A Variational Approach to Continuous Time Dynamic Models

Hannes Meinlschmidt and Meike Sons and Mark Stemmler

Abstract In this study we introduce a new approach to parameter estimation in continuous time modeling in the spirit of variational data assimilation or machine learning. This is a purely time-continuous approach relying on the theory of optimization for dynamical systems. We complement the proposed algorithm with a practical example, comparing the results of this approach to those obtained via Continuous Time Structural Equation Modeling (`ctsem`). To this end, we assess the reciprocal relationship between satisfaction with health and satisfaction with work using data from the German Socio-Economic Panel. It turns out the proposed algorithm determines a drift matrix whose the principle directions (eigenvectors) are qualitatively equivalent to the ones estimated via `ctsem`, but the associated eigenvalues differ substantially, leading to quantitatively different conclusions.

1 Introduction

Continuous time (CT) models are a well established concept to extract information about the evolution of psychological processes from longitudinal data as it arises

Hannes Meinlschmidt
FAU Erlangen-Nürnberg, Department of Mathematics, Chair for Dynamics, Control, Machine Learning and Numerics (AvH-professorship), Cauerstraße 11, 91058 Erlangen (Germany)
e-mail: `meinlschmidt@math.fau.de`

Meike Sons
Bayerisches Landesamt für Gesundheit und Lebensmittelsicherheit, Arbeitsmedizinisches Institut für Schulen (AMIS-Bayern), Frauenplatz 4, 96049 Bamberg (Germany)
e-mail: `meike.sons@lgl.bayern.de`

Mark Stemmler
FAU Erlangen-Nürnberg, Institut für Psychologie, Lehrstuhl für psychologische Diagnostik, Methodenlehre & Rechtspsychologie, Nägelsbachstraße 49c, 91052 Erlangen (Germany)
e-mail: `mark.stemmler@fau.de`

e.g. in long-term and large-scale panel studies ([5, 17, 22, 25]). The basic idea is that the moods, feelings and, generally, the behavior of sentient beings evolve in a continuous manner in time, so that a model which describes (a particular part of) these traits should naturally also be of continuous-time form. With this point of view, change and interaction occurs continuously, and data collection in studies is merely an evaluation of the state of the traits under consideration at a given point in time. A conceptual model is illustrated in Fig. 1. CT models are complemented by discrete time (DT) models which are the “classical” viewpoint taken by many researchers in the past, in which the temporal evolution of the traits is modeled to occur from one time point to the next in a discrete manner. This represents the point of view of change occurring when we measure it.

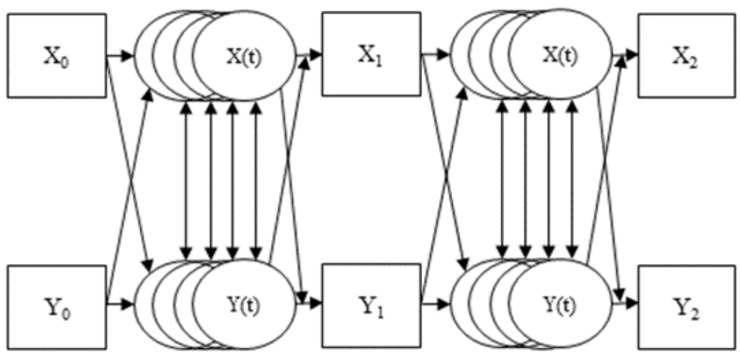

Fig. 1 Conceptual model. A continuous time perspective on a cross-lagged panel model. The first three of the several time points for the two-process continuous time model are illustrated. Occasions at which no observations are made are represented by latent variables (circles). The processes influence one another continuously over time.

CT models offer a few practical advantages over DT models which are all related to the fundamental difference of CT models being structurally independent of the underlying time intervals used in data collection; for example, different lengths of time intervals between panel evaluations are naturally incorporated into CT models whereas for DT models, particular care has to be taken in order to not obtain results which depend explicitly on the time intervals in data collection ([26]). This in particular also concerns the case where data may be missing for individuals at certain collection time points ([20]). In this sense, CT models are ideally suited to handle big and complex, imperfect data like panel data ([18]). Also, the study of dynamic relationships between variables such as cross-lagged panel effects is easily possible within the framework of CT models. Examining how cross-effects evolve and vary as a function of the time interval between evaluations leads to a more dynamic view on the underlying processes and avoids oversimplification ([6]). We do however not go into a detailed comparison between CT and DT models here, but only refer to e.g. [22, 25] and the references there for a comprehensive exposition.

In fact, we rather take the CT modeling approach for longitudinal data for granted, but propose a different way of deriving and calculating the actual model parameters which disposes of the connection to DT models nearly completely. Indeed, a most common approach ([18,19,25]) to CT modeling of panel data as we want to consider goes via a Structural Equation Model (SEM) and the Exact Discrete Model (EDM), although there of course also exist other approaches such as filter methods (e.g. [19]. In the SEM approach, one supposes that the underlying model is of continuous-time and sets up a large SEM by relating the CT model at time evaluation points with an associated DT model, the EDM. More precisely, if $\mathbf{x}$ represents the, say, V-dimensional vector of quantities of interest, evolving over time, and $t_0, t_1, t_2, \ldots$ are the time points of data collection, then a linear DT model in state-space form would be given by the difference equation

$$\mathbf{x}(t_i) = \mathbf{A}(\Delta t_i)\mathbf{x}(t_{i-1}) + \mathbf{b}(\Delta t_i) + \mathbf{w}(\Delta t_i), \tag{1}$$

where $\Delta t_i = t_i - t_{i-1}$ is the time interval between data collection time points, $\mathbf{A}(\Delta t_i)$ is the V×V transition matrix from time point t_{i-1} to t_i, and $\mathbf{b}(\Delta t_i)$ and $\mathbf{w}(\Delta t_i)$ are the V-dimensional intercept vector and random innovations depending on Δt_i ([25, eq. (13)]). The sought-for quantity here is $\mathbf{A}(\Delta t_i)$. The associated linear CT model would be described by the continuous-time (stochastic) differential equation[1]

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + b + G\frac{\mathrm{d}W(t)}{\mathrm{d}t} \tag{2}$$

with the time derivative $\dot{\mathbf{x}}$, the *drift matrix* A ($V \times V$), the continuous-time intercept b and a continuous-time error process (random walk, Wiener process) induced by the $V \times V$ lower triangle matrix G ([25, (14)]). Relating the explicit solution to (2) with the DT model (1), one obtains the EDM

$$\mathbf{x}(t_i) = e^{A\Delta t_i}\mathbf{x}(t_{i-1}) + A^{-1}\left[e^{A\Delta t_i} - I\right]b + \mathbf{w}(\Delta t_i),$$

with the matrix exponential $e^{A\cdot}$ and the $V \times V$ identity matrix I and the covariance

$$\operatorname{cov}\mathbf{w}(\Delta t_i) = \int_{t_{i-1}}^{t_i} e^{A(t_i-s)}GG^\top e^{A^\top(t_i-s)}\,\mathrm{d}s.$$

In particular, we have $\mathbf{A}(\Delta t_i) = e^{A\Delta t_i}$ which is the fundamental link between $\mathbf{A}$ and the drift matrix A estimated in a CT model. The foregoing identities for $\mathbf{x}(t_i)$ are then used to set up a SEM whose parameters are highly nonlinear in the sought-for quantities, most obviously A which is modulated by the matrix exponential. The SEM parameters are fitted to the given data by e.g. a maximum-likelihood estimator ([25, p. 184]). Nowadays there are sophisticated packages which automatically take care

[1] We stick with boldface notation $\mathbf{x}$ for the *trajectories*; otherwise, matrices and vectors related to the DT model are written in boldface, too, whereas the corresponding quantities in the CT model are in regular font.

of all the mathematical setup such as `ctsem` ([7, 8]) for `R`. Still, the SEM approach fundamentally links back to the discrete-time setup in the way it is set up and fitted.

We next propose a purely continuous-time, dynamical systems-type approach to CT modeling. This conforms to the following point of view for CT models as in [25, p. 179]:

> One can also conceive of such a continuous time model as a dynamical system. At each point in time, the system has a specific configuration, but time itself never acts as an explanatory variable in the model. Instead, the model itself is an explanatory model. Arguably, this is often a more realistic view of the world compared with models that include time explicitly as a predictor (e.g., HLMs or LGMs). For example, when using an LGM to study learning, we typically do not assume that time "causes" learning, although we mathematically model it this way.

This is done in the next section. We follow up with a preliminary comparison between the results obtained by `ctsem` and the proposed algorithm, respectively, for a practical example relating satisfaction with health and satisfaction with work from the German Socio-Economic Panel (SOEP, [27]).

2 Machine learning type approach to Continuous Time Modeling

We next describe a direct way of estimating the continuous-time matrix A in a CT model given by (2) by a direct mathematical optimization problem in terms of a variational problem. The idea follows the basic paradigm of continuous-time modeling by assuming that the behaviour of a subject in the respective data set follows a predetermined pattern over time whose properties can be captured by a linear continuous-time differential equation such as (2) above. There will be no reformulation in terms of the DT values in a SEM or the likes, but we use the way how the individual subjects' trajectories depend on A *directly*. This will prove to work in a very direct and mathematically fully tractable manner.

To make the idea clear, we consider only the simplest version of (2) and do not include intercepts or stochastic noise at first. We come back to these extensions later. That is, the CT model is now given by

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t). \tag{3}$$

For the sake of exposition, we further explain the approach first for one individual subject $\mathbf{z}$. Say we have $(k+1)$ evaluations of panel data $z_0, z_1, \ldots, z_k$ for this subject (vectors of a fixed length), each obtained at the respective times $t_0, t_1, \ldots, t_k$ which we assume to be in increasing order. The solution to the linear continuous-time differential equation (3) for $\mathbf{z}$ starting from the initial value z_0 at time t_0 is given by $\mathbf{z}(t) = e^{A(t-t_0)}z_0$, using the matrix exponential function. We can imagine $t \mapsto \mathbf{z}(t)$ as the trajectory of the initial starting point z_0 induced by the linear differential equation with matrix A. Thus, under the foregoing assumption, we would expect that the panel data points z_i each lie on that trajectory, so $z_i = \mathbf{z}(t_i) = e^{A(t_i-t_0)}z_0$ for every

i. Of course, due to random factors, we cannot expect this equality to hold precisely, but only in an approximating sense. From this point of view, we are looking for a matrix A such that the discrepancy between the panel data $z_1, z_2, \ldots, z_k$ and the trajectory evaluations $e^{A(t_1-t_0)}z_0, e^{A(t_2-t_0)}z_0, \ldots, e^{A(t_k-t_0)}z_0$ is minimal, for instance in a least-squares sense. That means that A should be a solution to the minimization problem

$$\min_A \sum_{i=1}^{k} \left\| e^{A(t_i-t_0)}z_0 - z_i \right\|_2^2, \tag{4}$$

where $\|\cdot\|_2$ is the standard Euclidean norm. (We do not include $i = 0$ in the minimization problem by construction: the trajectory $t \mapsto z(t)$ always starts at z_0 at time t_0.) In an optimal world without perturbations or random effects, we could hope to be able to determine A exactly such that the foregoing minimization problem (4) has an optimal value of 0. But in reality, there will be discrepancies between z_i and the trajectory point $\mathbf{z}(t_i) = e^{A(t_i-t_0)}z_0$. Moreover, (4) is a nonconvex optimization problem such that we can only expect a *local* solution instead of a global one. But still, generally, by (locally) solving (4), we determine A such that the trajectory evaluations $\mathbf{z}(t_1), \mathbf{z}(t_2), \ldots, \mathbf{z}(t_k)$ fit the observed data $z_1, z_2, \ldots, z_k$ in a (locally) *best possible way* with the given data. This is the basis of a *variational formulation* of continuous-time modeling of cross-lagged panel data. From this point of view, we do not require any implicit or explicit statistical assumptions about probability distributions of random elements or perturbations in the given panel data. Mathematically, this approach is similar to variational techniques in so-called data assimilation [1].

Moreover, if we view this task as one of fitting values given by a *model function*, here represented by the parameters in A (its coefficients) and solving the linear differential equation (3), to a given set of *measurements*, then we can interpret this approach exactly as a form of machine learning. Note that A only indirectly creates the model values: we use it to obtain time-continuous trajectories in dependence on A, which are then fitted to the observed data or measurements $z_1, z_2, \ldots, z_k$ at the time instants t_i.

The above description was given for one single subject. Next we formulate the above mentioned apporach for a full panel set. Suppose that we have N total subjects and for each subject, indexed by j, we have a number $(k_j + 1)$ data points $z_0^j, z_1^j, \ldots, z_{k_j}^j$ associated to time points $t_0^j, t_1^j, \ldots, t_{k_j}^j$ which we again suppose to be ordered in an increasing fashion for each j. (The time points of evaluation are allowed to vary with each subject j.) Then the associated full minimization problem becomes

$$\min_A \sum_{j=1}^{N} \omega_j \sum_{i=1}^{k_j} \left\| \mathbf{z}^j(t_i^j) - z_i^j \right\|_2^2 \tag{5}$$

where each function $\mathbf{z}^j$ is the solution to the time-continuous linear differential equation (3) for $\mathbf{z}^j$ with initial value z_0^j at time t_0^j, so $\mathbf{z}^j(t_0^j) = z_0^j$. We could specify that this means that $\mathbf{z}^j$ is given by

$$\mathbf{z}^j(t) = e^{A(t-t_0^j)} z_0^j$$

and insert this formula in (5) to get the analogue to (4). It will however be convenient for later considerations to retain the abstract description for z^j as the solution to (3) with initial value $\mathbf{z}^j(t_0^j) = z_0^j$. In the full formulation (5), we moreover have included a weighting parameter $\omega_j \geq 0$ for each subject. This parameter could be used, for example, to normalize each outer summand in (5) (indexed by j) with respect to the number of subject-specific data points k_j, which we could do by choosing $\omega_j = 1/k_j$ for this case. However, for example it may also be reasonable to choose these parameters ω_j as 1 for each j in which case a subject with a higher number of data points would indirectly be considered more important to be well fitted than subjects with a lower number of data points. (In principle, it would also be possible to have a parameter ω_i^j for every individual data point. This could be used if extensive prior knowledge about individual data points was available.)

Intercepts and trait variables. So far, the optimization approach to CT modeling of cross-lagged panel data was described for the simplest case of the underlying differential equation (3). It is also possible to include continuous-time intercepts b (possibly with a selector $b = Bu$ for subgroups of subjects) and continuous-time trait variables κ^j (for each subject j) in these considerations in a very straightforward way. We refer to [17] for a comprehensive overview of these aspects. The modification would be to keep the minimization problem (5), but now $\mathbf{z}^j$ is defined to be the solution to

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + b + \kappa^j \tag{6}$$

with initial value $\mathbf{z}(t_0^j) = z_0^j$, and in addition to minimizing with respect to A in (5), we also minimize with respect to b (or u in $b = Bu$) and the collection $\kappa^1, \kappa^2, \ldots, \kappa^N$. Regarding the subgroup selectors $b = Bu$, it would be interesting to consider recursive partitioning techniques which were very recently also transferred to continuous-time models [2]. We note that also equation (6) admits a unique solution satisfying the given initial value, and there is an explicit formula (the *variation of constants* or *Duhamel* formula) which we could use if so desired.

Stochastic differential equation formulation. It seems natural to include also a stochastic error term as in the CT model (2) in the linear differential equations (3) or (6), and then also to optimize with respect to the matrix G. Then $\bar{\mathbf{z}}^j$ is the solution to the continuous-time *stochastic* differential equation (2) with initial value $\bar{\mathbf{z}}(t_0^j) = z_0^j$. (We use the bar in $\bar{\mathbf{z}}^j$ as a visual cue that $\bar{\mathbf{z}}^j$ is supposed to be a solution to the *stochastic* differential equation.) The trajectories $\bar{\mathbf{z}}^j$ are now random variables since they arise as solutions to a stochastic differential equation. Accordingly, we can not optimize the trajectories directly but only stochastic indicators such as moments. We choose the expected value, and the associated minimization problem then becomes

$$\min_{A,G} \sum_{j=1}^{N} \omega_j \sum_{i=1}^{k_j} \mathbb{E}\Big[\big\|\bar{\mathbf{z}}^j(t_i^j) - z_i^j\big\|_2^2\Big]. \tag{7}$$

However, since (2) is a *linear* stochastic differential equation, we know that the relationship between the white noise induced by G and the solutions $\bar{\mathbf{z}}^j$ is linear. Using this and a few more properties of the Wiener process W when expanding the square in (7), it turns out that (7) is equivalent to

$$\min_{A,G} \sum_{j=1}^{N} \omega_j \sum_{i=1}^{k_j} \left[\left\| \mathbf{z}^j(t_i^j) - z_i^j \right\|_2^2 + \int_{t_0^j}^{t_i^j} \left\| e^{A(t_i^j - s)} G \right\|_{\mathrm{Fr}}^2 \mathrm{d}s \right], \tag{8}$$

where each $\mathbf{z}^j$ is again the solution to the *deterministic* (so, *non-stochastic*!) linear differential equation (6) with initial value $\mathbf{z}(t_0^j) = z_0^j$. Here, $\|\cdot\|_{\mathrm{Fr}}$ denotes the Frobenius norm of a matrix. We derive this in the appendix. But in (8), we see that any optimal pair for the variables A and G will necessary be of the form $(\bar{A}, 0)$, since G *only* occurs in a sum of non-negative terms in the objective function to be minimized, so it will always be optimal to choose $G = 0$. In this sense, we postulate that we can always consider the deterministic problem setup (5), which is exactly the resulting problem with $G = 0$, in order to determine the dynamics of the problem, that is, A and possible trait vectors, without loss of generality. The situation would change if G was given and fixed and thus not subject to minimization in (7), but it is not clear how one would construct G from given data. Vice versa, given or having optimized for the dynamics, we expect that one could devise a second-stage optimization problem to estimate G or $GG^T = Q$ from the data. We leave this for future work.

Measurement equation and error. There is also a straightforward extension of the method described so far to also include a measurement equation and associated measurement error. Then, we assume that the trajectories $\mathbf{z}^j$ or $\bar{\mathbf{z}}^j$, respectively, are in fact *latent* and we only directly observe a, say, linear function of them at the time points t_i^j, induced by a matrix C, with an error w_i^j and an offset d^j:

$$y_i^j = C\mathbf{z}^j(t_i^j) + d^j + w_i^j.$$

Then we would replace $\mathbf{z}^j(t_i^j)$ in (5) and (8) by y_i^j. Since this is an affine-linear function in $\mathbf{z}^j(t_i^j)$, this modification gives rise to only marginal mathematical changes in the approach. In particular, assuming that the expected value of the measurement errors is zero, and that they are uncorrelated with respect to the random error in the stochastic differential equation (2), the derivation of (8), and the conclusion that it suffices to consider the deterministic counterpart of the respective formulation, stays correct.

Particularities of this approach. The approach as laid out so far seems promising to be very effective mainly due to three reasons.

First, minimizing (5) with respect to the drift matrix A will certainly involve solving the differential equation (6) for each $\mathbf{z}^j$ many times, at least as often as the current value of the function to be minimized need be evaluated. For a large number N of trajectories, this will amount to a huge amount of solves. On the other hand, numerical methods for differential equations are extremely well researched and highly efficient solvers are ubiquitous in nearly every programming language which we can make use of; in particular because, typically, the dimension V of the state vectors will not be large at all.

Second, while (5) is highly nonlinear in the optimization variable A, the latter's number of variables V^2 will rather be very low for standards in numerical optimization. Similarly to the situation as for the differential equation, this allows to use the full width of available theory for nonlinear optimization problems and associated efficient solvers. If we include subject-specific trait variables, the number of optimization variables increases quite drastically by $V \times N$; however, while the overall dynamics induced by A are coupled across all subjects, subject-specific traits only influence one particular trajectory, effectively leading to a certain decoupling between optimization for each subject-specific trait. (For the same reason, we do not expect problems regarding over-fitting, neither in this regard nor without subject-specific traits, since then the number of degrees of freedom V^2 is very small compared to the number of subjects.)

Still, there is the caveat, that if we denote the function to be minimized in (5)—the so-called *objective function*—by $f(A)$, then we will require at least the gradient $\nabla f(A)$ of f to use more efficient optimization algorithms. That is, we need to understand and derive for which directions H (this is a matrix!) the objective function f will infinitesimally decrease when going from A towards $A + H$. This requires either some more or less involved mathematics to derive an analytic expression for the gradient $\nabla f(A)$, or usage of *automatic differentiation* tools, which are ubiquitous nowadays, to derive the gradient automatically on a machine level. The latter could however become a bottleneck in performance, such that the former is preferable in general, if feasible. Having an explicit description of the gradient $\nabla f(A)$ also allows to optimize the efficiency with which it is calculated. Note that there are also results on an explicit description of the gradients for the maximum likelihood estimation which is done in `ctsem` [23], but it seems that these have been mostly superseded by automatic differentiation tools these days. We also mention that the particular form of the objective function to be minimized in (5) or (8), respectively, is susceptible to so-called stochastic optimization algorithms such as the nowadays ever present SGD, Stochastic Gradient Descent, which could provide further performance speed ups.

Third, and most importantly, there is room for generalization. Indeed, the whole approach does not at all rely on an explicit formula for the solution to the differential equation (6). It is thus straightforward to generalize the whole approach to e.g. time-dependent drift matrices $A(t)$. Here, the amount of parameters to be optimized would increase drastically; formally, it would be uncountable, but practically, it would become $V^2 \times$ (the amount of time steps used in a numerical discretization of the differential equation).

Recently, regularization for continuous-time models, leading to sparsity in the parameters, that is,less complex models, has been considered by Orzek, Arnold and Völkle in [14, 15]. Extending the minimization problems (5) or (8) with such regularizations is very natural and could be considered without further ado. (In the present formulation such a regularization approach would only make sense with trait variables and intercepts, since the parameters in A are already few enough.) This would require to use adapted optimization algorithms to actually solve the optimization problems, but these are well researched and very efficient nowadays. We refer to the references mentioned before.

Going even further, one could switch to *nonlinear* models of the form

$$\dot{\mathbf{x}}(t) = F_\theta(\mathbf{x}(t))$$

with a nonlinear function parametrized by a vector of parameters θ such as for example a Neural Network. In this case, a stochastic differential equation formulation would also be much more difficult to handle. While such a generalization would not at all be straightforward, it nevertheless seems very promising to investigate. We leave this for future work.

Finally, for full disclosure, we point out again that the present approach is inherently deterministic so far. In particular, by design, there will be no confidence intervals for the estimated parameters, and we do not yet estimate the diffusion matrix Q; only the actual dynamics induced by A will be estimated.

2.1 Empirical Example

Work, pressures, strains, and stresses within the workplace, which occupy most people's waking time have been identified as being a potentially important health factor ([9, 11]). Numerous theories now exist, developed from a wide range of perspectives, postulating a direct link between organisational/workplace stress and personal well-being. Meta-analytic results show that an increase in work satisfaction would be associated with improved health satisfaction. Furthermore, there is evidence for a reciprocal effect from health satisfaction to job satisfaction ([9, 11]). We tested cross-lagged effects between satisfaction with health and satisfaction with work using Continuous Time Structural Equation Modeling (`ctsem`). While many questions might be asked using this approach, the questions we will address here are the following: What are the continuous temporal dynamics of satisfaction with health and satisfaction with work? How much variation does there exist in such dynamics? And, in particular, how do the results obtained using a `ctsem` approach compare with those yielded by the above variational algorithm, and what are possible implications for model fitting?

Sample In this example, we used data from the German Socio-Economic Panel (SOEP, [27]), an ongoing survey of German households and persons which has been conducted annually since 1984. More precisely, we chose the measurements from 1984 to 2021.

Measures We are interested in the association between satisfaction with health and satisfaction with work. In the SOEP participants were asked to rate their satisfaction with various aspects of their lives on a 11-point scale, ranging from 0 (totally

Table 1 Continuous time parameter estimates: satisfaction with health and satisfaction with work

	Satisfaction Health (OV)		Satisfaction Work (OV)	
Parameter	Estimate	[95% CI][a]	Estimate	[95% CI][a]
Drift matrix (A)				
Auto-effects	-0.457*	[-0.462; -0.452]	-0.578*	[-0.583; -0.572]
Cross-effects	0.262*	[0.258; 0.267]	0.376*	[0.371; 0.382]
Diffusion matrix (Q)				
q	4.105*	[4.082; 4.129]	5.399*	[5.367; 5.432]
$qJ = qS$	0.388*	[0.384; 0.408]		
Baseline ($t0$)				
M_{t0}	2.468*	[2.453; 2.484]	2.307*	[2.29; 2.324]
var_{t0}	4.42*	[4.375; 4.467]	5.272*	[5.218; 5.328]
cov_{t0}	1.462 *	[1.425; 1.5]		
Model indices				
-2LL	3496829			
degrees of freedom	2837			

[a] Confidence intervals for many parameters in CT models may not be symmetric around the point estimate.
* $p < 0.05$.
Note: Continuous time parameter estimates for auto-effects are always negative, but correspond to positive effects in discrete time.

unhappy) to 10 (totally happy). The question in this regard is as follows: "How satisfied are you today with the following areas of your life?" The chosen items are "With your health" and "With your work".

Model We used two different approaches to obtain a model for the foregoing aspects in the given data.

First, we fitted a SEM-based continuous time (CT) cross-lagged panel model ([6, 7, 25, 26]) to our data. We fit the model using the open-source software `R` (version 4.2.2; [21]) and the package `ctsem` (version 3.7.2, [7, 8], which interfaces to `OpenMx` ([13]) via `ctsemOMX` (version 1.0.4). The precise setup of the model in `R` can be found in the appendix.

Second, we used the variational approach to Continuous Time Modeling as it was explained before. The code used to generate the results was written from scratch in `python` and can be found on `github` [10].

Since this is a first glimpse and a trial run for the variational type approach to continuous-time modeling, we do not consider time-continuous intercepts or individual traits in the empirical example.

2.2 Results

`ctsem`: The results of the continuous time models estimated by `ctsem` are summarized in Table 1. Estimates of the auto- and cross-effects for both processes are displayed in the drift matrix which resulted to

$$A_{\text{ctsem}} = \begin{pmatrix} -0.457024 & 0.262225 \\ 0.376226 & -0.5775 \end{pmatrix}.$$

The time-continuous auto-effects on the diagonal of the drift matrix reflect the processes' temporal stability; negative values for auto-effects are expected in the continuous time approach. Looking at the derived auto-effects, so the diagonal entries of the matrix $\mathbf{A}_{\text{ctsem}}(\Delta t) = e^{A_{\text{ctsem}}\Delta t}$ for $0 < \Delta t < 37$ years (see Fig. 2), the average within-person stability of satisfaction with health for a discrete time interval of 1 year ($\Delta t = 1$) was 0.66. As expected, stability coefficients decreased with increasing time. The stability of satisfaction with work was lower with 0.59 for a discrete time interval of 1 year ($\Delta t = 1$) and asymptotically approaches zero with increasing time. Of primary interest for our research question are the cross-effects of the drift matrix, which reflect the inter-relatedness of the processes (e.g., an increase in X leads to a decrease/increase in Y). We analyse the significance of the effects using 95% likelihood-based confidence intervals. The drift coefficients make it possible to estimate the auto-regressive and cross-effects for any chosen time interval Δt between 0 and 37 years. In order to predict the relationships between variables over a given time period, the continuous time parameters are used to calculate the discrete time parameters [6]. The cross-effects, so the non-diagonal entries of $\mathbf{A}_{\text{ctsem}}(\Delta t) = e^{A_{\text{ctsem}}\Delta t}$ as a function of the time interval between observations ($0 < \Delta t < 37$ years) are shown in Fig. 3. Regarding the relationship between satisfaction with health and satisfaction with work we found that both the positive effect of satisfaction with health on satisfaction with work (coef. = 0.24 for $\Delta t = 1$ year) and the positive effect of satisfaction with work on satisfaction with health (coef. = 0.16 for $\Delta t = 1$ year) reach their respective maxima at a time length between two and three years and decreases for longer times, although the maximum is decisively smaller than for the positive effect of satisfaction with health on satisfaction with work. (Indeed the maxima must occur at the same time for mathematical reasons related to the matrix exponential.)

Variational approach We next present results for the SOEP panel data with regards to the relationship between satisfaction with health and satisfaction with work, as described earlier.

Before we present the actual results, a few words on the computations and methodology. We have used the normalization $\omega_j = 1/k_j$. The minimization was solved by a Quasilinear Newton Method (BFGS) which was supplied with the objective function to be minimized as in (5) and its gradient with respect to the parameters in A. The code for these calculations was written in Python whose SciPy package offers the aforementioned BFGS optimization method. For the initialization of A, we used the matrix which was obtained from `ctsem` as in Table 1.

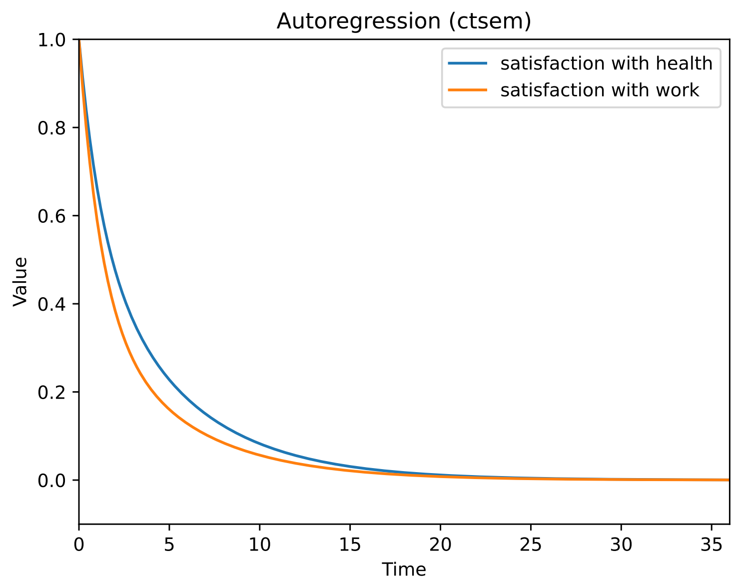

Fig. 2 Auto-effects derived from the continuous time model for satisfaction with health and satisfaction with work. The plot shows the auto-effects on the diagonal of $\mathbf{A}_{\text{ctsem}}(\Delta t)$ as a function of a time interval $0 < \Delta t < 37$ years. The stability of satisfaction with health (blue) and the stability of satisfaction with work (orange).

The two-dimensional data points were set up such that the first variable corresponds to "satisfaction with health" and, accordingly, the second one corresponds to "satisfaction with work". Then the optimization result, so the matrix A minimizing—at least locally—problem (5) was found to be

$$A_{\text{learn}} = \begin{pmatrix} -0.258455 & 0.220716 \\ 0.386343 & -0.49353 \end{pmatrix}.$$

Let us note that, in order to validate our expectation that the possible phenomenon of over-fitting should not occur due to the low number $V^2 = 4$ of degrees of freedom in this example, the algorithm was run several times with a reduced data set, where only 50% or 25% of the participants in the SOEP study were chosen randomly and used to optimize for A_{learn}, and the resulting matrix A_{learn} was indeed stable.

Looking at the matrix values for the discrete-time drift matrix as derived from the continuous-time one, $\mathbf{A}_{\text{learn}}(\Delta t) = e^{A_{\text{learn}}\Delta t}$ as depicted in Figures 4 (auto-effects) and 5 (cross-effects), we find the expected behavior decreasing values for the auto-effects. The average within-person stability of satisfaction with health for a discrete time interval of 1 year ($\Delta t = 1$) was 0.8, whereas the stability of satisfaction with work was lower with 0.63 for the same discrete time interval. The cross-effects again exhibit a maximum after which they decline, however, this time the maximum occurs

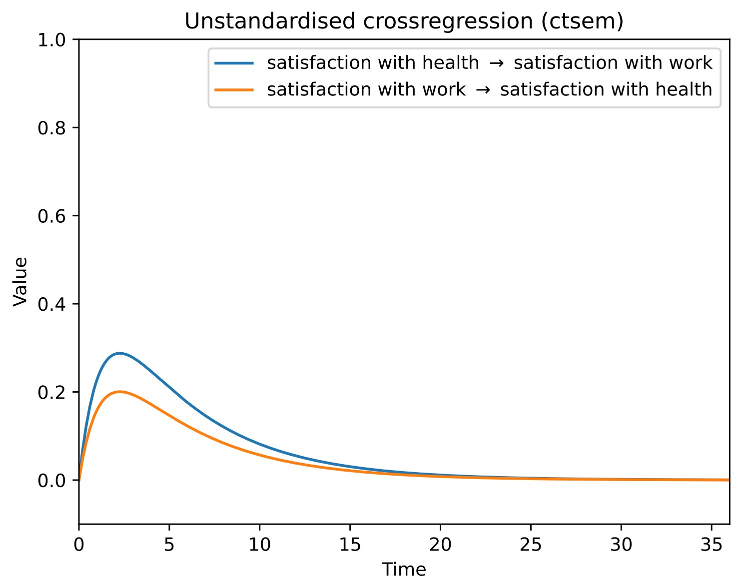

Fig. 3 Cross-effects derived from the continuous time model for satisfaction with health and satisfaction with work. The plot shows the cross-lagged parameters on the off-diagonal of $\mathbf{A}_{\mathrm{ctsem}}(\Delta t)$ as a function of a time interval of $0 < \Delta t < 37$ years. Satisfaction with health predicting subsequent changes in satisfaction with work (blue) and changes in satisfaction with work predicting subsequent changes in satisfaction with health (orange).

at about $\Delta t = 4$ years. Again, the positive influence of satisfaction with health on satisfaction with a coefficient of 0.27 at $\Delta t = 1$ year is more substantial than the reverse one with a coefficient of 0.15 at $\Delta t = 1$ year, as seen in the graphs.

2.3 Discussion

We next compare the different results A_{ctsem} and A_{learn} for the time-continuous drift matrices and associated consequences for the time-discrete drift matrices $\mathbf{A}_{\mathrm{ctsem}}$ and $\mathbf{A}_{\mathrm{learn}}$.

Firstly, we remark that if f denotes the function to be minimized in the optimization problem (5), then $f(A_{\mathrm{learn}})$ is 12% less (so: better) than $f(A_{\mathrm{ctsem}})$ which is a substantial but not completely outrageous difference. Note that A_{ctsem} was the starting point for the numerical optimization algorithm used to solve the minimization problem (5) and the algorithm moved away to find A_{learn}, so A_{ctsem} is not a local minimum for (5). Thus, A_{ctsem} is a reasonable first approximation of a matrix minimizing (5), but not an actual optimal solution. Of course this comparison is a bit unfair since `ctsem` actually estimates—read: optimizes for—more parameters, in

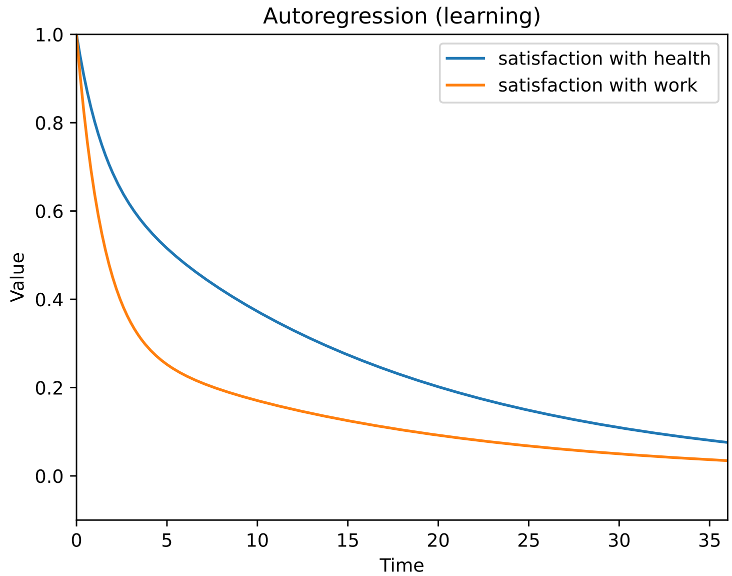

Fig. 4 Auto-effects derived from the continuous time model for satisfaction with health and satisfaction with work via the machine-learning approach. The plot shows the auto-effects on the diagonal of $\mathbf{A}_{\text{learn}}(\Delta t)$ as a function of a time interval $0 < \Delta t < 37$ years. The stability of satisfaction with health (blue) and the stability of satisfaction with work (orange).

particular the diffusion matrix $Q = GG^T$. However, if our main interest is in finding a matrix A such that the present panel data is best described by the dynamics induced by the associated linear differential equation, then it makes sense to compare the performances of A_{learn}, so A_{ctsem} only with regard to the optimization problem (5).

Comparing the actual parameters in A_{ctsem} and A_{learn}, we observe that the non-diagonal entries in each matrix are reasonably well matched, whereas the diagonal entries differ more substantially. However, the qualitative properties of a matrix A are better represented by its spectral properties, so its eigenvalues $\lambda(A)$ and eigenvectors $v(A)$. Here we find (rounded and ordered)

$$\lambda(A_{\text{ctsem}}) = (-0.19, -0.84), \quad v(A_{\text{ctsem}}) = \left(\begin{pmatrix} 0.71 \\ 0.7 \end{pmatrix}, \begin{pmatrix} -0.57 \\ 0.82 \end{pmatrix}\right)$$

and

$$\lambda(A_{\text{learn}}) = (-0.06, -0.69), \quad v(A_{\text{learn}}) = \left(\begin{pmatrix} 0.74 \\ 0.67 \end{pmatrix}, \begin{pmatrix} -0.45 \\ 0.89 \end{pmatrix}\right).$$

While the eigenvalues are again rather different, it becomes apparent that the eigenvectors, so the principal directions of either A along which the solution to the differential equation (3) will evolve, are well matched. We can also observe that in

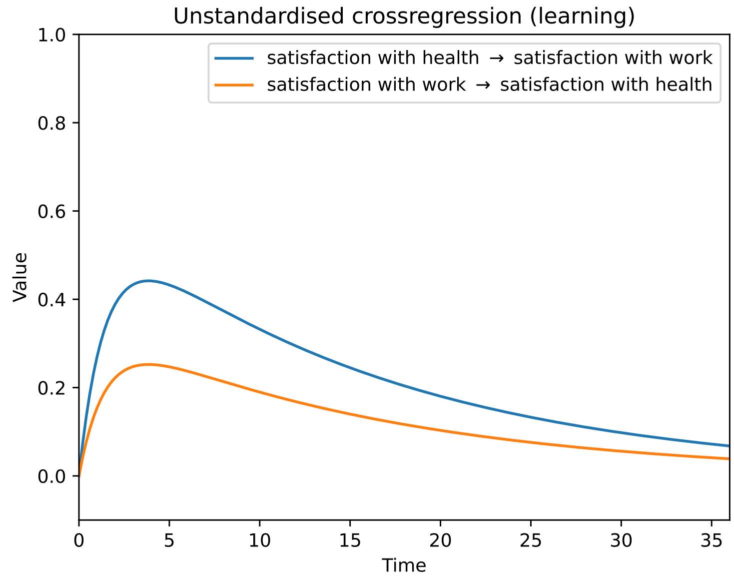

Fig. 5 Cross-effects derived from the continuous time model for satisfaction with health and satisfaction with work via the machine-learning approach. The plot shows the cross-lagged parameters on the off-diagonal of $\mathbf{A}_{\text{learn}}(\Delta t)$ as a function of a time interval of $0 < \Delta t < 37$ years. Satisfaction with health predicting subsequent changes in satisfaction with work (blue) and changes in satisfaction with work predicting subsequent changes in satisfaction with health (orange).

Fig. 6 where the trajectories of a few randomly selected individuals in the panel study under the linear differential equation (3) are shown for $A = A_{\text{ctsem}}$ and $A = A_{\text{learn}}$, respectively. For $A = A_{\text{ctsem}}$, the dashed red lines indicate the eigenvectors $v(A_{\text{ctsem}})$ scaled by their respective (inverted) eigenvalue to show how the trajectory evolution aligns with the eigenvectors, making them a most important quantity of interest in this context.

With respect to the questions of study for the SOEP data, so the influence of satisfaction with health on satisfaction with work and vice versa, we again refer to Figures 2–5. It is apparent that the decay in the auto-effects as Δt increases is much slower for $\mathbf{A}_{\text{learn}}(\Delta t)$ than in $\mathbf{A}_{\text{ctsem}}(\Delta t)$; the mathematical reason behind this is that the eigenvalues $\lambda(A_{\text{learn}})$ of A_{learn} are less negative than those of A_{ctsem} in $\lambda(A_{\text{ctsem}})$. This means that the eigenvalues $\lambda(\mathbf{A}_{\text{learn}}(\Delta t = 1)) = e^{\lambda(A_{\text{learn}})}$ of $\mathbf{A}_{\text{learn}}(\Delta t = 1)$ are closer to 1 than those of $\mathbf{A}_{\text{ctsem}}$. On the other hand, the cross-effects exhibit quite similar behavior for both results, with the respective peaks being a bit higher and also for larger values of Δ for A_{learn} or $\mathbf{A}_{\text{learn}}(\Delta t)$, respectively, obtained from the learning approach.

We conclude that *qualitatively*, the results A_{ctsem} and A_{learn} of the different approaches agree, while *quantitatively*, they yield different conclusions.

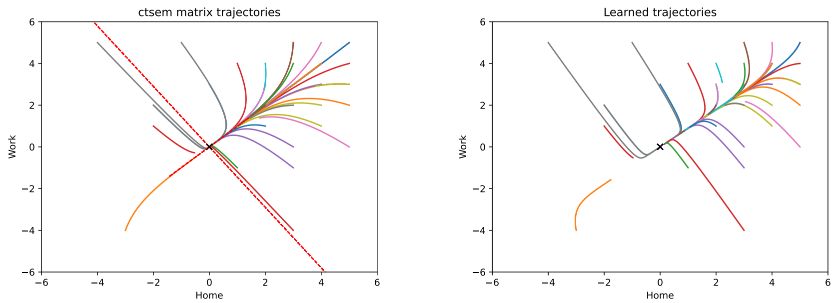

Fig. 6 Trajectories of a few randomly selected individuals in the panel study under the linear differential equation (3) for $A = A_{\text{ctsem}}$ (left) and $A = A_{\text{learn}}$ (right). For $A = A_{\text{ctsem}}$, the dashed red lines indicate the eigenvectors $v(A_{\text{ctsem}})$ scaled by their respective inverted eigenvalue.

Appendix

`ctsem` settings

The following `R` snippet was used to perform the `ctsem` model estimation:

Program Code

```
data <- read.csv("soep_panel_file.csv")

data_selection <- data.frame(
    data$pid,
    data$syear,
    data$plh0171,
    data$plh0173
)

names(data_selection) <- c("pid","syear","hsat","wsat")

library("ctsem")
library("ctsemOMX")

wide <- ctLongToWide(
    datalong = data_selection,
    id = "pid",
    time = "syear",
    manifestNames = c("hsat", "wsat")
)

wide_int <- ctIntervalise(
    datawide = wide,
    Tpoints = 37,
    n.manifest = 2,
    manifestNames = c("hsat", "wsat"),
    individualRelativeTime = TRUE,
    imputedefs = FALSE
)

model <- ctModel(
```

```
    type = 'omx',
    n.latent = 2,
    n.manifest = 2,
    Tpoints = 37,
    manifestNames = c("hsat", "wsat"),
    latentNames = c("hsat", "wsat"),
    LAMBDA = diag(2),
    MANIFESTMEANS = matrix(data=0, nrow=2, ncol=1),
    MANIFESTVAR = matrix(data=0, nrow=2, ncol=2),
    DRIFT ="auto",
    CINT = matrix(data=0, byrow=T, nrow=2, ncol=1),
    DIFFUSION = "auto",
    TRAITVAR = NULL,
    T0TRAITEFFECT = NULL,
    MANIFESTTRAITVAR = NULL,
    startValues = NULL
)

set.seed(1)

model_fit <- ctFit(
    dat = wide_int,
    ctmodelobj = model,
    transformedParams = T
)

summary(model_fit, verbose=T)
```

Reformulation of the variational stochastic approach

Let $\bar{\mathbf{x}}(t)$ be the solution to the linear stochastic differential equation (2) with initial value $\bar{\mathbf{x}}(t_0) = x_0$. Then $\bar{\mathbf{x}}(t)$ is given by ([3, Cor. 8.2.4])

$$\bar{\mathbf{x}}(t) = e^{A(t-t_0)}x_0 + \int_{t_0}^{t} e^{A(t-s)}b\,\mathrm{d}s + \int_{t_0}^{t} e^{A(t-s)}G\,\mathrm{d}W(s). \tag{9}$$

Set further

$$\mathbf{x}(t) = e^{A(t-t_0)}x_0 + \int_{t_0}^{t} e^{A(t-s)}b\,\mathrm{d}s,$$

which is the solution to the non-stochastic linear differential equation (6) with initial value $\mathbf{x}(t_0) = x_0$ and with $\kappa^j = 0$, and let some time $T > t_0$ and datum x_T be given. Then, expanding the square,

$$\begin{aligned}\left\|\bar{\mathbf{x}}(T) - x_T\right\|_2^2 = \left\|\mathbf{x}(T) - x_T\right\|_2^2 + 2\Big(\mathbf{x}(T) - x_T\,, \int_{t_0}^{T} e^{A(T-s)}G\,\mathrm{d}W(s)\Big)\\ + \left\|\int_{t_0}^{T} e^{A(T-s)}G\,\mathrm{d}W(s)\right\|_2^2,\end{aligned}$$

where in the second term on the right, we mean the inner product. We next consider the expectation of the foregoing equation. From [3, Thm. 4.4.14], the expectation of the stochastic Itô integral in (9) is zero, that is,

$$\mathbb{E}\left[\int_{t_0}^{T} e^{A(T-s)} G \,\mathrm{d}W(s)\right] = 0,$$

and we have the *Itô isometry*

$$\mathbb{E}\left\|\int_{t_0}^{T} e^{A(T-s)} G \,\mathrm{d}W(s)\right\|_2^2 = \int_{t_0}^{T} \mathbb{E}\left\|e^{A(T-s)} G\right\|_{\mathrm{Fr}}^2 \mathrm{d}s = \int_{t_0}^{T} \left\|e^{A(T-s)} G\right\|_{\mathrm{Fr}}^2 \mathrm{d}s.$$

Accordingly,

$$\mathbb{E}\left\|\bar{\mathbf{x}}(T) - x_T\right\|_2^2 = \left\|\mathbf{x}(T) - x_T\right\|_2^2 + \int_{t_0}^{T} \left\|e^{A(T-s)} G\right\|_{\mathrm{Fr}}^2 \mathrm{d}s.$$

Every summand in (7) is precisely of the foregoing form, so the last identity yields (8).

References

1. Asch, M., Bocquet, M., Nodet, M. (2016) *Data Assimilation: Methods, Algorithms, and Applications.* SIAM, Philadelphia, PA, USA. doi:10.1137/1.9781611974546
2. Arnold, M., Cáncer, P.F., Estrada, E., & Voelkle, M. (2023). Score-Guided Recursive Partitioning of Continuous-Time Structural Equation Models.
3. Arnold, L. (1974) *Stochastic Differential Equations: Theory and Applications.* Dover Publications.
4. Angraini, Y., Toharudin, T., Folmer, H., & Oud, J. H. (2014). The relationships between individualism, nationalism, ethnocentrism, and authoritarianism in Flanders: A continuous time-structural equation modeling approach. Multivariate Behavioral Research, 49, 41–53. doi:10.1080/00273171.2013.836621
5. Boker, S. M. (2002). Consequences of continuity: The hunt for intrinsic properties within parameters of dynamics in psychological processes. Multivariate Behavioral Research, 37, 405–422. doi:10.1207/S15327906MBR3703_5
6. Dormann, C., Guthier, C., & Cortina, J. M. (2019). Introducing continuous time meta-analysis (CoTiMA). Organizational Research Methods, 23(4), 620–650. doi:10.1177/1094428119847277
7. Driver, C. C., Oud, J., & Voelkle, M. (2017). Continuous time structural equation modeling with `R` package `ctsem`. Journal of Statistical Software, 77, 1–35. doi:10.18637/jss.v077.i05
8. Driver C. C., Voelkle, M. C. (2018). Hierarchical Bayesian Continuous Time Dynamic Modeling. Psychological Methods, 23(4), 774–799. doi:10.1037/met0000168.
9. Faragher, E. B., Cass, M., & Cooper, C. L. (2013). The relationship between job satisfaction and health: a meta-analysis. In: Cooper, C. L. (eds), *From Stress to Wellbeing Volume 1* (pp. 254–271). Palgrave Macmillan, London. doi:10.1057/9781137310651_12
10. Meinlschmidt, H. (2023) github.com/hannesmeinlschmidt/cautious-carnival
11. Judge, T. A., & Watanabe, S. (1993). Another look at the job satisfaction-life satisfaction relationship. Journal of Applied Psychology, 78, 939–948. doi:10.1037/0021-9010.78.6.939

12. Kuiper, R. M., & Ryan, O. (2018). Drawing conclusions from cross-lagged relationships: Re-considering the role of the time-interval. Structural Equation Modeling, 25, 809–823. doi:10.1080/10705511.2018.1431046
13. Neale, M. C., Hunter, M. D., Pritikin, J. N. *et al.* (2016). OpenMx 2.0: Extended structural equation and statistical modeling. Psychometrika, 81, 535–549. doi:10.1007/s11336-014-9435-8
14. Orzek, J. H., Arnold, M., & Voelkle, M. C. (2023). Striving for Sparsity: On Exact and Approximate Solutions in Regularized Structural Equation Models. Structural Equation Modeling: A Multidisciplinary Journal, 1-18. doi:10.1080/10705511.2023.2189070
15. Orzek, J. H., & Voelkle, M. C. (2023). Regularized continuous time structural equation models: A network perspective. Psychological Methods. doi:10.1037/met0000550
16. Oud, J. H. L. (2002). Continuous time modeling of the cross-lagged panel design. Kwantitatieve Methoden, 23, 1–26.
17. 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. H. L. Oud, & A. Satorra (Eds.), *Longitudinal research with latent variables* (pp. 201—244). New York, NY: Springer. doi:10.1007/978-3-642-11760-2_7
18. 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
19. 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
20. 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
21. R Core Team (2022). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/
22. Ryan, O., Kuiper, R. M., & Hamaker, E. L. (2018). A continuous time approach to intensive longitudinal data: What, why and how? In K. van Montfort, J. H. L. Oud, & M. C. Voelkle (Eds.), *Continuous time modeling in the behavioral and related sciences* (pp. 27–54). New York: Springer. doi:10.1007/978-3-319-77219-6_2
23. Singer, H. (1990). *Parameterschätzung in zeitkontinuerlichen dynamischen Systemen.* Konstanzer Dissertationen–282.
24. Voelkle, M. C., Gische, C., Driver, C.C., & Lindenberger, U. (2018). The role of time in the quest for understanding psychological mechanisms. Multivariate Behavioral Research, 53, 782–805, doi:10.1080/00273171.2018.1496813
25. 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(2), 176–192. Erratum in: Psychological Methods (2012) 17(3), 384. doi:10.1037/a0027543
26. Voelkle, M. C., & Oud, J. H. L. (2013). Continuous time modelling with individually varying time intervals for oscillating and non-oscillating processes. British Journal of Mathematical and Statistical Psychology, 66, 103–126. doi:10.1111/j.2044-8317.2012.02043.x
27. Sozio-oekonomisches Panel (SOEP), Daten der Jahre 1984–2018. doi:10.5684/soep-core.v35