

Time series analysis of intensive longitudinal data in psychosomatic research: A methodological overview

Sigert Ariens^{*,1}, Eva Ceulemans^{*}, and Janne K. Adolf^{*}

^{*}KU Leuven, Quantitative Psychology and Individual Differences, Faculty of Psychology and Educational Sciences, Leuven, 3000, Belgium

¹sigert.ariens@kuleuven.be

ABSTRACT

Time-series analysis (TSA) of intensive longitudinal data (ILD) provides the psychological literature with a powerful tool for assessing how psychological processes evolve through time. Recent applications in the field of psychosomatic research have provided insights in the dynamical nature of the relationship between somatic symptoms, physiological measures, and emotional states. These promising results highlight the intrinsic value of employing TSA, although application comes with some important challenges. This paper aims to present an approachable, non-technical overview of the state of the art on these challenges and the solutions that have been proposed, with emphasis on application towards psychosomatic hypotheses. Specifically, we elaborate on issues related to measurement intervals, the number and nature of the variables used in the analysis, modeling stable and changing processes, concurrent relationships, and extending TSA to incorporate the data of multiple individuals. We also briefly discuss some general modeling issues, such as lag-specification, sample size and time series length, and the role of measurement errors in the discussed models. We hope to arm applied researchers with an overview from which to select appropriate techniques from the ever growing variety of TSA approaches.

Keywords: Time-series analysis, Intensive longitudinal data, Vector autoregressive modeling

1 INTRODUCTION

Although psychological and somatic variables are often closely related within individuals (e.g. Balon, 2006; Jensen, Hoffman, & Cardenas, 2005; Tunks, Crook, & Weir, 2008), reviews of the literature aimed at clinicians almost exclusively discuss between-person research (El Sount et al., 2019; Sharp & Keefe, 2006). This discrepancy is problematic since findings between individuals, such as correlations computed across individuals, are generally not informative of how phenomena/processes evolve over time within individuals (e.g. Hamaker, 2012; Molenaar, 2004; Molenaar & Campbell, 2009). Moreover, recent research assessing these individual-level processes have highlighted that their precise nature can display considerable heterogeneity over individual cases (Frumkin, Haroutounian, & Rodebaugh, 2020; Kelly, Weigard, & Beltz, 2020; van Giels, Emerencia, Bos, & Rosmalen, 2020; Zink et al., 2020).

In the domain of emotion research, the between vs within individuals debate has prompted a paradigm shift from between-subjects analysis at single, or limited time points, to investigating dynamical relationships between affective processes by employing intensive longitudinal sampling schemes, where dozens to thousands of data are collected over time, within individuals (e.g. Kuppens, 2015). When subsequently analysing this intensive longitudinal data (ILD) with statistical tools sensitive towards the serial dependence assumed to underlie the time series of data, a rich description of the temporal unfolding of emotional processes becomes available.

The field of psychosomatic research also appears particularly well-suited to embrace this novel

paradigm and its tools: Many hypotheses in the domain imply some form of dynamics or temporal profile which could be assessed rigorously when appropriate study designs and statistical methodologies are used. For example, Ghiggia et al. (2017) hypothesize that the prevalence of psychological distress in patients suffering from fibromyalgia is the result of a vicious cycle, or dynamical relationship, between psychological symptoms and psychosomatic manifestations of chronic pain. von Leupoldt, Riedel, and Dahme (2006) call for an assessment of the relationship between asthma exacerbations and affective states in naturalistic settings. Researchers have also emphasized the need to assess the time course of intervention and treatment strategies (Wild et al., 2010), an example of which is the temporal profile of central sensitization processes in relation to different treatment and intervention strategies for chronic pain (McGreevy, Bottros, & Raja, 2011). Here, a dynamic paradigm would provide concrete ways to assess patient-tailored approaches to therapy and intervention (e.g. Rosmalen, Wenting, Roest, de Jonge, & Bos, 2012), often called for by researchers in the field.

As many others (e.g. Kucyi & Davis, 2015; Rosmalen et al., 2012; van der Krieke et al., 2017), we believe that time is ripe for committing to a truly dynamic psychosomatic research agenda. However, adequately applying the dynamic paradigm and its methods requires informed decisions on challenges related to data analysis. Therefore, this paper offers an approachable overview of five core challenges and the solutions that have been proposed in the methodological literature.

The structure of this paper is as follows. In the first section, properties of ILD will be discussed. The second section will serve as a didactic introduction in the dominant modeling techniques for time series analysis of ILD. In the third section, we elaborate on five identified topics, the proposed solutions, and provide directions for implementation. Specifically, we will discuss issues related to measurement intervals, the variables used in analysis, modeling stable and changing processes, concurrent relationships, and extending TSA to incorporate the data of multiple individuals. The discussion consists of a summary of the five discussed methodological problems and solutions, and an elaboration on some additional issues regarding model specification and data requirements.

2 INGREDIENTS OF THE DYNAMIC PARADIGM

2.1 Intensive longitudinal data

ILD consist of a large amount of repeated measures within individuals over time. They are most often obtained by means of repeated self-reports in ecologically valid contexts, often termed experience sampling (ESM), ecological momentary assessment (EMA), or real time data collection (RTDC) (Chun, 2016; Ebner-Priemer & Trull, 2009; Nehrkorn-Bailey, Reardon, & Hicks Patrick, 2018; Scollon, Prieto, & Diener, 2009; Stone & Broderick, 2007; Trull & Ebner-Priemer, 2009). However, ILD can also be collected in experimental studies (e.g. Neubauer, Lerche, & Voss, 2018), and may also involve other types of data than self-reports, such as physiological (Nakamura, Kiyono, Wendt, Abry, & Yamamoto, 2016) or observational (Henrie, Bodily, Manwaring, & Graham, 2015) data. The distinction between ILD and classical longitudinal designs is a matter of degree. Classical longitudinal designs have a limited number of measurement occasions, providing only a rough assessment of time dependent processes (Molenaar, De Gooijer, & Schmitz, 1992), while ILD emphasizes a large number of repeated measures per individual (Schafer, 2006), allowing dynamical relationships to be investigated in far more detail. Finally, the choice of which variables are included in an ILD study should primarily rest on theory and hypotheses about the processes under investigation.

As with more traditional designs, study design and data-analytic considerations are often closely related. This paper is primarily intended to elucidate data-analytic issues when conducting ILD studies, but we refer readers to Bolger, Davis, and Rafaeli (2003), Burton, Weller, and Sharpe (2007), Collins (2006), Ebner-Priemer and Trull (2009), Trull and Ebner-Priemer (2020) for an overview of and guidelines for design-related issues. To illustrate how closely design and analysis are related we discuss a concrete example. Chronic pain in individuals suffering from spinal chord injury has been

found to remain relatively stable over time (Jensen et al., 2005). Over minutes however, the level of experienced chronic pain may fluctuate around a stable mean in tandem with fluctuating attention towards nociceptive stimuli (Kucyi & Davis, 2015). Investigating whether the fluctuations are centered around a relatively stable mean, or displays trends or cycles, requires models which translate these theoretical assumptions, and to study such rapidly fluctuating processes demands measurement on a fine temporal scale. While ESM designs may be well suited to study the time course of chronic pain following spinal chord injury over days or weeks, experimental designs may excel at revealing the dynamics between chronic pain and attention, allowing the inclusion of neuro-imaging techniques and explicit attentional manipulations. Other variables included in the study can be considered stable, fixed effects, such as the site at which pain is experienced, or dynamic, such as attention or psychological covariates such as pain catastrophizing.

2.2 Time series analysis

Many time series methods have in common the assumption that multiple variables display (mutual) dependence through time. That is, a variable X_1 at time t can to some extent be predicted by that same variable at time $t - 1$, or by another variable X_2 at time $t - 1$. For example, rumination at time $t - 1$ may be predictive of itself and future social anxiety at time t . In turn, also social anxiety may affect itself, and rumination over time. Often, such relationships are of prime interest from a theoretical perspective, yet many researchers opt to apply classical analysis approaches which discount the very temporal dependence which endows ILD with its appeal (Windt et al., 2018).

By far the most prevalent methods used to capture these temporal dependencies are those belonging to the family of *vector autoregressive models* (VAR; see for instance Hamilton, 1994; Lütkepohl, 2013, for more technical detail). They do not only match our above example on rumination and anxiety, but also form the basis of many other methods.

VAR models are extensions of univariate autoregressive (AR) models to the multivariate case. In a general form, the univariate AR model is given by:

$$y_t = \delta + \sum_{k=1}^K \phi_k y_{t-k} + \varepsilon_t \quad (1)$$

The variable y_t thus depends on the K previous variables $y_{t-1}, y_{t-2}, \dots, y_{t-K}$ via the autoregressive effects $\phi_1, \phi_2, \dots, \phi_K$. K thus reflects the number of lags, or the order of the formalized AR process, and the associated model is denoted an AR(K) model. Note that this model could also be conceived as a special variant of a conventional multiple regression model. In the case of a single lag, we get an AR(1) model, which can be written as:

$$y_t = \delta + \phi y_{t-1} + \varepsilon_t \quad (2)$$

An AR process is assumed to fluctuate around its mean value, which depends on the intercept term δ . The fluctuations originate from the residual or error term ε_t , which follows a zero-mean normal distribution with variance σ_ε^2 , and does not display any dependencies over time.

In the time series literature, this residual term is called innovation, reflecting the notion that ε_t provides a perturbation or shock to the system at every time t . In the psychological literature, innovations are often assumed to reflect a collection of internal or external events influencing the system at a given time. In practice, these could be all sorts of situational factors that influence psychosomatic processes, such as various social interactions or stressful events, but also thoughts and appraisals (Hamaker, 2012). Perturbations due to the innovations are then carried forward in the process of interest, to the extent the auto-regressive relationships are different from zero. This property distinguishes innovations from

conventional error terms such as measurement error, the impact of which is assumed to be restricted to individual time points (Schuurman, Houtveen, & Hamaker, 2015). The amount of overall fluctuations characterizing the process thus results from the residual variation, and the autoregressive effects.

At this point, we briefly mention autoregressive moving average (ARMA; for a more comprehensive overview of these classes of models, see Hamaker & Dolan, 2009; Jebb, Tay, Wang, & Huang, 2015; Scargle, 1982; Torre, Delignieres, & Lemoine, 2007) as an extension of AR models. ARMA models regress a process variable not only on lagged versions of itself, but also on lagged versions of the innovation term. We will come back to them in later parts of the paper.

Extending the univariate AR model to the multivariate case is possible by using vector autoregressive models (VAR). Rather than single variables, VAR models feature a *set* of variables that are regressed on themselves and each other over time, yielding auto- and crossregressive effects. AR models can thus be seen as special cases of VAR models, where only single variables are modeled, and we again would like to highlight the parallel with conventional multiple regression models.

The order of the model again denotes the number of lags that are to be taken into account. For illustrative purposes, we discuss an example of a bivariate VAR(1) model, linking two processes with a single lag:

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}}_{\mathbf{y}_t} = \underbrace{\begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}}_{\boldsymbol{\delta}} + \underbrace{\begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}}_{\boldsymbol{\Phi}} \underbrace{\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}}_{\mathbf{y}_{t-1}} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}}_{\boldsymbol{\varepsilon}_t} \quad (3)$$

with

$$\underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}}_{\boldsymbol{\varepsilon}_t} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \underbrace{\begin{bmatrix} \sigma_{\varepsilon_1}^2 & \sigma_{\varepsilon_1 \varepsilon_2} \\ \sigma_{\varepsilon_2 \varepsilon_1} & \sigma_{\varepsilon_2}^2 \end{bmatrix}}_{\boldsymbol{\Sigma}_\varepsilon} \right)$$

The process variables, intercepts, and innovations are contained in 2×1 vectors $\mathbf{y}_t / \mathbf{y}_{t-1}$, $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}_t$. The auto- and crossregressive effects reside in the diagonals and off-diagonals of the 2×2 transition matrix $\boldsymbol{\Phi}$ respectively. The innovations affecting each process are assumed to be drawn from a zero-mean Gaussian distribution, with zero-mean vector and covariance matrix $\boldsymbol{\Sigma}_\varepsilon$. If the off-diagonal elements of the innovation covariance matrix, $\sigma_{\varepsilon_1 \varepsilon_2}$ and $\sigma_{\varepsilon_2 \varepsilon_1}$, take on a non-zero value, innovations are correlated, reflecting a contemporaneous relationship between the two processes.

Different parameter values naturally imply different process trajectories, and Figure 1 illustrates the bivariate VAR(1) model and its behaviour in more detail. It contains a path diagram of the model (Panel B), and displays - for different parameter values - model-implied process trajectories (Panel C), and impulse response functions, i.e., how the VAR process would react to a single controlled impulse given to either of the two processes (Panel D).

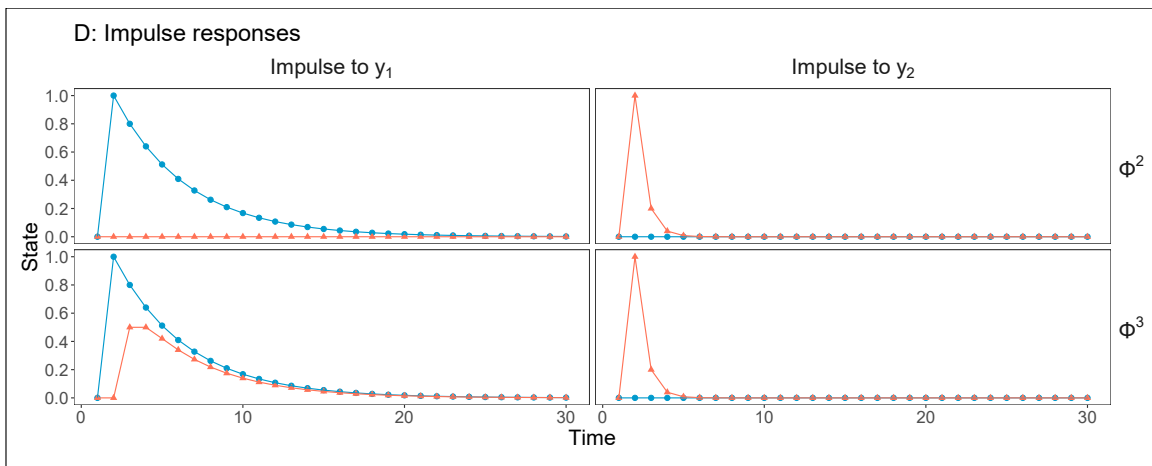
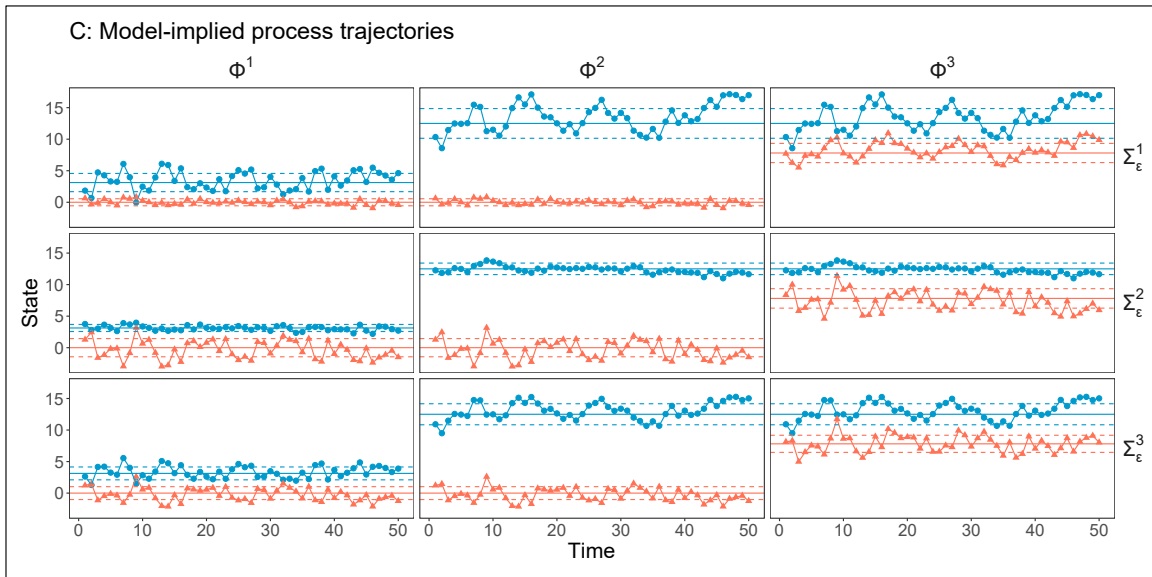
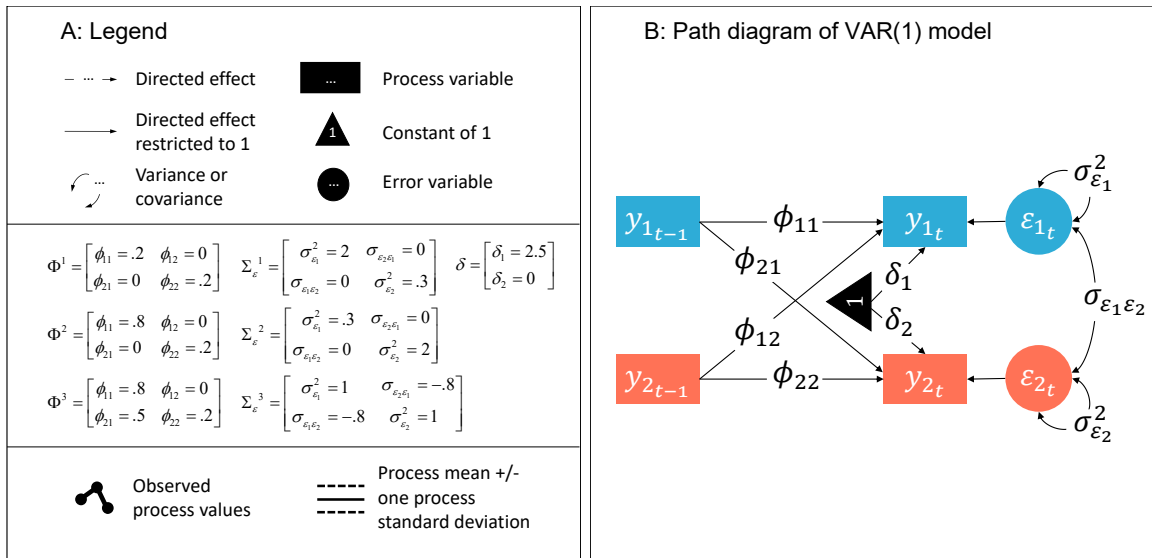


Figure 1. Setup and behavior of a bivariate VAR(1) model. Panel A holds the legend to read the figure, Panel B displays time series data generated from different model variants, and Panel C shows the model-implied impulse response functions. For the latter, we assume that a single unit perturbation at time $t = 2$ drives one of the two processes away from its mean. With regard to Panel C, one can observe how combinations of different parameter values imply (and would thus fit) different processes. For instance, as the autoregressive effect increases (left to center column), the time series produced by process y_1 becomes more structured over time, which manifests in a somewhat smoother trajectory with longer departures from the process mean. Introducing the crossregressive effect ϕ_{21} leads to y_2 at time t to be dependent on y_1 at time $t - 1$ (center to right column), with the resulting trajectory demonstrating clear cross-variable temporal dependence in that process two follows process one. The magnitude of the innovation variance (co-)determines how strongly the processes fluctuate around their mean levels, as can be seen directly by comparing the two process trajectories either within the top or the middle row of the left column. In addition, it can be seen that auto- and crossregressive effects contribute not only to the temporal pattern, but also the amount of overall fluctuations (e.g., process one in the top row of the left vs. middle column for the autoregressive effect, process two in the top row of the middle vs. right column for the crossregressive effect). When innovation covariances are introduced, the time series display contemporaneous relationships. This can best be seen in the absence of crossregressive effects (bottom row, left column). In this case the correlation is rather strong and negative, resulting in the two process trajectories almost being mirror images of each other. When both crossregressive and contemporaneous effects are included (bottom row, right column), both temporal and instantaneous relationships are blended together and thus harder to see in the resulting trajectory. The impulse response functions (Panel D) illustrate differences in dynamics more clearly. The impulse leads to a sudden increase to a level of one in the process it is administered to. Following that, we see mean-reverting values, with the return-rate being inversely related to the strength of the AR effect (left vs. right column). In addition, if present, cross-regressive effects can be seen to take effect over time (left column, bottom row). Note that it takes some time until the effect unfolds and the impulse fully reaches process two.

VAR models can be implemented and estimated in a variety of modeling frameworks, ranging from ordinary regression modeling (via multiple regression models), to complex multivariate modeling frameworks such as structural equation modeling (SEM; Hamaker, Dolan, & Molenaar, 2003; Voelkle, Oud, von Oertzen, & Lindenberger, 2012), state-space modeling (SSM; Chow, Ho, Hamaker, & Dolan, 2010; Harvey, 1989; Hunter, 2018), or Bayesian formulations (e.g., Asparouhov, Hamaker, & Muthén, 2018; Driver & Voelkle, 2018a). These different modeling and estimation frameworks all come with their own specifics (e.g., the possibility to evoke common and unique latent variables in SEM and SSM, the flexibility of distributional assumptions in Bayesian modeling), and we will in part revisit them in the following.

A type of causal reasoning invoked for inference in VAR models is Granger causality (see for instance Schuurman, Ferrer, de Boer-Sonnenschein, & Hamaker, 2016). Briefly, Granger causality is concerned with the crossregressive effects, and y_1 Granger-causes y_2 if using past values of y_1 to predict current values of y_2 improves predictive accuracy over just using past values of y_2 to predict current values of y_2 . However, Granger-causal inference requires in principle that one controls for, and thus includes in the model, any possible covarying processes (Hoover, 2005). We return to the problem of causal inference in the final discussion.

Finally, the models discussed so far assume stationarity. Throughout this paper, we use stationary in the sense of weakly stationary (Scargle, 1982), meaning that the mean, variance, and autocorrelations of the process under study do not change as a function of time. An important condition for VAR models to be stationary is that the temporal dependencies as captured by the auto- and crossregressive effects are not too strong, so that the perturbations due to the innovations "die out" over time, rather than accumulate (cf. Hamilton, 1994, p. 54). For an AR(1) model, for instance, this condition is given by $|\phi| < 1$.

3 PROBLEMS AND SOLUTIONS IN TIME SERIES ANALYSIS OF INTENSE LONGITUDINAL DATA

The VAR model has a number of important assumptions and implications which often do not hold for specific data sets, or do not correspond to the hypotheses of interest. Fortunately, many extensions have

been proposed to accommodate researchers in handling these challenges. We will primarily focus on five specific issues and associated solutions throughout the body of the text.

Firstly, conventional VAR models yield results that are dependent on the time interval between ILD measurements, and furthermore assume that the process under investigation is measured at equidistant intervals. One must therefore ensure that time interval information is taken into account when comparing VAR modeling results from multiple studies. Moreover, within studies, the time intervals between observations should in principle be equal across observations, although this is often difficult or impossible to impose in practice. The issue here is thus how to handle between and within studies differences in measurement intervals. *Secondly* in many cases a relatively large number of variables are included in a study, and preferably jointly modeled to reveal dynamical relationships. A critical advantage of VAR models is the ability to model the joint dynamics of multiple variables, yet issues of multicollinearity and over-fitting are an inevitable consequence in practical application. *Thirdly*, VAR models are appropriate when modeling stationary processes. Change however is often of critical interest in psychological research. In psychosomatic applications for example, effective intervention strategies imply some form of change in the trajectories of the processes of interest. We will discuss different options for studying such changes in processes. *A fourth issue* pertains to the emphasis on lagged relationships in VAR. Researchers may also have specific hypotheses about contemporaneous relationships between variables, yet taking these into account using traditional VAR frameworks is not straightforward. *Finally*, while TSA has traditionally been employed to investigate processes within single subjects, between-subjects analyses are often of key interest to psychological researchers. Multiple extensions for between-subjects analysis have been developed, each imposing different assumptions concerning the nature of these between-subject differences.

3.1 Handling varying measurement intervals between and within studies

ILD can be collected at different time scales, and hence with different time intervals between measurements. The sampling frequency of time series data should foremost depend on the frequency at which the underlying process is assumed to change (e.g. Bolger et al., 2003; Hamaker & Wichers, 2017; Trull, Lane, Koval, & Ebner-Priemer, 2015; Voelkle & Oud, 2013). For example, processes changing slowly over time, such as chronic pain, can be sampled at relatively large intervals. Others, such as mood swings and affective states, presumably vary more frequently and should be measured at shorter intervals. But additional theoretical and pragmatic considerations also influence the sampling scheme. For instance and trivially, participants of an ESM study need to sleep, necessitating night breaks. Also, intervals are often randomized, and hence vary in length during the day, to avoid expectancy effects. The decision about a sampling scheme is thus a complex one, to be made during the design phase of an ILD study. We come back to it here, because it has profound implications for data analysis, in the sense that both measurement intervals varying between studies and measurement intervals varying within studies can affect VAR modeling results, and can thus require data analytic responses.

We first consider the case of varying measurement intervals between studies. This becomes an issue if the same or similar phenomena are assessed across multiple studies, and modeling results are to be compared. The reason for this is that VAR models are of recursive nature, predicting present from past states, and the formalized relationships vary as a function of the interval between time points (and hence between states; Oud & Delsing, 2010). This is especially important in the case of auto- and crossregressive relationships, as these are often of prime interest: While auto-regressive effects normally show an exponential decay with increasing interval length, crossregressive effects are close to zero if the interval is close to zero, reach their maximum or minimum (or their maxima and minima, in case of more than two processes) at some larger interval, and then diminish again. The latter pattern reflects that it takes time for one variable to exert its influence on another (e.g., Deboeck & Preacher, 2016). In consequence, comparing the dynamics of one VAR process assessed at different time intervals naively (i.e., without taking time interval-information into account), can lead to confusing, and spuriously

inconsistent results. See Dormann and Griffin (2015), Kuiper and Ryan (2018), Oud and Delsing (2010), Voelkle, Gische, Driver, and Lindenberger (2018) for more detailed illustrations. Note, again, that the time interval-dependency of VAR models is not some sort of anomaly, but simply an inherent feature of such model structures. Since it is (in principle) known, it can be accounted for.

Before we elaborate on proposed solutions, we turn to measurement intervals varying within a given study. Also here, the specificity of auto- and crossregressive parameters to the interval over which they are estimated holds. A core assumption of (conventional) VAR models discussed so far is that data constituting the time series are sampled at equidistant intervals, and neglecting this assumption will bias parameter estimates (e.g., de Haan-Rietdijk, Voelkle, Keijsers, & Hamaker, 2017). Note that this is also the case if measurement intervals are varied randomly around some average interval, since the dependence of the VAR parameters on the time interval is highly non-linear, and aggregation is a linear operation.

A final issue is that researchers might find it difficult to formulate a process model in terms of discrete time points. The reason for this is that processes are typically assumed to be continuous phenomena, which should not be specific to a given time interval.

3.1.1 Proposed solutions

As mentioned above, there are well justified reasons for allowing variability in ILD sampling schemes, and therefore, it is no surprise that recent developments in dynamic modeling have sought to address this issue.

3.1.1.1 Continuous-time modeling

A general approach involves dynamic models that are formulated in continuous time (e.g., Boker, 2012; Chow, 2019; Oravecz, Tuerlinckx, & Vandekerckhove, 2009; Ou, Hunter, & Chow, 2017; Oud & Delsing, 2010; Ryan, Kuiper, & Hamaker, 2018; Singer, 2012; Voelkle, Oud, Davidov, & Schmidt, 2012). Unlike conventional discrete-time models that formalize change between discrete time points, continuous-time models formalize change in the process(es) of interest as change at any given time point. In more technical terms, continuous-time models feature differential equations, as compared to difference equations which underlie discrete-time models.

As such, continuous-time models are independent of a particular sampling interval (Boker, Tiberio, & Moulder, 2018; Ryan et al., 2018; Voelkle, Oud, Davidov, & Schmidt, 2012). Consequently, they readily handle unequal sampling intervals within and between studies, and corresponding discrete-time effects can be derived for any interval of interest. In addition, they offer a conceptually sound account of psychosomatic processes as continuously ongoing and smoothly changing (Boker, 2012; Deboeck, Nicholson, Kouros, Little, & Garber, 2015; Driver & Voelkle, 2018a; Karch, 2016). This not only allows for a direct mapping between (most) theory and model, but also opens up new ways of thinking about prominent problems such as effects of interventions (Driver & Voelkle, 2018b), mediation (Deboeck & Preacher, 2016), or optimal sampling rates (Adolf, Loossens, Tuerlinckx, & Ceulemans, 2019).

Continuous-time approaches are becoming more and more popular, resulting in increasing numbers of accessible didactic treatments (e.g., Ryan et al., 2018; Voelkle et al., 2018) and software packages aimed at facilitating their use (e.g., Driver, Oud, & Voelkle, 2017; Ou et al., 2017). It should however not be forgotten, that estimation of such models can be computationally intense to problematic, especially if more complex model variants are fit (Voelkle et al., 2018). Also, models are less straightforward than their discrete time counterparts on a technical level.

3.1.1.2 Discrete-time approaches

In addition to continuous-time modeling, there are also possible alternative strategies which can be employed while remaining in the discrete-time modeling framework. For instance, one can artificially refine the time scale by including missing values in the data, which are thus used to equalize the time intervals between observations up to a certain level of precision. On the modeling side, this then

requires the introduction of so-called phantom variables and/or the use of estimation techniques that handle missing values (e.g., Adolf, Voelkle, Brose, & Schmiedek, 2017; Asparouhov et al., 2018; Rindskopf, 1984).

The applicability of this solution depends on properties of the data. If measurement intervals show large variations, the percentage of missing values in a data set can rise considerably, rendering the approach infeasible (de Haan-Rietdijk et al., 2017; Oud & Voelkle, 2014). If the latter is the case, or if phantom variables and/or missing values cannot be handled by the preferred estimation framework, then the imputation of missing values, for instance through multiple imputation techniques (Van Buuren, 2018), can be an option.

3.2 Handling multiple possibly correlated variables

A single process can be indicated by a single, or multiple observed variables, and in addition multiple processes may mutually interact over time to form an intricate dynamic relationship. As mentioned previously, the set of variables included in analysis depends primarily on theory and hypotheses about the processes under study, but the interpretability of model solutions may become cumbersome as the number of variables included increases.

Specifically, including many variables may lead to three problems. Firstly, interpreting VAR coefficients becomes problematic when large numbers of variables are employed, because the auto- and crossregressive effects (and regression weights in general) pertain to the unique direct effects of specific variables on top of all the other variables, and do not account for shared effects (Bulteel, Tuerlinckx, Brose, & Ceulemans, 2016b; Cohen, West, & Aiken, 2014). Secondly, if variables are highly correlated over time, one may run into multicollinearity issues which lead to unstable VAR estimates (i.e., the bouncing beta problem in regression analysis, Cohen et al., 2014; Kiers & Smilde, 2007). Thirdly, VAR models include a considerable number of parameters, and this number exponentially increases with each additional variable included in the model (i.e., in case of J variables, J^2 lagged effects have to be estimated). Such large numbers of parameters render VAR estimation vulnerable to overfitting, leading to poor generalisability of model results (Babyak, 2004; McNeish, 2015), as recently demonstrated by means of cross-validation (Bulteel, Tuerlinckx, Brose, & Ceulemans, 2018). Importantly, insufficient numbers of observations can lead cross-validation techniques to prefer an AR model even when the process is actually generated from a VAR model, illustrating how prone VAR models are to overfitting, and emphasizing the importance of measuring the processes at a sufficiently high number of time points for highly parameterized models (Bulteel, Mestdagh, Tuerlinckx, & Ceulemans, 2018).

3.2.1 Proposed solutions

All three problems can be more or less solved by restricting the VAR model, so that less parameters have to be freely estimated and a sparser model is obtained. To achieve this, one can either collapse the variables into a smaller number of principal components or common factors and estimate the VAR model on the component/factor level, which vastly reduces the dimensionality of the VAR problem, or use techniques which limit complexity by forcing some estimates to zero. The former strategy uses dimension reduction methods, the latter uses regularization methods.

3.2.1.1 Dimension reduction techniques

To collapse the variables into components or factors one can either use a confirmatory approach or an exploratory one. So far, the confirmatory setting has received the most attention. In this setting, it is decided beforehand, usually based on theory, which variables are indicators of the same latent construct and therefore can be taken together in a factor. To this end, so-called dynamic factor analysis (DFA, Ram, Brose, Molenaar, et al., 2013) and related approaches have been put forward. These translate the typically cross-sectional (confirmatory) factor analysis framework to the case of time series data and combine it with VAR modeling. Associated software has become available (e.g., Asparouhov et al., 2018; Driver et al., 2017; Hunter, 2018). Moreover, some of these model combinations come with fit

indices that allow to assess whether the hypothesized structure in the variables seems to hold up. In the exploratory setting, where the variables are collapsed in a data-driven way, approaches were developed that build on principal component analysis (Jolliffe, Trendafilov, & Uddin, 2003). Here, PC-VAR is a recent and intuitive proposition, that consist of two steps: First, perform principal component analysis on the variables and rotate them to simple structure, second, perform VAR on the rotated component scores (Bulteel, Tuerlinckx, et al., 2018). This technique has been demonstrated to outperform lasso-based VAR, a typical regularization method, with regard to predictive accuracy (Bulteel, Tuerlinckx, et al., 2018). In addition, several exploratory dynamic factor analysis approaches (Browne & Nesselroade, 2005; Molenaar et al., 1992) have been proposed.

3.2.1.2 Regularization methods

Regularization methods solve the above mentioned problems by penalizing properties of the estimated VAR parameters. A tuning or regularization parameter indicates how strongly this penalty is imposed when fitting the model. The most popular penalty in VAR analysis is the lasso penalty (Hsu, Hung, & Chang, 2008) which equals the sum of the absolute values of the regression weights (Epskamp & Fried, 2018; Wild et al., 2010). The higher the value of the tuning parameter, the more estimates will be forced to zero, yielding a sparser model. Whereas this approach helps to minimize overfitting, it does not fully solve the interpretation and collinearity issues. Specifically, if two very strongly correlated variables are included, the lasso penalty will force one of the associated effects to zero. Yet, as demonstrated by Bulteel, Tuerlinckx, et al. (2018) and discussed in general by Bondell and Reich (2008), Tibshirani (2011), Zou and Hastie (2005), which of both parameters is set to zero is largely arbitrary.

3.3 Modeling changing processes

A core assumption of conventional VAR models, and other models using autoregressive structures to quantify temporal effects (e.g. DFA), is that the processes under investigation are stationary. That means that their means, (co-)variances and auto-covariances do not change over time, thus, how such processes evolve over time is subject to stable characteristics. However, change - also in how phenomena evolve - is often of key interest in psychopathology and psychology in general (e.g., Molenaar, 2004; Scholz, 2019; van de Leemput et al., 2014). How do individuals, for example, develop a depressive episode, how do they achieve remission, how do treatments affect the dynamics of the system? To answer these questions, statistical approaches are needed that make it possible to assess time-dependent processes, whose characteristics change.

3.3.1 Proposed solutions

The models discussed in this section thus deal with non-stationary time series, where change rather than stability of process characteristics is of interest. We present techniques as falling into three categories: Models that capture non-stationarities via time-varying parameters, integrated models, and data-driven approaches. This last category can be of interest for more explorative research questions, or for the pre-processing of time series data if (certain forms of) non-stationarity are not of substantive interest but should be corrected for. In addition, we offer some distinctions between models in how they handle non-stationarities, or assume these non-stationarities to arise. We chose for these distinctions in order to serve as a practical guide for researchers trying to select from the diverse range of TSA models for non-stationary data, but do not assume that this taxonomy is substantive, nor exhaustive. Specifically, we will remark on, first, the degree to which models treat the sources of non-stationarity as unknown vs known, or in other words, the degree to which they incorporate uncertainty about such sources, and, second, whether changes are incorporated as smooth and continuous or as discrete and abrupt (cf. Adolf et al., 2017).

Typically non-stationary models will include more parameters than their stationary counterparts to capture the more complex temporal patterns in the data. A noteworthy consequence is that this increased model complexity usually implies that more data is needed for estimation relative to stationary

processes.

3.3.1.1 VAR models with time-varying parameters

In psychological applications where changing processes are of interest, VAR models with time-varying parameters are widely used. A first popular modeling option is given by so-called regime-switching models (e.g., Hamaker & Grasman, 2012; Hamilton, 2010; Kim, Nelson, et al., 1999; Ou et al., 2017). As the name suggests, such models assume that the process of interest switches abruptly back and forth between a few regimes, during which it is governed by distinct parameters and hence displays distinct dynamics. Regime-switching may occur if people transition into and out of depressive episodes, or if they switch between different cognitive strategies when solving a problem (Hamaker, Grasman, & Kamphuis, 2010). The switching itself is usually governed by a Markov chain, a stochastic autoregressive-type process that captures temporal dependencies among discrete-valued variables, in this case among the regimes. Since the parameter changes are produced by a stochastic process, whose parameters - next to the regime-specific parameters - are unknown and estimated from the data, regime-switching models can handle quite some uncertainty about non-stationarity.

Alternative modeling approaches, also allowing parameter changes to arise from unknown stochastic processes, assume continuous-valued autoregressive-type parameter processes (e.g., Chow, Zu, Shifren, & Zhang, 2011; Molenaar, Sinclair, Rovine, Ram, & Corneal, 2009). As such, they can thus capture gradually - rather than suddenly - changing processes, which might be appropriate to formalize stress-dependent affect regulation (Chow et al., 2011) or long-term adjustments to major environmental changes (Boker, 2015). In addition, there are also approaches that incorporate smooth parameter changes drawing upon techniques from nonparametric statistics (Bringmann et al., 2017; Haslbeck, Bringmann, & Waldorp, 2017).

Quite a different approach is taken by fixed moderated VAR and related models (Adolf et al., 2017; Bringmann et al., 2013), which imply that we are more certain about parameter changes. Specifically, these models require that we know or hypothesize that specific time-varying covariates are associated with parameter changes in a VAR model. The parameter changes, and in particular their timing, are then determined via the observed changes in the covariates, while the amount of change is freely estimated. This applies to situations where an external event is likely to trigger a change in dynamics, such as major life events during and ESM protocol, the administration of medicine or treatments, or the experience of certain stimuli during an experimental procedure. Different shapes of change (e.g., gradual, abrupt, trending) can then flexibly be incorporated by using covariates of according format. Fixed moderated models are obviously more restrictive than the initially presented ones, but, if appropriate, they require less data for accurate model estimation. Note that it is of course also possible to include observed covariates into, for example, regime-switching models in order to aid the identification of parameter changes, but unlike in fixed moderated VAR, covariates are not needed for model estimation.

Finally, threshold AR models might be seen as bridging the gap between regime-switching models and VAR models with fixed (discrete) moderators, as both could be conceptualized as instances of this model family (see Tong, 2011).

3.3.1.2 Integrated models

We briefly introduced ARMA models in section 2, and discuss them in more detail here. When non-stationarities in the form of mean trends or cyclically recurring patterns are apparent in the data, an additional model parameter, d , is added to an ARMA model describing the order of differencing (difference scores between observations are used in the model, rather than the original scores) needed to result in a stationary series (e.g., Hamaker & Dolan, 2009). The resulting autoregressive *integrated* moving average (ARIMA) model thus, like an ARMA model, indexes short range AR and/or MA dependencies through the strength of these respective parameters. In addition, the value of the integrating parameter indicates the strength of trends in the data.

A more general variant is the autoregressive fractionally integrated moving average (ARFIMA) model,

which is identical to the ARIMA model save for the fact that the differencing parameter is allowed to take on non-integer values. The result is a very slow decrease in serial dependence over time, which is why the magnitude of this parameter is often interpreted to reflect long memory, or long term dependence (Torre et al., 2007; Wagenmakers, Farrell, & Ratcliff, 2004, 2005).

In experimental settings, ARFIMA models have been employed to quantify long term serial dependencies arising in cognitive tasks (Wagenmakers et al., 2004). Long term dependence in physiological time series such as gait variability (Delignières & Torre, 2009; Hunt, McGrath, & Stergiou, 2014), heartbeat rhythm (Peng et al., 1993), and respiration (Peng et al., 2002) have been considered indicative of healthy functioning, with deviations of these dynamics considered to be indicative of pathology. Different types of ARFIMA specifications correspond to different types of serial dependencies in the data and can thus be used to test hypotheses, with information criteria measures typically being employed for model selection (e.g., Torre et al., 2007). ARFIMA can be extended to multivariate cases, termed fractionally integrated vector autoregression (FIVAR) (Chung, 2001; Do, Brooks, & Treepongkaruna, 2013), although to the best of our knowledge this has not been applied to psychological data thusfar. Despite the promising and innovative contribution that the study of long term dependence in human behaviour could have, there is much discussion on the extent to which it can be distinguished from multiple short-memory processes (Wagenmakers et al., 2004), and why specifically these patterns have indicative properties is a topic of debate (Farrell, Wagenmakers, & Ratcliff, 2006; Wagenmakers et al., 2005). In addition, ARFIMA and related methods model such behavior are demanding, since many observations are needed to adequately estimate slow decay in the autocorrelation function (Chung & Baillie, 1993; Torre et al., 2007)).

3.3.1.3 Data-driven approaches

As indicated above, researchers might find themselves in a situation where they want to explore non-stationarities in the data without having to specify and fit a full time series model. Also, if non-stationarities are present but not of primary relevance to the research question, it might be more convenient or feasible to correct for them a priori to the actual analysis. In such cases, data-driven approaches that pinpoint the presence of non-stationarities can be useful.

One example is a recently developed method to screen time series for abrupt changes by performing non-parametric multivariate kernel change point detection (KCP-RS). Building on work by Arlot, Celisse, and Harchaoui (2012), this method was presented and validated by Cabrieto and colleagues (Cabrieto, Adolf, Tuerlinckx, Kuppens, & Ceulemans, 2018; Cabrieto, Tuerlinckx, Kuppens, Hunyadi, & Ceulemans, 2018; Cabrieto, Tuerlinckx, Kuppens, Wilhelm, et al., 2018) and comes with an associated R-package (Cabrieto & Meers, 2019). The method can be used to signal and locate abrupt, long-lasting changes in means, variances, (auto-)correlations, or any other summary statistic of interest. As such, KCP-RS can be used for exploratory purposes (e.g., to find early warning signs of a relapse into depressions; Ceulemans et al., 2018), or to assess deviations from stationarity assumptions. For example, if evidence is found for a mean change, this implies that the time series can be split into multiple phases that are characterized by different means. Researchers can use these findings to detrend the data, run separate analyses for the different phases or to inform a fixed moderated analysis.

3.4 The issue of concurrent relationships

As mentioned above, VAR techniques estimate unique direct effects of the lagged variables. They do however not specify explicitly how the different variables under study impact one another concurrently. Such information would be highly informative from a clinical point of view. To illustrate the issue, imagine running a classical VAR on a bivariate time series encompassing social anxiety and rumination. If no crossregressive relationships are found, one may be tempted to conclude that any intervention targeting one of the variables would leave the other variable unaffected. Indeed, although these variables may display some innovation covariance, this interrelation is usually understood as a common response to external perturbations and not as an 'inherent' direct relationship between the variables. The question

here is whether or how the latter type of interpretation can be made, since this is often desirable in psychology. In this section, we aim to briefly summarize what has already been said about this admittedly difficult and technical topic, which however has major interpretational ramifications.

At this point, it is useful to mention that VAR modeling was initially developed in econometrics. In this field, researchers are usually not interested in interpreting the VAR regression weights, but rather aim to forecast future observations by means of obtained VAR solutions. However, when interpretation is of central interest, a distinction has been made between two types of VAR models. These are statistically equivalent and would thus yield equal forecasts, but differ in how they handle contemporaneous relations. The first type is the VAR version that we have introduced in section 2.2 (and worked with throughout); this version is also called the reduced form (Gottschalk, 2001; Molenaar, 2017). The second type is called structural VAR (SVAR). SVAR includes contemporaneous unique direct effects among the variables. As a consequence, the innovations are not allowed to covary anymore. For a mean-centered time series the SVAR model of order 1 is formulated as follows (SVAR, Sims, 1980):

$$y_t = Ay_t + By_{t-1} + \varepsilon_t \quad (4)$$

Where B again holds estimates of auto- and crossregressive effects and the covariance matrix of the innovations is now diagonal. The additional regression weight matrix A captures contemporaneous relationships. Importantly, for each pair of variables only one of these concurrent regression weights is allowed differ from zero, which indicates in principle which of both variables drives the other.

Including contemporaneous effects this way may sound appealing. However, it comes with the drawback, in that SVAR solutions are not unique from a statistical point of view (for a non-technical elaboration, see Gottschalk, 2001). Any single (reduced-form) VAR model corresponds to multiple possible SVAR models, and each of the latter may imply that very different dynamics underlie the data. Indeed, these possible SVAR models will often not only differ in the estimates of contemporaneous relationships, but also in the estimates of lagged effects, rendering the interpretation of VAR regression weights problematic (Molenaar, 2017). To obtain a unique solution, theoretical restrictions on the possible contemporaneous relationships are needed, indicating which of the corresponding regression weights can be set to zero.

3.4.1 Proposed solutions

The statistical equivalence of VAR and multiple SVAR solutions thus poses a significant problem for straightforward interpretation of VAR coefficients as reflecting dynamics of a psychologically meaningful system. To handle this issue, techniques have been proposed recently, which either aim to attenuate the problem, or to at least expose the possible non-unique solutions, allowing a researcher to make informed decisions on grounds of theory-based hypotheses about the regression weights. One such technique is the GIMME framework (Gates, Molenaar, Iyer, Nigg, & Fair, 2014). It estimates SVAR models directly, rather than indirectly via the reduced form VAR (see section 3.5 for more information on how GIMME also allows to model multiple time series) and allows to output all statistically equivalent solutions (Beltz & Molenaar, 2016). A second proposition is hybrid VAR, which, as indicated by its name, tries to find a compromise between VAR and SVAR by allowing some of the innovation covariances to differ from zero while simultaneously including the contemporaneous relationships (Molenaar, 2017).

3.5 Integrating multiple individuals

In their original formulations, the methods discussed here are applicable to data from single individual cases. Often, however, data from multiple individuals are gathered, and the aim is to learn not only about intra-individual processes and dynamics, but also about how they behave over individuals. Such situations render some sort of integration of data and/or modeling results across cases imperative.

Which integration strategy one selects depends on the kind and degree of unifying assumptions one deems appropriate to make across individuals (cf. Hamaker, Ceulemans, Grasman, & Tuerlinckx, 2015). This in turn depends on substantive as well as statistical considerations.

From a substantive perspective, one might expect intra-individual processes to behave highly individual-specific or highly uniform across individuals. For instance, the relationship between the experience of emotional and physical pain seems to be highly heterogeneous across individual chronic pain patients (Frumkin et al., 2020). On the other hand, some low-level psychological processes such as those related to visual perception may be rather homogeneous across individuals (i.e., the finding that visual orientation judgements are biased towards previously seen orientations, Fischer and Whitney (2014)). Of course, in between these two extreme positions lies a whole continuum (e.g., Adolf & Fried, 2019; Voelkle, Brose, Schmiedek, & Lindenberger, 2014), and it may well be that the to be described phenomena are situated somewhere on this continuum.

From a statistical perspective, a main question is whether modeling solutions gain or lose accuracy if one integrates across individuals. Here, arguments revolve around trading different sources of inaccuracy (bias vs. variance), which can be of different importance depending on the goal of an analysis (for instance, explanation or prediction, see Yarkoni & Westfall, 2017). Whereas unifying assumptions across individuals might be overly restrictive and hence bias individual modeling solutions and inferences based on them, they might also prevent imprecise modeling solutions, under-powered tests and imprecise predictions if the amount of data per individual is limited (Bulteel, Tuerlinckx, et al., 2018; Ram, Brinberg, Pincus, & Conroy, 2017). The challenge thus lies in weighing between the group and the individual.

3.5.1 Proposed solutions

3.5.1.1 Modeling individuals (in)dependently

There exist multiple ways in which data and/or modeling solutions from multiple individuals can be combined. Matching the endpoints of the above construed continuum are the extreme positions (cf. Hamaker et al., 2015) of either modeling the data of each person independently of others, and possibly subsequently drawing comparisons, or modeling the pooled data in a way that ignores potential inter-individual differences completely, for instance by treating different individuals as replicates of each other and fitting one VAR model to their data.

Whereas the latter strategy is acknowledged as inappropriate in light of plausible and actual differences between individuals in intra-individual processes and dynamics (e.g., Fisher, Medaglia, & Jeronimus, 2018; Hamaker, 2012; Molenaar, 2004; Molenaar & Campbell, 2009; Schmiedek, Lövdén, von Oertzen, & Lindenberger, 2019), the first strategy allows for a maximally unconstrained analysis of such differences. It can therefore be viewed as enabling "unbiased descriptions" of inter-individual differences and commonalities in intra-individual phenomena (Lindenberger & von Oertzen, 2006, p.300), which might, on the one hand, lead to "informed aggregations" across individuals and the articulation of "lawful relationships" (Nesselrode, 2010, p.211). On the other hand, it can also capture highly individual-specific phenomena that do not permit (more than very abstract) generalizations between individuals (Ram et al., 2017). Clearly, however, fitting a model for each person separately requires sufficient amounts of data within individuals to produce accurate results. If this requirement is not met, or if a process is assumed to unfold in a somewhat homogeneous way in multiple individuals, modeling data from multiple individuals as to some extent pooled is an option. Here, one can distinguish approaches that incorporate inter-individual differences as continuous and quantitative and approaches that group individuals into homogeneous subgroups that differ in more categorical or qualitative ways.

3.5.1.2 Modeling individual differences as quantitative

A prototypical quantitative integration strategy, also among VAR models, is to use hierarchical or multi-level or mixed effects model extensions (e.g., Bringmann et al., 2013; Oud & Delsing, 2010). These comprise an average intra-individual process model (via so-called fixed effects) and an inter-individual

model capturing individual deviations from the average structure via continuous distributions (via so-called random effects). Hence, the same model structure applies to all individuals, with individuals differing gradually in the strengths of the models' parameters. Traditional model formulations and implementations usually rely on normal distributions to describe individual differences, but some flexibility has been gained, also with respect to which parameters can be included as random effects, especially due to the rise of simulation-based estimation in the Bayesian modeling framework (e.g., Asparouhov et al., 2018; Driver & Voelkle, 2018a; Schuurman, Grasman, & Hamaker, 2016).

Another approach drawing upon multiple individuals to estimate the parameters of a VAR-type model is given by so-called group iterative multiple model estimation (GIMME; Gates & Molenaar, 2012; Gates et al., 2014). This recently developed method also represents individual differences as quantitative, but not subject to distributional assumptions. Instead, a data-driven strategy is employed. In a first iterative group-level step, the procedure decides which dynamic effects should be estimated for everyone based on modification indices (i.e., the expected improvement in model fit) associated with each effect for each person. Effects that are significant for a majority of the individuals are selected. Next, it proceeds with an iterative individual-level step, in which estimates of the selected effects are obtained for each individual separately and one or more of the other effects can be added until the time series of that individual is optimally fitted. GIMME thus yields estimates that are allowed to differ across persons, but the order in which effects are added to the person-specific models is based on group-level information. In addition, it features not the conventional VAR model, but the structural model variant, further decomposing contemporaneous relationships among the modelled processes.

3.5.1.3 Modeling individual differences as qualitative

Instead of continuous, quantitative differences, cluster-based approaches assume discrete differences between individuals, which are often interpretable as qualitative or categorical. Individuals are thereby grouped according to their full profile of auto- and crossregressive effects, such that cases that belong to the same cluster have similar intra-personal dynamics, whereas cases that are assigned to different clusters differ in at least some dynamic aspects. To our knowledge, four types of clustering approaches have been proposed so far.

Firstly, one can use a two-step approach, where a (V)AR model is initially fitted to the data of each individual separately. Subsequently, one computes dissimilarities (e.g., using the Euclidean distance) between the effect profiles for each pair of individuals and feeds these dissimilarities to a hierarchical clustering technique (Zheng, Wiebe, Cleveland, Molenaar, & Harris, 2013).

Secondly, Bulteel, Tuerlinckx, Brose, and Ceulemans (2016a) proposed an extension of the well-known K-means approach, which they called clusterwise VAR, as the method clusters on the VAR estimates. This method simultaneously partitions the individuals into clusters and fits a VAR model within each cluster.

Thirdly, Ernst, Timmerman, Jeronimus, and Albers (2019) built on mixture modeling and developed a probabilistic version of clusterwise VAR. One benefit of this method is that it yields an estimate of how probable the assignment of an individual to a cluster is, rather than this assigning being either true or false.

Finally, a subgrouping version of GIMME, called S-GIMME, has been presented that allows to subgroup individuals as well (Gates, Lane, Varangis, Giovanello, & Guiskewicz, 2017; Lane, Gates, Pike, Beltz, & Wright, 2019). To this end, a subgroup-level step is added in between the group-level and individual-level steps as described in the previous paragraph, implying that specific effects are added for subgroups of persons. An important distinction between S-GIMME and the previous approaches, is that within each obtained subgroup or cluster, the effects are ultimately still estimated per individual, implying that estimates can still differ within each subgroup.

4 DISCUSSION

The benefits of employing intensive within-subjects paradigms are clear, and the complex relationships between mind and body appears well suited for analysis from a dynamical perspective. It is also clear, however, that the methodological challenges associated with employing these promising techniques are far from trivial, and have direct implications for the accuracy, interpretability, replicability, and generalizability of findings. Here, we briefly recapitulate the most important issues, and highlight three additional concerns we have not explicitly mentioned in the other sections. Specifically, we discuss the importance of the lag employed, data requirements such as sample size and time series length, and the role of measurement error in the discussed frameworks. We close with a note on causal inference in the dynamic paradigm.

4.1 Varying measurement intervals

Most conventional VAR models operate in discrete time and thus assume that intervals between observations are equal. Also, the parameter values recovered from such models are only specific to the time interval over which they are estimated. Within a study, parameter estimates will be biased when differences in time intervals are neglected. In addition, if time interval differences between studies are ignored, cross-study comparisons can confound actual differences between effects with time-interval induced ones. Conflicting findings may therefore pertain to differences regarding methodology rather than substance. The problem can be mitigated to some extent by artificially refining the time scale of the data. Alternatively, continuous time methods offer an appealing solution, but are paired with their own difficulties on both conceptual and statistical levels.

4.2 Including multiple possibly covarying variables

We have highlighted that VAR is essentially a linear regression technique, and thus the problems associated with multivariate linear regression are applicable to the VAR framework too. Multicollinearity issues may be apparent in many practical applications, and can result in unstable VAR estimates. In addition, the fact that VAR is intrinsically highly parameterized can quickly lead to issues concerning overfitting. Techniques aimed at limiting this complexity such as those employing dimension reduction or regularization could provide relatively straightforward solutions for these issues, and overfitting of models due to excessive parameterization can be assessed by employing cross-validation techniques.

4.3 The issue of concurrent effects

The VAR methodology has seen extensive use and development in econometrics, where economic predictions or forecasts rather than the structure of the economic system are typically of interest. This distinction may be more problematic for psychological research, where the underlying psychological dynamics and effects of variables on each other are often of key interest. VAR models do not explicitly model contemporaneous relationships between variables. To handle this issue, the structural VAR model has been proposed, but using this model will typically not result in a unique solution, in that any single VAR model is statistically equivalent to multiple structural VAR models. This conundrum can be solved by invoking theory to select the most plausible solution. However, given the fact that the dynamic paradigm in psychosomatic research is in its infancy, theoretical advances are also necessary in this regard.

4.4 Changing processes

TSA approaches should in principle directly translate the assumptions of the researcher. Assumptions on process change can be translated by employing models capable of capturing which features of the dynamical system change (e.g. changes in mean or variance over time), whether these changes occur abruptly, or as smooth transitions, and whether the sources of these changes are known or unknown. If little is known about the time-course of a process, tools have been developed to screen time

series for temporal changes, and tests of stationarity can be applied. For psychosomatic applications, change may be of key interest in understanding how pathological mind-body relationships form and perpetuate. Theory should aim to make hypotheses about these temporal profiles concrete, allowing one to determine which methodological approaches are most appropriate.

4.5 Integrating multiple individuals

Regarding the integration of multiple individuals, significant recent methodological advances have been made, in that multiple approaches have been developed with different perspectives on the nature of inter-individual differences between individual-level processes. These perspectives can be situated on a continuum ranging from those assuming that each individual constitutes a unique system, to those assuming that all individual systems are equal with regard to the processes under investigation. The type of method(s) employed should ideally depend on the assumptions researchers have about the processes under investigation. The research already conducted in the psychosomatic domain appears to imply that heterogeneity may be a common issue. If little is known about how a specific psychosomatic process behaves across individuals, models that make less strong unifying assumptions at the population level should be preferred, because such models are less likely to impose group-level structures on individual-level solutions. At the same time, integrating over individuals can be beneficial for the accuracy of model solutions, and thus choices regarding this integration will always be a balancing act between these different aspects.

4.6 The importance of the lag employed

Typical psychological applications of VAR models have focused on lag-1 relationships, where y_t is predicted from y_{t-1} . There are no substantive reasons for excluding other lags and thus using higher order TSA models. The selection of the appropriate lag structure required to model the data is thus an issue which pertains to all models discussed throughout the paper, and incorrect specification will lead to biased parameter estimates (Braun & Mittnik, 1993). Indeed, any inference made from model results is specific to the employed lag. The question thus boils down to how to select the lag structure which best fits the data. Including additional lags increases the parameterisation of the VAR model, which means that criteria used to select the most appropriate lag specification must be sensitive towards model complexity. Solutions therefore typically involve comparing information criterion fit indexes, such as the Akaike or Bayesian information criterion (AIC and BIC), or a combination of criteria (Hacker & Hatemi-J, 2008; Hatemi-j, 2003), which are run on all possible lag specified models, such that an optimal selection can be made. In small samples specifically (e.g. $T = 40$), Hacker and Hatemi-J (2008) recommend usage of either AIC or the Schwarz–Bayesian criterion (SBC). As another alternative, Jacobson, Chow, and Newman (2019) recently developed the differential time-varying effect model (DTVEM), which employs a combination of exploratory modeling and confirmatory model selection to gauge the optimal lag structure. Although specifically tailored to accommodate multi-subject ILD, the framework currently imposes the same lag structure over all individuals, and can become computationally demanding in some cases. However, the technique is straightforward to implement in R, and offers an accessible tool for exploratory analysis.

4.7 Sample size and time series length

We have throughout this paper refrained from making concrete recommendations concerning sample size requirements, since no unitary answer to the question of 'how much data is sufficient' exists. The only general statement one can make in this regard is noting that more data is always better for purposes of statistical estimation and inference (e.g., Solanas, Manolov, & Sierra, 2010), and that this applies equally so to the methods discussed throughout this paper. Note that the effectivity of the above-mentioned techniques for deriving the appropriate lag structure, which is fundamental for the usefulness of model results, also increases as a function of the length of the time series (Hacker &

Hatemi-J, 2008; Hatemi-j, 2003; Jacobson et al., 2019).

While we can not contribute more than reiterate and emphasize that sufficient data are needed, one heuristic is that more parameters require more data to be estimated with equal precision (Deng, Yang, & Marcoulides, 2018). Researchers investigating non-stationary processes, models with high dimensionality (Bulteel, Mestdagh, et al., 2018), or processes with large lag orders, should be especially considerate to include sufficient observations within individuals, or as we have illustrated above, think of appropriate ways to integrate information over individual cases (e.g., Liu, 2017). In addition, the sampling rate of time series data can affect estimation accuracy of VAR models (for an example of an AR(1) model in continuous time, see Adolf et al., 2019).

4.8 The role of measurement error

Innovations play an important role in VAR models and extensions thereof. They affect the process over time and therefore are part of the dynamics of the system. Measurement error in contrast is only assumed to arise from imperfect measurement of a construct at time t , lacking the temporal informative content of innovations. In conventional VAR models, measurement error is not taken into account, implying that each construct in question is measured perfectly. Many may justly consider this to be a problematic assumption, especially in psychological applications. In fact, neglecting measurement error has been found to lead to underestimations of autoregressive parameters in AR models (Schuurman et al., 2015). Schuurman et al. (2015) also note that what is considered measurement error, and what is considered innovation, can vary depending on the sampling interval. If large/small intervals are used to assess the same process, any effects which did carry over to subsequent observations in the small interval case, and are thus considered innovations, will be considered measurement errors in the large interval case.

Disentangling process from measurement effects can be done in a variety of ways, but these have in common that the difference is explicitly incorporated into the model. The state-space modeling framework and structural equation modeling framework employ this distinction, include measurement error as a random variable in the measurement model, and process innovation as a random variable in the process model. In multilevel extensions of VAR, a well known problem is that variability between individuals must be disentangled from variability within individuals (Curran & Bauer, 2011), but Schuurman and Hamaker (2019) emphasize that variability within individuals will also consist of variability due to the process and variability due to measurement. These authors propose a multilevel model that allows one to specify whether measurement error should be common to all individuals, or whether measurement error is also person-specific.

4.9 Causal inference

A final remark we wish to make concerns causal inference. Multiple authors have made a case that adopting a dynamic multivariate approach to psychological and psychosomatic phenomena may aid the identification of causal effects (e.g., Boker, 2002; Driver & Voelkle, 2018b; Schuurman, Ferrer, et al., 2016; Voelkle et al., 2018). The presumed assets are threefold. Firstly, intra-individual change is formalized, and especially so in terms of interacting processes, instead of as a function of time. Secondly, the introduction of time and temporal order allows one to disentangle (statistical) antecedents and consequences, and to account for confounding variables within time points. Thirdly, the dynamic paradigm provides the opportunity to control for associations between variables across individuals as important confounding factors.

However, causal inference - which itself is not a unitary concept and might be approached from different theoretical standpoints with different techniques - remains a complicated problem even in the context of a dynamic research paradigm. For instance, since observational data often comprise only a limited set of variables, potentially measured with errors, information for causal inference is likely insufficient. Also, that the targeted phenomena are often poorly understood does not exactly help in formalizing

causal mechanisms. Proposals to further improve causal inference under the dynamic paradigm mainly draw upon the idea of bringing the targeted phenomena partly under experimental (and statistical) control (e.g. Koval, Pe, Meers, & Kuppens, 2013; McCall, Hildebrandt, Hartmann, Baczkowski, & Singer, 2016; Neubauer et al., 2018; Schmiedek & Neubauer, 2019).

5 CONCLUSION

In summary, if we wish to safeguard the promising possibilities associated with the dynamic paradigm, methodological issues must be taken seriously. Concrete steps can be taken to ensure the accuracy, interpretability, replicability, and generalizability of findings. While it may not always be possible to meet each modeling assumption from a pragmatic point of view, studies should nevertheless be transparent in reporting details associated with the sampling scheme, included variables, and reasons for employing certain modeling techniques. In this way one may more easily disentangle methodological issues from conflicting theoretical findings.

6 ACKNOWLEDGEMENTS

This work was supported by a research fellowship from the German Research Foundation (DFG) awarded to Janne Adolf, and by a research grant from the Fund for Scientific Research-Flanders (FWO, Project No. G.074319N) awarded to Eva Ceulemans by the Research Council of KU Leuven (C14/19/054).

7 DECLARATIONS OF INTEREST

The authors declare the absence of any financial, intellectual, or other conflicts of interest which may have biased any aspect of this manuscript.

REFERENCES

- Adolf, J. K., & Fried, E. I. (2019). Ergodicity is sufficient but not necessary for group-to-individual generalizability. *Proceedings of the National Academy of Sciences*, 201818675.
- Adolf, J. K., Loossens, T., Tuerlinckx, F., & Ceulemans, E. (2019). Optimal fixed sampling rates for reliable continuous-time first-order autoregressive modeling.
- Adolf, J. K., Voelkle, M. C., Brose, A., & Schmiedek, F. (2017). Capturing context-related change in emotional dynamics via fixed moderated time series analysis. *Multivariate behavioral research*, 52(4), 499–531.
- Arlot, S., Celisse, A., & Harchaoui, Z. (2012). Kernel change-point detection. *arXiv preprint arXiv:1202.3878*, 6.
- Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(3), 359–388.
- Babiyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic medicine*, 66(3), 411–421.
- Balon, R. (2006). Mood, anxiety, and physical illness: Body and mind, or mind and body? *Depression and anxiety*, 23(6), 377–387.
- Beltz, A. M., & Molenaar, P. C. (2016). Dealing with multiple solutions in structural vector autoregressive models. *Multivariate behavioral research*, 51(2-3), 357–373.
- Boker, S. M. (2002). Consequences of continuity: The hunt for intrinsic properties within parameters of dynamics in psychological processes. *Multivariate Behavioral Research*, 37(3), 405–422.
- Boker, S. M. (2012). Dynamical systems and differential equation models of change. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol 3: Data analysis and research publication*. (pp. 323–333). doi:10.1037/13621-016
- Boker, S. M. (2015). Adaptive equilibrium regulation: A balancing act in two timescales. *Journal for person-oriented research*, 1(1-2), 99.
- Boker, S. M., Tiberio, S. S., & Moulder, R. G. (2018). Robustness of time delay embedding to sampling interval misspecification. In *Continuous time modeling in the behavioral and related sciences* (pp. 239–258). Springer.
- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual review of psychology*, 54(1), 579–616.
- Bondell, H. D., & Reich, B. J. (2008). Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with oscar. *Biometrics*, 64(1), 115–123.
- Braun, P. A., & Mittnik, S. (1993). Misspecifications in vector autoregressions and their effects on impulse responses and variance decompositions. *Journal of Econometrics*, 59(3), 319–341.
- Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2017). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological methods*, 22(3), 409.
- Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., . . . Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PloS one*, 8(4), e60188.
- Browne, M. W., & Nesselroade, J. R. (2005). Representing psychological processes with dynamic factor models: Some promising uses and extensions of autoregressive moving average time series models.
- Bulteel, K., Mestdagh, M., Tuerlinckx, F., & Ceulemans, E. (2018). Var (1) based models do not always outpredict ar (1) models in typical psychological applications. *Psychological methods*, 23(4), 740.

- Bulteel, K., Tuerlinckx, F., Brose, A., & Ceulemans, E. (2016a). Clustering vector autoregressive models: Capturing qualitative differences in within-person dynamics. *Frontiers in psychology*, 7, 1540.
- Bulteel, K., Tuerlinckx, F., Brose, A., & Ceulemans, E. (2016b). Using raw var regression coefficients to build networks can be misleading. *Multivariate behavioral research*, 51(2-3), 330–344.
- Bulteel, K., Tuerlinckx, F., Brose, A., & Ceulemans, E. (2018). Improved insight into and prediction of network dynamics by combining var and dimension reduction. *Multivariate behavioral research*, 53(6), 853–875.
- Burton, C., Weller, D., & Sharpe, M. (2007). Are electronic diaries useful for symptoms research? a systematic review. *Journal of psychosomatic research*, 62(5), 553–561.
- Cabrieto, J., Adolf, J. K., Tuerlinckx, F., Kuppens, P., & Ceulemans, E. (2018). Detecting long-lived autodependency changes in a multivariate system via change point detection and regime switching models. *Scientific reports*, 8(1), 15637.
- Cabrieto, J., & Meers, K. (2019). Kcprs-package: Kcp on the running statistics.
- Cabrieto, J., Tuerlinckx, F., Kuppens, P., Hunyadi, B., & Ceulemans, E. (2018). Testing for the presence of correlation changes in a multivariate time series: A permutation based approach. *Scientific reports*, 8(1), 769.
- Cabrieto, J., Tuerlinckx, F., Kuppens, P., Wilhelm, F. H., Liedlgruber, M., & Ceulemans, E. (2018). Capturing correlation changes by applying kernel change point detection on the running correlations. *Information Sciences*, 447, 117–139.
- Ceulemans, E. et al. (2018). An objective, comprehensive and flexible statistical framework for detecting early warning signs of mental health problems. *Psychotherapy and psychosomatics*, 1–3.
- Chow, S.-M. (2019). Practical tools and guidelines for exploring and fitting linear and nonlinear dynamical systems models. *Multivariate Behavioral Research*, 54(5), 690–718.
- Chow, S.-M., Ho, M.-h. R., Hamaker, E. L., & Dolan, C. V. (2010). Equivalence and differences between structural equation modeling and state-space modeling techniques. *Structural Equation Modeling*, 17(2), 303–332.
- Chow, S.-M., Zu, J., Shifren, K., & Zhang, G. (2011). Dynamic factor analysis models with time-varying parameters. *Multivariate Behavioral Research*, 46(2), 303–339.
- Chun, C. A. (2016). The expression of posttraumatic stress symptoms in daily life: A review of experience sampling methodology and daily diary studies. *Journal of Psychopathology and Behavioral Assessment*, 38(3), 406–420.
- Chung, C.-F. (2001). Calculating and analyzing impulse responses for the vector arfima model. *Economics Letters*, 71(1), 17–25.
- Chung, C.-F., & Baillie, R. T. (1993). Small sample bias in conditional sum-of-squares estimators of fractionally integrated arma models. *Empirical Economics*, 18(4), 791–806.
- Cohen, P., West, S. G., & Aiken, L. S. (2014). *Applied multiple regression/correlation analysis for the behavioral sciences*. Psychology Press.
- Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. *Annu. Rev. Psychol.* 57, 505–528.
- Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual review of psychology*, 62, 583–619.
- de Haan-Rietdijk, S., Voelkle, M. C., Keijsers, L., & Hamaker, E. L. (2017). Discrete-vs. continuous-time modeling of unequally spaced experience sampling method data. *Frontiers in psychology*, 8, 1849.
- Deboeck, P. R., Nicholson, J., Kouros, C., Little, T. D., & Garber, J. (2015). Integrating developmental theory and methodology: Using derivatives to articulate change theories, models, and inferences. *Applied developmental science*, 19(4), 217–231.

- Deboeck, P. R., & Preacher, K. J. (2016). No need to be discrete: A method for continuous time mediation analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(1), 61–75.
- Delignières, D., & Torre, K. (2009). Fractal dynamics of human gait: A reassessment of the 1996 data of hausdorff et al. *Journal of Applied Physiology*, 106(4), 1272–1279.
- Deng, L., Yang, M., & Marcoulides, K. M. (2018). Structural equation modeling with many variables: A systematic review of issues and developments. *Frontiers in psychology*, 9, 580.
- Do, H. X., Brooks, R. D., & Treepongkaruna, S. (2013). Generalized impulse response analysis in a fractionally integrated vector autoregressive model. *Economics Letters*, 118(3), 462–465.
- Dormann, C., & Griffin, M. A. (2015). Optimal time lags in panel studies. *Psychological methods*, 20(4), 489.
- Driver, C. C., Oud, J. H., & Voelkle, M. C. (2017). Continuous time structural equation modeling with r package ctsem.
- Driver, C. C., & Voelkle, M. C. (2018a). Hierarchical bayesian continuous time dynamic modeling. *Psychological methods*, 23(4), 774.
- Driver, C. C., & Voelkle, M. C. (2018b). Understanding the time course of interventions with continuous time dynamic models. In *Continuous time modeling in the behavioral and related sciences* (pp. 79–109). Springer.
- Ebner-Priemer, U. W., & Trull, T. J. (2009). Ecological momentary assessment of mood disorders and mood dysregulation. *Psychological assessment*, 21(4), 463.
- El Sount, C. R.-O., Windthorst, P., Denking, J., Ziser, K., Nikendei, C., Kindermann, D., . . . Junne, F. (2019). Chronic pain in refugees with posttraumatic stress disorder (ptsd): A systematic review on patients' characteristics and specific interventions. *Journal of psychosomatic research*, 118, 83–97.
- Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological methods*.
- Ernst, A. F., Timmerman, M. E., Jeronimus, B. F., & Albers, C. J. (2019). Insight into individual differences in emotion dynamics with clustering. *Assessment*, 1073191119873714.
- Farrell, S., Wagenmakers, E.-J., & Ratcliff, R. (2006). 1/f noise in human cognition: Is it ubiquitous, and what does it mean? *Psychonomic bulletin & review*, 13(4), 737–741.
- Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature neuroscience*, 17(5), 738.
- Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, 115(27), E6106–E6115.
- Frumkin, M., Haroutounian, S., & Rodebaugh, T. (2020). Examining emotional pain among individuals with chronic physical pain: Nomothetic and idiographic approaches. *Manuscript under review*.
- Gates, K. M., Lane, S. T., Varangis, E., Giovanello, K., & Guiskewicz, K. (2017). Unsupervised classification during time-series model building. *Multivariate behavioral research*, 52(2), 129–148.
- Gates, K. M., & Molenaar, P. C. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage*, 63(1), 310–319.
- Gates, K. M., Molenaar, P. C., Iyer, S. P., Nigg, J. T., & Fair, D. A. (2014). Organizing heterogeneous samples using community detection of gimme-derived resting state functional networks. *PloS one*, 9(3), e91322.
- Ghiggia, A., Torta, R., Tesio, V., Di Tella, M., Romeo, A., Colonna, F., . . . Castelli, L. (2017). Psychosomatic syndromes in fibromyalgia. *Clin Exp Rheumatol*, 105(3), 106–111.
- Gottschalk, J. (2001). *An introduction into the svar methodology: Identification, interpretation and limitations of svar models*. Kiel working paper.

- Hacker, S. R., & Hatemi-J, A. (2008). Optimal lag-length choice in stable and unstable var models under situations of homoscedasticity and arch. *Journal of Applied Statistics*, 35(6), 601–615.
- Hamaker, E. L. (2012). Why researchers should think” within-person”: A paradigmatic rationale.
- Hamaker, E. L., Ceulemans, E., Grasman, R., & Tuerlinckx, F. (2015). Modeling affect dynamics: State of the art and future challenges. *Emotion Review*, 7(4), 316–322.
- Hamaker, E. L., & Dolan, C. V. (2009). Idiographic data analysis: Quantitative methods—from simple to advanced. In *Dynamic process methodology in the social and developmental sciences* (pp. 191–216). Springer.
- Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. (2003). Arma-based sem when the number of time points t exceeds the number of cases n : Raw data maximum likelihood. *Structural Equation Modeling*, 10(3), 352–379.
- Hamaker, E. L., Grasman, R. P., & Kamphuis, J. H. (2010). Regime-switching models to study psychological processes.
- Hamaker, E. L., & Grasman, R. (2012). Regime switching state-space models applied to psychological processes: Handling missing data and making inferences. *Psychometrika*, 77(2), 400–422.
- Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science*, 26(1), 10–15.
- Hamilton, J. D. (1994). *Time series analysis*. Princeton New Jersey.
- Hamilton, J. D. (2010). Regime switching models. In *Macroeconometrics and time series analysis* (pp. 202–209). Springer.
- Harvey, A. C. (1989). *Forecasting, structural time series models and the kalman filter*. Cambridge university press.
- Haslbeck, J., Bringmann, L. F., & Waldorp, L. J. (2017). How to estimate time-varying vector autoregressive models? a comparison of two methods. *arXiv preprint arXiv:1711.05204*.
- Hatemi-j, A. (2003). A new method to choose optimal lag order in stable and unstable var models. *Applied Economics Letters*, 10(3), 135–137.
- Henrie, C. R., Bodily, R., Manwaring, K. C., & Graham, C. R. (2015). Exploring intensive longitudinal measures of student engagement in blended learning. *International Review of Research in Open and Distributed Learning*, 16(3), 131–155.
- Hoover, K. D. (2005). Automatic inference of the contemporaneous causal order of a system of equations. *Econometric Theory*, 21(1), 69–77.
- Hsu, N.-J., Hung, H.-L., & Chang, Y.-M. (2008). Subset selection for vector autoregressive processes using lasso. *Computational Statistics & Data Analysis*, 52(7), 3645–3657.
- Hunt, N., McGrath, D., & Stergiou, N. (2014). The influence of auditory-motor coupling on fractal dynamics in human gait. *Scientific reports*, 4, 5879.
- Hunter, M. D. (2018). State space modeling in an open source, modular, structural equation modeling environment. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(2), 307–324.
- Jacobson, N. C., Chow, S.-M., & Newman, M. G. (2019). The differential time-varying effect model (dtvem): A tool for diagnosing and modeling time lags in intensive longitudinal data. *Behavior research methods*, 51(1), 295–315.
- Jebb, A. T., Tay, L., Wang, W., & Huang, Q. (2015). Time series analysis for psychological research: Examining and forecasting change. *Frontiers in psychology*, 6, 727.
- Jensen, M., Hoffman, A., & Cardenas, D. (2005). Chronic pain in individuals with spinal cord injury: A survey and longitudinal study. *Spinal cord*, 43(12), 704.
- Jolliffe, I. T., Trendafilov, N. T., & Uddin, M. (2003). A modified principal component technique based on the lasso. *Journal of computational and Graphical Statistics*, 12(3), 531–547.
- Karch, J. (2016). A machine learning perspective on repeated measures.
- Kelly, D., Weigard, A., & Beltz, A. (2020). How are you doing? the person-specificity of daily links between neuroticism and physical health. *Manuscript under review*.

- Kiers, H. A., & Smilde, A. K. (2007). A comparison of various methods for multivariate regression with highly collinear variables. *Statistical Methods and Applications*, 16(2), 193–228.
- Kim, C.-J., Nelson, C. R. et al. (1999). State-space models with regime switching: Classical and gibbs-sampling approaches with applications. *MIT Press Books*, 1.
- Koval, P., Pe, M. L., Meers, K., & Kuppens, P. (2013). Affect dynamics in relation to depressive symptoms: Variable, unstable or inert? *Emotion*, 13(6), 1132.
- Kucyi, A., & Davis, K. D. (2015). The dynamic pain connectome. *Trends in neurosciences*, 38(2), 86–95.
- Kuiper, R. M., & Ryan, O. (2018). Drawing conclusions from cross-lagged relationships: Re-considering the role of the time-interval. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(5), 809–823.
- Kuppens, P. (2015). It's about time: A special section on affect dynamics. *Emotion Review*, 7(4), 297–300.
- Lane, S. T., Gates, K. M., Pike, H. K., Beltz, A. M., & Wright, A. G. (2019). Uncovering general, shared, and unique temporal patterns in ambulatory assessment data. *Psychological methods*, 24(1), 54.
- Lindenberger, U., & von Oertzen, T. (2006). *Of book: Lifespan cognition: Mechanisms of change*. Oxford University Press.
- Liu, S. (2017). Person-specific versus multilevel autoregressive models: Accuracy in parameter estimates at the population and individual levels. *British Journal of Mathematical and Statistical Psychology*, 70(3), 480–498.
- Lütkepohl, H. (2013). *Introduction to multiple time series analysis*. Springer Science & Business Media.
- McCall, C., Hildebrandt, L. K., Hartmann, R., Baczkowski, B. M., & Singer, T. (2016). Introducing the wunderkammer as a tool for emotion research: Unconstrained gaze and movement patterns in three emotionally evocative virtual worlds. *Computers in Human Behavior*, 59, 93–107.
- McGreevy, K., Bottros, M. M., & Raja, S. N. (2011). Preventing chronic pain following acute pain: Risk factors, preventive strategies, and their efficacy. *European journal of pain supplements*, 5(2), 365–376.
- McNeish, D. M. (2015). Using lasso for predictor selection and to assuage overfitting: A method long overlooked in behavioral sciences. *Multivariate Behavioral Research*, 50(5), 471–484.
- Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, 2(4), 201–218.
- Molenaar, P. C. (2017). Equivalent dynamic models. *Multivariate behavioral research*, 52(2), 242–258.
- Molenaar, P. C., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current directions in psychological science*, 18(2), 112–117.
- Molenaar, P. C., De Gooijer, J. G., & Schmitz, B. (1992). Dynamic factor analysis of nonstationary multivariate time series. *Psychometrika*, 57(3), 333–349.
- Molenaar, P. C., Sinclair, K. O., Rovine, M. J., Ram, N., & Corneal, S. E. (2009). Analyzing developmental processes on an individual level using nonstationary time series modeling. *Developmental psychology*, 45(1), 260.
- Nakamura, T., Kiyono, K., Wendt, H., Abry, P., & Yamamoto, Y. (2016). Multiscale analysis of intensive longitudinal biomedical signals and its clinical applications. *Proceedings of the IEEE*, 104(2), 242–261.
- Nehrkorn-Bailey, A. M., Reardon, M. S., & Hicks Patrick, J. (2018). Some methodological and analytical issues related to real-time data capture studies. *Translational Issues in Psychological Science*, 4(4), 349.
- Nesselroade, J. R. (2010). On an emerging third discipline of scientific psychology.

- Neubauer, A. B., Lerche, V., & Voss, A. (2018). Interindividual differences in the intraindividual association of competence and well-being: Combining experimental and intensive longitudinal designs. *Journal of personality*, *86*(4), 698–713.
- Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2009). A hierarchical ornstein–uhlenbeck model for continuous repeated measurement data. *Psychometrika*, *74*(3), 395–418.
- Ou, L., Hunter, M. D., & Chow, S.-M. (2017). What's for dynr: A package for linear and nonlinear dynamic modeling in r. *Journal of Statistical Software*.
- Oud, J. H., & Delsing, M. J. (2010). Continuous time modeling of panel data by means of sem. In *Longitudinal research with latent variables* (pp. 201–244). Springer.
- Oud, J. H., & 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*(6), 3271–3288.
- Peng, C.-K., Mietus, J., Hausdorff, J., Havlin, S., Stanley, H. E., & Goldberger, A. L. (1993). Long-range anticorrelations and non-gaussian behavior of the heartbeat. *Physical review letters*, *70*(9), 1343.
- Peng, C.-K., Mietus, J. E., Liu, Y., Lee, C., Hausdorff, J. M., Stanley, H. E., . . . Lipsitz, L. A. (2002). Quantifying fractal dynamics of human respiration: Age and gender effects. *Annals of biomedical engineering*, *30*(5), 683–692.
- Ram, N., Brinberg, M., Pincus, A. L., & Conroy, D. E. (2017). The questionable ecological validity of ecological momentary assessment: Considerations for design and analysis. *Research in human development*, *14*(3), 253–270.
- Ram, N., Brose, A., Molenaar, P. C., et al. (2013). Dynamic factor analysis: Modeling person-specific process. *The Oxford handbook of quantitative methods*, *2*, 441–457.
- Rindskopf, D. (1984). Using phantom and imaginary latent variables to parameterize constraints in linear structural models. *Psychometrika*, *49*(1), 37–47.
- Rosmalen, J. G., Wenting, A. M., Roest, A. M., de Jonge, P., & Bos, E. H. (2012). Revealing causal heterogeneity using time series analysis of ambulatory assessments: Application to the association between depression and physical activity after myocardial infarction. *Psychosomatic medicine*, *74*(4), 377–386.
- Ryan, O., Kuiper, R. M., & Hamaker, E. L. (2018). A continuous-time approach to intensive longitudinal data: What, why, and how? In *Continuous time modeling in the behavioral and related sciences* (pp. 27–54). Springer.
- Scargle, J. D. (1982). Studies in astronomical time series analysis. ii-statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, *263*, 835–853.
- Schafer, T. A. W. J. L. (2006). *Models for intensive longitudinal data*. Oxford University Press.
- Schmiedek, F., Lövdén, M., von Oertzen, T., & Lindenberger, U. (2019). *Within-person structures of daily cognitive performance cannot be inferred from between-person structures of cognitive abilities*. PeerJ Preprints.
- Schmiedek, F., & Neubauer, A. B. (2019). Experiments in the wild: Introducing the within-person encouragement design. *Multivariate behavioral research*, 1–21.
- Scholz, U. (2019). It's time to think about time in health psychology. *Applied Psychology: Health and Well-Being*.
- Schuurman, N. K., Ferrer, E., de Boer-Sonnenschein, M., & Hamaker, E. L. (2016). How to compare cross-lagged associations in a multilevel autoregressive model. *Psychological methods*, *21*(2), 206.
- Schuurman, N. K., Grasman, R., & Hamaker, E. (2016). A comparison of inverse-wishart prior specifications for covariance matrices in multilevel autoregressive models. *Multivariate Behavioral Research*, *51*(2-3), 185–206.

- Schuurman, N. K., & Hamaker, E. L. (2019). Measurement error and person-specific reliability in multilevel autoregressive modeling. *Psychological methods*, 24(1), 70.
- Schuurman, N. K., Houtveen, J. H., & Hamaker, E. L. (2015). Incorporating measurement error in $n=1$ psychological autoregressive modeling. *Frontiers in psychology*, 6, 1038.
- Scollon, N. C., Prieto, C.-K., & Diener, E. (2009). Experience sampling: Promises and pitfalls, strength and weaknesses. In E. Diener (Ed.), *Assessing well-being: The collected works of ed diener* (pp. 157–180). doi:10.1007/978-90-481-2354-4_8
- Sharp, J., & Keefe, B. (2006). Psychiatry in chronic pain: A review and update. *Focus*, 7(4), 213–580.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, 1–48.
- Singer, H. (2012). Sem modeling with singular moment matrices part ii: MI-estimation of sampled stochastic differential equations. *The Journal of Mathematical Sociology*, 36(1), 22–43.
- Solanas, A., Manolov, R., & Sierra, V. (2010). Lag-one autocorrelation in short series: Estimation and hypotheses testing. *Psicológica*, 31(2), 357–381.
- Stone, A. A., & Broderick, J. E. (2007). Real-time data collection for pain: Appraisal and current status. *Pain Medicine*, 8, S85–S93.
- Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: A retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3), 273–282.
- Tong, H. (2011). Threshold models in time series analysis—30 years on. *Statistics and its Interface*, 4(2), 107–118.
- Torre, K., Delignieres, D., & Lemoine, L. (2007). Detection of long-range dependence and estimation of fractal exponents through arfima modelling. *British Journal of Mathematical and Statistical Psychology*, 60(1), 85–106.
- Trull, T. J., & Ebner-Priemer, U. W. (2009). Using experience sampling methods/ecological momentary assessment (esm/ema) in clinical assessment and clinical research: Introduction to the special section.
- Trull, T. J., & Ebner-Priemer, U. W. (2020). Ambulatory assessment in psychopathology research: A review of recommended reporting guidelines and current practices. *Journal of Abnormal Psychology*, 129(1), 56.
- Trull, T. J., Lane, S. P., Koval, P., & Ebner-Priemer, U. W. (2015). Affective dynamics in psychopathology. *Emotion Review*, 7(4), 355–361.
- Tunks, E. R., Crook, J., & Weir, R. (2008). Epidemiology of chronic pain with psychological comorbidity: Prevalence, risk, course, and prognosis. *The Canadian Journal of Psychiatry*, 53(4), 224–234.
- Van Buuren, S. (2018). *Flexible imputation of missing data*. Chapman and Hall/CRC.
- van de Leemput, I. A., Wichers, M., Cramer, A. O., Borsboom, D., Tuerlinckx, F., Kuppens, P., ... Aggen, S. H., et al. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences*, 111(1), 87–92.
- van der Krieke, L., Blaauw, F. J., Emerencia, A. C., Schenk, H. M., Slaets, J. P., Bos, E. H., ... Jeronimus, B. F. (2017). Temporal dynamics of health and well-being: A crowdsourcing approach to momentary assessments and automated generation of personalized feedback. *Psychosomatic medicine*, 79(2), 213–223.
- van Giels, A., Emerencia, A., Bos, E., & Rosmalen, J. (2020). Exploring interrelationships among worrying, anxiety, and somatic symptoms using time-series analysis. *Manuscript under review*.
- Voelkle, M. C., Brose, A., Schmiedek, F., & Lindenberger, U. (2014). Toward a unified framework for the study of between-person and within-person structures: Building a bridge between two research paradigms. *Multivariate Behavioral Research*, 49(3), 193–213.
- 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(6), 782–805.

- Voelkle, M. C., & Oud, J. H. (2013). Continuous time modelling with individually varying time intervals for oscillating and non-oscillating processes. *British Journal of Mathematical and Statistical Psychology*, *66*(1), 103–126.
- Voelkle, M. C., Oud, J. H., Davidov, E., & Schmidt, P. (2012). An sem approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological methods*, *17*(2), 176.
- Voelkle, M. C., Oud, J. H., von Oertzen, T., & Lindenberger, U. (2012). Maximum likelihood dynamic factor modeling for arbitrary n and t using sem. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*(3), 329–350.
- von Leupoldt, A., Riedel, F., & Dahme, B. (2006). The impact of emotions on the perception of dyspnea in pediatric asthma. *Psychophysiology*, *43*(6), 641–644.
- Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2004). Estimation and interpretation of $1/f$ α noise in human cognition. *Psychonomic bulletin & review*, *11*(4), 579–615.
- Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2005). Human cognition and a pile of sand: A discussion on serial correlations and self-organized criticality. *Journal of Experimental Psychology: General*, *134*(1), 108.
- Wild, B., Eichler, M., Friederich, H.-C., Hartmann, M., Zipfel, S., & Herzog, W. (2010). A graphical vector autoregressive modelling approach to the analysis of electronic diary data. *BMC medical research methodology*, *10*(1), 28.
- Windt, J., Arden, C. L., Gabbett, T. J., Khan, K. M., Cook, C. E., Sporer, B. C., & Zumbo, B. D. (2018). Getting the most out of intensive longitudinal data: A methodological review of workload–injury studies. *BMJ open*, *8*(10), e022626.
- Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122.
- Zheng, Y., Wiebe, R. P., Cleveland, H. H., Molenaar, P. C., & Harris, K. S. (2013). An idiographic examination of day-to-day patterns of substance use craving, negative affect, and tobacco use among young adults in recovery. *Multivariate behavioral research*, *48*(2), 241–266.
- Zink, J., Nicolo, M., Imm, K., Ebrahimian, S., Yu, Q., Lee, K., . . . Belcher, B. (2020). Interstitial glucose and subsequent affective and physical feeling states: A pilot study combining continuous glucose monitoring and ecological momentary assessment in adolescents. *Manuscript under review*.
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, *67*(2), 301–320.