Recovering Within-Person Dynamics from Psychological Time Series

Jonas Haslbeck*†1 and Oisín Ryan*‡2

[1]Psychological Methods Group, University of Amsterdam
[2]Department of Methodology and Statistics, Utrecht University

This manuscript has been published open access in *Multivariate Behavioral Research*
Haslbeck, J.M.B. Ryan, O. (2021). Recovering Within-Person Dynamics from Psychological Time Series. *Multivariate Behavioral Research*. DOI: 10.1080/00273171.2021.1896353

Abstract

Idiographic modeling is rapidly gaining popularity, promising to tap into the within-person dynamics underlying psychological phenomena. To gain theoretical understanding of these dynamics, we need to make inferences from time series models about the underlying system. Such inferences are subject to two challenges: time series models will arguably always be misspecified, meaning it is unclear how to make inferences to the underlying system; and second, the sampling frequency must be sufficient to capture the dynamics of interest. We discuss both problems with the following approach: we specify a toy model for emotion dynamics as the true system, generate time series data from it, and then try to recover that system with the most popular time series analysis tools. We show that making straightforward inferences from time series models about an underlying system is difficult. We also show that if the sampling frequency is insufficient, the dynamics of interest cannot be recovered. However, we also show that global characteristics of the system can be recovered reliably. We conclude by discussing the consequences of our findings for idiographic modeling and suggest a modeling methodology that goes beyond fitting time series models alone and puts formal theories at the center of theory development.

1 Introduction

Idiographic modeling is rapidly gaining popularity, both in response to concerns about the validity of inferences from cross-sectional data to within-person processes (Fisher, Medaglia, & Jeronimus, 2018; Hamaker, 2012; Molenaar, 2004), and due to the increased availability of intensive longitudinal data (Miller, 2012). A central promise of idiographic models is that they allow us to tap into the system of within-person dynamics underlying psychological phenomena (e.g., Fisher et al., 2018; Hamaker & Wichers, 2017; Wichers, 2014). With this aim in mind, many studies have used statistical time series models, such as the Vector Autoregressive (VAR) model, to investigate psychological and psychiatric phenomena (e.g., Bak, Drukker, Hasmi, & van Os, 2016; Bringmann et al., 2013; Curtiss, Fulford, Hofmann, & Gershon, 2019; Fisher, Reeves, Lawyer, Medaglia, & Rubel, 2017; Groen et al., 2019; Hasmi et al., 2017; Klippel et al., 2017, 2018; Kroeze et al., 2016; Lee et al., 2017; Pe et al., 2015; Snippe et al., 2017; van der Krieke et al., 2017; van Winkel et al., 2017; Vrijen, Hartman, Van Roekel, De Jonge, & Oldehinkel, 2018; J. Wigman et al., 2015).

However, the final goal of this type of research is typically not to fit a statistical time series model. Instead, it is to further our theoretical understanding of the within-person dynamics underlying the phenomenon at hand, which allows us to explain, predict and control the dynamics

*Both authors contributed equally to this work.
†jonashaslbeck@gmail.com | www.jonashaslbeck.com
‡ryanoisin@gmail.com | https://ryanoisin.github.io
Note that this is a revised version of the manuscript "Recovering Bistable Systems from Psychological Time Series" (https://psyarxiv.com/kcv3s)

(e.g., Haslbeck, Ryan, Robinaugh, Waldorp, & Borsboom, 2020). This raises the question of how to make inferences from time series models to an underlying system of within-person dynamics. Such inferences are subject to two fundamental challenges. First, the true system will most likely be more complicated than the time series model at hand, which means that the latter is *misspecified*. Misspecification is a problem, because it implies that it is generally unclear how to make inferences from the parameters of time series models to characteristics of the true system. Second, if we are to make any inferences about within-person dynamics the *sampling frequency* of measurements has to match the time scale at which those dynamics operate. If it is too low, the dynamics are not captured in the data and consequently cannot be inferred from a time series model. Of course, misspecification and insufficient sampling frequency are problems that are pervasive in all situations in which one's aim is to identify an underlying system from data. This generality makes it difficult to derive results that are true in all situations. Instead, results likely depend on the particular class of system and the particular class of time series model at hand. While this makes it difficult to study these inference problems, understanding them is essential to successfully constructing theories of within-person dynamics.

The goal of this paper is to illustrate these two fundamental problems to the growing community of applied researchers that aims to study within-person dynamics with intensive longitudinal data. To do so, we adopt the following simulation approach: we define a simple but non-trivial within-subjects model as the true system, and then try to recover this system with the methodology typically employed in the psychological time series literature. The idea behind this approach is that if we run into problems with this simple system, then these problems are unlikely to go away when studying more complicated systems. This approach is similar to the one of Lazebnik (2002) who studied whether the methods of biologists allow them to fix a radio, and Jonas and Kording (2017) who studied whether the methods of neuroscientists allow them to understand a microprocessor. These papers led to re-evaluations of the methods used to recover systems in these disciplines, and our hope is that our paper can contribute to a similar discussion in the idiographic modeling literature in psychology.

To study the problems of misspecification and low sampling frequency in this paper, we use a toy model for emotion dynamics, which can switch between two emotional states (cf. van de Leemput et al., 2014). We chose this system for two reasons: First, bistability is a frequently theorized property of psychological phenomena (e.g., Borsboom, 2017; Cramer et al., 2016; Cramer, Waldorp, Van Der Maas, & Borsboom, 2010; Kalisch et al., 2019; Nelson, McGorry, Wichers, Wigman, & Hartmann, 2017; van de Leemput et al., 2014; Wichers, Schreuder, Goekoop, & Groen, 2019; Wichers, Wigman, & Myin-Germeys, 2015); and second, although it is relatively simple, it is complex enough to render the most common time series models misspecified. We introduce this system in Section 2. Our general approach to investigate the problems of misspecification and insufficient sampling frequency is to simulate time series data from our true system and attempt to infer characteristics of the system from statistical time series models.

We tease apart the problems caused by misspecification and low sampling frequency by using a time series with a very high sampling frequency in Section 3, which allows us to investigate the extent to which one can make inferences from parameters of time series models (such as the VAR model) to the characteristics of an underlying system, if the time series models are misspecified. In Section 4, we reduce the sampling frequency to a level that is typical for studies using the Experience Sampling Method (ESM) and discuss the consequences this has for making inferences from time series models about the true system. We find that misspecifaction is a fundamental barrier for making straightforward inferences from time series models to the underlying system. We also show that if the sampling frequency is too low, the underlying system cannot be recovered in principle. However, we also show that some aspects of the global behavior of the system can be recovered despite misspecification and insufficient sampling frequency. In Section 5 we discuss the consequences of our findings for idiographic modeling and suggest to adopt a more general modeling methodology that puts formal theories at the center of theory development.

2 A Bistable Dynamical System for Emotion Dynamics

We begin by introducing the dynamical systems model which we will use as the true system throughout the paper. We will describe the dynamics of this system, how we can generate data from it, and finally the characteristics of this system which we would hope to infer from statistical time series models.

2.1 Model Specification

The system we will study in this paper is a bistable toy model of emotion dynamics. We chose a bistable system because this class of system has received considerable attention in the psychological literature (e.g., Borsboom, 2017; Cramer et al., 2016; Kalisch et al., 2019; Nelson et al., 2017; van de Leemput et al., 2014; Wichers et al., 2019; Wichers, Wigman, & Myin-Germeys, 2015). Bistable systems have two stable states, which can be interpreted as different psychological states such as "healthy" or "unhealthy" (e.g., depressed). These types of system are often formalized within the framework of differential equations, (e.g., Hirsch, Smale, & Devaney, 2012; Strogatz, 2015), a formalization we will adopt in the current paper. Differential equations describe dynamics in terms of the derivative $\frac{dx_i}{dt}$, that is, the rate-of-change of each variable x_i with respect to time. For an introduction to the interpretation of differential equation models in a psychological setting, see for example Boker (2002), Boker, Montpetit, Hunter, and Bergeman (2010) or Ryan, Kuiper, and Hamaker (2018). To produce bistable behaviour these differential equation models must contain non-linear terms, for example in the form of (product) interaction effects between variables (Strogatz, 2015).

Specifically, we choose a four-variable generalization of the classic Lotka-Volterra model for competing species (e.g., Freedman, 1980), which has been used previously as the basis for a toy model for emotion dynamics by van de Leemput et al. (2014). The variables in the system represent two emotions with positive valence (Cheerful (x_1) and Content (x_2)) and two emotions with negative valence (Anxious (x_3) and Sad (x_4)) with a value of zero interpreted as the absence of that emotion. The dynamics of the system are defined by the stochastic differential equations shown in the left panel of Figure 1:

$$\frac{dx_i}{dt} = 1.6 + x_i + \sum_{j=1}^{4} C_{ij} x_j x_i + \sigma \frac{dW_i}{dt}$$

$$W_{i,t} \sim \mathcal{N}(0, \Delta t), \quad \sigma = 4.5$$

$$\boldsymbol{C} = \begin{bmatrix} -0.2 & 0.04 & -0.2 & -0.2 \\ 0.04 & -0.2 & -0.2 & -0.2 \\ -0.2 & -0.2 & -0.2 & 0.04 \\ -0.2 & -0.2 & 0.04 & -0.2 \end{bmatrix}$$

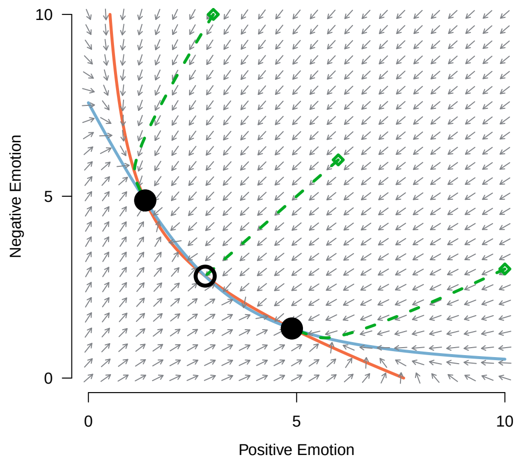

Figure 1: Left panel: Model equations and parameters of the bistable emotion system. Right panel: the vector field defined by the parameters in the left panel. Solid points indicate stable fixed points and the empty point indicates an unstable fixed point. The solid lines indicate the values at which derivative of positive emotion (orange) and negative emotion (light blue) is equal to zero. At the points at which the two lines meet, both derivatives are equal to zero and the system remains in this (stable) state. The green dashed lines illustrate three trajectories that the system can take through the vector field.

This equation defines for each emotion variable how it changes (dx_i) over a infinitesimal time step (dt). In other words, the differential equations specify the local or micro-dynamics of the system. The parameters defining the rate-of-change $\frac{dx_i}{dt}$ can be interpreted similarly to a standard regression model. The constant term 1.6 ensures that the emotion variables (with high probability) take on only positive values. The second term defines the linear main effect of an emotion x_i on its own rate-of-change, with a regression parameter of 1. The matrix $\boldsymbol{C}$ represents the dependencies between emotions in the form of interaction effects. Emotions of the same valence reinforce each other, with for example $C_{12} = 0.04$ indicating that the rate-of-change of x_1 (Cheerful) depends on the product of x_1 and x_2 (Content) weighted by 0.04. Emotions of different valence suppress each other, with $C_{13} = -0.2$ indicating that the rate-of-change of x_1 depends on the product of x_1 and x_3 (Anxious) weighted by -0.2. The diagonal elements are quadratic effects, indicating that the rate-of-change of x_i depends on x_i^2 weighted by -0.2. Finally, the term $\sigma \frac{dW_i}{dt}$ represents the stochastic part of the model in the form of a so-called Wiener process. This is essentially a differential-equation version of a standard Gaussian perturbation term with independent increments, representing the short-term fluctuations in emotions caused by the environment the system interacts with. The size of the stochastic input to the system is determined by the time-interval between realizations of the process Δt and σ. For example, if we were to generate data from this model with a step size of $\Delta t = 1$, we would have Gaussian noise with $\mathcal{N}(0, \sigma^2)$.

Why did we choose this specific model to illustrate the problems of misspecification and insufficient sampling frequency? First, it is misspecified with respect to the time-series models used by substantive researchers, yet it is similar in form to those models. For instance, similar to the popular VAR model, it specifies dynamic relationships in terms of time-lagged relationships, though here in differential-equation form, akin to the continuous-time version of the VAR model (cf. Ryan & Hamaker, 2020; Voelkle & Oud, 2013). However, it is also misspecified with respect to the VAR model as it contains additional interaction terms. We hope that the familiarity of many readers with the VAR model will help them to understand this true system, and the similarity of the true model to the VAR model allows us keep discussion of the effects of misspecification concise. Second, our true model exhibits bistability, which is a property that is interesting theoretically to many psychological researchers.

To get an idea about the qualitative dynamics of the system we first focus on the deterministic part of the differential question, that is, the equation in Figure 1 without the stochastic term $\sigma \frac{dW_i}{dt}$. Since the matrix $\boldsymbol{C}$ is symmetric, we are able to collapse the system into a 2-dimensional system consisting of only positive emotions (PE) and negative (NE) emotions. This allows us to study the qualitative dynamics of the system using a 2-dimensional *vector field*, in which arrows indicate how the system is expected to change given any combination of positive and negative emotion values at a particular point in time (see right panel of Figure 1). Perfectly horizontal arrows in the vector field indicate where we would expect no change in the negative emotion dimension at the next time point $dNE/dt = 0$, indicated by the blue *solution line*, while perfectly vertical arrows indicate no change in the positive emotion dimension $dPE/dt = 0$, designated by the orange solution line. The locations at which these lines cross indicate *fixed points* (also known as *resting states* or *equilibrium positions*) of the system. At these locations, both derivatives are equal to zero, which means that the values of the variables in the system will not change anymore once they have reached this location.

From Figure 1 we can see that the system exhibits three fixed points: Two are *stable*, located at ($PE = 1.36$, $NE = 4.89$) and ($PE = 4.89$, $NE = 1.36$), which could be characterized as unhealthy (low positive, high negative emotion) and healthy (high positive, low negative emotion) stable states of the system. If the system takes on any value above the diagonal ($NE > PE$) it will eventually return to the unhealthy fixed point, and if it takes on any value below the diagonal ($NE < PE$) it will return to the healthy fixed point. The third fixed point, ($PE = 2.80$, $NE = 2.80$) represents an *unstable* fixed point: If the system starts exactly on the diagonal of the vector field ($PE = NE$) it will return to this fixed point, but any deviation will cause the trajectory to veer off towards one of the stable fixed points. The behavior of the system can be read off the vector field by starting at a given point and following its arrows. The three green lines in Figure 1 show the trajectory of the system for three different starting points.

We have seen that in the deterministic system, no matter what value the emotion variables take on initially, the system eventually moves to one of the fixed points and stays there indefinitely. However, this is not a realistic model, since the emotions of a person are continuously changing over time. Rather than staying at a particular fixed point, they are likely to move around the two fixed points. To allow our model to show such behavior, we add the stochastic noise term $\sigma \frac{dW_i}{dt}$ to the model. This stochastic version of the bistable system will generally fluctuate around either the healthy or unhealthy fixed point, but occasionally the noise term will be large enough to "push" the system from one stable fixed point to another, that is, the noise will cause the system to switch from the healthy to unhealthy state or vice versa. The frequency with which this switching occurs is a function of the distance between the two fixed points, the form of the vector field in the areas between the two fixed points, and the variance of the Gaussian noise process, as determined by the σ parameter. If the noise variance is low, the probability of a noise draw that is large enough to "push" the system to the other fixed point is small, and consequently the frequency of switching is low. In contrast, if the variance is high, the probability of a large enough noise draw to switch to the other fixed point is high, and consequently the switching frequency is high. Here we choose $\sigma = 4.5$ to give a relatively high number of switches in the time series data we generate from this model. We describe this data generation in the following section, which will also allow us to visualize the behaviour of the full stochastic system.

2.2 Generating time series

We now show how to generate time series data from the model specified above, allowing us to illustrate the behavior of the system of differential equations including noise. We generated data by computing the numerical solution to the model described in the left-hand panel of Figure 1 on the interval $[0, 20160]$. We interpret a unit of time $t = 1$ as one minute, and therefore the time series spans two weeks ($60 \times 24 \times 14 = 20160$). To generate the data we use Euler's method (e.g., Atkinson, 2008), with a step size of $\Delta t = 0.01$. We chose this step size to limit computational cost and disk space, however the system shows qualitatively the same behavior for smaller step sizes. We obtain a time series dataset by sub-sampling the numerical solution obtained via Euler's method 10 times per minute (that is, every six seconds). We therefore obtain a time series dataset with $20160 \times 10 = 201600$ measurements, which is shown in Figure 2. The code to generate data and reproduce all results and figures shown in this paper can be found at `https://github.com/jmbh/RecoveringWithinPersonDynamics`.

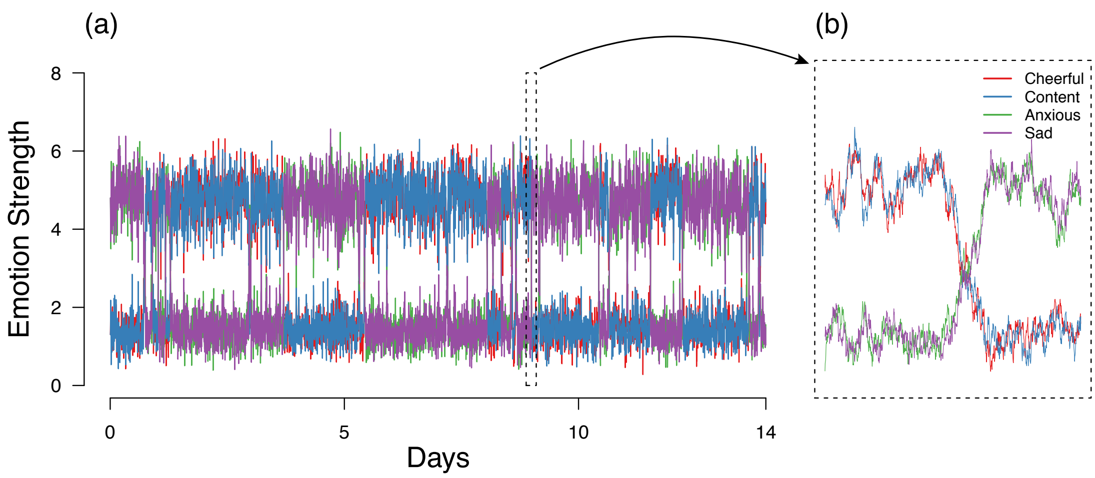

Figure 2: Panel (a) shows the "ideal" time series of the four emotion variables Cheerful, Content, Anxious and Sad. We see that the system switches 17 times between healthy and unhealthy state. Panel (b) displays the twelfth switch, which is a transition from the unhealthy to the healthy state, which occurs on day 9.

This time series can be seen as an "ideal" time series since it has been measured with an extremely high sampling frequency of every six seconds, a continuous response scale, no measurement error or missing values, and frequent switches. This implies that sampling variation plays essentially no role when estimating models based on this time series dataset. From Figure 2, we can see that the configuration of the toy model we have chosen ensures that we obtain a time series which appears to switch approximately 17 times between both stable fixed points in the two week interval. This represents again an idealistic scenario, as observing this many switches between states gives us the best possible chance to correctly characterize the full system. We use these "ideal" time series data in Section 3 to study to what extent one can make inferences from misspecified statistical time series models to the characteristics of the true model if the sampling frequency is high enough. In Section 4 on insufficient sampling frequency we will use a time series that is measured with a much lower sampling frequency.

2.3 Qualitative Characteristics of the True System

In Section 2.1 we defined the true system with a set of four differential equations. These equations define the rate-of-change of each variable x_i given the state of all variables (including x_i itself) and therefore represent the local or micro-dynamics of the system. In the remainder of this paper we aim to use statistical time series models to make inferences about the true system. These statistical time series models are misspecified and we therefore already know that it is impossible to recover the exact local dynamics of the true system. However, it is conceivable that one can make inferences about qualitative characteristics of the true system that are less specific than the exact local dynamics. To be able to evaluate to which extent this is indeed possible, we define a number of general characteristics that describe the true system. We divide those general characteristics in local and global characteristics:

Local Characteristics

1. Suppressing effects between valences, reinforcing effects within valences
2. Relative size of suppressing/reinforcing effects
3. All parameters are independent of time and independent of variables outside the model

Global Characteristics

4. Bistability (two stable fixed points)
5. Position of stable fixed points
6. Variability around fixed points
7. Frequency of transitions

The local characteristics describe qualitative properties of the local dynamics: The first characteristic is that emotions of the same valence reinforce each other, while emotions of different valence suppress each other. The second characteristic is the fact that the size (absolute value) of the reinforcing effects (0.04) are smaller than the suppressing effects (-0.20). The third characteristic is that all parameters in the system of differential equations are independent of time and independent of variables outside the model. That is, the parameters of the model given in Figure 1 remain fixed over the window of observation.

We describe the remaining four characteristics as global in the sense that they do not refer to a specific property of the local dynamics, but to the overall qualitative behavior of the system (which are, of course, produced by the local dynamics). The first global characteristic is bistability, which means that the data generating mechanism exhibits two stable fixed points. The second characteristic is the position of the stable fixed points, which are at (PE = 4.89, NE = 1.36) for the healthy fixed point, and (PE = 1.36, NE = 4.89) for the unhealthy fixed point. Third, we consider the variability around the different fixed points. Figure 2 shows that, for both fixed points, the variability of the emotions with lower values is smaller than the variability of the emotions with larger values. The final characteristic is the frequency of transitions between the area around the healthy fixed point and the area around the unhealthy fixed point. In the time series shown in Figure 2 we see that the system switches around 17 times in a two week period.

3 The Problem of Misspecification

We study the problem of misspecification by trying to recover the characteristics of the true system with the most popular and some more advanced analysis strategies for time series used in psychological research. We begin by using simple descriptive statistics and data visualizations, before analyzing the data with the well-known Hidden Markov Model (HMM) and Vector Autoregressive (VAR) model, and finally a less commonly used regime-switching extension of the VAR model known as the Threshold VAR (TVAR) model. We chose the VAR model, because it has become an extremely popular model for intensive longitudinal data (see Appendix E). HMMs are also a popular class of time series models in psychological research (Asparouhov, Hamaker, & Muthén, 2017; de Haan-Rietdijk et al., 2017; Neale, Clark, Dolan, & Hunter, 2016; Visser, 2011) and the TVAR model is an interesting extension of the VAR model (De Haan-Rietdijk, Gottman, Bergeman, & Hamaker, 2016; Hamaker, Grasman, & Kamphuis, 2010, 2016; Hamaker, Zhang, & van der Maas, 2009), which allows us to discuss how theoretical input can mitigate the problem of misspecification. Each of these models is misspecified, which means that the true system is not a special case of the model at hand. For each model, we examine whether it is possible to recover the core characteristics of the dynamical system. To study the problem of misspecification independently of the problem of insufficient sampling frequency, we use the high sampling frequency time series described in Section 2.2. This means that we have essentially the best data one could hope for to recover within-person dynamics, and therefore any problems with recovery have to be due to misspecification.

3.1 Descriptive Statistics and Data Visualization

We begin by analyzing the data with simple descriptive statistics and data visualizations. As a first step we inspect the histograms of each variable which are shown in the top panel of Figure 3. We see that each of the variables is bimodal, appearing roughly like a mixture of two normally distributed variables. These distributions are roughly centered around the values 1.6 and 4.8, and the variance of the higher-valued distribution seems to be slightly larger. This kind of analysis can be made more sophisticated in many ways. For example, separating the overall density into two distributions would allow us to calculate means and variances, as well as calculating how frequently observations appear in each of the two distributions. We will formalize these ideas with a statistical model in the following subsection, but for now we study the underlying system with additional visualizations.

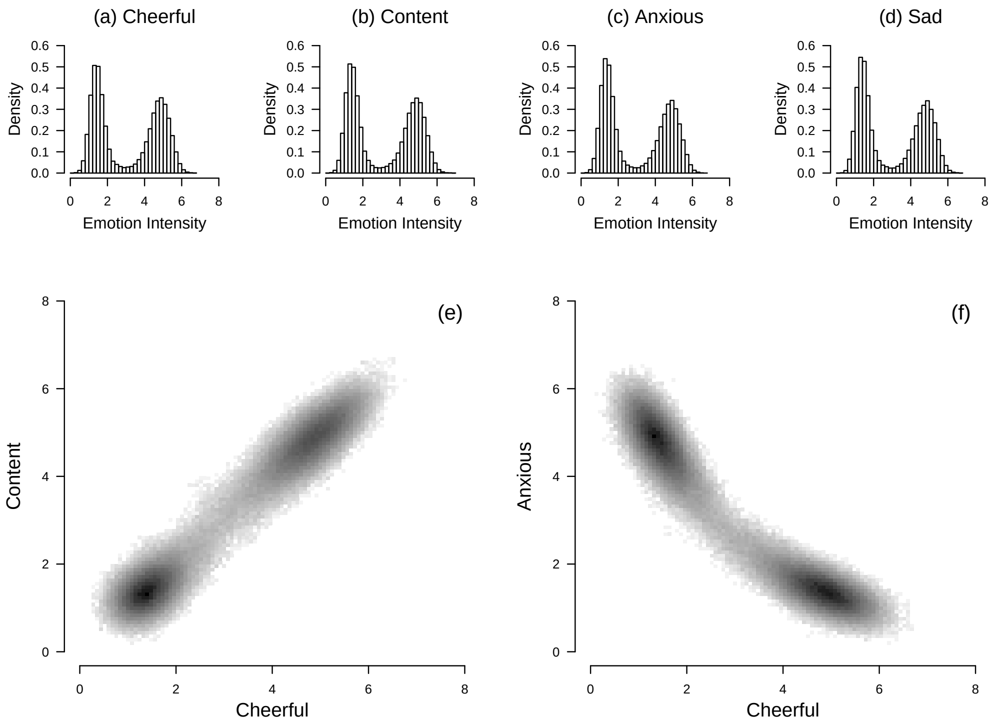

Figure 3: Top panel: The histograms of the four emotion variables Cheerful, Content, Anxious and Sad. Lower panel: The bivariate relationships between Content and Cheerful, and between Anxious and Cheerful. All other bivariate relationships are similar to the ones in panel (e) and (f), due to the symmetry in the true system.

Going beyond the univariate distributions, we inspect the bivariate relationships between emotion variables at the same time point. Panel (e) shows the relationship between Cheerful and Content, two emotions with the same (positive) valence. We see that most density is concentrated at values in which both emotions are either high, or low. This means that the system is mostly in a state in which Cheerful and Content are both high, or both low. Some observations are also in-between the two states, which suggests that the system is occasionally switching between those two states. Panel (f) shows the relationship between Cheerful and Anxious, two emotions with different valence. We see that the system is mostly in a state in which Cheerful is high and Anxious is low, or the other way around. These two clusters clearly indicate that the system is bistable, and enable us to get a rough indication of the locations and variance around both fixed points. If we plot the data as a function of time as in Figure 2 we can additionally gauge how frequently the system switches between the states at which the system stays most of the time.

Taking all results together, which characteristics of the true system did we recover? We clearly recovered the fact that the system is bistable and we got a rough idea about the location and variance of the distributions at the two fixed points. By inspecting Figure 2 we were also able to gauge the frequency of switches. To provide a more quantitative picture of the fixed points and the variance around them, in the following subsection we fit a Hidden Markov Model (HMM).

3.2 Mean Switching Hidden Markov Model

In the previous subsection we eyeballed the location and variance of distributions around fixed points, their relative frequency, and how often the system appeared to switch from one distribution to the other. We now obtain numerical estimates of these quantities by employing a more rigorous approach. Specifically, we will fit a Hidden Markov Model (HMM). In an HMM observations are drawn from a mixture of distributions, where at each time step there is some estimated probability of switching from one distribution component to another, encoded by a transition matrix. To fit the HMM, we must specify a particular type of distribution, and the number of components we believe to make up the mixture.

Here we chose a Gaussian distribution with diagonal covariance matrix and $K = 2$ components. For more details on this type of model see Zucchini, MacDonald, and Langrock (2017). We fit the model using the R-package *depmixS4* (Visser & Speekenbrink, 2010) and obtained the following estimates for the means and standard deviations for the two components k_1 and k_2 and the transition matrix $\hat{\boldsymbol{M}}$:

$$\hat{\boldsymbol{\mu}}^{(1)} = \begin{pmatrix} 1.47 \\ 1.46 \\ 4.71 \\ 4.71 \end{pmatrix}, \hat{\boldsymbol{\sigma}}^{(1)} = \begin{pmatrix} 0.41 \\ 0.40 \\ 0.63 \\ 0.62 \end{pmatrix}, \hat{\boldsymbol{\mu}}^{(2)} = \begin{pmatrix} 4.75 \\ 4.76 \\ 1.45 \\ 1.45 \end{pmatrix}, \hat{\boldsymbol{\sigma}}^{(2)} = \begin{pmatrix} 0.63 \\ 0.62 \\ 0.40 \\ 0.40 \end{pmatrix}, \hat{\boldsymbol{M}} = \begin{matrix} \\ k_1 \\ k_2 \end{matrix} \begin{matrix} \begin{matrix} k_1 \quad\quad & k_2 \end{matrix} \\ \begin{pmatrix} 0.9996 & 0.0004 \\ 0.0004 & 0.9996 \end{pmatrix} \end{matrix}.$$

The entries in the parameter vectors refer to the four variables Cheerful, Content, Anxious and Sad. When inspecting the means $\hat{\boldsymbol{\mu}}^{(1)}, \hat{\boldsymbol{\mu}}^{(2)}$ of the two components, we see that the HMM picked up two states: one in which the positive emotions are low and negative emotions are high (State 1); and one in which the reverse is true (State 2). The HMM also picked up that the variance is lower for variables that are in a state in which their values are lower. The HMM identified the same qualitative features as eyeballing the visualizations in the previous section. However, it provides a more principled way to estimate the means and variances of the components. Indeed, the mean estimates of the two components are very close to the true fixed points of the system (see Section 2.1).

In addition, the HMM predicts for each data point to which state (or components) it belongs and also provides a transition matrix between states $\hat{\boldsymbol{M}}$. Figure 4 displays the full time series together with the predicted state for each time point (grey/white shading). When inspecting the predicted components visually, it seems that the HMM captured the switches well. Next to the larger blocks in which the system stays in the same state, it also identifies switches in which the system switches back and forth within only a few time points. These switches might have been missed when merely eye-balling the time series. Since the system mostly stays in the same state and switches only occasionally, the probabilities on the diagonal of the transition matrix $\hat{\boldsymbol{M}}$ are much larger than on the off-diagonal elements.

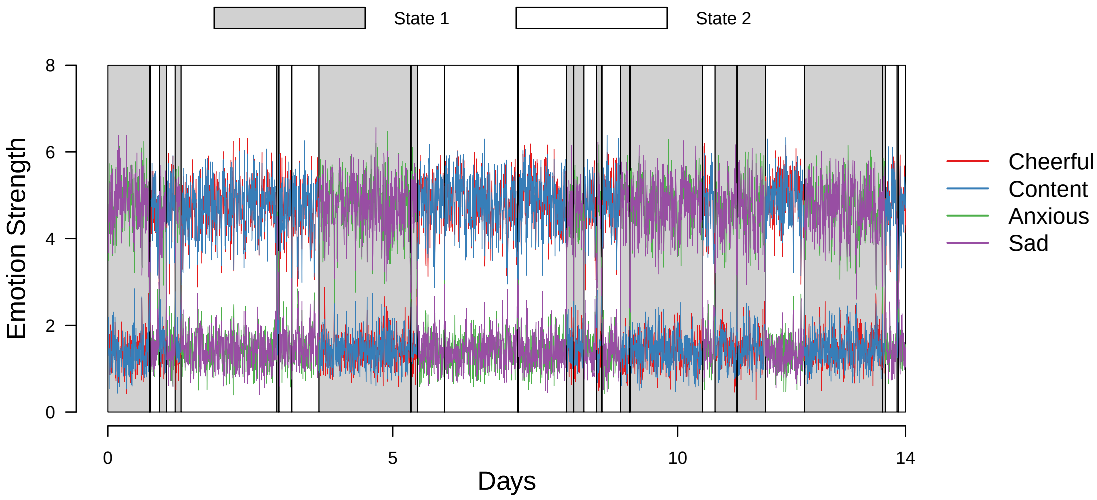

Figure 4: Time series of the four emotion variables, also shown in panel (a) of Figure 2, with background color indicating whether a given time point is assigned to state 1 or state 2 of the mean-switching HMM.

Taking all results together, which characteristics of the true system did we recover? Qualitatively, we reached the same conclusions as with the visualizations in the previous sections. However, by going beyond mere eyeballing and fitting a HMM, we obtained estimates of the location and variance of the distributions around fixed points, we obtained a transition matrix, and we were able to predict for each time point to which component it belongs. This means that we successfully captured the global characteristics, however we did not capture any local characteristics. We turn to those next.

3.3 Lag-1 Relationships & VAR Model

We now turn to the local dynamics describing the temporal relations between the four emotion variables at a very small time scale. We begin by visualizing the temporal relationships between pairs of variables over two subsequent time steps, which are in the present dataset six seconds apart. These relationships are also modeled by the VAR model, with the only difference that in the VAR model these relationships are conditional on all other variables at the previous time point. Figure 5 displays this temporal relationship between Cheerful_{t-1} and Content_t in panel (a), and the relationship between Content_{t-1} and Anxious_t in panel (b). We first consider the lagged relationship between Cheerful_{t-1} and Content_t in panel (a). We see that most of the density is where both variables have either high or low values. Note that the densities look quite similar to the ones of the lag-0 relationships shown in the lower panels of Figure 3. This is because the sampling variance is extremely high and therefore observations do not change much over one time step. Fitting a linear regression model we find a strong positive relationship between these two lagged variables $\rho = 0.98$ (red line).

What can we conclude from this strong positive lagged relationship? It might be tempting to conclude that Cheerful has a positive linear effect on the rate-of-change of Content. However, this would be incorrect, because the relationships in the true system include product interaction terms. Instead, this parameter represents the best linear *approximation* to the lagged relationship, so perhaps we can at least use it to infer a more coarse local characteristic such as whether the relationship between the two variables is suppressing or reinforcing. For this bivariate relationship, it turns out that we would correctly infer the local characteristic of the true system that emotions of the same valence have a reinforcing effect on each other. The same argument applies to the relationship between Content_{t-1} and Anxious_t, except that the best linear approximation is negative, and that we would correctly conclude that emotions with different valences suppress each other. However, the problem with this type of inference is that we do not know when it actually turns out to be correct. We illustrate this problem with the popular lag-1 Vector Autoregressive (VAR) model.

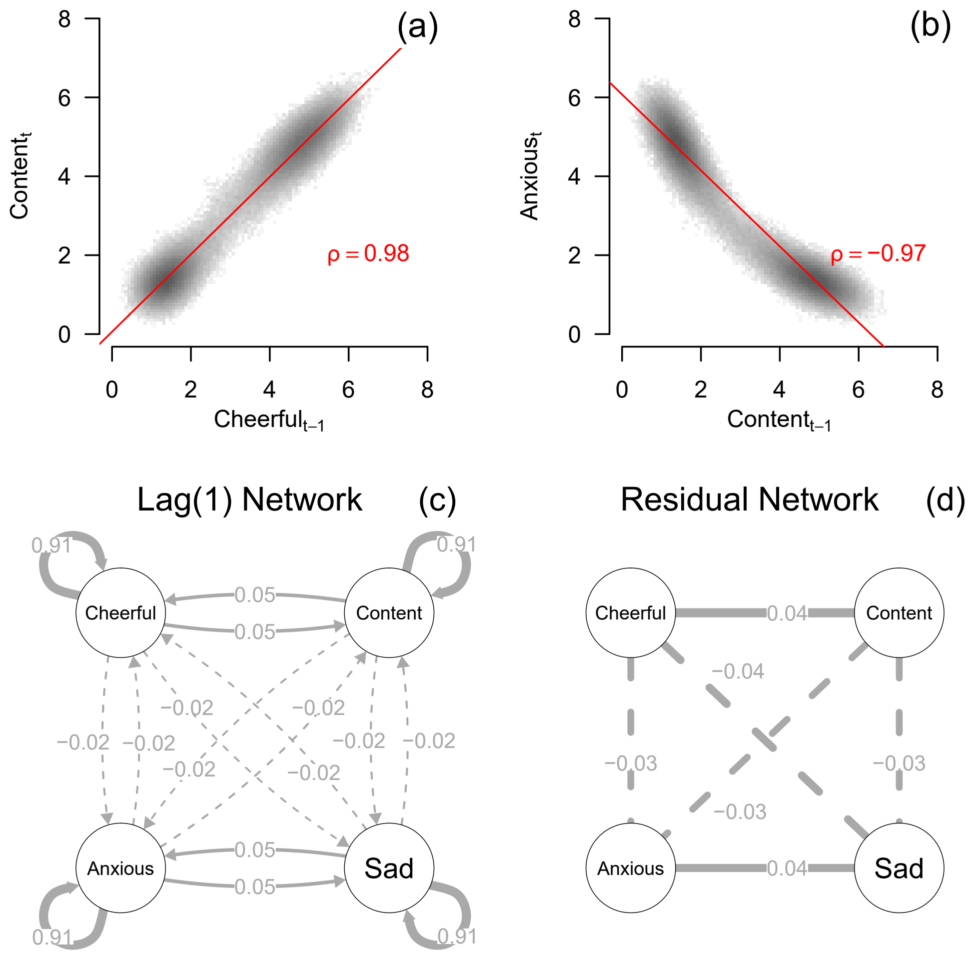

Figure 5: Panel (a) shows the relationship between Content and Cheerful, two emotions with the same valence, spaced one time point apart (at a lag of one). The red line indicates the best fitting regression model. Similarly, panel (b) shows the relationship between Anxious and Content, two emotions with different valence, at a lag of one. Panel (c) displays the matrix of lagged regression parameters, estimated from a VAR(1) model, as a network, and panel (d) displays the partial correlation matrix of the residuals of the VAR(1) model as a network. This latter network is often referred to as the contemporaneous network.

The VAR(1) model predicts each variable X_i at time t by a linear combination of all variables (including X_i) at the previous time point, written in vector form as

$$\boldsymbol{X}_t = \boldsymbol{\alpha} + \boldsymbol{\Phi}\boldsymbol{X}_{t-1} + \boldsymbol{\varepsilon}_t \ , \tag{1}$$

where $\boldsymbol{\alpha}$ is a vector of intercepts, $\boldsymbol{\Phi}$ is a matrix containing the auto-regressive (ϕ_{ii}) and cross-lagged ($\phi_{ij}, i \neq j$) effects, and $\boldsymbol{\varepsilon}_t$ is a vector of normally distributed residuals $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, \boldsymbol{\Psi})$, which are independent across time, with residual variance-covariance matrix $\boldsymbol{\Psi}$.

Panel (c) of Figure 5 displays the estimated parameters in the $\boldsymbol{\Phi}$-matrix, and the estimated intercepts are $\boldsymbol{\alpha} = \{0.27, 0.28, 0.26, 0.26\}$. We see that there are strong auto-regressive effects which could be taken as evidence that each variable has a strong effect on its own rate-of-change, which is indeed the case in the true model. In addition, we see again that there are positive linear lagged effects between emotions with the same valence, and negative lagged effects between emotions with a different valence. We also see that the effects within valence are about twice as large as the effects between valence, which could lead us to conclude that the reinforcing effects in the true model are stronger than the suppressing effects. However, we know that this is incorrect. Indeed, in the true model the reverse is true. This illustrates the fact that there is no guarantee that naive inferences from the parameter estimates in misspecified time series models to characteristics of the true model need to be correct. The source of misspecification in this case is the presence of non-linear relationships in the true system, relationships which the VAR model approximates with linear effects. This becomes clear when comparing the VAR model in equation (1) with the true bistable system defined on the left-hand side of Figure 1: while the VAR model does contain a constant term, linear self-dependency and Gaussian perturbation variance term, it fails to incorporate the non-linear product term $C_{ij}x_jx_i$. Note that, although there are well known complications to making inferences from (discrete-time) VAR parameters to differential equation models (Kuiper & Ryan, 2018; Ryan & Hamaker, 2020; Voelkle, Gische, Driver, & Lindenberger, 2018), fitting the continuous-time equivalent to the VAR model will not overcome the issue of model misspecification in this case (for details see Appendix B).

Panel (d) of Figure 5 shows the partial correlations between the residuals of the model. In practice, dependencies between residuals may be brought about due to multiple factors, such as the presence of unobserved common cause variables, dependencies between processes operating at a different time-scale, differences in time-intervals between observed measurement occasions, or misspecification of the functional form of the lagged dependencies. In this case, we know that misspecification of the functional form is the source of these residual dependencies, since the other issues we have listed here are absent from the true system and data generation scheme. Although the residual partial correlations succeed in flagging model misspecification, it is difficult to interpret the exact values of the partial correlations in this setting. In practice, applying a reasonable interpretation of residual structures will likely require assumptions about the absence of one or more potential sources: For example, if we are willing to accept that the data-generating mechanism is linear and first-order, and observations are equally spaced, residual partial correlations may flag the presence of unobserved common cause variables.

An additional consequence of the fact that the VAR model only includes linear effects, but the true system includes non-linear effects, is that the VAR model cannot generate data that has the same global characteristics of the data generated by the true system. The reason is that the VAR model exhibits only a single fixed point which is equal to its mean vector $\boldsymbol{\mu} = (\boldsymbol{I} - \boldsymbol{\Phi})^{-1}\boldsymbol{\alpha}$, where $\boldsymbol{I}$ is the identity matrix (Hamilton, 1994). Thus, the histograms of data generated from a VAR model will show a uni-modal distribution, which means that none of the four global characteristics can be reproduced. We illustrate this in Figure 14 in Appendix C.1 by generating two weeks of data from the VAR model estimated in this section. Checking the characteristics of the empirical data against the characteristics of the model-generated data is a valuable tool for model evaluation. This is especially the case for high-frequency data, because variables do not change much from one time point to the next. In such a situation, a VAR model can fit the data very well even though it recovers none of the characteristics of the true system.

Careful researchers may detect that the VAR model is misspecified in this type of situation, for instance by comparing empirical and model-generated data as we have here, or more generally by using diagnostic tools (such as correlation and partial correlation plots) to check whether the assumptions of the VAR model are met. This type of model exploration is crucial: At a minimum, it would indicate in this case that the VAR model parameters should be interpreted with caution, thus mitigating the possibility of drawing incorrect conclusions about the true system. Ideally, such model exploration would lead the researcher to choose a model which is less misspecified, that is, "closer" to the true system. However, we would caution that there is likely no guarantee that such a procedure will succeed. In Appendix D we show that standard application of time-series diagnostics can easily lead one to, for instance, mistakenly remove a (seasonal) trend from the data, obtaining a VAR model whose assumptions appear to be met in this transformed dataset. Although the resulting model appears to fit well to the data, it still does not help us to correctly characterize the true system. Thus, diagnostics and assumption checks should not primarily used to motivate data transformation that ensure that the assumptions of the misspecified model are met, but rather to choose models that allow us to characterize the key features of the system under investigation.

To summarize, the true system is not a special case of the VAR model, and therefore the VAR model is misspecified. In the present case the main source of misspecification is the fact that the true system includes non-linear effects, while the VAR model does not. This misspecification has the effect that we cannot make reliable inferences from parameters of statistical models to the local characteristics of the true system.

3.4 Threshold VAR

In the previous section we showed that the VAR model did not allow us to recover most key characteristics of the dynamical system. The main barrier to using the VAR model was that the true dynamics in our system are non-linear. That is, the effect of X_i now on X_j later depends on the current value of X_i. This means that a) it is difficult to infer local characteristics such as which relationships are stronger or weaker than others, and b) the VAR model in principle cannot reproduce the global characteristics related to multiple stable states. Here, we go beyond the VAR model by considering *regime-switching* VAR models, which allow observations to be drawn from two or more distinct states with different sets of parameters (Tong & Lim, 1980), potentially allowing us to overcome these limitations.

Regime-switching VAR models differ from one another in how they specify the mechanism that allows the system to switch from one state to another. For example, the Markov-Switching VAR combines the HMM and VAR models, modeling state-switching behaviour as governed by a latent Markov process. For the dynamical system we study in the current paper, the threshold VAR (TVAR) model would most closely approximate the behaviour of our true system. The TVAR describes a system which switches between states depending on the value of a thresholding variable z_t:

$$\begin{aligned} \boldsymbol{X}_t &= \boldsymbol{\alpha}^{(1)} + \boldsymbol{\Phi}^{(1)}\boldsymbol{X}_{t-1} + \boldsymbol{e}_t^{(1)} && \text{if } z_t \leqslant \tau \\ \boldsymbol{X}_t &= \boldsymbol{\alpha}^{(2)} + \boldsymbol{\Phi}^{(2)}\boldsymbol{X}_{t-1} + \boldsymbol{e}_t^{(2)} && \text{if } z_t > \tau \end{aligned}$$

where τ is known as the threshold value, a parameter which must be estimated, and $\boldsymbol{\alpha}^{(i)}, \boldsymbol{\Phi}^{(i)}$ are the intercept vector and the matrix of lagged regression coefficients in state i respectively. Notably, the thresholding variable z_t must be specified a-priori, either as a time-varying covariate, or as one of the X variables in the model itself, and, similar to the HMM, the number of states expected must be specified a-priori or tested through model selection. Here we choose to fit a TVAR model with a single threshold value, that is, two states, and use *Cheerful* ($X_{1,t-1}$) as the thresholding variable. We estimate the TVAR model using the R-package *tsDyn* (Fabio Di Narzo, Aznarte, & Stigler, 2009).

Figure 6 displays the main results from the estimated TVAR model, in which the threshold is estimated as $\hat{\tau} = 2.811$. In panel (a) of Figure 6 we show the time series and the shading indicates which observations of *Cheerful* are below (grey) or above (white) the threshold. We see that the estimated threshold does well in separating the time series into periods in which the system is in an unhealthy state (based on *Cheerful* values below the threshold) and a healthy state (*Cheerful* values above the threshold). Inspecting the lagged networks for each regime in panels (b) and (c) of Figure 6 we see that the auto-regressive effects and the within-valence cross-lagged effects are quite similar across both regimes. However, this is not the case for the cross-lagged effects between variables of opposite valence. In the healthy regime, negative valence emotions have much stronger cross-lagged effects on positive emotions ($\hat{\phi}_{13}^{(2)} = \hat{\phi}_{14}^{(2)} = \hat{\phi}_{23}^{(2)} = \hat{\phi}_{24}^{(2)} = -0.08$), and vice versa for the unhealthy regime ($\hat{\phi}_{31}^{(1)} = \hat{\phi}_{41}^{(1)} = \hat{\phi}_{32}^{(1)} = \hat{\phi}_{42}^{(1)} = -0.08$). Applying the standard way of calculating model-implied means for the VAR models above and below the threshold separately, we obtain $\hat{\boldsymbol{\mu}}_2 = \{4.74, 4.75, 1.45, 1.46\}$ for the healthy state and $\hat{\boldsymbol{\mu}}_1 = \{1.49, 1.48, 4.69, 4.69\}$ for the unhealthy state.

There are two striking aspects of the TVAR model results. The first is that the TVAR threshold value picks up on many of the global dynamics present in the system: The threshold correctly separates the time series based on whether the system is close to the healthy or unhealthy fixed point, and captures the approximate position of those two fixed points. In part, this is due to the specific parameterization of our bistable dynamical system. Due to the symmetries in the true system, the value of Cheerful happens to be a very good indicator of the multivariate position of the system: If Cheerful is high (i.e. above the unstable fixed point) Content is high and Anxious and Sad are low (below the unstable fixed point) and vice versa. Thus, although the switching behaviour in the true system is dependent on the position of the system in multivariate space, in the simple system used in this paper a univariate threshold is sufficient to capture this to a high degree of accuracy.

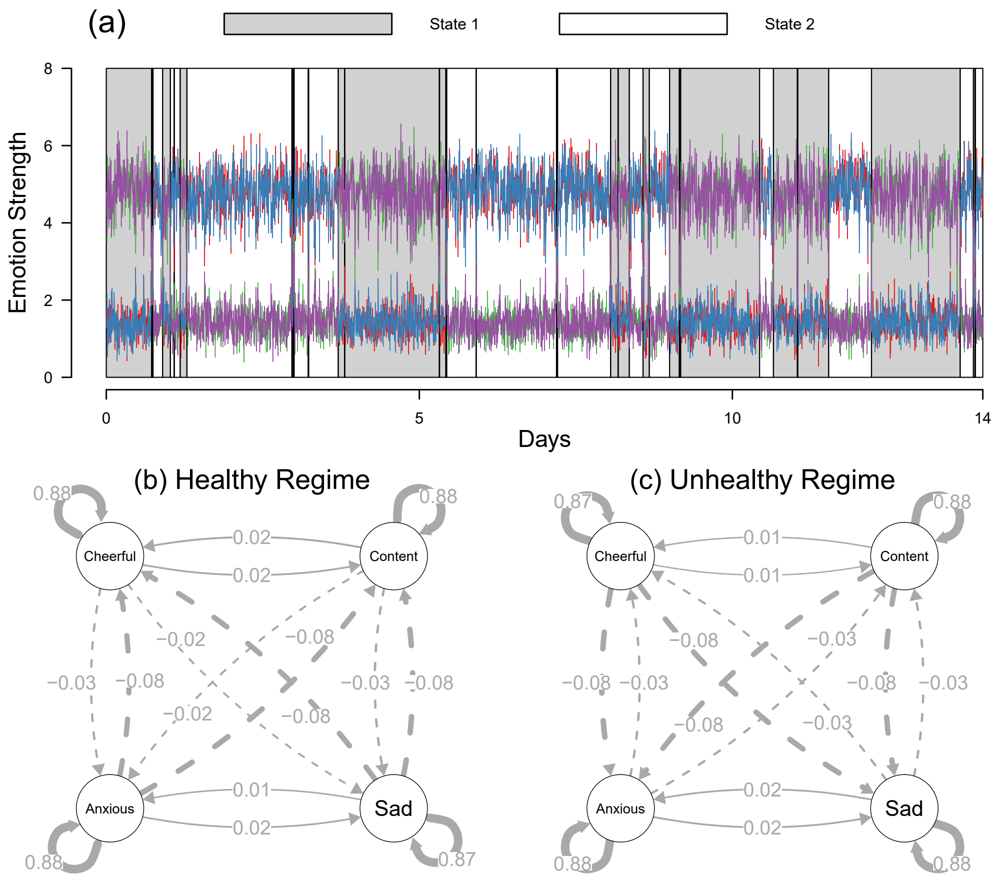

Figure 6: Panel (a) shows the two weeks of the time series, with observations shaded in either grey or white as a function of whether Cheerful_{t-1} is above or below the threshold $\hat{\tau} = 2.811$. Panels (b) and (c) show the estimated VAR(1) parameters as lagged networks in the healthy (white) and unhealthy (grey) regimes respectively.

Second, the TVAR recovers two separate asymmetric matrices of lagged parameters. These asymmetries are present because the TVAR model is approximating the continuous non-linear relationships in the true system with a step-function along the variable *Cheerful*. That is, we fail to capture that each lagged relationship is an interaction effect, dependent on the exact value of the variables involved. However, since thresholding on Cheerful succeeds in separating the two states in our system, the TVAR lagged parameters correctly pick up that the strength of each relationship differs depending on whether the system is near the healthy or unhealthy fixed point. We saw that this was indeed the case in the left hand-panel of Figure 5 (b), where the marginal relationship between Cheerful_{t-1} and Anxious_t is clearly more strongly negative near the unhealthy fixed (top left quadrant) point and less strongly negative near the healthy fixed point (bottom right quadrant). Reflecting this, in Figure 6 we see that in the unhealthy regime Cheerful has a stronger negative lagged effect on Anxious $\phi_{14}^{(2)} = -0.08$ than in the Healthy regime $\phi_{14}^{(1)} = -0.03$. In Appendix C.2 we show that these asymmetric lagged parameter matrices (in combination with the state-dependent residual covariance matrices) also succeed in reproducing the global characteristic that the variability of the emotion variables differ in each regime.

In summary, the TVAR model reproduces two regimes or states which are determined by a univariate threshold. The lagged parameters in each regime are asymmetric, showing that the strength of the lagged relationships between variables of opposite valence (e.g. Cheerful_{t-1} and Anxious_t) is dependent on which state the system is in. With knowledge of the true system we could explain that the univariate threshold is capable of approximating the position of the unstable fixed point in multivariate space, and that the asymmetric lagged parameters are produced by the continuous non-linear relationships present in the system. However, *without* theoretical knowledge about the true system, one might incorrectly conclude that, for instance, the true parameters of the system change over time, or that they are dependent only on the *Cheerful* variable. This shows that the use of a misspecified time series model to make direct inferences about the system is highly challenging even under ideal conditions: Although the global dynamics of the system were recovered in this case, any inferences about local dynamics were heavily dependent on theoretical insight about the system at hand.

With those caveats in mind, these results can be interpreted with cautious optimism. The more complex the model we were able to fit, and the more theoretical knowledge we had with which to guide model choice, the more characteristics of this system we were able to recover. However, there is no guarantee that for other dynamic systems, or even variants of the present dynamic system with different parameter values, that the TVAR model would succeed in allowing any such accurate inferences. Rather, our analysis suggests that strong theoretical ideas about the underlying system are necessary to be able to make even the most basic inferences from common time series models.

4 The Problem of Insufficient Sampling Frequency

We now turn to the problem of sampling the system with a sampling frequency that is insufficient for it to be recovered. To study the effects of insufficient sampling frequency, we create a time series with measurements every 90 minutes, which is a sampling frequency that is typical for ESM studies. We create this time series by taking every 900th measurement from the original time series shown in Figure 2. The resulting sub-sampled time series is shown in Figure 7.

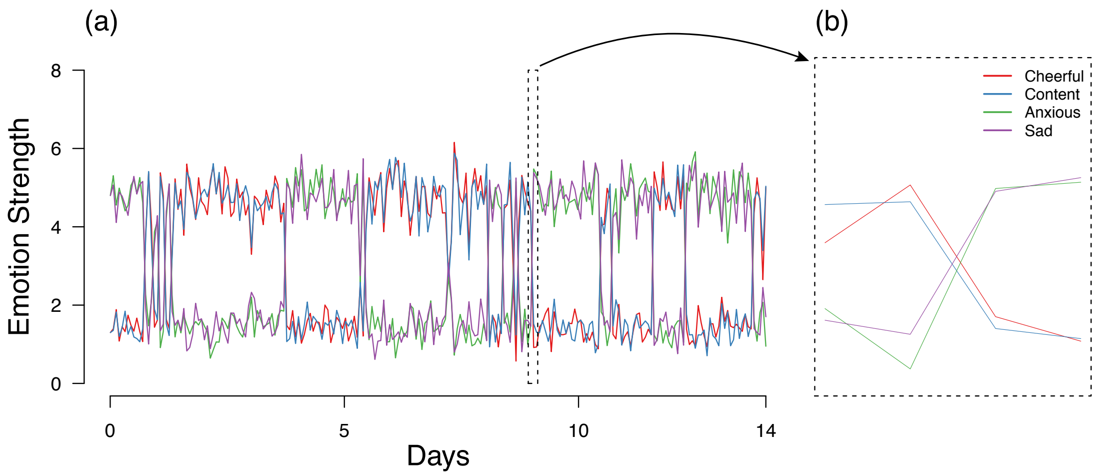

Figure 7: Panel (a) shows the ESM time series which was obtained by taking measurements snapshots every 90 minutes. Panel (b) displays the information the ESM data captures about the twelfth switch in the system, as depicted in Figure 2. Note that the ESM time series we analyze in Section 4 is much longer (1800 weeks) than the 14 day ESM time series shown here.

The ESM time series appears less dense, which is what we would expect since it contains only 1/900 of the time points of the ideal time series in Figure 2. However, we see that the system is still bistable and that the location of and variance around the fixed points is largely the same. In this section we will use this emulated ESM time series to try to recover the true bistable system using the same array of methods as in Section 3, in which we analyze the ideal time series. To ensure that results reflect only the effect of lowering the sampling frequency, and not the effect of lowering the overall sample size, we increase the period of the ESM time series to 1800 weeks, resulting in 201600 observations spaced at 90 minute time-intervals, equal to the sample size used in the previous section.[1]

[1]Note that, to ensure comparability, we use the same Euler step size of $\Delta t = 0.01$ to generate the ESM data as was used to generate the ideal time series. However, this makes generating a single time series of the required length prohibitively time-consuming. In order to make data generation more feasible, we instead generate 900 sub-sampled time series in parallel, where the first observation of each series is a random draw from a Gaussian distribution centered around the healthy fixed point and with unit variance.

4.1 Descriptive Statistics, Data Visualization & HMM

We again begin by analyzing the data with descriptive statistics and data visualizations. We see that both the histograms and the bivariate relationships at the same time point look exactly the same as the ones of the high sampling frequency time series shown in Figure 3. We again see that the system is bistable, and get a rough idea of the location and variance and the frequency of the two states.

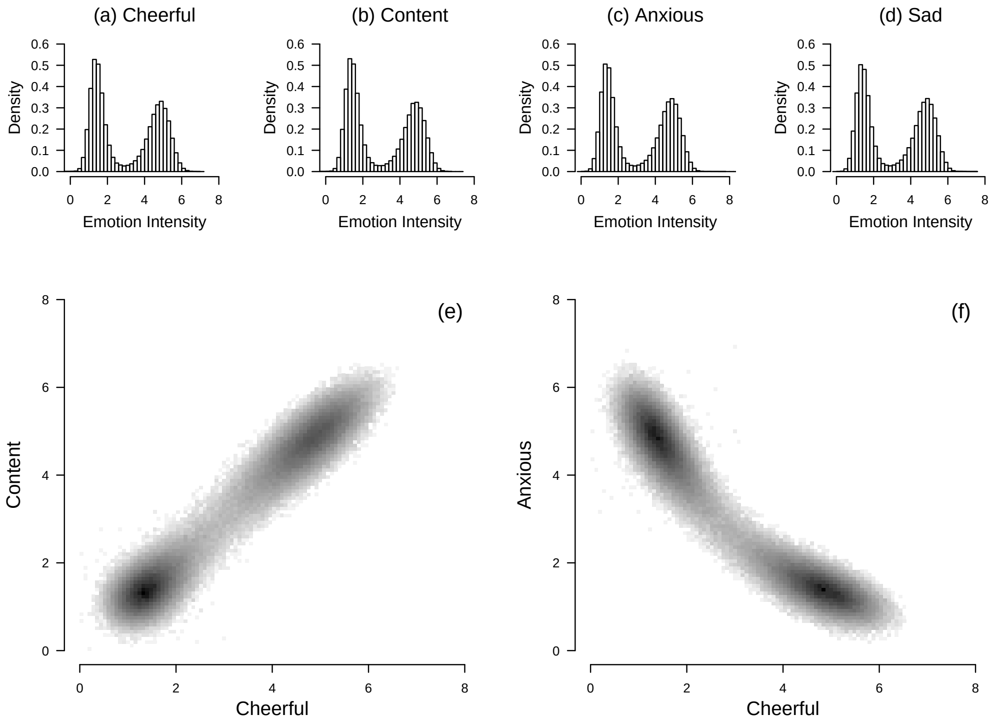

Figure 8: Top panel: The histograms of the four emotion variables Cheerful, Content, Anxious and Sad in the ESM-time series. Lower panel: The bivariate relationships between Content and Cheerful, and between Anxious and Cheerful in the ESM-time series. The red line indicates the best fitting regression line. Note that all other bivariate relationships are similar to the two shown ones, due to the symmetry in the true system.

To obtain numerical estimates for those quantities, we fit the Mean Switching HMM to the ESM-time series and obtain the following estimates:

$$\hat{\boldsymbol{\mu}}^{(1)} = \begin{pmatrix} 1.47 \\ 1.46 \\ 4.71 \\ 4.71 \end{pmatrix}, \; \hat{\boldsymbol{\sigma}}^{(1)} = \begin{pmatrix} 0.41 \\ 0.41 \\ 0.63 \\ 0.63 \end{pmatrix}, \; \hat{\boldsymbol{\mu}}^{(2)} = \begin{pmatrix} 4.71 \\ 4.71 \\ 1.47 \\ 1.47 \end{pmatrix}, \; \hat{\boldsymbol{\sigma}}^{(2)} = \begin{pmatrix} 0.64 \\ 0.64 \\ 0.41 \\ 0.41 \end{pmatrix}, \; \hat{\boldsymbol{M}} = \begin{matrix} \\ k_1 \\ k_2 \end{matrix} \begin{matrix} \begin{matrix} k_1 & \quad k_2 \end{matrix} \\ \begin{pmatrix} 0.915 & 0.085 \\ 0.090 & 0.910 \end{pmatrix} \end{matrix}.$$

Again, we see a very similar pattern of results as obtained from the HMM fit to the ideal time series in Section 3.2, with the means and standard deviations of state 1 and state 2 reflecting the unhealthy and healthy states respectively. The only difference is that the off-diagonal elements are larger than in the high sampling frequency time series. The reason is that there is still the same amount of switches in a given time period, however, there are much fewer observations between any two switches. Similarly to the analysis of the high frequency time series, we could predict the state for each time point. We display the predicted states for the first two weeks of the time series in Figure 11 in Appendix A, where we can see that the model again succeeds in capturing which observations are close to the healthy or unhealthy fixed point.

Taken together, we were still able to recover bistability, the location and variance around fixed points, and the frequency of switches. That is, the recovery of global dynamics was unaffected by reducing the sampling frequency. This makes sense, because the global characteristics used here are not defined with respect to dependencies across time. Therefore, and because we kept the sample size constant, the data visualizations and HMM model show results that are similar to the ones obtained in the high sampling frequency time series. This is good news, because it shows that one can learn some characteristics of a system, even if measurements are taken at a frequency that is unlikely to tap into the local dynamics of the phenomena of interest.

4.2 Lag-1 Relationships & VAR Model

We now turn to dependencies across time. Again, we begin by inspecting bivariate dependencies between $t-1$ and t. Panel (a) of Figure 9 displays the within-valence dependency between Cheerful at $t-1$ and Content at t. We see that the density looks very different from the one of the high frequency time series shown in Figure 5: First, there seem to be four density clusters instead of two. And second, each of the density clusters Cheerful_{t-1} and Content_t has a circular shape or is oriented along one of the two axis, which implies that the variables are locally uncorrelated.

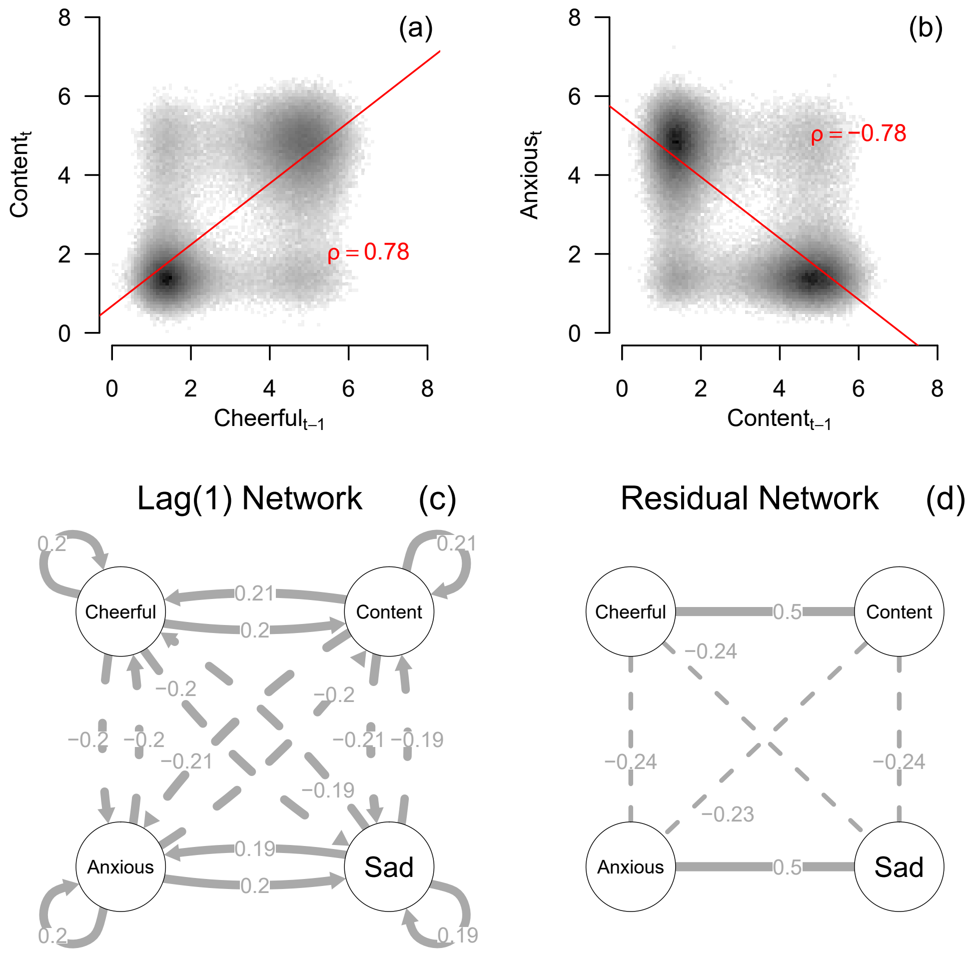

Figure 9: Panel (a) shows the relationship between Content and Cheerful, two emotions with the same valence, spaced one measurement occassion apart (i.e., at a lag of one but with 90 minutes between measurements) for the ESM time series. The red line indicates the best fitting regression model. Similarly, panel (b) shows the relationship between Anxious and Content, two emotions with different valence, at a lag of one. Panel (c) displays the matrix of lagged regression parameters, estimated from a VAR(1) model, as a network, and panel (d) displays the partial correlation matrix of the residuals of the VAR(1) model as a network. This latter network is often referred to as the "contemporaneous" network.

The two additional density clusters are explained by the fact that in the ESM time series it is possible that a switch occurs *between* two subsequent measurements. For example, at $t-1$ the system is in the healthy state, and at t the system is in the unhealthy state. This would lead to a point in the cluster in the bottom right corner. Such large "jumps" are possible in the ESM data, because in-between every pair of ESM measurement we skip 900 measurements in the high frequency time series. In the latter, such jumps are not possible, because the process changes relatively slowly from one time step to the next.

The lowered sampling frequency also explains the absence of dependency in each separate density. The differential equations of the true system specify a local dependency between variables, essentially from one infinitely small time step to the next. However, we now omitted 900 time points that are present in the ideal dataset between any two measurement points in the ESM dataset. Since independent noise is injected at a rate of every six seconds, there is hardly any dependency left between measurements taken 90 minutes apart. This is very similar to simulating 900 measurements from a VAR model, and trying to predict the 900th observation from the first. The density of the relationship between Content at $t-1$ and Anxious at t shown in panel (b) of Figure 9 is explained analogously.

We now know how to explain the densities of the bivariate relationships between lagged variables. How does this influence how we interpret the parameters of the VAR model fit to these variables? Above, we were struggling with the fact that the linear relationships were an approximation of non-linear relationships. However, now we are dealing with the additional problem that there are essentially no dependencies anymore between subsequent time points, except the dependency implied by the global characteristics. While the VAR parameters still reflect the global characteristics, it would be a mistake to interpret its parameters as "moment-to-moment" interactions at a short time scale. In fact, the remaining relationships between pairs of variables can be summarized in a simple 2×2 table which shows that emotions of the same valence are usually both high or both low, and emotions with different valence are hardly ever both high or both low. If such relations hold between any pair of variable, then each variable is equally predictive of any other variable. This is why all parameters in the VAR model in panel (c) have roughly the same value.

To summarize, we showed that, in the present case study, subsampling the data to a sampling frequency typical for an ESM-study removed essentially all local dependency from the time series. This means that it is impossible to recover them with any type of model. However, the linear lagged effects still provided some information about the global characteristics. This shows that important information can still be obtained from time series models even if the data is subsampled to the extent that all local dependencies are lost. However, it is crucial to carefully interpret these time series models. Interpreting them directly as reflecting local dynamics would clearly be incorrect. The example of the VAR model illustrated this: The relations do describe how the variables are related, however, they provide an extremely poor estimate of the local dynamics of the true system.

4.3 Threshold VAR

Finally, we examine the degree to which the threshold VAR model is able to recover characteristics of the bistable system based on data with a low sampling frequency. Figure 10 displays the main results of the estimated TVAR model. As we might expect based on the previous analyses, the TVAR model is able to recover some global characteristics of the system. The estimated threshold value (again using *Cheerful* as the thresholding variable) is $\hat{\tau} = 2.796$, quite close to the position of the unstable fixed point position. Panel (a) of Figure 10 shows that this threshold value succeeds in separating the time series into healthy and unhealthy states respectively, since the identification of these states is largely driven by the means.

However, the recovery of local characteristics by the TVAR model runs into the same fundamental problems as we encountered in the VAR model: The sampling frequency is so low that no local dependencies are present in the data, and introducing a threshold into our model is unable to solve this problem. Panels (b) and (c) of Figure 10 display the lagged parameters for each regime, as estimated by the TVAR model. Although we again recover asymmetric between-valence relationships, as we did in Section 3.4, there are a number of differences in the lagged parameter estimates based on ESM data, such as the positive relationship of Cheerful_{t-1} with Sad_t and Anxious_t in the healthy regime, and the large within-valence cross-lagged relationships of Anxious and Sad in the healthy regime and between Cheerful and Content in the unhealthy regime, respectively. As was the case for the VAR model above, these lagged parameter estimates are produced by the global characteristics of this system and should not be interpreted as reflecting local characteristics directly.

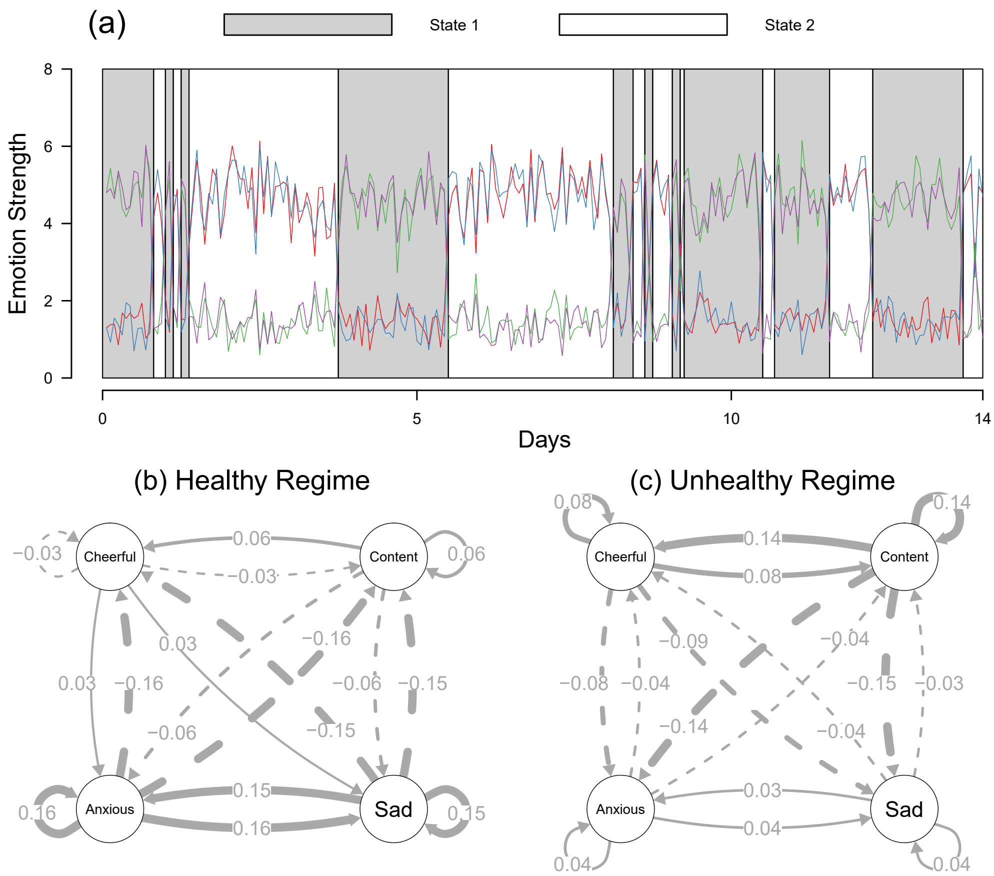

Figure 10: Panel (a) shows the first two weeks of the time series, with observations shaded in either grey or white as a function of whether Cheerful_{t-1} is above or below the threshold $\hat{\tau} = 2.796$. Panels (b) and (c) show the estimated VAR(1) parameters as lagged networks in the healthy (white) and unhealthy (grey) regimes respectively.

To summarize, we again have seen that sub-sampling the data to a sampling frequency typical for an ESM-study removes all local dynamics from the time series. However, even with a lowered sampling frequency, the threshold VAR model was still able to recover global characteristics, again separating the time series appropriately into the healthy and unhealthy regimes. This suggests that time series models which are appropriately informed by theory, though still misspecified, can be used to gain insights into global characteristics of the system at hand, even when the sampling frequency is to low to recover local characteristics directly.

5 Discussion

In this paper we discussed the two fundamental challenges of misspecification and insufficient sampling frequence in recovering within-person dynamics from psychological time series. To be able to illustrate these general problems, we assumed a theoretically interesting true system and attempted to recover this system using popular time series analysis tools. First, we studied the problem of misspecification. We were able to obtain global characteristics of the true system, such as that the system is bistable, the rough nature of the two stable states and an estimate of the switching frequency. However, we also showed that we cannot make straightforward inferences from the parameters of misspecified models to the local or even the more coarse global characteristics of the true system. Next, we investigated the problem of insufficient sampling frequency. We showed that if the sampling frequency is too low, the temporal dependencies in the true system are not captured anymore in the data. Consequently, no statistical model was able to recover the local characteristics of the true system. However, the global characteristics could still be recovered, since they are not based on dependencies across time.

We used our example system to illustrate that misspecification and insufficient sampling frequency present fundamental challenges to making inferences from time series models about within-person dynamics. This raises the questions of how time series models can be used, despite those two fundamental challenges, to learn about within-subjects dynamics and to develop formal theories of such dynamics.

5.1 Consequences for Analyzing Psychological Time Series

We have demonstrated that making inferences from time series models about the characteristics of an unspecified and more complicated model is highly problematic. For instance, for our toy system we found that in the VAR model the cross-lagged effects across emotions with the same valence were stronger than the cross-lagged effects across emotions with a different valence (see Figure 5). However, the perhaps intuitive interpretation of these parameters as evidence for stronger reinforcing effects within-valence compared to the suppressing effects between-valence in the data-generating model would be incorrect. From the true system we know that the suppressing effects are in fact stronger (see Figure 1). Since all models are misspecified in practice, any kind of direct inference from a time series model to some unknown underlying system is potentially problematic.

Critically, however, this is not a criticism of the time series models themselves, but rather the manner in which researchers attempt to interpret or make inferences based on their parameters. Since in the current paper we have essentially no sampling error due to the large sample sizes, in a sense all statistical models we have estimated here are correct, in that they correctly capture some *implication* of the data-generating model: For instance, that at short time intervals, conditional linear dependencies within-valence are stronger than those between-valence (Figure 5) but that at very long time intervals those linear dependencies are essentially equal in value (Figure 9). These statements are true, and although it may in some cases be a somewhat complicated process, they could be derived from the data-generating model without any need to simulate data as we have done here. The problems arise when we try to reverse this process, to take an implication of a model and attempt to infer from it the model itself or other characteristics of it. We have shown here the extreme difficulty of such inferences, and highlighted the caution with which they must be approached. To twist an oft-misused quote in the statistics literature (cf. Box, 1976): All statistical models are right, but most aren't useful (or at least not in the way you might think).

So how can we hope to improve our ability to make inferences from misspecified times series models? This problem can be mitigated to some extent by informing the choice of time series model theoretically, and thereby reducing the degree of misspecifiction. For example, if we assume that the true system has several stable states, and if we take time series measurements from the system while it remains around a single stable state, then we could fit a VAR model to obtain a linear approximation of the dynamics close to that stable state. If we have other time series capturing other states, taken from the same or similar individuals, we are possibly able to piece together these separate models to obtain a characterization of the overall dynamics of the system. If such local models are allowed to vary over time, it might even be possible to detect critical transitions in the stability landscape of the system, even though we only observe a part of the landscape (e.g., Snippe et al., 2017; Wichers, 2016). The crucial point here is that if we use time series in this way, we do not make naive inferences from parameters of the time series models about characteristics of the true system. Instead, we use our theoretical expectations to choose the right time series model to extract a specific characteristic of the true system. Eventually, all these characteristics should be integrated in a formal theory of the true system, which is what we turn to in the next subsection.

In the second part of the paper we illustrated the problem of insufficient sampling frequency. We reduced the sampling frequency from every six seconds to 90 minutes and showed that all micro-dynamics were removed from the time series data. Of course, to illustrate the problem of insufficient sampling frequency we chose the time scale of our system such that the sampling frequency of a typical ESM study would be insufficient to recover the true system. However, this is not necessarily always the case. For example, when investigating slow changing variables such as mood, body weight, psychiatric symptoms, or processes that occur at regular intervals such as sleep, then a few measurements a day may be sufficient. However, we think that for many psychological dynamics this is likely not the case. For example, it seems unlikely to appropriately capture the dynamics of emotions, which are defined on a time scale of seconds and minutes (Houben, Van Den Noortgate, & Kuppens, 2015), by sampling every 1.5 hours. While we suspect that this problem occurs often in practice, a quick survey of the literature reveals that the time scale of target processes is often unspecified in empirical research. In Appendix E we provide a review of 43 ESM studies within the network perspective of psychopathology, which shows that the time scale of variables is typically not clearly defined and that in all studies at least one variable plausibly changes at a time scale of minutes. This suggests that it is important to characterize more clearly on which time scale the variables of interest change, and to adapt the study design accordingly. While intuitively we may consider the highest possible sampling frequency to be optimal, in practice questions of optimal design will depend on the number of observations we are able to make, the dynamics of the underlying system (such as the timescale of the target processes), and our analysis goals, such as the characteristics we wish to learn about and the way in which the design should be optimal (to see the variety of different discussions on optimal design, see for instance Adolf, Loossens, Tuerlinckx, & Ceulemans, 2019; Adolph, Robinson, Young, & Gill-Alvarez, 2008). Furthermore, while study designs can be adapted to some extent, for example with higher sampling frequencies or measurement bursts (Adolph et al., 2008; Nesselroade, 1991), there are of course limits to how often a day individuals can be queried with an ESM questionnaire. Then the question becomes how we can still make use of time series models based on data that was sampled with an insufficient frequency to allow, at least in principle, direct recovery of the microdynamics. We think that this problem can be addressed by adopting a way of constructing theories that puts formal theories at the center of theory development.

5.2 Constructing Formal Theories of Within-person Dynamics

We showed that we can use global characteristics to obtain a rough description of the dynamics of interest; and we suggested that, if guided by theory, relatively simple time series models like the VAR model could be used to describe different aspects of the within-person dynamics of interest. However, the goal of idiographic modeling should not only be the accumulation of empirical facts. Instead, we think that the eventual goal should be to construct a formal theory of the within-person dynamics which explains those facts.

We recently proposed a framework for how to construct such formal theories, which consists of a three-step methodology of generating an initial theory, developing the theory, and testing the theory (Haslbeck et al., 2020). Here we only focus on the theory development step, since it is most relevant for the two problems studied in this paper. An initial theory is developed by generating time series data from it, and fitting a time series model that captures the aspect of the formal theory that should be tested with empirical data. This gives rise to a theory-implied time series model. Next, we collect the corresponding empirical time series data and fit the same time series models to them. Finally, we compare the theory-implied and empirical time series model. If they are similar, we take this as tentative evidence for the adequacy of our current formal theory; if not, we use the discrepancy between the two to devise adaptations to the current theory, such that it implies a time series model that is closer to the empirical one. This procedure is similar to the idea of predictive checks in Bayesian analysis, where data are simulated from the fitted model, and checks are performed on summary statistics of the simulated data (Gelman et al., 2013; Gelman & Hill, 2006).

While this approach clearly does not solve the fundamental problems of misspecification and insufficient sampling frequency, it does allow us to deal with those problems in a more flexible way. For example, our current formal theory may predict a strong negative cross-lagged effect between variables X_t and Y_{t+1}. We then collect the corresponding empirical data and estimate this cross-lagged effect. If we find a strong negative cross-lagged effect, we take this as evidence that our theory correctly accounts for this empirical phenomenon; if not, we try to think of ways to change the formal theory such that it predicts the cross-lagged effect we find empirically. Critically, in this approach the VAR model including the cross-lagged effect between X_t and Y_{t+1} does not serve as a plausible true system, and we do not make naive inferences from its parameters to the structure of the true system. Instead, we consider it as a *descriptive* model, which captures patterns in the data that need to be reproduced by a successful theory. This provides a much clearer process of theory development: instead of having to make inferences about a completely unknown and unspecified structure, we have a formal theory which we can adapt to account for more and more empirical phenomena captured by time series models and other statistical models. Similarly, if we have an insufficient sampling frequency to recover the entire true system, we can still use the above approach. We now downsample the time series generated from the formal theory to the empirical sampling frequency, and create an implied time series model on the level of this sampling frequency. While there is clearly a loss in information due to the insufficient sampling frequency, it is still possible to make predictions with the formal theory on that level, and make adaptions to the theory if its predictions do not line up with the empirical time series model.

One way in which this approach could be applied is to use global characteristics to select a class of dynamical systems that is known to produce these characteristics. For example, van der Maas, Molenaar, et al. (1996) list a number of global characteristics, which they refer to as "catastrophe flags", which indicate that a dynamical system based on catastrophe theory could have generated the data. Next, one could use additional data to narrow down the possibilities to a smaller class of models, such as the cusp catastrophe model. Finally, some parameters of this model could be estimated directly from data (Grasman, van der Maas, Wagenmakers, et al., 2009). Another example is the formal theory of panic disorder developed by Robinaugh et al. (2019). This theory was developed by listing the core empirical phenomena of panic disorders and then using existing verbal cognitive-behavioral theories to construct a formal theory of the within-person dynamics that lead to the development of panic disorder. In Haslbeck et al. (2020) we provide a detailed discussion for how to develop this method with the above outlined approach both with time series data and cross-sectional data.

We argued that we need to adopt a more flexible modeling approach to obtain formal theories of within-person dynamics, which goes beyond fitting statistical time series models alone. While such an approach cannot resolve the fundamental problems of misspecification and insufficient sampling frequency, it allows us to deal more flexibly with these problems. For a detailed account of the above proposed approach to constructing formal theories of within-person dynamics see Haslbeck et al. (2020). Our approach is only one of several recently proposed approaches that have identified the use of formal theories and computational models as critical to developing theoretical understanding of psychological phenomena (e.g., Borsboom, van der Maas, Dalege, Kievit, & Haig, 2020; Burger et al., 2020; Guest & Martin, 2020; Robinaugh, Haslbeck, Ryan, Fried, & Waldorp, 2020; van Rooij & Baggio, 2020).

5.3 Limitations of our Approach

In the present paper we investigated to which extent one can infer the characteristics of our bistable system of emotion dynamics from a specific set of time series models. Strictly speaking, we thereby only provided evidence that misspecification and insufficient sampling frequency are a problem in this specific setting. However, we argue that it is reasonable to assume that the psychological systems we are interested in are more complicated than the simple bistable system used in this paper (for a discussion see Haslbeck et al., 2020). And our intuition is that the problem of misspecification will not disappear when studying a more complicated system with the same time series models.

If the true system is more complex, why don't we simply fit more complex time series models? Much progress has been made in broadening the dynamic modeling toolbox in psychology in recent years, with new tools allowing for the estimation of a variety of more complex auto-regressive and moving average of time series models (e.g Asparouhov, Hamaker, & Muthén, 2018) and linear as well as non-linear differential equation models (e.g. Driver, Oud, & Voelkle, 2017; Ou, Hunter, & Chow, 2019) directly from data. These developments allow researchers to overcome many practical issues typically faced in psychological time series studies, such as accounting for the presence of unequal time-intervals between measurement occasions (for a discussion, see Kuiper & Ryan, 2018; Ryan & Hamaker, 2020; Voelkle & Oud, 2013, amongst many others) and allowing for the use of multi-level models to regularize parameter estimates. However, the use of more complex statistical models to recover within-person dynamics is still subject to the same fundamental problems we have highlighted in this paper. First, even with more complex time series models we will never know when and to what extent the model will be misspecified, and consequently, we cannot trust straightforward inferences from the time series model to the measured system. Second, it is unclear whether in psychological settings we ever have the data to fit such models. Either because the sampling frequency is too low, or because the time series is too short. Although the use of multilevel models may enable us to estimate time series models using less observations per person, the barrier of attaining sufficient sampling frequency is more difficult to overcome. While integral solution based (Voelkle & Oud, 2013) and Kalman filter based (Chow, Ferrer, & Nesselroade, 2007) methods of estimating differential equation models (such as implemented in the *dynr* and *ctsem* packages Carpenter et al., 2017; Driver et al., 2017; Ou et al., 2019) may require less frequent observations than difference-based methods (e.g. Boker, Deboeck, Edler, & Keel, 2010), these methods rely still have the fundamental limit that, to recover dynamics at a certain frequency, information on those dynamics must still be present in the dataset at hand.

In addition, we choose a parameter configuration for our toy model such that we observe a high number of switches between states in the simulated datasets. The premise of our approach here is that, if we wish to recover the *full* bistable system, or as many characteristics of that system as possible, having a dataset in which we observe multiple switches is ideal. However, it may be the case that for many psychological phenomena, it is not possible to collect single-individual time series data in which a large number of transitions are observed. For example, depressive states may last a number of months or years and transitions from healthy to depressive episodes or vice versa may only occur once in a lifetime (American Psychiatric Association, 2013; Freeman, 1996; Post, 1992). With this in mind, many recent ESM studies in psychology have begun to collect such long time series in the hope of capturing those transitions (e.g., Helmich et al., 2020; Kuranova et al., 2020; Smit, Snippe, & Wichers, 2019). From a theoretical viewpoint, many discussions of bistable systems as they relate to psychological phenomena indeed focus on so-called cusp-catastrophe models, where the system exhibits one stable state for a long period of time, before the system becomes a bistable system (Cramer et al., 2016; Wagenmakers, Molenaar, Grasman, Hartelman, & Van Der Maas, 2005; Wichers, Wigman, & Myin-Germeys, 2015). While we believe that it is possible in principle to learn about bistable systems from datasets where, for instance, we only observe a single state per individual, or a single transition for some individuals, we suspect that this is more difficult than attempting to make inferences from the type of dataset we study in the current paper. This would likely require additional assumptions about complicating factors such as the homogeneity or heterogeneity of individuals and counterfactual behaviours which may not be directly captured in the dataset.

Throughout the paper we used a very large sample size of $n = 201600$ to be able to study the effects of misspecification and insufficient sampling frequency without confounding them with the effects of sampling variance. Typical sample sizes in applied research are of course much lower and one has to deal with sampling variance, which poses an additional challenge to recovering within-person dynamics from time series data. Another idealization in our setting was that we were able to sample all variables in the true system. This means that there are no spurious statistical relationships that can be explained by the presence of unobserved common causes. In addition, we were able to measure all variables directly and without any distortion in the measurement process. Both idealizations are unrealistic, which renders system recovery more challenging in practice.

5.4 Conclusions

Idiographic modeling of within-person dynamics is an exciting area in psychological science, both because it avoids known problems associated with making inferences to within-person processes based on cross-sectional data and because of the rapidly increasing availability of within-person time series data. However, in this paper we showed that it is difficult to make direct inferences from time series models to underlying within-person systems. The reasons are that arguably all time series models are misspecified and that the sampling frequency is possibly insufficient. To deal with these problems we suggest to adopt a framework to construct theories of within-person dynamics that goes beyond estimating statistical time series models alone and puts formal theories at the center of theory development.

Acknowledgements

We would like to thank Denny Borsboom, Tessa Blanken, Julian Burger, Fabian Dablander, Ellen Hamaker, Han van der Maas, Alessandra Mansueto, Don Robinaugh and Lourens Waldorp for valuable discussions and feedback on earlier versions of this manuscript. JMBH has been supported by the European Research Council Consolidator Grant no. 647209. OR has been supported by the Netherlands Organisation for Scientific Research (NWO) Onderzoekstalent Grant 406-15-128.

References

Aalbers, G., McNally, R. J., Heeren, A., De Wit, S., & Fried, E. I. (2019). Social media and depression symptoms: A network perspective. *Journal of Experimental Psychology: General*, *148*(8), 1454. doi: 10.1037/xge0000528

Adolf, J., Loossens, T., Tuerlinckx, F., & Ceulemans, E. (2019, Nov). Optimal sampling rates for reliable continuous-time first-order autoregressive and vector autoregressive modeling. Retrieved from psyarxiv.com/5cbfw doi: 10.31234/osf.io/5cbfw

Adolph, K. E., Robinson, S. R., Young, J. W., & Gill-Alvarez, F. (2008). What is the shape of developmental change? *Psychological review*, *115*(3), 527. doi: 10.1037/0033-295X.115.3.527

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed. ed.). Washington, DC: Autor.

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2017). Dynamic latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(2), 257–269. doi: 10.1080/10705511.2016.1253479

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 359–388. doi: 10.1080/10705511.2017.1406803

Atkinson, K. E. (2008). *An introduction to numerical analysis.* John wiley & sons.

Bak, M., Drukker, M., Hasmi, L., & van Os, J. (2016). An n= 1 clinical network analysis of symptoms and treatment in psychosis. *PloS one*, *11*(9), e0162811. doi: 10.1371/journal.pone.0162811

Beck, E. D., & Jackson, J. J. (2020). Consistency and change in idiographic personality: A longitudinal esm network study. *Journal of Personality and Social Psychology*, *118*(5), 1080. doi: 10.1037/pspp0000249

Bernstein, E. E., Curtiss, J. E., Wu, G. W., Barreira, P. J., & McNally, R. J. (2019). Exercise and emotion dynamics: An experience sampling study. *Emotion*, *19*(4), 637. doi: 10.1037/emo0000462

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., Deboeck, P. R., Edler, C., & Keel, P. K. (2010). Generalized local linear approximation of derivatives from time series. In *Statistical methods for modeling human dynamics: An interdisciplinary dialogue* (pp. 161–178). Routledge.

Boker, S. M., Montpetit, M. A., Hunter, M. D., & Bergeman, C. S. (2010). Modeling resilience with differential equations. In *Individual pathways of change: Statistical models for analyzing learning and development.* (pp. 183–206). American Psychological Association.

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*(1), 5–13. doi: 10.1002/wps.20375

Borsboom, D., & Cramer, A. O. (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annual review of clinical psychology*, *9*, 91–121.

Borsboom, D., Cramer, A. O., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. *PloS one*, *6*(11), e27407.

Borsboom, D., van der Maas, H., Dalege, J., Kievit, R., & Haig, B. (2020, Feb). *Theory construction methodology: A practical framework for theory formation in psychology.* PsyArXiv. Retrieved from psyarxiv.com/w5tp8 doi: 10.31234/osf.io/w5tp8

Box, G. E. (1976). Science and statistics. *Journal of the American Statistical Association*, *71*(356), 791–799.

Box, G. E., & Jenkins, G. M. (1976). *Time series analysis, forecasting and control.* San Francisco, CA: Holden Day.

Bringmann, L. F., Pe, M. L., Vissers, N., Ceulemans, E., Borsboom, D., Vanpaemel, W., ... Kuppens, P. (2016). Assessing temporal emotion dynamics using networks. *Assessment*, *23*(4), 425–435.

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.

Brose, A., Wichers, M., & Kuppens, P. (2017). Daily stressful experiences precede but do not succeed depressive symptoms: Results from a longitudinal experience sampling study. *Journal of Social and Clinical Psychology*, *36*(3), 196–220. doi: 10.1521/jscp.2017.36.3.196

Burger, J., van der Veen, D. C., Robinaugh, D. J., Quax, R., Riese, H., Schoevers, R. A., & Epskamp, S. (2020). Bridging the gap between complexity science and clinical practice by formalizing idiographic theories: a computational model of functional analysis. *BMC medicine*, *18*, 1–18.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., . . . Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1). doi: 10.18637/jss.v076.i01

Chow, S.-M., Ferrer, E., & Nesselroade, J. R. (2007). An unscented kalman filter approach to the estimation of nonlinear dynamical systems models. *Multivariate Behavioral Research*, *42*(2), 283–321. doi: 10.1080/00273170701360423

Clasen, P. C., Fisher, A. J., & Beevers, C. G. (2015). Mood-reactive self-esteem and depression vulnerability: Person-specific symptom dynamics via smart phone assessment. *PLoS One*, *10*(7), e0129774. doi: 10.1371/journal.pone.0129774

Cramer, A. O., van Borkulo, C. D., Giltay, E. J., van der Maas, H. L., Kendler, K. S., Scheffer, M., & Borsboom, D. (2016). Major depression as a complex dynamic system. *PloS one*, *11*(12), e0167490. doi: 10.1371/journal.pone.0167490

Cramer, A. O., Waldorp, L. J., Van Der Maas, H. L., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behavioral and brain sciences*, *33*(2-3), 137–150. doi: 10.1017/S0140525X09991567

Cristóbal-Narváez, P., Sheinbaum, T., Rosa, A., Ballespí, S., de Castro-Catala, M., Peña, E., . . . Barrantes-Vidal, N. (2016). The interaction between childhood bullying and the fkbp5 gene on psychotic-like experiences and stress reactivity in real life. *PLoS one*, *11*(7), e0158809. doi: doi.org/10.1371/journal.pone.0158809

Curtiss, J., Fulford, D., Hofmann, S. G., & Gershon, A. (2019). Network dynamics of positive and negative affect in bipolar disorder. *Journal of affective disorders*, *249*, 270–277. doi: 10.1016/j.jad.2019.02.017

De Haan-Rietdijk, S., Gottman, J. M., Bergeman, C. S., & Hamaker, E. L. (2016). Get over it! a multilevel threshold autoregressive model for state-dependent affect regulation. *Psychometrika*, *81*(1), 217–241. doi: 10.1007/s11336-014-9417-x

de Haan-Rietdijk, S., Kuppens, P., Bergeman, C. S., Sheeber, L., Allen, N., & Hamaker, E. (2017). On the use of mixed markov models for intensive longitudinal data. *Multivariate behavioral research*, *52*(6), 747–767. doi: 10.1080/00273171.2017.1370364

De Vos, S., Wardenaar, K. J., Bos, E. H., Wit, E. C., Bouwmans, M. E., & De Jonge, P. (2017). An investigation of emotion dynamics in major depressive disorder patients and healthy persons using sparse longitudinal networks. *PLoS One*, *12*(6), e0178586. doi: 10.1371/journal.pone.0178586

Driver, C. C., Oud, J. H. L., & Voelkle, M. C. (2017). Continuous time structural equation modeling with R package ctsem. *Journal of Statistical Software*, *77*(5), 1–35. doi: 10.18637/jss.v077.i05

Fabio Di Narzo, A., Aznarte, J. L., & Stigler, M. (2009). tsDyn: Time series analysis based on dynamical systems theory [Computer software manual]. Retrieved from https://cran.r-project.org/package=tsDyn/vignettes/tsDyn.pdf (R package version 0.7)

Fang, L., Marchetti, I., Hoorelbeke, K., & Koster, E. H. (2019). Do daily dynamics in rumination and affect predict depressive symptoms and trait rumination? an experience sampling study. *Journal of behavior therapy and experimental psychiatry*, *63*, 66–72. doi: 10.1016/j.jbtep.2018.11.002

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*, 201711978. doi: 10.1073/pnas.1711978115

Fisher, A. J., Reeves, J. W., Lawyer, G., Medaglia, J. D., & Rubel, J. A. (2017). Exploring the Idiographic Dynamics of Mood and Anxiety via Network Analysis. *Journal of Abnormal Psychology*, *126*(8), 1044–1056. doi: 0.1037/abn0000311

Freedman, H. I. (1980). *Deterministic mathematical models in population ecology* (Vol. 57). Marcel Dekker Incorporated.

Freeman, H. L. (1996). *Interpersonal factors in the origin and course of affective disorders.* RCPsych Publications.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Geschwind, N., Peeters, F., Drukker, M., van Os, J., & Wichers, M. (2011). Mindfulness Training Increases Momentary Positive Emotions and Reward Experience in Adults Vulnerable to Depression: A Randomized Controlled Trial. *Journal of Consulting and Clinical Psychology*, *79*(5), 618–628. doi: 10.1037/a0024595

Grasman, R., van der Maas, H. L., Wagenmakers, E.-J., et al. (2009). Fitting the cusp catastrophe in R: A cusp package primer. *Journal of Statistical Software*, *32*(i08). doi: 10.18637/jss.v032.i08

Greene, T., Gelkopf, M., Epskamp, S., Fried, E., et al. (2018). Dynamic networks of ptsd symptoms during conflict. *Psychological Medicine*, *48*(14), 2409–2417. doi: 10.1017/S0033291718000351

Groen, R. N., Snippe, E., Bringmann, L. F., Simons, C. J., Hartmann, J. A., Bos, E. H., & Wichers, M. (2019). Capturing the risk of persisting depressive symptoms: A dynamic network investigation of patients' daily symptom experiences. *Psychiatry Research*, *271*, 640–648. doi: 10.1016/j.psychres.2018.12.054

Guest, O., & Martin, A. E. (2020, Feb). *How computational modeling can force theory building in psychological science.* PsyArXiv. Retrieved from psyarxiv.com/rybh9 doi: 10.31234/osf.io/rybh9

Hamaker, E. L. (2012). Why researchers should think "within-person": A paradigmatic rationale. In M. R. Meehl & T. S. Conner (Eds.), (pp. 43–61). New York, NY: Guilford.

Hamaker, E. L., Grasman, R. P., & Kamphuis, J. H. (2010). Regime-switching models to study psychological process. In P. C. M. Molenaar & K. M. Newell (Eds.), *Individual pathways of change: Statistical models for analyzing learning and development* (p. 155-168). Washington, DC: American Psychological Association.

Hamaker, E. L., Grasman, R. P., & Kamphuis, J. H. (2016). Modeling bas dysregulation in bipolar disorder: Illustrating the potential of time series analysis. *Assessment*, *23*(4), 436–446. doi: 10.1177/1073191116632339

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. doi: 10.1177/0963721416666518

Hamaker, E. L., Zhang, Z., & van der Maas, H. L. (2009). Using threshold autoregressive models to study dyadic interactions. *Psychometrika*, *74*(4), 727. doi: 10.1007/s11336-013-9366-9

Hamilton, J. D. (1994). *Time series analysis* (Vol. 2). Princeton, NJ: Princeton University Press.

Haslbeck, J., Ryan, O., Robinaugh, D., Waldorp, L., & Borsboom, D. (2020). Modeling psychopathology: From data models to formal theories. Retrieved from https://psyarxiv.com/jgm7f

Hasmi, L., Drukker, M., Guloksuz, S., Menne-Lothmann, C., Decoster, J., van Winkel, R., ... others (2017). Network approach to understanding emotion dynamics in relation to childhood trauma and genetic liability to psychopathology: replication of a prospective experience sampling analysis. *Frontiers in psychology*, *8*, 1908.

Hasmi, L., Drukker, M., Guloksuz, S., Viechtbauer, W., Thiery, E., Derom, C., & Van Os, J. (2018). Genetic and environmental influences on the affective regulation network: a prospective experience sampling analysis. *Frontiers in psychiatry*, *9*, 602. doi: 10.3389/fpsyt.2018.00602

Helmich, M. A., Wichers, M., Olthof, M., Strunk, G., Aas, B., Aichhorn, W., ... Snippe, E. (2020). Sudden gains in day-to-day change: Revealing nonlinear patterns of individual improvement in depression. *Journal of Consulting and Clinical Psychology*, *88*(2), 119. doi: 10.1037/ccp0000469

Hirsch, M. W., Smale, S., & Devaney, R. L. (2012). *Differential equations, dynamical systems, and an introduction to chaos*. Academic press.

Hoorelbeke, K., Van den Bergh, N., Wichers, M., & Koster, E. H. (2019). Between vulnerability and resilience: A network analysis of fluctuations in cognitive risk and protective factors following remission from depression. *Behaviour research and therapy*, *116*, 1–9. doi: 10.1016/j.brat.2019.01.007

Hosenfeld, B., Bos, E. H., Wardenaar, K. J., Conradi, H. J., van der Maas, H. L., Visser, I., & de Jonge, P. (2015). Major depressive disorder as a nonlinear dynamic system: bimodality in the frequency distribution of depressive symptoms over time. *Bmc psychiatry*, *15*(1), 1–9. doi: https://doi.org/10.1186/s12888-015-0596-5,doi={10.1037/a0038822}

Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, *141*(4), 901. doi: 10.1037/a0038822

Jonas, E., & Kording, K. P. (2017). Could a neuroscientist understand a microprocessor? *PLoS computational biology*, *13*(1). doi: 10.1371/journal.pcbi.1005268

Kalisch, R., Cramer, A. O., Binder, H., Fritz, J., Leertouwer, I., Lunansky, G., ... Van Harmelen, A.-L. (2019). Deconstructing and reconstructing resilience: a dynamic network approach. *Perspectives on Psychological Science*, *14*(5), 765–777. doi: 10.1177/1745691619855637

Klippel, A., Myin-Germeys, I., Chavez-Baldini, U., Preacher, K. J., Kempton, M., Valmaggia, L., ... others (2017). Modeling the interplay between psychological processes and adverse, stressful contexts and experiences in pathways to psychosis: an experience sampling study. *Schizophrenia bulletin*, *43*(2), 302–315.

Klippel, A., Viechtbauer, W., Reininghaus, U., Wigman, J., van Borkulo, C., MERGE, ... Wichers, M. (2018). The cascade of stress: a network approach to explore differential dynamics in populations varying in risk for psychosis. *Schizophrenia Bulletin*, *44*(2), 328–337.

Kroeze, R., Van Veen, D., Servaas, M. N., Bastiaansen, J. A., Oude Voshaar, R., Borsboom, D., & Riese, H. (2016). Personalized feedback on symptom dynamics of psychopathology: a proof-of-principle study. *Journal Person-Orient Research*. doi: 10.17505/jpor.2017.01

Kroeze, R., Van Veen, D., Servaas, M. N., Bastiaansen, J. A., Oude Voshaar, R., Borsboom, D., & Riese, H. (2017). Personalized feedback on symptom dynamics of psychopathology: A proof-of-principle study. *Journal for Person-Oriented Research*, *3*(1), 1–10. doi: 10.17505/jpor.2017.01

Kuiper, R. M., & Ryan, O. (2018). Drawing conclusions from cross-lagged relationships: Reconsidering the role of the time-interval. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(5), 809–823. doi: 10.1080/10705511.2018.1431046

Kuranova, A., Booij, S. H., Menne-Lothmann, C., Decoster, J., van Winkel, R., Delespaul, P., ... others (2020). Measuring resilience prospectively as the speed of affect recovery in daily life: a complex systems perspective on mental health. *BMC medicine*, *18*(1), 1–11. doi: 10.1186/s12916-020-1500-9

Lazebnik, Y. (2002). Can a biologist fix a radio?—or, what i learned while studying apoptosis. *Cancer cell*, *2*(3), 179–182. doi: 10.1016/S1535-6108(02)00133-2

Lee, Y.-S., Jung, W.-M., Jang, H., Kim, S., Chung, S.-Y., & Chae, Y. (2017). The dynamic relationship between emotional and physical states: an observational study of personal health records. *Neuropsychiatric disease and treatment*, *13*, 411. doi: 10.2147/NDT.S120995

Levinson, C. A., Vanzhula, I., & Brosof, L. C. (2018). Longitudinal and personalized networks of eating disorder cognitions and behaviors: Targets for precision intervention a proof of concept study. *International Journal of Eating Disorders*, *51*(11), 1233–1243. doi: 10.1002/eat.22952

Lutz, W., Schwartz, B., Hofmann, S. G., Fisher, A. J., Husen, K., & Rubel, J. A. (2018). Using network analysis for the prediction of treatment dropout in patients with mood and anxiety disorders: A methodological proof-of-concept study. *Scientific reports*, *8*(1), 1–9. doi: 10.1038/s41598-018-25953-0

Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on psychological science*, *7*(3), 221–237. doi: 10.1177/1745691612441215

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. doi: 10.1207/s15366359mea0204_1

Neale, M. C., Clark, S. L., Dolan, C. V., & Hunter, M. D. (2016). Regime switching modeling of substance use: Time-varying and second-order markov models and individual probability plots. *Structural equation modeling: a multidisciplinary journal*, *23*(2), 221–233. doi: 10.1080/10705511.2014.979932

Nelson, B., McGorry, P. D., Wichers, M., Wigman, J. T., & Hartmann, J. A. (2017). Moving from static to dynamic models of the onset of mental disorder: a review. *JAMA Psychiatry*, *74*(5), 528–534. doi: 10.1001/jamapsychiatry.2017.0001

Nesselroade, J. R. (1991). The warp and the woof of the developmental fabric. In (pp. 213–240). Lawrence Erlbaum Associates, Inc.

Oreel, T. H., Borsboom, D., Epskamp, S., Hartog, I. D., Netjes, J. E., Nieuwkerk, P. T., ... Sprangers, M. A. (2019). The dynamics in health-related quality of life of patients with stable coronary artery disease were revealed: a network analysis. *Journal of clinical epidemiology*, *107*, 116–123. doi: 10.1016/j.jclinepi.2018.11.022

Ou, L., Hunter, M. D., & Chow, S.-M. (2019). dynr: Dynamic modeling in r [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=dynr (R package version 0.1.14-9)

Pavani, J.-B., Le Vigouroux, S., Kop, J.-L., Congard, A., & Dauvier, B. (2017). A network approach to affect regulation dynamics and personality trait-induced variations: Extraversion and neuroticism moderate reciprocal influences between affect and affect regulation strategies. *European Journal of Personality*, *31*(4), 329–346.

Pe, M. L., Kircanski, K., Thompson, R. J., Bringmann, L. F., Tuerlinckx, F., Mestdagh, M., ... others (2015). Emotion-network density in major depressive disorder. *Clinical Psychological Science*, *3*(2), 292–300.

Poerio, G. L., Kellett, S., & Totterdell, P. (2016). Tracking potentiating states of dissociation: an intensive clinical case study of sleep, daydreaming, mood, and depersonalization/derealization. *Frontiers in psychology*, *7*, 1231. doi: 10.3389/fpsyg.2016.01231

Post, R. (1992). Transduction of psychosocial stress into the neurobiology of recurrent affective disorder. *The American journal of psychiatry*, *149*(8), 999-1010. doi: 10.1176/ajp.149.8.999

Robinaugh, D., Haslbeck, J., Waldorp, L., Kossakowski, J., Fried, E. I., Millner, A., ... others (2019). Advancing the network theory of mental disorders: A computational model of panic disorder.

Robinaugh, D., Haslbeck, J. M. B., Ryan, O., Fried, E. I., & Waldorp, L. (2020, Mar). *Invisible hands and fine calipers: A call to use formal theory as a toolkit for theory construction.* PsyArXiv. Retrieved from psyarxiv.com/ugz7y doi: 10.31234/osf.io/ugz7y

Ryan, O., & Hamaker, E. (2020, Sep). *Time to intervene: A continuous-time approach to network analysis and centrality.* PsyArXiv. Retrieved from psyarxiv.com/2ambn doi: 10.31234/osf.io/2ambn

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.

Schmittmann, V. D., Cramer, A. O., Waldorp, L. J., Epskamp, S., Kievit, R. A., & Borsboom, D. (2013). Deconstructing the construct: A network perspective on psychological phenomena. *New ideas in psychology*, *31*(1), 43–53. doi: 10.1016/j.newideapsych.2011.02.007

Smit, A. C., Snippe, E., & Wichers, M. (2019). Increasing restlessness signals impending increase in depressive symptoms more than 2 months before it happens in individual patients. *Psychotherapy and psychosomatics*, *88*(4), 249–251. doi: 10.1159/000500594

Snippe, E., Jeronimus, B. F., aan het Rot, M., Bos, E. H., de Jonge, P., & Wichers, M. (2018). The reciprocity of prosocial behavior and positive affect in daily life. *Journal of Personality*, *86*(2), 139–146. doi: 10.1111/jopy.12299

Snippe, E., Viechtbauer, W., Geschwind, N., Klippel, A., De Jonge, P., & Wichers, M. (2017). The impact of treatments for depression on the dynamic network structure of mental states: Two randomized controlled trials. *Scientific Reports*, *7*, 46523.

Spanakis, G., Weiss, G., Boh, B., & Roefs, A. (2015). Network analysis of ecological momentary assessment data for monitoring and understanding eating behavior. In *Icsh* (pp. 43–54). doi: 10.1007/978-3-319-29175-8_5

Strogatz, S. H. (2015). *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering.* Reading, MA.

Tong, H., & Lim, K. (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society: Series B (Methodological)*, *42*(3), 245–268. doi: 10.1111/j.2517-6161.1980.tb01126.x

Uink, B., Modecki, K. L., Barber, B. L., & Correia, H. M. (2018). Socioeconomically disadvantaged adolescents with elevated externalizing symptoms show heightened emotion reactivity to daily stress: An experience sampling study. *Child Psychiatry & Human Development*, *49*(5), 741–756. doi: 10.1007/s10578-018-0784-x

van de Leemput, I. A., Wichers, M., Cramer, A. O., Borsboom, D., Tuerlinckx, F., Kuppens, P., ... others (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. doi: 10.1097/PSY.0000000000000378

van der Maas, H. L., Molenaar, P. C., et al. (1996). Catastrophe analysis of discontinuous development. In (pp. 77–105). Academic Press.

van der Velden, R. M., Mulders, A. E., Drukker, M., Kuijf, M. L., & Leentjens, A. F. (2018). Network analysis of symptoms in a parkinson patient using experience sampling data: An n= 1 study. *Movement Disorders*, *33*(12), 1938–1944. doi: 10.1002/mds.93

van Os, J., Lataster, T., Delespaul, P., Wichers, M., & Myin-Germeys, I. (2014). Evidence that a psychopathology interactome has diagnostic value, predicting clinical needs: an experience sampling study. *PLoS One*, *9*(1), e86652. doi: 10.1371/journal.pone.0086652

van Rooij, I., & Baggio, G. (2020, Feb). *Theory before the test: How to build high-verisimilitude explanatory theories in psychological science.* PsyArXiv. Retrieved from psyarxiv.com/7qbpr doi: 10.31234/osf.io/7qbpr

van Winkel, M., Wichers, M., Collip, D., Jacobs, N., Derom, C., Thiery, E., ... Peeters, F. (2017). Unraveling the role of loneliness in depression: The relationship between daily life experience and behavior. *Psychiatry*, *80*(2), 104–117.

Visser, I. (2011). Seven things to remember about hidden markov models: A tutorial on markovian models for time series. *Journal of Mathematical Psychology*, *55*(6), 403–415. doi: 10.1016/j.jmp.2011.08.002

Visser, I., & Speekenbrink, M. (2010). depmixS4: An R package for hidden markov models. *Journal of Statistical Software*, *36*(7), 1–21. Retrieved from http://www.jstatsoft.org/v36/i07/ doi: 10.18637/jss.v036.i07

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. doi: 10.1080/00273171.2018.1496813

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. doi: 10.1111/j.2044-8317.2012.02043.x

Vrijen, C., Hartman, C. A., Van Roekel, E., De Jonge, P., & Oldehinkel, A. J. (2018). Spread the joy: How high and low bias for happy facial emotions translate into different daily life affect dynamics. *Complexity*, *2018*. doi: 10.1155/2018/2674523

Wagenmakers, E.-J., Molenaar, P. C., Grasman, R. P., Hartelman, P. A., & Van Der Maas, H. L. (2005). Transformation invariant stochastic catastrophe theory. *Physica D: Nonlinear Phenomena*, *211*(3-4), 263–276. doi: 10.1016/j.physd.2005.08.014

Wichers, M. (2014). The dynamic nature of depression: a new micro-level perspective of mental disorder that meets current challenges. *Psychological medicine*, *44*(7), 1349–1360. doi: 10.1017/S0033291713001979

Wichers, M. (2016). Critical slowing down as a personalized early warning signal for depression wichers, marieke; groot, peter c.; psychosystems; esm grp; ews grp. *Psychotherapy and psychosomatics*, *85*(2), 114–116.

Wichers, M., Kasanova, Z., Bakker, J., Thiery, E., Derom, C., Jacobs, N., & Van Os, J. (2015). From affective experience to motivated action: Tracking reward-seeking and punishment-avoidant behaviour in real-life. *PloS one*, *10*(6), e0129722. doi: 10.1371/journal.pone.0129722

Wichers, M., Schreuder, M. J., Goekoop, R., & Groen, R. N. (2019). Can we predict the direction of sudden shifts in symptoms? transdiagnostic implications from a complex systems perspective on psychopathology. *Psychological medicine*, *49*(3), 380–387. doi: 10.1017/S0033291718002064

Wichers, M., Wigman, J., & Myin-Germeys, I. (2015). Micro-level affect dynamics in psychopathology viewed from complex dynamical system theory. *Emotion Review*, *7*(4), 362–367. doi: 10.1177/1754073915590623

Wigman, J., Van Os, J., Borsboom, D., Wardenaar, K., Epskamp, S., Klippel, A., ... others (2015). Exploring the underlying structure of mental disorders: cross-diagnostic differences and similarities from a network perspective using both a top-down and a bottom-up approach. *Psychological medicine*, *45*(11), 2375–2387.

Wigman, J. T., Collip, D., Wichers, M., Delespaul, P., Derom, C., Thiery, E., ... others (2013).

Altered transfer of momentary mental states (atoms) as the basic unit of psychosis liability in interaction with environment and emotions. *PLoS One*, *8*(2), e54653. doi: 10.1371/journal.pone.0054653

Wigman, J. T., van Os, J., Thiery, E., Derom, C., Collip, D., Jacobs, N., & Wichers, M. (2013). Psychiatric diagnosis revisited: towards a system of staging and profiling combining nomothetic and idiographic parameters of momentary mental states. *PLoS One*, *8*(3), e59559. doi: 10.1371/journal.pone.0059559

Wolf, L. D., Davis, M. C., Yeung, E. W., & Tennen, H. A. (2015). The within-day relation between lonely episodes and subsequent clinical pain in individuals with fibromyalgia: Mediating role of pain cognitions. *Journal of psychosomatic research*, *79*(3), 202–206. doi: 10.1016/j.jpsychores.2014.12.018

Yang, X., Ram, N., Gest, S. D., Lydon-Staley, D. M., Conroy, D. E., Pincus, A. L., & Molenaar, P. C. (2018). Socioemotional dynamics of emotion regulation and depressive symptoms: A person-specific network approach. *Innovation in Aging*, *2*(suppl_1), 15–16. doi: 10.1093/geroni/igy023.053

Zucchini, W., MacDonald, I. L., & Langrock, R. (2017). *Hidden markov models for time series: an introduction using r*. CRC press.

A Predicted States for Mean Switching HMM for ESM time series

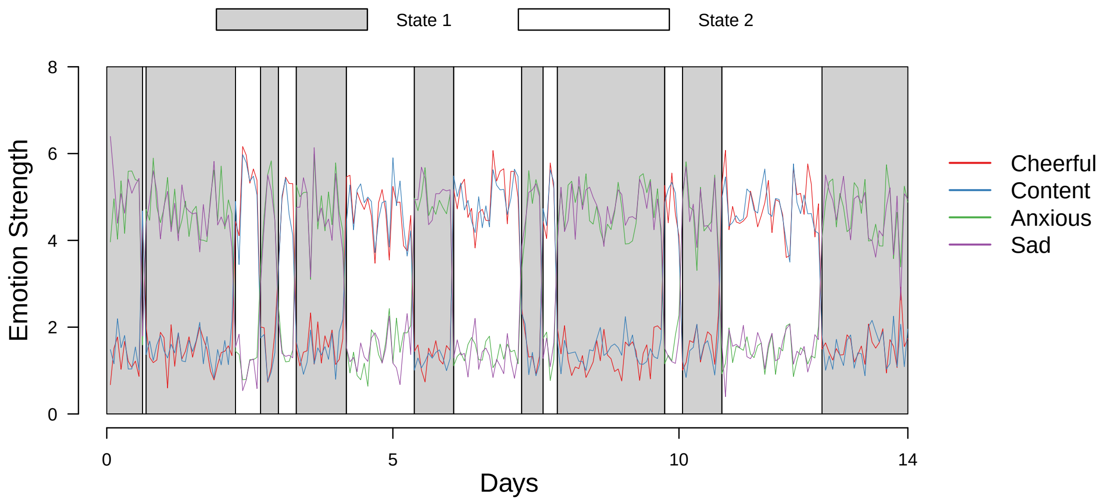

Figure 11: Time series of the four emotion variables, also shown in panel (a) of Figure 2, with background color indicating whether a given time point is assigned to the first or second component of the mean-switching HMM estimated from the ESM dataset.

B Continuous-Time VAR(1) results

Multiple researchers have pointed out that when data are generated from a differential equation model, the signs, sizes and relative orderings of lagged parameters of VAR model parameters depend on the sampling frequency in the dataset (Kuiper & Ryan, 2018; Ryan & Hamaker, 2020; Voelkle et al., 2018). In practice this means that direct inferences from VAR parameters to differential equation parameters is challenging due purely to this time-interval dependency. In the main text we have demonstrated in section 3.3 that it is difficult to make inferences about local characteristics such as the relative size of suppressing and reinforcing effects from the VAR model parameters due to the problem of model misspecification. Primarily, the VAR model is misspecified in that it allows only for linear relationships between processes, but strictly speaking, we could also say that the VAR model is misspecified in the sense that it is a discrete-time autoregressive model rather than a continuous-time differential equation model, of which the true system is a special case.

In order to remove this latter source of model misspecification, one can instead fit a continuous-time VAR model to data. This model is the integral form of (and thus equivalent to) a first-order stochastic differential equation with only linear relationships between variables, and so is still critically misspecified with respect to the true system. The parameters estimated by fitting a continuous-time VAR model fit to the ideal data, using the *stan* functionality in the R-package *ctsem*, (Driver et al., 2017), are displayed in network form in Figure 12 (cf. Ryan & Hamaker, 2020). Inspecting the drift matrix in panel (a), that is, the linear dependencies in the estimated differential equation model, we can see that we reach very similar conclusions as we did when interpreting the parameters of the DT-VAR model: The model yields negative moment-to-moment dependencies of approximately equal strength between-valence, with stronger positive dependencies, also of approximately equal strength, within valence. The diffusion matrix in panel (b), which is the continuous-time version of a residual covariance matrix, shows weak negative residual dependencies between each process, similar to the discrete-time residual dependencies (for more detail on the interpretation of this model see Ryan & Hamaker, 2020; Ryan et al., 2018; Voelkle & Oud, 2013). The similarity of the CT- and DT-VAR models in this case is likely due to the high sampling frequency present in the ideal data: As was the case for the discrete-time model, the problematic aspect of model misspecification in the current situation is the approximation of non-linear relationships with linear dependencies, a problem which cannot be solved by fitting a linear continuous-time model to the data.

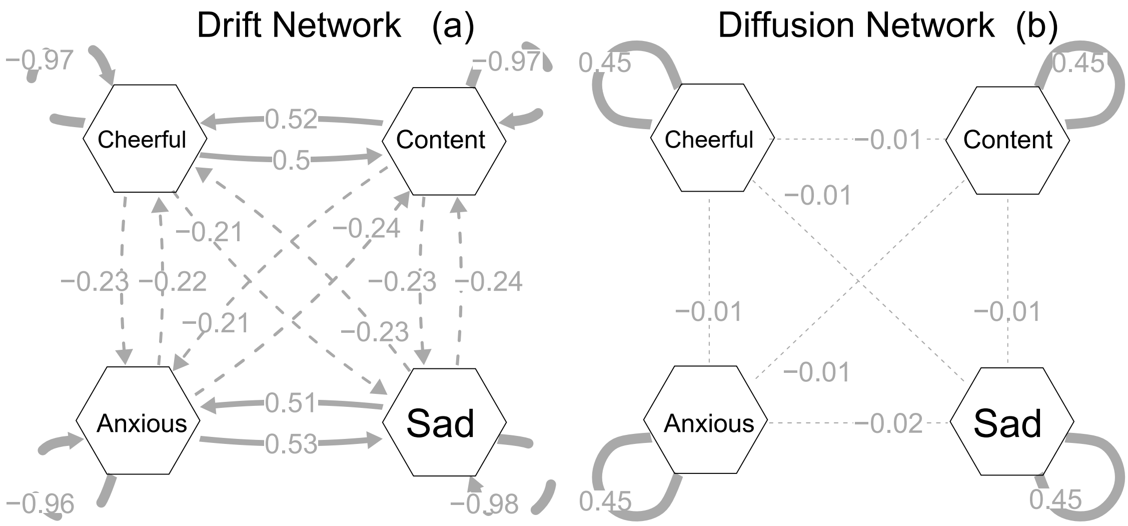

Figure 12: Parameter estimates from the CT-VAR model fit to the ideal time series. Panel (a) shows the estimated drift matrix parameters, the linear relationships between in the first-order differential equation model. Panel (b) shows the diffusion matrix, the residual covariance matrix of the continuous-time model

One may also wonder whether continuous-time models could help to overcome the problem of insufficient sampling frequency discussed in Section 4. In principle, continuous-time models can aid in making parameter estimates from different studies comparable, by mapping lagged dependencies at different time-intervals back to the same moment-to-moment dependencies in a differential equation model (Kuiper & Ryan, 2018; Voelkle & Oud, 2013). However, this will only work if the sampling frequency is sufficiently high such that some information on the moment-to-moment dependencies is present. To explore this, we also fit the continuous-time VAR model to the ESM time series described in Section 4. The estimated parameters are shown in Figure 13.

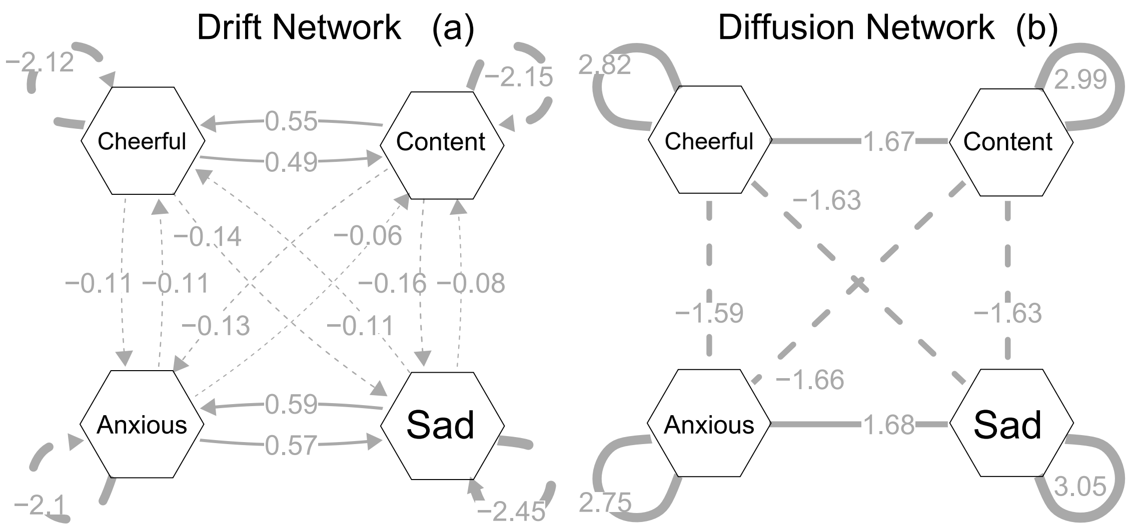

Figure 13: Parameter estimates from the CT-VAR model fit to the ESM time series. Panel (a) shows the estimated drift matrix parameters, the linear relationships between in the first-order differential equation model. Panel (b) shows the diffusion matrix, the residual covariance matrix of the continuous-time model

As we can see from Figure 13 (a) the estimated drift matrix parameters are quite different to those obtained from the ideal time series. Although we again obtain negative between-valence and positive within-valence relationships, the latter have an even greater difference in absolute value than in the ideal case. Furthermore, there is quite some variation in the size of the between-valence relationships, with the smallest effect (from Anxious to Content -0.06) more than half the size of the largest effect (from Content to Sad -0.16). Furthermore, although the qualitative patterns of relationships obtained here are somewhat different from those obtained from the discrete-time VAR model (Figure 9), no greater clarity or insight into the underlying dynamics is obtained from the continuous-time model.

That no direct inferences can be made from the continuous-time VAR parameters at a longer time-interval is hardly surprising, since the model is still misspecified. What is instructive however is that the continuous-time VAR parameters here do not capture the same linear approximation of the moment-to-moment dynamics as the continuous-time VAR parameters obtained from the high-sampling frequency dataset. In this sense, misspecified continuous-time models do not allow us to overcome the time-interval dependency problem in this setting. This may be entirely due to model misspecification, but we also suspect that this can be explained by the extremely low sampling frequency chosen for the latter dataset: Even if the continuous-time model was to be correctly specified, there is likely to be some limit to how large the time-intervals between observations can be in order to still enable model recovery in practice, since at long enough time-intervals the implied dependency between observations approaches zero for most systems.

C Simulated Data from Estimated Time Series Models

In this appendix we presented data generated from on models estimated on the ideal dataset in Section 3. This kind of simulation based model checking can be used to check the appropriateness of a model by allowing is to visually inspect which characteristics of the original time series a particular model can reproduce.

C.1 First-Order Vector Autoregressive Model

Figure 14 displays a time series of two weeks generated from the VAR(1) model in Section 3.3:

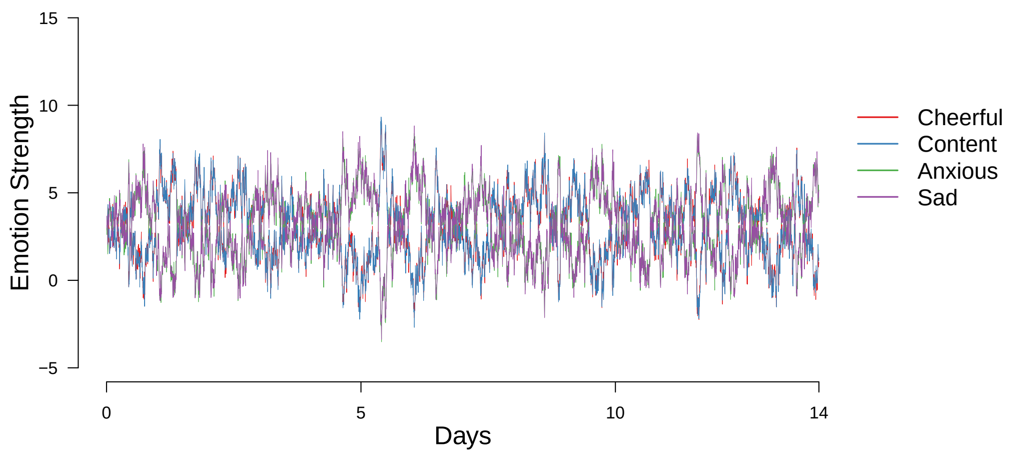

Figure 14: A time series of two weeks generated from the VAR(1) model estimated in Section 3.3

The generated data does not show bistability, which is expected because the VAR(1) model exhibits only a single fixed point. What looks approximately like oscillating behaviour is a result of the high auto-regressive effects present in the estimated VAR(1) model: given a stochastic input, the high auto-regressive effects ensure that the system is slow to eventually return to equilibrium. This oscillating behaviour is also evident in the eigenvalues of $\mathbf{\Phi}$, which consist of one complex conjugate pair (Strogatz, 2015).

C.2 Threshold Vector Autoregressive Model

Figure 15 displays a time series of two weeks generated from the threshold VAR(1) model in Section 3.4:

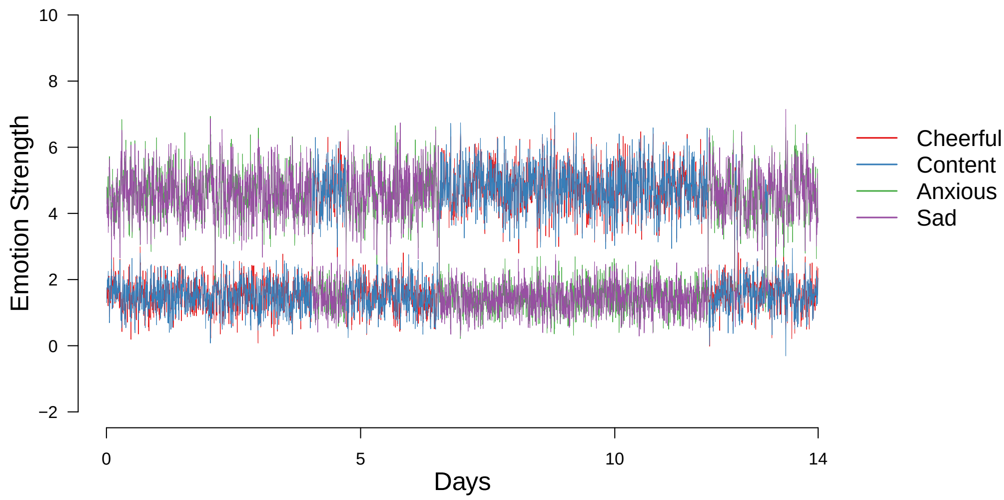

Figure 15: A time series of two weeks generated from the Threshold VAR(1) model estimated in Section 3.4

As expected, the generated data shows bistability as governed by a univariate threshold. We can also see that the combination of lagged regression parameters and residual covariance matrices unique to each state succeeds in reproducing the characteristic that emotions show a lower variability when they are in the low state (i.e., the healthy state for negative emotions, unhealthy state for positive emotions) than when they are in the high state (i.e., the unhealthy state for negative emotions, healthy state for positive emotions).

D Assumption Checking, Diagnostics and the Box-Jenkins Approach

In this appendix we present the results of applying standard time-series diagnostics and model assumption checking to the ideal dataset discussed in Section 3. Our analysis here is based on applying what is known as the Box-Jenkins approach (Box & Jenkins, 1976; Hamilton, 1994), a standard approach to investigate and tackle model misspecification in a time-series setting. The main tools used in this approach are the correlation and partial correlation functions, which allow us to visualize the marginal $(\rho(X_t, X_t), \rho(X_t, X_{t+1}) \ldots \rho(X_t, X_{t+T}))$ and partial correlations $(\rho(X_t, X_{t+1}), \rho(X_t, X_{t+2} \mid X_{t+1}), \ldots \; \rho(X_t, X_{t+T} \mid X_{t+1} \ldots X_{t+T-1}))$ in the time series respectively, as a function of the lag between measurements. The auto correlation and partial correlation functions depict the dependency of a time series variable on itself, and the *cross*-correlation and partial correlations depict the dependencies between variables. Due to the symmetries in the true system, we will here consider only two variables (as we did in Section 2.1), since the patterns observed between those two variables generalize to all within-valence and between-valence relationships.

Essentially, the Box-Jenkins approach involves using time-series plots and the correlation functions in an iterative way in order to a) establish the need to remove a trend or seasonal effect from the data such that the transformed data is stationary, b) choose an appropriate ARIMA model to fit to the (possibly transformed) data, and c) check the residuals of the chosen model for evidence of model misspecification. We begin by plotting the correlation and partial correlation functions based on the raw data, shown in Figure 16. In panel (a) we can see that the auto-correlation function of the Cheerful variable decays to zero very slowly, and oscillates between somewhat strong negative and positive values at longer lags. An identical pattern occurs for the cross-correlation function of the opposite-valence variable (in this case Anxious). In panel (b) however we see that the partial auto-correlation and partial cross-correlations are non-zero at a lag of 1, but are approximately zero at longer lags: When we control for the value of X_{t+1}, there is no longer any relationship between X_t and X_{t+2} or X_{t+3} and so forth. Note that although we plot the partial correlations only for a lag of 10, no strongly non-zero partial correlation appears at longer lags.

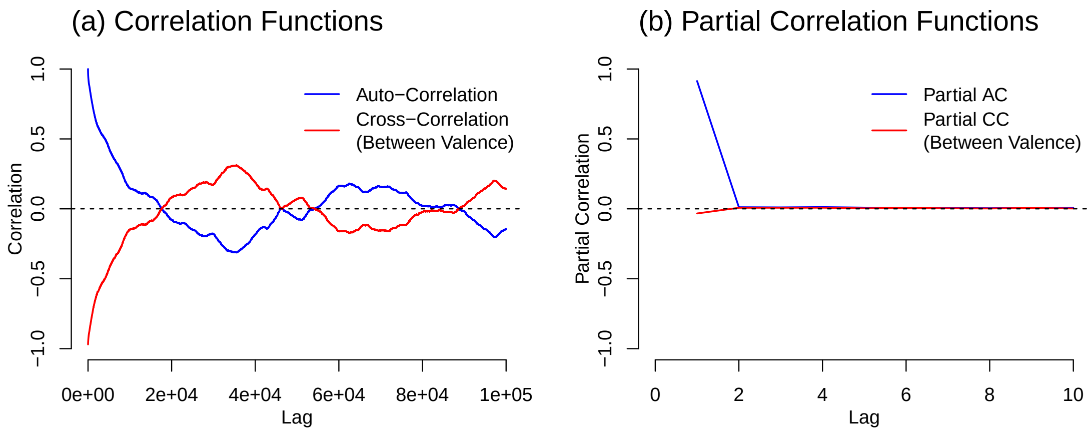

Figure 16: Correlation and Partial Correlation Functions for the raw time series. Here we plot only the auto-(partial)-correlations for Cheerful and the (partial) cross-correlation functions of Cheerful with Anxious.

Taken together, what might we infer from these diagnostics? Usually, a correlation function that oscillates between positive and negative values is taken to be suggestive of either a) some kind of trend or seasonal effect, or b) an indication of a higher-order autoregressive process. However, this second possibility is ruled out by the partial correlation function: If a higher-order lagged relationship was present, we would see this with some non-zero partial correlation at that longer lag. With this in mind, inspecting the raw time series (Figure 2) we may conclude that there is indeed some kind of cyclic or seasonal effect present. The time-series overall appears to be non-stationary, in that the mean of each process is not constant across our window of observation. In the true system, we know that this is the case because the system switches between two stable fixed points, a key characteristic produced by the microdynamics of the system. From a Box-Jenkins perspective, however, this suggests that, in order to fit an appropriate ARIMA model, we must first transform the data by removing this apparent source of non-stationarity.

Given this conclusion, the best we could hope to achieve would be to center the observed time series around a time-varying mean which exactly captures the position of the nearest stable fixed point. This produces a new transformed time series which is Gaussian, mean-stationary and appears to be at least approximately (co)variance-stationary, displayed in Figure 17 (a).

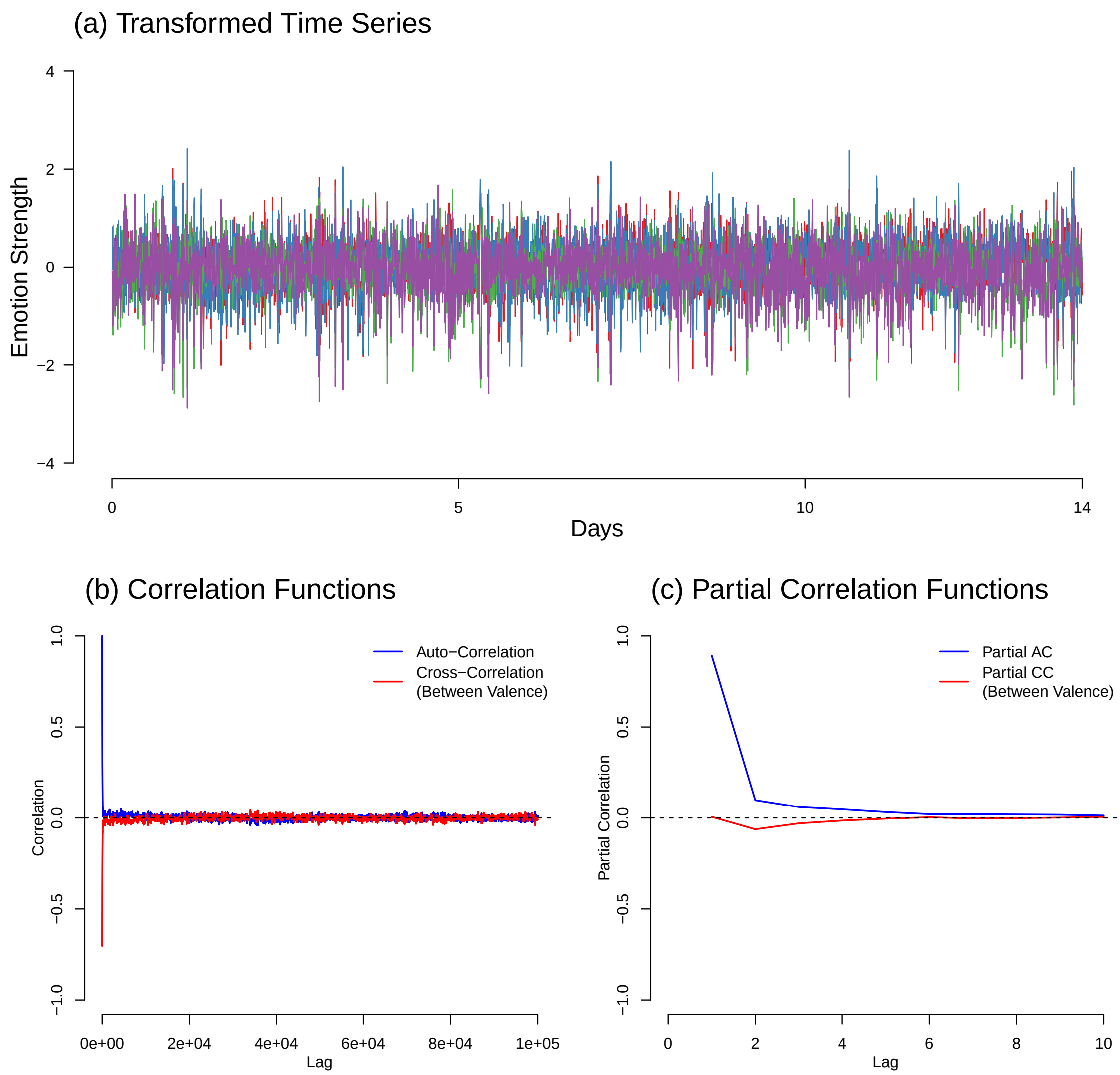

Figure 17: Panel (a) shows the ideal time series of the four emotion variables Cheerful, Content, Anxious and Sad, centered around the position of the nearest stable fixed point. This represents the best case scenario for a combination of de-trending and removing seasonal effects from the time series that we could hope to achieve. Panels (b) and (c) show the correlation and partial correlation functions for the transformed dataset, with the auto-correlation based on the Cheerful variable, and the cross-correlation based on Cheerful and Anxious.

The correlation and partial correlation functions calculated from this transformed data are shown in Figure 17(b) and (c) respectively. We can see from this that the auto-correlation function now decays quickly to zero at longer lags, and no longer shows the oscillatory pattern of the raw data. Instead, the auto- and cross-correlations look similar to what we would expect a first order (lag 1) system to produce. The partial correlations look qualitatively similar to those produced by the raw data. Taken together, these diagnostics (correctly) suggest that a lag-1 model would be appropriate for the transformed data. As such, a natural choice of model based on these diagnostics is the first order vector autoregressive (VAR) model.

Figure 18 displays the lagged relationships and VAR model obtained from the centered time series.

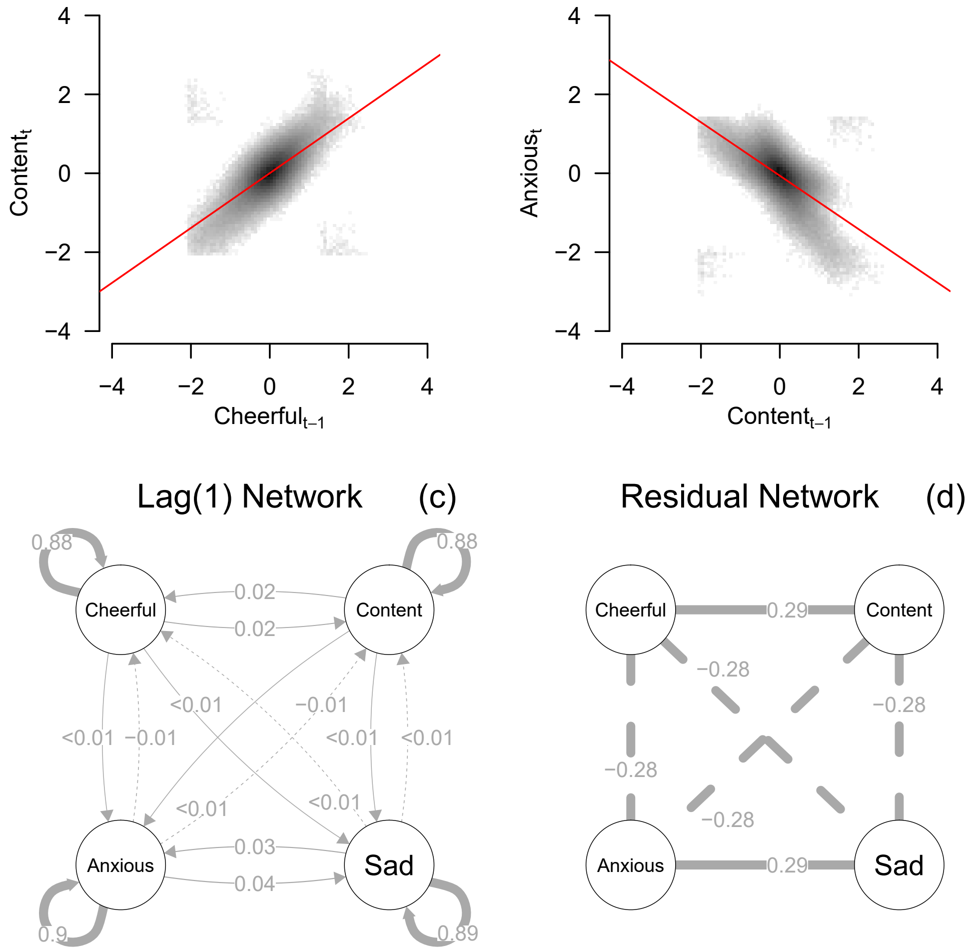

Figure 18: This figure replicates the VAR analysis in the main text using the centered time series. Panel (a) shows the relationship between Content and Cheerful, two emotions with the same valence, spaced one time point apart (at a lag of one). The red line indicates the best fitting regression model. Similarly, panel (b) shows the relationship between Anxious and Content, two emotions with different valence, at a lag of one in the centered time series. Panel (c) displays the matrix of lagged regression parameters, estimated from a VAR(1) model, as a network, and panel (d) displays the partial correlation matrix of the residuals of the VAR(1) model as a network. This latter network is often referred to as the contemporaneous network.

As we can see, although the VAR model appears to meet the necessary assumptions based on the transformed data, the resulting model gives us no greater insight into the underlying system than obtained previously. For instance, we would still conclude that the within-valence relationships are stronger than the between-valence relationships. In this model, we would even be unsure whether any between-valence relationships exist at all due to their small absolute value. Thus, although we have followed standard procedure and obtained a model whose assumptions appear to be met, we do not obtain an appropriate insight into the underlying mechanism of interest. Of course, the Box Jenkins method can be continued in an iterative way from this point forward, though our analysis of this approach ends here. Researchers may investigate the considerable autocorrelation still present in the residuals of this model, and use this to guide further model building: For instance, by including lag-2 effects, a different functional form for the lagged relationships, or a different time-series model entirely. Of course, some of these model building choices may eventually lead to the choice of an appropriate model, such as the threshold VAR discussed in Section 3.4. We suspect however a large number of models which could result from this process (such as any model which involves transforming the data to adjust for non-stationarity) would appear to be more or less appropriate based on these model diagnostics while failing to capture the key characteristics of the true system.

E Literature Review of ESM studies within the network approach to psychopathology

The goal of the literature review was to obtain an overview of the time scales of variables and the primary time series analyses used by empirical studies using ESM data that operate within the network approach to psychopathology. To do so, we compiled all empirical papers that analyzed ESM data and cited at least one of the following eight key papers on the network approach to psychopathology according to Google Scholar in May 2019: Borsboom (2017); Borsboom and Cramer (2013); Borsboom, Cramer, Schmittmann, Epskamp, and Waldorp (2011); Bringmann et al. (2013); Cramer et al. (2016); Hosenfeld et al. (2015); Schmittmann et al. (2013); Wichers, Wigman, and Myin-Germeys (2015). This procedure yielded the 43 papers shown in Table 1:

Paper	M/day	Main Analysis	Target Process	Seconds	Minutes	Hours	Days	Weeks
Wichers (2016)	10	PCA + VAR	Momentary mental states	1	1	0	0	0
Vrijen et al. (2018)	3	ML VAR	Emotions	1	1	0	0	0
Lee et al. (2017)	6	ML regression	Emotional, pain, sleep	1	1	0	0	0
van Winkel et al. (2017)	10	ML VAR	Loneliness, social contact, appraisal	0	1	1	1	0
Bringmann et al. (2013)	10	ML VAR	Positive/Negative Emotions	1	1	0	0	0
J. Wigman et al. (2015)	10	ML VAR	Momentary mental states	1	1	0	0	0
Bak et al. (2016)	10	VAR in subsets	Psychotic symptoms, emotions	0	1	1	1	0
Snippe et al. (2017)	10	ML VAR	Momentary affect and cognitions	1	1	0	0	0
Klippel et al. (2017)	10	ML mod/mediation	Emotions and psychotic experience	1	1	0	0	0
De Vos et al. (2017)	3	ML VAR	Emotions	1	1	0	0	0
Hasmi et al. (2017)	10	ML VAR	Emotions	1	1	0	0	0
Oreel et al. (2019)	9	ML VAR	Emotions, physical symptoms	0	1	1	1	0
Wigman et al. (2013a)	10	Lagged effects	Momentary mental states	1	1	0	0	0
Wigman et al. (2013b)	10	Hypothesis tests	Momentary mental states, psychosys liability	1	1	1	0	0
Kroeze et al. (2017)	5	(lagged) correlations	Mood, physical activity, social context	1	1	1	0	0
Geschwind et al. (2011)	10	Hypothesis tests	Momentary positive emotions	1	1	0	0	0
Levinson et al. (2018)	4	ML VAR	Eating disorder cognitions and behaviors	0	1	1	0	0
Pe et al. (2015)	7	ML VAR	Positive/negative emotions	0	1	1	0	0
Pavani et al. (2017)	5	moderated ML VAR	Positive/negative affect, rumination, appraisal	1	1	1	0	0
Lutz et al. (2018)	4	ML VAR	Momentary affective states	1	1	0	0	0
Beck et al. (2020)	4	VAR models	Personality traits at state level	0	1	1	1	0
Poerio et al. (2016)	3	Mediation model	Daydreaming, mood, and dissociative symptoms	0	1	1	0	0
Klippel et al. (2018)	10	ML VAR	affect, daily stressors, psychotic experiences	1	1	0	0	0
Greene et al. (2018)	2	ML VAR	PTDS related intrusions, avoidance, mood	1	1	1	1	0
Yang et al. (2018)	6	VAR models	Emotions, depression symptoms	1	1	1	1	1
Bringmann et al. (2016)	10	ML VAR	Emotions, stress	1	1	1	0	0
Aalbers et al. (2019)	7	ML VAR	Momentary depr. symptoms, social media usage	0	1	1	0	0
Fisher et al. (2017)	4	ML VAR	Mood, anxiety	0	1	1	1	0
van de Leemput et al. (2014)	10	AR(1) and Var	Emotions	1	1	0	0	0
Clasen et al. (2015)	5	Lagged effects	Feeling well, self-esteem	1	1	1	0	0
Brose et al. (2017)	10	Cross-lagged model	Indicators of micro-level stress	0	1	1	0	0
Hoorelbeke et al. (2019)	6	ML VAR	Positive affect, rumination, positive appraisal	1	1	1	0	0
Snippe et al. (2018)	3	ML VAR	Prosocial behavior, positive affect	1	1	1	0	0
Bernstein et al. (2019)	5	ML VAR	Present emotion, physical activity	1	1	1	0	0
Spanakis et al. (2015)	10	ML VAR	Emotion, eating behavior	1	1	1	0	0
Fang et al. (2019)	8	State space analysis	Affect, rumination	1	1	0	0	0
Van der Velden et al. (2018)	10	VAR	Motor symptoms, mood states	0	1	1	0	0
Hasmi et al. (2018)	10	ML VAR	Emotions	1	1	0	0	0
Van Os et al. (2014)	10	ML regression	Psychotic experiences, affect, psychotic symptoms	1	1	1	1	1
Wichers, Kasanova, et al. (2015)	10	Lagged effects	Affect, appraisal of social context, physical activity	0	1	1	0	0
Cristóbal-Narváez et al. (2016)	8	ML regression	Psychotic-like experiences, stress reactivity	0	1	1	0	0
Wolf et al. (2015)	4	Lagged ML regression	Loneliness, clinical pain	0	1	1	0	0
Uink et al. (2014)	5	ML AR change model	Emotions, daily stress	1	1	1	0	0

Table 1: ESM-measurements per day, main analysis, target process and coded time scale of the target process for 43 empirical ESM studies working within the network approach to psychopathology. A one indicates that for at least one of the studied variables it seems reasonable that it evolves at the given time scale.

For each study we coded the main analysis and the main target process under investigation. We define the target process as the process that is captured by the ESM-questions. For example, Fang et al. (2019) capture daily dynamics in rumination and affect with ESM-questions, and then evaluate whether those dynamics predict depressive symptoms at a follow up. In such cases, we treat the variables captured by ESM as parts of the target process.

We then tried to categorize the target process in terms of its characteristic time scale, which we define to be smallest time scale at which meaningful change can be observed. We use the five categories: seconds, minutes, hours, days, and weeks. In none of the 43 papers we were able to clearly identify the time scale, and we therefore assigned all time scales that seemed plausible based on how those processes are typically discussed in the substantive field. In some studies the ESM questions include variables with clearly different time scales, like emotions and symptoms. In these cases we collapsed to coding of all variables into a single coding for all variables captured.

Many studies assert that they aim to study “moment-to-moment“ mental states or relations. While this seems to refer to a time scale of seconds, we are unsure whether authors take “moment” literally and therefore also assign the time scale of minutes to these studies. Whenever a study investigates emotion dynamics, we assign a time scale of seconds to minutes, since emotions are defined on such a time scale to differentiate them from mood (e.g., Houben et al., 2015). Next to the target process and the intuitively coded time scales we also report the type of main analysis how many measurements were taken each day.

A few things in Table 1 stand out. With respect to the time scale of variables, we were not able to identify a single time scale for the studied processes in any of the 43 papers. For 58% of the papers we assigned two time scales, and for the remaining three or more. In addition, we identified a time scale of minutes for all of the 43 papers. In relation to the problem of insufficient sampling frequency, the results of our small literature review are interesting for at least two reasons: first, it is often unclear on which time scale the processes of interest are evolving. And second, most processes seem to evolve at a time scale of seconds and minutes, which are possibly difficult to capture with ESM measurements with a frequency in the order of hours.

Many ESM studies (47%) use 10 measurements a day, but a lot of studies also use substantially less. While there was a large variety in analyses performed, the main analysis was usually a linear regression model at the same time point, or a lagged linear regression model, usually in a multilevel setting. The most ubiquitous main analysis was the multilevel VAR model, which was used in 49% of all papers as the main analysis.