A}\delta_d \right) \right] \text{ for } \delta_d = \frac{\Delta t}{D}, \tag{1.18A}$$

which converges much more rapidly than (1.18). It is based on the approximate discrete model (ADM), described by Oud and Delsing (2010), which just as the EDM goes back to Bergstrom (1984). Computation of the matrix exponential by (1.18) or (1.18A) has the advantage over the diagonalization method (Oud and Jansen 2000) that no assumptions with regard to the eigenvalues need to be made. Currently by most authors the Padé-approximation (Higham 2009) is considered the best computation method, which therefore is implemented in the most recent version of ctsem.

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 $a_{\Delta t} = e^{a\Delta t}$ 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 et al. 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 $\mathbf{A}_{\Delta t} = e^{\mathbf{A}\Delta t}$ 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 and 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 and Jansen 2000; Tómasson 2015; Voelkle and 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 has especially consequences for the initial time point. Suppose for the ARMA(2,1) model in state space form (1.19)–(1.20),

$$\begin{bmatrix} \mathbf{x}_{1,t-\Delta t} \\ \mathbf{x}_{1,t} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{F}_{t,t-2\Delta t} & \mathbf{F}_{t,t-\Delta t} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{1,t-2\Delta t} \\ \mathbf{x}_{1,t-\Delta t} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{t,t-\Delta t} \end{bmatrix} \mathbf{e}_{t-\Delta t},$$
$$\mathbf{x}_t \quad = \quad \mathbf{A}_{\Delta t} \quad\quad \mathbf{x}_{t-\Delta t} \quad + \quad \mathbf{G}_{\Delta t} \quad \mathbf{e}_{t-\Delta t}, \quad (1.19)$$

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{0} & \mathbf{I} \end{bmatrix} \mathbf{x}_t, \quad\quad\quad\quad\quad\quad (1.20)$$

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 $t_0 + \Delta t$. That would result in data loss, since the $\mathbf{0}$ in (1.20) simply eliminates the lagged part of $\mathbf{x}_{t-\Delta t}$ without any connection to the data. Similar remarks apply to the initial

derivative $d\mathbf{x}_1(t)/dt$

$$
\begin{bmatrix} \frac{d\mathbf{x}_1(t)}{dt} \\ \frac{d^2\mathbf{x}_1(t)}{dt^2} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{F}_0 & \mathbf{F}_1 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(t) \\ \frac{d\mathbf{x}_1(t)}{dt} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_0 \end{bmatrix} \frac{d\mathbf{W}(t)}{dt},
$$

$$
\frac{d\mathbf{x}(t)}{dt} = \mathbf{A} \quad \mathbf{x}(t) \; + \; \mathbf{G} \quad \frac{d\mathbf{W}(t)}{dt}, \qquad (1.21)
$$

$$
\mathbf{y}(t) = \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \mathbf{x}(t), \qquad\qquad (1.22)
$$

in (1.21)–(1.22), 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. One should 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 $d\mathbf{x}_1(t)/dt$. This lack of data and relative unreliability of estimates are the price one has to pay for choosing higher-order ARMA(p, q) and CARMA(p, q) models.

1.2.3 Extended Continuous Time Model

The extended continuous time state space model reads

$$
\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \boldsymbol{\gamma} + \mathbf{G}\frac{d\mathbf{W}(t)}{dt}, \qquad (1.23)
$$

$$
\mathbf{y}_{t_i} = \mathbf{C}\mathbf{x}(t_i) + \mathbf{D}\mathbf{u}(t_i) + \boldsymbol{\kappa} + \mathbf{v}_{t_i}. \qquad (1.24)
$$

In comparison to the basic model in (1.4)–(1.5), 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 $\mathbf{B}\mathbf{u}(t)$ and $\mathbf{D}\mathbf{u}(t)$ of fixed exogenous variables in 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 (1.23) lead to the following extended solution:

$$
\mathbf{x}(t) = \mathrm{e}^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0) + \int_{t_0}^{t} \mathrm{e}^{\mathbf{A}(t-s)}\mathbf{B}\mathbf{u}(s)\mathrm{d}(s)
$$
$$
+ \mathbf{A}^{-1}\big[\mathrm{e}^{\mathbf{A}(t-t_0)} - \mathbf{I}\big]\boldsymbol{\gamma} + \int_{t_0}^{t} \mathrm{e}^{\mathbf{A}(t-s)}\mathbf{G}\mathrm{d}\mathbf{W}(s). \tag{1.25}
$$

1.2.4 Exogenous Variables

We have seen that the basic model in the case of stability (all eigenvalues of $\mathbf{A}$ having negative real part) has $\mathbf{0}$ as stable equilibrium state. Exogenous effects $\mathbf{B}\mathbf{u}(t)$ 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 Eq. (1.26) 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.

$$
\mathbf{x}(t) = \mathrm{e}^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0) + \mathbf{A}^{-1}\left[\mathrm{e}^{\mathbf{A}(t-t_0)} - \mathbf{I}\right]\mathbf{B}\mathbf{u}(t_0)
$$
$$
+ \mathbf{A}^{-1}\left[\mathrm{e}^{\mathbf{A}(t-t_0)} - \mathbf{I}\right]\boldsymbol{\gamma} + \int_{t_0}^{t} \mathrm{e}^{\mathbf{A}(t-s)}\mathbf{G}\mathrm{d}\mathbf{W}(s). \tag{1.26}
$$

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 (1.26) 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.[3]

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, how the exogenous behavior function looks like in between observations. By endogenizing the exogenous variables, they are handled as random variables, added to the state vector, and in the same way as the other state variables related to their past values in an autoregressive fashion. Advantages of endogenizing are its nonapproximate 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 differentiating time points within subjects, exogenous variables also enable to differentiate 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 $\mathbf{A}^{-1}[e^{\mathbf{A}(t-t_0)} - \mathbf{I}]\mathbf{b}_d$ from the one of boys. This amount will be zero for $t - t_0 = 0$, but the regression-like analysis procedure applied 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 may impact 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.

1.2.5 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

---

[3]A better approximation than a step function is given by a piecewise linear or polygonal approximation (Oud and Jansen 2000; Singer 1992). Then we write $\mathbf{u}(t)$ in (1.23) as $\mathbf{u}(t) = \mathbf{u}(t_0) + (t - t_0)\mathbf{b}_{(t_0,t]}$ and (1.26) becomes:

$$\mathbf{x}(t) = e^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0) + \mathbf{A}^{-1}\big[e^{\mathbf{A}(t-t_0)} - \mathbf{I}\big]\mathbf{B}\mathbf{u}(t_0) + \Big\{\mathbf{A}^{-2}\big[e^{\mathbf{A}(t-t_0)} - \mathbf{I}\big] - \mathbf{A}^{-1}(t - t_0)\Big\}\mathbf{B}\mathbf{b}_{(t_0,t]}$$

$$+ \mathbf{A}^{-1}\big[e^{\mathbf{A}(t-t_0)} - \mathbf{I}\big]\boldsymbol{\gamma} + \int_{t_0}^{t} e^{\mathbf{A}(t-s)}\mathbf{G}d\mathbf{W}(s).$$

(1.26A)

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 toward (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 then after enough time is elapsed, the subject's trajectory in a stable model will be coinciding with its group trajectory.

From solution equation (1.25), 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, $t_0 \rightarrow -\infty$, the subject's expected trajectory is

$$E[\mathbf{x}(t)|\boldsymbol{\gamma}] = \int_{-\infty}^{t} e^{\mathbf{A}(t-s)}\mathbf{B}\mathbf{u}(s)\mathrm{d}(s) - \mathbf{A}^{-1}\boldsymbol{\gamma}, \tag{1.27}$$

which keeps a subject-specific distance $-\mathbf{A}^{-1}\boldsymbol{\gamma}$ from the subject's group mean trajectory $E[\mathbf{x}(t)] = \int_{-\infty}^{t} e^{\mathbf{A}(t-s)}\mathbf{B}\mathbf{u}(s)\mathrm{d}(s)$. As a result the subject's sample trajectory regresses toward 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.

It should be noted that the impact of the fixed and random effects $\mathbf{B}\mathbf{u}(t)$ and $\boldsymbol{\gamma}$ in the state equation (1.23) is quite different from that of $\mathbf{D}\mathbf{u}(t_i)$ and $\boldsymbol{\kappa}$ in the measurement equation (1.24). 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 Eq. (1.25) or (1.26). 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).

1.3   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 (1.28). In (1.28) $\otimes$ 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 $\mathbf{A}, \mathbf{B}, \mathbf{Q}$.

$$
\begin{aligned}
&\mathbf{A} &&\mathbf{A}_{\Delta t_{i,j}} = \mathrm{e}^{\mathbf{A}\Delta t_{i,j}} \\[2mm]
&\mathbf{B} &&\mathbf{B}_{\Delta t_{i,j}} = \mathbf{A}^{\circ}{}_{\Delta t_{i,j}}\mathbf{B} \quad \text{for} \quad \mathbf{A}^{\circ}{}_{\Delta t_{i,j}} = \mathbf{A}^{-1}(\mathrm{e}^{\mathbf{A}\Delta t_{i,j}} - \mathbf{I}) \\[2mm]
&\mathbf{Q} = \mathbf{GG'} &&\mathbf{Q}_{\Delta t_{i,j}} = \int_{t-\Delta t_{i,j}}^{t} \mathrm{e}^{\mathbf{A}(t-s)}\mathbf{Q}\mathrm{e}^{\mathbf{A}'(t-s)}\mathrm{d}(s) \\[2mm]
& &&\qquad\quad = \mathrm{irow}[\mathbf{A}_{\#}^{-1}(\mathrm{e}^{\mathbf{A}_{\#}\Delta t_{i,j}} - \mathbf{I})\mathrm{row}\,\mathbf{Q}] \\[2mm]
& &&\qquad\quad\ \ \text{for}\ \mathbf{A}_{\#} = \mathbf{A} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{A} \qquad\qquad (1.28)\\[2mm]
&\mathbf{C} &&\mathbf{C}_{t_{i,j}} = \mathbf{C} \\[2mm]
&\mathbf{D} &&\mathbf{D}_{t_{i,j}} = \mathbf{D} \\[2mm]
&\mathbf{R} &&\mathbf{R}_{t_{i,j}} = \mathbf{R}
\end{aligned}
$$

$$
\boldsymbol{\mu}_{\mathbf{x}(t_0)} = E[\mathbf{x}(t_0)] = \mathbf{B}_{t_0}\mathbf{u}(t_0)
$$

$$
\boldsymbol{\Phi}_{\mathbf{x}(t_0)} = E[(\mathbf{x}(t_0) - \boldsymbol{\mu}_{\mathbf{x}(t_0)})(\mathbf{x}(t_0) - \boldsymbol{\mu}_{\mathbf{x}(t_0)})']
$$

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 and Voelkle 2014).

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, $\mathbf{\Phi}_{\mathbf{x}(t_0)}$ 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 $\mathbf{\mu}_{\mathbf{x}(t)} = \mathbf{\mu}_{\mathbf{x}(t_0)}$ 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 additional gender or any other additional exogenous variables, differentiating subjects from each other but constant in time. The assumption of equilibrium, $\mathbf{\mu}_{\mathbf{x}(t)} = \mathbf{\mu}_{\mathbf{x}(t_0)}$,—possibly but not necessarily a stable equilibrium— leads to equilibrium value

$$\mathbf{\mu}_{\mathbf{x}(t)} = \mathbf{\mu}_{\mathbf{x}(t_0)} = -\mathbf{A}^{-1}\mathbf{B}\mathbf{u}_c, \tag{1.29}$$

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 $\mathbf{\Phi}_{\mathbf{x}(t)} = \mathbf{\Phi}_{\mathbf{x}(t_0)}$ is assumed to be in equilibrium, leading to equilibrium value

$$\mathbf{\Phi}_{\mathbf{x}(t)} = \mathbf{\Phi}_{\mathbf{x}(t_0)} = \text{irow}[-\mathbf{A}_{\#}^{-1}\text{row}\,\mathbf{Q}]. \tag{1.30}$$

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 (1.29) and (1.30). 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 (1.28) by SEM, we put all variables and matrices of the EDM into SEM model

$$\mathbf{\eta} = \boldsymbol{B}\,\mathbf{\eta} + \mathbf{\zeta} \quad \text{with} \quad \mathbf{\Psi} = E(\mathbf{\zeta}\mathbf{\zeta}'), \tag{1.31}$$

$$\mathbf{y} = \mathbf{\Lambda}\mathbf{\eta} + \mathbf{\epsilon} \quad \text{with} \quad \mathbf{\Theta} = E(\mathbf{\epsilon}\mathbf{\epsilon}'). \tag{1.32}$$

The SEM model consists of two equations, structural equation (1.31) and measurement equation (1.32), in terms of four vectors, $\mathbf{\eta}, \mathbf{\zeta}, \mathbf{y}, \mathbf{\epsilon}$, and four matrices, $\boldsymbol{B}, \mathbf{\Psi}, \mathbf{\Lambda}, \mathbf{\Theta}$. From Eqs. (1.31)–(1.32), one easily derives the model implied mean $\mathbf{\mu}$

and covariance matrix $\boldsymbol{\Sigma}$ and next the raw maximum likelihood equation (1.33) (see e.g., Bollen 1989)

$$\text{RML} = \sum_{j=1}^{N} \left[m_j \log(2\pi) + \log(|\boldsymbol{\Sigma}_j|) + (\mathbf{y}_j - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}_j^{-1} (\mathbf{y}_j - \boldsymbol{\mu}_j) \right]. \qquad (1.33)$$

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 $\mathbf{y}_j$, 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 $\boldsymbol{\eta}, \mathbf{y}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}$, and matrices, $\boldsymbol{B}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Theta}$ include the variables and matrices of the EDM. This is done in (1.34). 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 $\mathbf{u}_v(t)$. We abbreviate the dynamic errors $\int_{t_{i,j}-\Delta t_{i,j}}^{t_{i,j}} e^{\mathbf{A}(t_{i,j}-s)} \mathbf{G} d\mathbf{W}(s)$ to $\mathbf{w}(t_{i,j} - \Delta t_{i,j})$. 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 $\mathbf{x}(t_{0,j}) - \boldsymbol{\mu}_{\mathbf{x}(t_0)}$ 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 (1.28)) and for $\boldsymbol{\kappa}$ the $\mathbf{D}_{c,t_{i,j}}$ in $\boldsymbol{\Lambda}$ by $\mathbf{I}$.

$$\boldsymbol{\eta} = [\mathbf{x}'_j \, \mathbf{u}'_j]' \quad \text{with} \quad \mathbf{x}_j = [\mathbf{x}'(t_{0,j}) \, \mathbf{x}'(t_{1,j}) \, \cdots \, \mathbf{x}'(t_{T-1,j})]'$$

$$\text{and} \quad \mathbf{u}_j = [\mathbf{u}'_{c,j} \, \mathbf{u}'_v(t_{0,j}) \, \mathbf{u}'_v(t_{1,j}) \, \cdots \mathbf{u}'_v(t_{T-1,j})]',$$

$$\boldsymbol{\zeta} = [\mathbf{w}'_j \, \mathbf{u}'_j]' \quad \text{with} \quad \mathbf{w}_j = [\mathbf{x}'(t_{0,j}) - \boldsymbol{\mu}'_{\mathbf{x}(t_{0,j})} \, \mathbf{w}'(t_{1,j} - \Delta t_{1,j})$$

$$\cdots \, \mathbf{w}'(t_{T-1,j} - \Delta_{T-1,j})]',$$

$$\mathbf{y} = [\mathbf{y}'_j \, \mathbf{u}'_j]' \quad \text{with} \quad \mathbf{y}_j = [\mathbf{y}'(t_{0,j}) \, \mathbf{y}'(t_{1,j}) \, \cdots \, \mathbf{y}'(t_{T-1,j})]',$$

$$\boldsymbol{\epsilon} = [\mathbf{v}'_j \, \mathbf{0}']' \quad \text{with} \quad \mathbf{v}_j = [\mathbf{v}'(t_{0,j}) \, \mathbf{v}'(t_{1,j}) \, \cdots \, \mathbf{v}'(t_{T-1,j})]',$$