

Inference With Cross-Lagged Effects—Problems in Time

Charles C. Driver

Institute of Education, University of Zurich

Abstract

The interpretation of cross-effects from vector autoregressive models to infer structure and causality among constructs is widespread and sometimes problematic. I describe problems in the interpretation of cross-effects when processes that are thought to fluctuate continuously in time are, as is typically done, modeled as changing only in discrete steps (as in e.g., structural equation modeling)—zeroes in a discrete-time temporal matrix do not necessarily correspond to zero effects in the underlying continuous processes, and vice versa. This has implications for the common case when the presence or absence of cross-effects is used for inference about underlying causal processes. I demonstrate these problems via simulation, and also show that when an underlying set of processes are continuous in time, even relatively few direct causal links can result in much denser temporal effect matrices in discrete-time. I demonstrate one solution to these issues, namely parameterizing the system as a stochastic differential equation and focusing inference on the continuous-time temporal effects. I follow this with some discussion of issues regarding the switch to continuous-time, specifically regularization, appropriate measurement time lag, and model order. An empirical example using intensive longitudinal data highlights some of the complexities of applying such approaches to real data, particularly with respect to model specification, examining misspecification, and parameter interpretation.

Translational Abstract

The interpretation of causality in longitudinal modeling endeavors within psychology is widespread and sometimes problematic. I describe problems in the interpretation of causal effects when processes that are thought to fluctuate continuously in time are, as is typically done, modeled as changing only in discrete steps. I demonstrate these problems via simulation, and also show that when an underlying set of processes are continuous in time, even relatively few direct causal links can result in denser patterns of relationships in discrete-time. I demonstrate one solution to these issues, namely representing the system as a stochastic differential equation, and focusing inference on the continuous-time effects. I follow this with some discussion of issues regarding the switch to continuous-time, specifically regularization, appropriate measurement time lag, and model order. The article concludes with an empirical example addressing some of the complexities of applying these approaches in practice.

Keywords: continuous time, dynamic model, cross-lagged, vector autoregressive, stochastic differential equation

Supplemental materials: <https://doi.org/10.1037/met0000665.supp>

Predicting the future based on a set of past observations is, in at least some sense, a straightforward problem. We may be wrong to varying degrees, but the basic tools of statistics and computation have evolved to the point where standard recipes for the use of past data to predict

future data can often give quite reasonable results. In contrast, predicting how the future will be different if we intervene in a specific way, based on the same set of past observations, is in most cases far more difficult. Therein lies both the promise and the pain of much of statistical modeling. Discussing mere correlation is easy, causal interpretation under the assumption of having fit the “true” model is easy, but unfortunately, most science tends to be different—causality is of interest and models plus data are only approximations of a more complex reality. In this work, I address the representation of psychological constructs over time, specifically in regards to limitations and common pitfalls of inference when processes that are thought to operate continuously are modeled as if they only change at discrete moments in time. The implications of choosing between discrete and continuous time have been considered before (Aalen, 1987; Aalen et al., 2016; Deboeck & Preacher, 2016; Deboeck et al., 2018; Kuiper & Ryan, 2018; Ryan & Hamaker, 2022), but the incorrect view that the distinction is only important when time-intervals vary between observations appears pervasive in the field. The distinction has substantial

This article was published Online First July 18, 2024.

Sacha Epskamp served as action editor.

Charles C. Driver  <https://orcid.org/0000-0002-4174-2970>

This work was funded by the Swiss National Science Foundation. The author would like to thank Manuel Voelkle and Martin Tomasik for comments and critique and Ethan McCormick for the empirical data suggestion. Analysis scripts and generated data are available in the [online supplemental materials](#).

Charles C. Driver served as lead for the entire work.

Correspondence concerning this article should be addressed to Charles C. Driver, Institute of Education, University of Zurich, Freiestrasse 36, 8032 Zurich, Switzerland. Email: charles.driver2@uzh.ch

importance for the interpretation of temporal and system-noise-related parameters, particularly with respect to typical practices of fixing or regularizing cross-effect parameters to zero. After elaborating on this issue of what the different representations imply, I discuss some related considerations that arise when switching from discrete to continuous-time representations. These include the time-scale of the processes and how estimates of such interact with measurement error, model order and structure, as well as some discussion of model fit assessment with such models. Following this is an empirical example of the continuous-time approach, using intensive longitudinal data of affect and physical state (Fried et al., 2022). In this, questions of measurement, model order, individual differences, and other complexities are combined. The article concludes with a general discussion and outlook.

Causality and Time

In classical experimental settings such as a randomized controlled trial, the statistical model can sometimes be quite simple, and the interpretation seemingly clear—the prime inferential difficulty in such cases is usually mapping understanding of one or more very specific, controlled contexts, to the broader and messier reality we're usually interested in (some common concerns are highlighted by Dekkers et al., 2010; Rothwell, 2006). The use of observational data approaches typically aims to reduce some of these generalization issues (i.e., data better reflects the circumstances of interest; Degtiar & Rose, 2023), but from this arises what can be thought of as an inverse generalization issue—that of going from “what is likely to happen to y when x changes for any reason” to “what is likely to happen to y if x is directly manipulated.” For a tangible example, when the number of people with umbrellas on the street increases, ice cream sales likely decrease, so umbrella carrying may be a useful predictor of ice cream sales. However, taking peoples umbrellas from them during a downpour is unlikely to increase peoples desire for ice cream—umbrella carrying and ice cream sales are influenced by a common cause, the weather.

Considering the ice cream umbrella system described, a critical feature that could help a naive observer make sense of the causal relations in the system is the relationship of the three variables over time. Changes in umbrella carrying and ice cream buying will tend to happen after changes in the weather. Moreover, after accounting for the information contained in all past states of the weather, umbrella carrying will not contain useful information for predicting ice cream sales. Such ideas of temporal causality were formalized by Wiener (1956), made practical in linear model form by Granger (1969), and have since seen nonlinear generalizations (for an overview see Hlaváčková-Schindler et al., 2007). The general tenets of this approach are that causes happen prior to their effects, and causes contain unique information about future effects.

Within this approach, variables are sometimes considered “Granger causal”—this term implies that in the model of causality based on the approach of Granger, the variables appear causally linked, but that additional variables or model changes may alter this. This typical view of Granger causality appears then a very weak notion of causality. A more standard view of causality is that of the counterfactual perspective, wherein if we intervened directly on a causal variable we would expect to see the downstream changes predicted by the model (for some major works, see Lewis, 1973; Pearl, 2009; Rubin, 2005; Spirtes et al., 2000). Imagine that we do not include

the weather variable in an analysis of our earlier system, then umbrella carrying is useful information to predict future ice cream sales, since it is related to the weather, and might be considered “Granger causal” according to common parlance. However, if we intervene on umbrella carrying at a given point in time, then of course no specific change in later ice cream sales is to be expected. Granger (1980) did discuss at some length the information necessary to include in an analysis of causality, and this can be broadly understood as ensuring there are no unmeasured confounders (i.e., common causes) (Shojaie & Fox, 2022). The approach of Granger can be seen as a search for genuine causal effects, but conclusions should always be tempered by one's confidence in the many assumptions of the models.

In the psychological domain, models that are commonly used (often implicitly) in the Granger causality framework include the cross-lagged panel (Campbell, 1963), vector autoregressive (VAR) (Orcutt & Irwin, 1948), latent change score (McArdle, 2001), as well as many extensions and combinations (e.g., Asparouhov et al., 2018; Bringmann et al., 2017; Epskamp et al., 2018; Gates & Molenaar, 2012; Zyphur et al., 2020). Ruissen et al. (2022) provided a recent overview contrasting some of the different approaches and software. All of these models contain (or can be rewritten to contain, e.g., Voelke & Oud, 2015) some form of parameter that can be understood as a cross-lagged effect—the (potentially causal) effect of one variable at an earlier time, on another variable at a later time. The common, first-order (lag one) version of a VAR can be written as:

$$\mathbf{y}_u = \mathcal{A}\mathbf{y}_{u-1} + \mathcal{B} + \mathcal{G}\mathbf{z}_u, \quad \mathbf{z} \sim N(0, 1), \quad (1)$$

where $\mathbf{y}$ is a d length vector of process values (possibly but not necessarily data), u is an index of measurement occasion, $\mathcal{A}$ is a matrix of temporal regression coefficients, $\mathcal{B}$ is an intercept, and $\mathcal{G}$ is the effect matrix of the d length system noise vector $\mathbf{z}$, where $\mathbf{z}$ contains independent and identically distributed deviations with zero mean. Many extensions and variations are in use, nonlinear forms may be written more generically as functions, but the issues discussed herein, while broadly applicable, are simplest to understand using the basic form shown.

Cross-lagged effects in longitudinal models have for long offered a tantalizing vision of causal inference in observational data, but their interpretation has for just as long been problematic and controversial, as interpretations often go far beyond what can reasonably be said given the data and modeling. Shojaie and Fox (2022) pointed out a number of important assumptions of the standard VAR model for causal inference. These core assumptions (quoted verbatim from Shojaie & Fox, 2022) are:

1. Continuous-valued series: All series are assumed to have continuous-valued observations. However, many interesting data sources—such as social media posts or health states of an individual—are discrete-valued.
2. Linearity: The true data-generating process, and correspondingly, the causal effects of variables on each other, is assumed to be linear. In reality, many real-world processes are nonlinear.
3. Discrete time: The sampling frequency is assumed to be on a discrete, regular grid matching the true causal time lag. If the data acquisition rate is slower or otherwise irregular, causal effects may not be identifiable. Likewise, the analysis of point processes or other continuous-time processes is precluded.

4. Known lag: The (linear) dependency on a history of lagged observations is assumed to have a known order. Classically, the order was not estimated and was taken to be uniform across all series.
5. Stationarity: The statistics of the process are assumed time invariant, whereas many complex processes have evolving relationships (e.g., brain networks vary by stimuli and user activity varies over time and context).
6. Perfectly observed: The variables need to be observed without measurement errors.
7. Complete system: All relevant variables are assumed to be observed and included in the analysis—that is, there are no unmeasured confounders. This is a stringent requirement, especially given that early approaches for Granger causality focused on the bivariate case—that is, they did not account for any potential confounders.

All of these assumptions can be relaxed to varying degrees by different model extensions and approaches. While there are undoubtedly difficulties and scope for improvement with respect to all listed items, in this article, I focus on the discrete-time assumption, as although there is a known solution, there does not appear to be widespread understanding of the scope of the issue in the field.

The discrete-time assumption is essentially that the processes in question only change and interact when observations occur—only at these moments do the processes suddenly experience random shocks and influence from other processes, with supposedly no change happening in the time period between observations. Although this assumption is not always a problem for causal inference in first-order bivariate scenarios, with anything beyond first-order dynamics, or greater than two variables in the system, inferences can become invalid. While it is true that this problem is simply a specific variant of the general problem that misspecified models usually provide inaccurate estimates, the particular form of misspecification in this case is somewhat unusual and many researchers may not have considered it before.

Violations of the discrete-time assumption can have substantial implications for both classical hypothesis testing approaches, and data-driven approaches seeking the temporal structure of processes. The consideration of these two, via numerical example as well as simulation, constitutes the next section of this article. In addition, it becomes apparent when resolving problems with the discrete-time assumption that one cannot observe the system at every moment of change—this makes the known lag assumption important to address in the subsequent section.

The Discrete-Time Assumption and Misspecified Temporal Structure

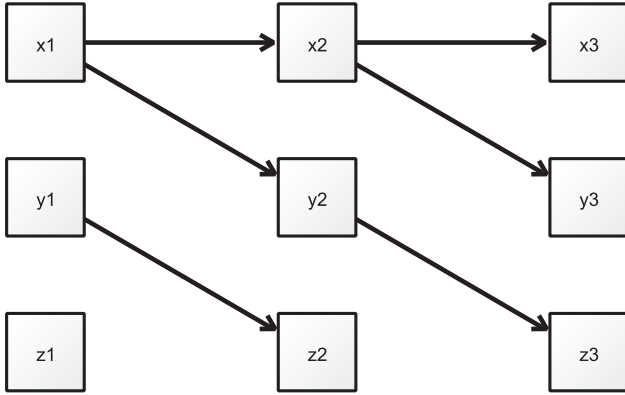
A common scenario where causal inferences are made using VAR-type models is when competing models are created based on theory, and their performance compared. A simple example of such could be the earlier question of whether umbrella carrying predicts ice cream sales—a model including an effect of umbrella carrying on sales would be compared to a model where this effect is fixed to zero. In contrast to the theory-driven approach, increases in both data availability and computation in recent years have seen the development of data-driven approaches. In this case, formal hypotheses are not explicitly specified but instead a model structure

is estimated by including a preference for sparsity in the estimation approach—cross-lagged coefficients of zero are preferred. This sparsity preference is achieved by forms of regularization (e.g., the L1 penalty or Lasso of Tibshirani, 1996), where additional terms are included in the estimation routine that preference smaller and or zero values of cross-lagged coefficients. Without such a sparsity preference, even when the effect of umbrella carrying on sales is truly zero, the estimated coefficient would not be exactly zero due to a limited observation size. Both classical and data-driven inference approaches to cross-lagged coefficients are regularly misapplied in psychological research. Cross-lagged panel models, multivariate change-scores, VAR models, and similar approaches all implicitly assume processes evolve in discrete jumps in time, when in reality researchers tend to think of the underlying processes as continually in flux. While the need for continuous-time systems for inference with stochastic processes has already been discussed somewhat, (see e.g., Aalen, 1987; Aalen et al., 2016; Deboeck & Preacher, 2016; Deboeck et al., 2018; Kuiper & Ryan, 2018; Ryan & Hamaker, 2022), discrete-time cross-regressive models are still the dominant paradigm for inference in such settings. To explicitly highlight the problem, in this section, I provide some examples in the domains of both hypothesis testing, and structural discovery via regularization. The take-home message will be that even in relatively sparse causal systems, the temporal regression matrix between observations at different times is very likely to be dense (contain no zeroes), or at least less sparse (contain fewer zeroes). In some cases, there can also be similar sparsity (i.e., number of coefficients with value zero) to the true system but in different locations. The fundamental problem is that this discrepancy invalidates attempts to infer structure or direct causal paths via the specification of, or regularization to, zeroes in the temporal regression coefficients. This issue arises due to a difference in how the systems are typically expected to behave, which is continuously (or very frequently) interacting, and how they are modeled, which is interacting only when measured. When the underlying system is continuously changing, this discrepancy can be resolved by directly estimating a continuous-time temporal effects matrix, and specifying or regularizing this to zero—the corresponding discrete-time regression weights and covariances can then be computed based on the continuous-time temporal effects, or “drift,” matrix, and the time interval between observations. Important to note, is that this all applies whether or not time intervals vary between observations or subjects.

To facilitate understanding, I first consider the temporal structure issue in a purely discrete-time context, where change can genuinely be represented as occurring at specific times. When a cross-lagged model represents a hypothesis that (a) the processes truly evolve in discrete jumps, and (b) the measurement occasions u capture each and every one of these discrete jumps, then there is no problem and we can have a perfect representation. To consider such a system, we could imagine three people exchanging information by post, and the post is delivered daily. Person x writes to person y , and person y in turn passes this information on to person z . If we know that person x has received new information, we can expect y to become aware of this the following day, while it will take an additional day before z knows it. In this case, we can precisely represent the causal pathways using a temporal regression matrix (or path diagram) that represents change over a time interval of 1 day or one postal delivery, as shown in Figure 1. Here, y is entirely dependent on x for information, and z

Figure 1

Representation of Daily Packets of Information Flow Between Persons x , y , and z , Across 3 Days



is dependent on y .

$$\begin{matrix} & x & y & z \\ \begin{matrix} x \\ y \\ z \end{matrix} & \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \end{matrix} \quad (2)$$

Using such a representation, our model will perform optimally in terms of prediction, where we can use the knowledge of what information each individual has on a particular day to predict future days. Furthermore, we can also use the structure to understand what will happen if the structure changes—if for instance, the link between y and z is removed, we know that z will not receive any information from x .

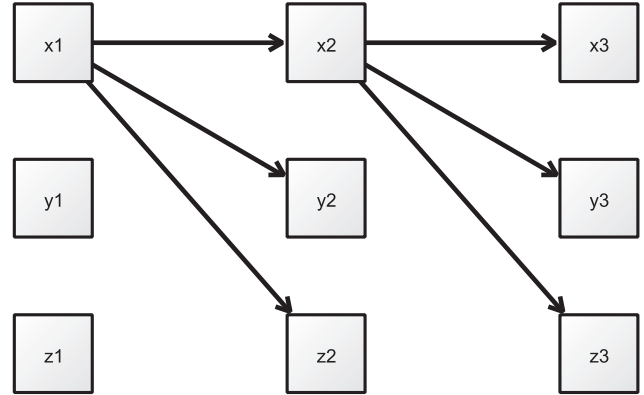
What happens to our representation if we instead shift to an observation time interval of 2 days or two deliveries though? As soon as we deviate from the “true” time step over which information flow occurs, our representation no longer matches the data-generating process, or causality, inherent to the system. For an interval of 2 days, the temporal regression matrix would look as follows, and the path diagram as shown in Figure 2. When the new time interval is simply an integer τ multiple of the original, the new temporal regression matrix can be obtained by matrix multiplying the original matrix by itself $\tau - 1$ times.

$$\begin{matrix} & x & y & z \\ \begin{matrix} x \\ y \\ z \end{matrix} & \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \end{matrix} \quad (3)$$

In this case, we have accurately represented how information flows through the system at time intervals of 2 days, in that once a change in x occurs, 2 days later we see changes in both y and z . So, we can still use the representation to make accurate predictions on a 2-day interval. However, we clearly have not represented the actual causal mechanism in the system. That is, our system does not accurately represent the way in which information flows through

Figure 2

Representation of Daily Packets of Information Flow Between Persons x , y , and z , Across 6 Days, When the Time Interval Is 2 Days



the system at other moments in time, and thereby can lead to incorrect estimates of causal effects under certain (hypothetical) interventions. Some causal effects still turn out accurate, as we see that the effect of a change in person x on person z 2 days later is correct. However, one might also be tempted to infer from this 2-day representation that person y is unnecessary for the information flow between x and z , but, in fact, if we were to intervene on the system and permanently block the route from y to z , we would see that no mail reaches z —person y is required if any information from x is to reach z .

Considering causality in such a way, where we also think of adding or removing paths between variables, may seem slightly unusual to those familiar with the consideration of causal effects by intervening on variables rather than paths. Under such a variable-oriented view, expectations of causal effects can be computed by intervening and setting one of the variables in the system (i.e., x , y , or z) to a specific value, either for a moment or for some length of time (Ryan & Hamaker, 2022). It is important to note that paths between variables may also be included in a system as variables themselves (e.g., via nonlinear state-space modeling).

This discrepancy between the true causal system (or generating model), and the information flow given an arbitrary unit of time, captures a problem inherent to many longitudinal modeling endeavors. Such a discrepancy is precisely why discrete-time cross-regression parameters cannot be naively used as the basis of hypothesis tests regarding direct causal paths in the underlying process, or as indicators of some generative structure when achieved via regularization. The causal conclusions that result can be invalid, even when the true model is some form of VAR and plenty of data is available—unless every possible occasion of change (whether observed or not) is included in the model.

As we can see from the example, problems related to hypothesis tests and structure determination with discrete-time models occur even when the true data-generating function is a discrete time function, whenever more than one step of change has occurred between observations. This poses something of a problem, because rather than interacting only when measured, most psychological processes are thought of as continuously existing and interacting with each other (Chow & Zhang, 2008; Voelkle et al., 2019), and regular longitudinal structural equation modeling and network modeling

approaches cannot represent such a data-generating process. While this is not inherently problematic when models are used only for predictive purposes—in the postal example, we could still make accurate 2-day forward predictions using the 2-day temporal regressions—it is a problem whenever the underlying causality (i.e., the mechanisms that bring about change), is of interest.

Continuous Time

Instead of modeling processes as a sequence of discrete jumps, differential equations—the mathematics of continuously changing processes—have been widely used in the sciences for hundreds of years. While the fields of psychology have recently shown more awareness of differential equation approaches, the full ramifications with respect to understanding and testing causal structure still appear little understood. Hansen and Sokol (2014) provided a detailed overview of causal interpretations of continuous-time models, and Maclaren and Nicholson (2021) highlighted how stochastic-differential equations can be incorporated into an expanded concept of the typical causal inference frameworks in use. While some of the mathematical computations required for software to solve the equations and fit continuous-time models can appear complex, an intuitive understanding of the approach can be sufficient to resolve a range of time-related confusions (for more on such confusions, see Voelkle et al., 2019).

In discrete-time systems, the matrix of temporal effects represents regression strengths between two points in time, as in the 1- or 2-day examples discussed in the postal example. Continuous-time approaches can be intuitively thought of in much the same way, but simply compressing the time interval to a “very small” value. This allows for usage of a conceptually similar temporal effects matrix, the only difference being that the matrix represents the influence of the current state of the system on the direction of change in the processes (analogous to the approach used in change-score structural equation models, where an autoregression of 1.00 is always applied in addition to any estimated self and cross feedback parameters). This contrasts to the discrete-time form where an earlier state of the system is used to predict the new level of the processes. A continuous-time form of the VAR discussed is:

$$dy(t) = (Ay(t) + b)dt + GdW(t). \quad (4)$$

This looks similar to the discrete-time form, but instead of telling us the new value of y given one step forward in time, it tells us how y is changing *at the moment*. Some mathematical complications due to the nature of stochastic differential equations are present: the dt on the right-hand side, which can be thought of as a very small step in time, and the $dW(t)$, which represents white noise in continuous-time. To compute y at some time point given an earlier value, one needs to solve the system. This is numerically involving and described in detail by Voelkle et al. (2012), however, approximations using linear extrapolation are a basic building block and can be done by hand, also aiding intuition:

$$y_u \approx y_{u-1} + \Delta t(Ay_{u-1} + b) + Gz_u \quad z \sim N(0, \Delta t). \quad (5)$$

Essentially, one computes an expected slope at observation occasion u using A , b , and the earlier value of y at observation occasion $u-1$. This is then multiplied by the length of the desired time step Δt to attain the expected change. The white noise element z , with variance

Δt , combines with G , the effect of the white noise, to give a system noise (co)variance of $GG^T \Delta t$. Shorter steps are (up to a point) more accurate, but require more computations. %Many refinements are possible – this only represents the basic idea, which is called the Euler-Maruyama method in the context of stochastic differential equations. The linear system shown also allows an ‘exact’ one-step solution. While this is described in more detail elsewhere \citep[e.g.]{contains the equations and code for plotting}{driver2018hierarchicala}, the basic component involves the matrix exponential, with temporal regression coefficients for particular time intervals given by $e^{A\Delta t}$.

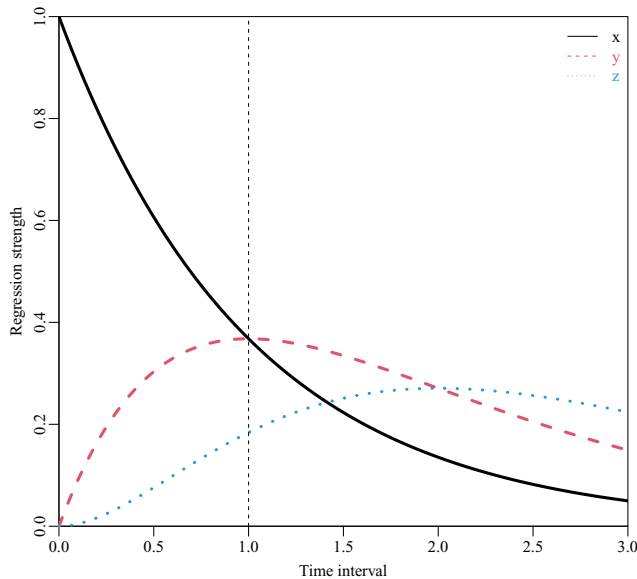
The continuous-time approach becomes important when the underlying system is assumed to be continuously changing and interacting. The exact proportion of psychological variables where this is the case may be an open question, but there are certainly many cases where a continuous assumption is more plausible than a discrete assumption. To illustrate this we can consider a similar system to the postal example, but where interactions occur continuously. In this case, we have x (motivation) affecting y (exercise) which in turn affects z (fitness). The continuous-time temporal matrix is:

$$\begin{matrix} & x & y & z \\ \begin{matrix} x \\ y \\ z \end{matrix} & \begin{bmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \end{matrix}. \quad (6)$$

The negative coefficients on the diagonal tell us that increases in any of the variables exert a downward pressure on that same variable in the future, which is typical of systems that fluctuate around a certain range rather than exploding out to negative or positive infinity (as can occur with positive diagonals). The positive cross-effects show where a change in one variable (determined by the column) leads to positive change in another (determined by the row). Figure 3 illustrates the implied discrete-time coefficients for the effect of a change in x on all three-system variables at later time points, as the time interval changes. This is computed using $e^{A\Delta t}$. The value of x at time zero (i.e., when the change in x occurs) becomes less predictive of later values of x , so the coefficient reduces. Initially, the time zero value of x is not predictive of either y or z , but as time goes by the change in x exerts influence on y , and then via y on z , so the regression coefficients rise. At some point, once the initial change in x has worn off to a sufficient degree, the coefficient for x predicting y reaches a peak and starts to reduce, and similarly for z . Note that while the cross-regression strength depends on the time interval, the underlying continuous-time cross-effect is constant over time, with the plot showing how continuous-time effects accumulate over time.

From Figure 3, we can see that in such a continuous system, the presence of nonzero discrete-time coefficients at any particular point in time does not imply a direct causal link between the variables. If we have a system of motivation to exercise (x), daily exercise levels (y), and fitness (z), a not unreasonable causal hypothesis would look like the structure of Figure 1, where motivation leads to exercise which leads to fitness, or $x \rightarrow y \rightarrow z$. Consider then a yearly panel study and a typical discrete-time modeling approach, where x at the first time point affects y at the second, which in turn affects z at the third. Mapping the underlying causal structure to the yearly temporal coefficients in such a way implies that changes in motivation take a year to affect exercise, and then more importantly a further year to affect fitness at

Figure 3
Temporal Regression Coefficients Between x at an Earlier Point in Time and x , y , z



Note. The coefficient between x and z is positive, even though there is no causal effect. A time interval of 1 unit is marked—the same temporal matrix structure in a discrete-time approach would still show 0.00 for the influence on z at this point. See the online article for the color version of this figure.

all. So, our model has encoded in it not only the assumptions that $x \rightarrow y \rightarrow z$, but also that x has no effect on z until 2 years have passed. If we instead had observations every 6 months, the very same model structure would imply that changes in x have no effect on z for only 1 year instead of two, so the implications of what look like the same model structure are heavily dependent on the observation timing. While Figure 3 is no doubt still inaccurate (because humans are highly complex and it is a made-up example using a linear system), it gets closer to what we would expect—an initial increase in motivation (x) starts to increase daily exercise levels (y), which in turn start to increase fitness levels (z).

So, nonzero coefficients in a discrete-time temporal matrix represent the idea that at some (potentially arbitrary) point in time, one variable is predictive of another. As we see in the exercise example, however, this does not readily lead to understanding the direct causal paths in the system, as even in the absence of a direct effect, nonzero coefficients are to be expected. In contrast, modeling the continuous-time system directly allows one to turn off (i.e., fix to zero) certain causal paths of the temporal effects **A** or system noise effects **G**. This point is easy to miss because there are substantial equivalences between discrete and continuous time models when temporal effects are freely estimated, in that with free parameters, equivalent predictions are (in many cases) arrived at. The important distinction comes about when restrictions are imposed on the system—zeroes in discrete-time temporal effects at one particular sampling rate do not necessarily translate to zeroes in discrete-time effects at a different sampling rate, or to zeroes in continuous-time temporal effects. For continuously evolving systems, it is often the continuous-time effects that are of interest, as these represent how change propagates through the system, while

the discrete-time effects represent total change over a specific length of time.

Inference on a Continuous-Time System—Simulation

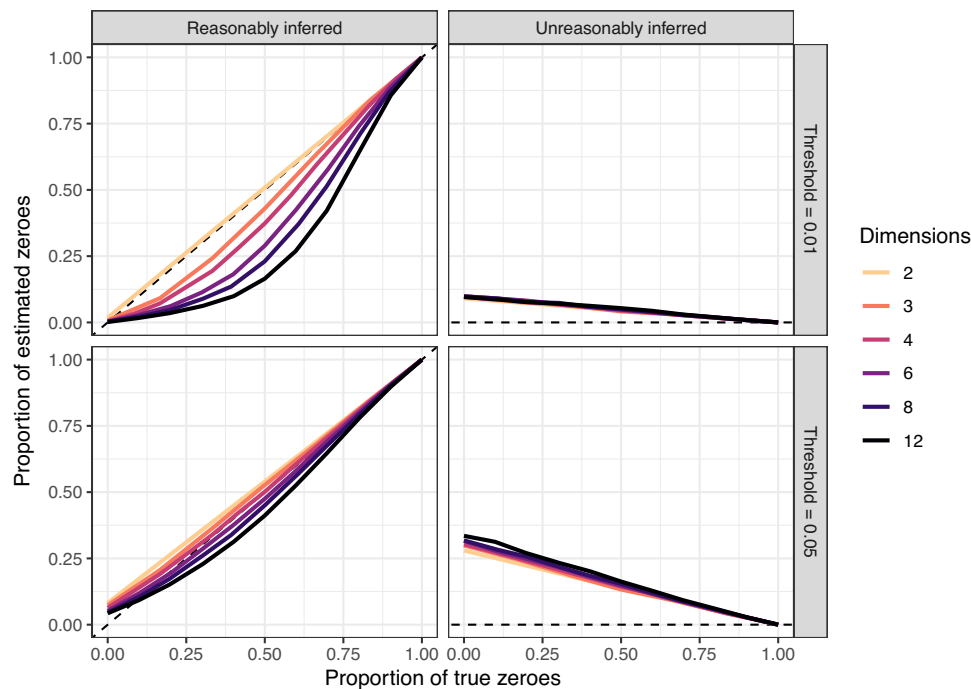
To highlight the potential for invalid inferences when using discrete-time approaches to represent continuous-time systems, I simulated inference under a variety of conditions. I based the simulation on temporal-effects matrices of continuously interacting systems with a range of dimensions and levels of sparsity, observed at time intervals of 1.00, and examined inferences likely to be made when using discrete-time approaches. A fundamental aspect this simulation will demonstrate is that more parameters are often required to represent the same underlying system when discrete-time representations are used, compared to when continuous-time systems are used. A corollary of this is that cross-effect parameters may be needed between variables that have no direct causal path in the generating system.

To setup the simulation, self-feedback effects (i.e., the diagonals) were randomly generated based on the negative exponent of a standard normal distribution, with cross effects drawn from a normal distribution with mean of 0 and standard deviation of 0.5. According to the specified level of sparsity, a number of cross-effects were then randomly selected and set to 0. The correct discrete-time temporal regression matrices were computed for each of these systems, and those with explosive tendencies (i.e., where one or more regression coefficients continue increasing over time due to strong feedback loops) were dropped from the simulation, as systems where the variance increases over time were deemed less likely to reflect common psychological scenarios. At this point, the computed discrete-time coefficients were checked for inferential consistency regarding zeroes, as zeroes represent the case of no direct effect between variables. Perfect consistency is found when the pattern of zeroes in the discrete-time matrix exactly matches the pattern of the generative continuous-time system. To reflect the fact that values near zero are often assumed to be, or are regularized to, zero, two arbitrary threshold conditions were used, a stricter 0.01 and more liberal 0.05. Coefficients that were closer to zero than these thresholds were assumed zero. All choices in this simulation are necessarily somewhat arbitrary, but the point is found in the general patterns, not the absolute values. The commented code provided in the [online supplemental materials](#) allows for changes to these decisions if desired. The left column of Figure 4 shows the proportion of reasonably inferred zeroes (i.e., the correct absence of a causal path) when using discrete-time matrices for inference. Note that for some cases the proportion reasonably inferred exceeds the target “perfect” level, shown as a dashed line, because of the threshold. The right column shows the proportion of unreasonably inferred zeroes (i.e., the incorrect absence of a causal path). For those interested, the same plot but for inference based on the continuous-time system is available in the [online supplemental materials](#)—results are, as should be expected, close to perfect besides some inflation of the reasonably estimated zeroes, due to the use of a nonzero threshold.

From Figure 4, we see that the proportion of unreasonably inferred zeroes increases in dense, highly connected systems. Moreover, as dimensions increase the proportion of zeroes in the discrete-time matrices is too low, except for systems with either mostly zeroes, or no zeroes. This means that when using discrete-time models to represent continuous-time systems (a) more parameters are often

Figure 4

Simulation Results, Showing How the Use of Discrete-Time Models for Inference About Continuously Interacting Systems Can Result in Invalid Conclusions, Across Different Dimensions, Sparsity, and Threshold Criteria



Note. The dashed line indicates what would be “perfect” inferences, though note that due to the use of a threshold (below which effects are assumed zero), “Reasonable” inferences about the existence of zero coefficients are in some cases higher than this. As the dimensionality rises, so too does the propensity for incorrect inference about a given coefficient. See the online article for the color version of this figure.

necessary to estimate the same causal system, and (b) cross-effect parameters may be needed between variables that have no direct causal path in the generating system.

Where does this difference in sparsity between continuous and discrete-time forms arise from? Is it perhaps simply an artifact of the simulation conditions? The dense(r) temporal networks from the discrete-time form arises because the continuous-time contains only the direct causal paths in the system, while the discrete-time form represents the predictiveness (or “total” effect) of each variable after some particular length of time. Such predictiveness can come about due to both the direct paths, but also any indirect paths as in mediation effects. If we consider again the motivation, exercise, and fitness example, there are direct paths from motivation to exercise, and from exercise to fitness. Once these relations are considered over some length of time, the direct paths still exist, in that earlier motivation predicts later exercise, and earlier exercise predicts later fitness. Additionally, earlier motivation predicts later fitness, because of the indirect path via exercise. Because the discrete-time form represents this expansion over some length of time, it inevitably contains some mixture of these direct and indirect effects, resulting in generally denser temporal relation matrices. It is possible to construct specific cases where this is not so, but they require a very specific balance of effects that are unlikely to be found in reality. Deboeck et al. (2018) and Ryan and Hamaker (2022) described the relationship between direct, indirect, and total effects in continuous

and discrete-time systems in much greater detail, from the perspective of path tracing in structural equation modeling.

Regularization

Having highlighted the problems of relying on discrete-time matrices as a basis for inference when systems evolve continuously, it is hopefully clear that for many confirmatory modeling and hypothesis testing endeavors, one should instead parameterize the continuous-time temporal effects matrix. From that point, regular tools of hypothesis testing and model comparison can be used as normal. Moving away from confirmatory modeling, however, it has become popular within some fields to use regularization to zero (e.g. using the Lasso of Tibshirani, 1996) as the basis for inferring causal structure of a system. Some such approaches include the graphical VAR with regularization (Epskamp et al., 2018; Yuen et al., 2018), regularized multivariate latent change scores (Jacobucci & Grimm, 2018b), and group iterative multiple model estimation (Gates & Molenaar, 2012). J. J. Park et al. (2020) contrasts some of the different approaches mentioned for the consideration of affect and personality dynamics, while some other example applications include relations between well-being and sleep (Dickens, 2020), and anxiety and borderline personality (Ellison et al., 2020). Historically, such approaches have tended to neglect the possibility of measurement error (potentially quite problematic, discussed below), though this is changing of late

(Bulteel et al., 2018; Epskamp, 2020; Gates et al., 2020; Liang et al., 2023).

The approaches typically used for data-driven discovery have primarily used discrete-time. As is hopefully clear by now, this means that regularization to zero is being applied to the temporal regression coefficients for a specific time interval, and that these do not necessarily correspond to zeroes in an underlying continuous-time process. This means that while sparsity may be achieved for the time interval in question, this does not imply sparsity at any other time interval, and the recovered sparse structure does not necessarily represent the underlying generative structure. In addition, given that discrete-time temporal matrices are likely to have more nonzero coefficients than continuous-time, resultant structures are less likely to achieve sparsity goals for comparable levels of regularization. A natural step at this point, assuming the generative structure is of interest, is to regularize the underlying continuous-time temporal-effects matrix instead. The ctsem software (Driver & Voelkle, 2018) offers the ability to specify Bayesian priors on the system parameters, which, when estimated using the default approach of optimization, are equivalent to typical penalized likelihood approaches used for regularization (for more elaborated discussion on this see Jacobucci & Grimm, 2018a; T. Park & Casella, 2008). An alternative approach by Orzek and Voelkle (2023), regCtsem, builds on the original ctsem software based on OpenMx (Neale et al., 2016), now available as ctsemOMX. This makes it more limited in terms of model flexibility, but it includes more emphasis on regularization, specifically, if and how standardization of variables or parameters is performed.

The default prior in ctsem (when priors are enabled) is a Gaussian distribution, which can be interpreted as L2, or ridge, and regression. L2 penalization does not lead to sparse matrices, because coefficients will not, in general, be pushed all the way to zero. In addition to this, large coefficients are heavily penalized. A common fix for these issues is to replace the Gaussian prior with a Laplace distribution (interpretable as L1 penalization or Lasso regression), ensuring that small coefficients will be pushed to zero, and large coefficients are not affected as heavily. As discussed by Orzek et al. (2023), a smoothed form of L1 penalization is implemented in ctsem removing typical optimization difficulties with standard L1 penalization.

A comparison between L1 regularization on a discrete versus continuous time representation is shown in Figure 5. The underlying system of four processes is a causal chain of the sort already discussed, where the first process causes the second, and so on. Random noise affects the processes independently with the same magnitude. These processes do not evolve in discrete-jumps, but change and interact continuously. The continuous (A) and corresponding discrete (e^A) time temporal relations matrices looks like:

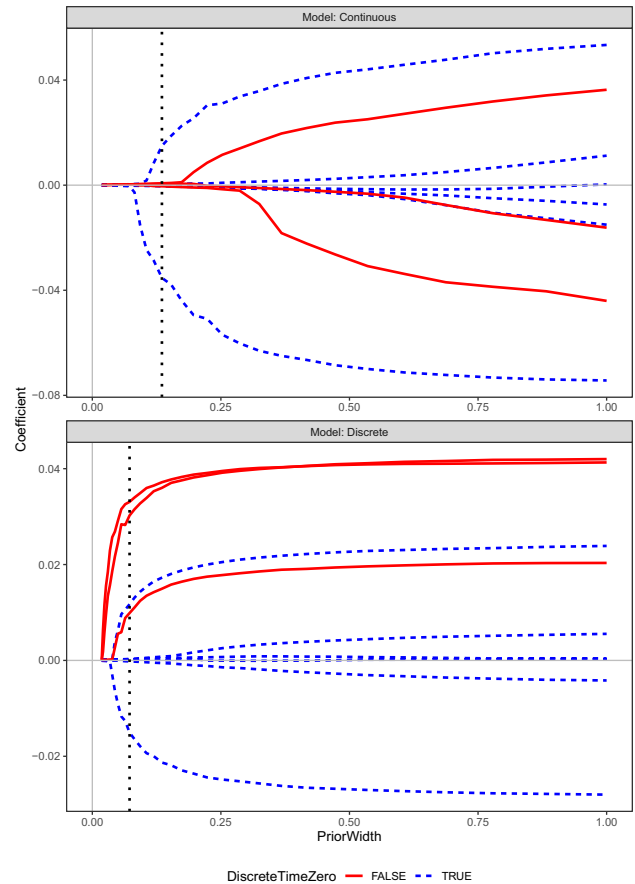
$$A = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0.5 & -1 & 0 & 0 \\ 0 & .5 & -1 & 0 \\ 0 & 0 & .5 & -1 \end{bmatrix} \quad (7)$$

$$e^A = \begin{bmatrix} 0.37 & 0 & 0 & 0 \\ 0.18 & 0.37 & 0 & 0 \\ 0.05 & .18 & 0.37 & 0 \\ 0.01 & 0.05 & .18 & 0.37 \end{bmatrix}$$

Data were generated for 10 subjects with 100 observation occasions each, sampled at time intervals of 1.0. Figure 5 shows how

Figure 5

Estimated Temporal Coefficients in a Four-Dimensional System, Using L1 (Lasso/Laplace Prior) Penalization With Different Penalties (Prior Widths)



Note. Both continuous and discrete-time models were estimated. All coefficients shown were zero in the continuous-time generating model, those that were nonzero in the discrete-time representation with a time step of 1.0 are shown in solid red. Best out-of-sample performance for each model is indicated by the vertical black dotted line. Note the (expected) presence of more nonzero coefficients in the discrete-time estimation. See the online article for the color version of this figure.

estimated temporal coefficients of the system change as the penalty (i.e., width of the prior) changes, for both a discrete and continuous-time estimation approach. Only the cross-effects that are zero in the generating model are shown (for simplicity). The penalty at which the best out-of-sample (subject-wise) predictive performance of the model was achieved is shown by the black dotted line. Note that because the scale of the discrete and continuous coefficients are not necessarily the same, the ideal penalty for out-of-sample prediction for each model will also not necessarily be the same. From the figure, we see the expected shrinkage of coefficients toward zero. At the point of best out-of-sample performance, the discrete-time model has more nonzero coefficients. This result is unsurprising, and should be expected based on the true discrete-time matrix shown above—nevertheless, this hopefully gives a more tangible sense of the problem to researchers in the field considering these approaches.

Nondirected Relations

Throughout the article so far I have primarily focused on directed relationships, but in many circumstances researchers may instead be interested in covariance, correlation, or partial-correlation parameters. The same general pattern of results already described holds, wherein zeroes in an underlying continuous system do not necessarily translate to zeroes in a discrete-time estimate. I will not treat this in depth, but for one example consider a two-dimensional system with direct positive feedback effects between processes, and uncorrelated system noise:

$$A = \begin{bmatrix} -1 & 0.4 \\ 0.4 & -1 \end{bmatrix}, \quad G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (8)$$

The discrete-time VAR for a time interval of 1.0 in this case has a system-noise correlation of 0.27, even though the generating system noise is uncorrelated. Such issues are part of the motivation behind “structural” VAR approaches that include both directed and nondirected relations for the system noise in discrete-time—see [Chow et al. \(2023\)](#) for the link between continuous-time and structural VAR approaches.

“True” Time Lag and the Importance of Measurement Error

A common question from researchers designing a study is at what speed the system is changing, what is the “true” time lag of the system. As highlighted by [Shojaie and Fox \(2022\)](#), this is a critical consideration if using a standard discrete-time VAR type model to represent a discrete-time system—the time interval of the observation procedure needs to match the frequency at which the system changes (i.e., the true time lag). This has led to works that focus on specifying the most appropriate time interval (e.g., [Dormann & Griffin, 2015](#); [Hopwood et al., 2022](#)), as well as those that aim to specify the correct lag structure (e.g., [Jacobs & Otter, 2008](#); [Jacobson et al., 2019](#)). This need to observe the system every time it changes may seem quite problematic for a system that changes continually, as we can generally not observe phenomena continuously, but take snapshots of the system state at specific moments in time. Fortunately, although the sampling rate needs to be timed to ensure there is information about the changes of interest, for estimating continuous-time systems there is no strong dependency on getting the sampling rate exactly right—so long as measurement error is accounted for.

The importance of accounting for measurement error in discrete-time VAR-type models is well documented ([Schuurman & Hamaker, 2019](#); [Schuurman et al., 2015](#)), as ignoring measurement error can lead to downwards biased auto-regression coefficients and potentially spurious cross-regression relationships. This all holds true for the continuous-time case also. A related issue that does not seem widely understood is that when measurement error is ignored, the apparent time scale of the underlying processes (inferred from estimates) can depend strongly on the frequency at which observations are obtained. This dependence can make the selection of sampling frequency (e.g., weekly vs. monthly) seem a crucial design decision. Fortunately, once measurement error is accounted for the choice of sampling rate is less crucial, and becomes instead a matter of maximizing information gain per observation, rather than the observation design being inherently wrong and inevitably leading to the wrong conclusions. For more

detail on the information gain aspect and various tradeoffs regarding target parameters and the relation to sampling rate, see [Adolf et al. \(2021\)](#) and [Timmons and Preacher \(2015\)](#).

A common approach to handling measurement error is to include a Gaussian measurement model of the processes. In most basic form this might look like:

$$\mathbf{x}_u = \mathbf{y}_u + \boldsymbol{\epsilon}_u, \quad \text{where } \boldsymbol{\epsilon}(t) \sim N(0, \boldsymbol{\Theta}), \quad (9)$$

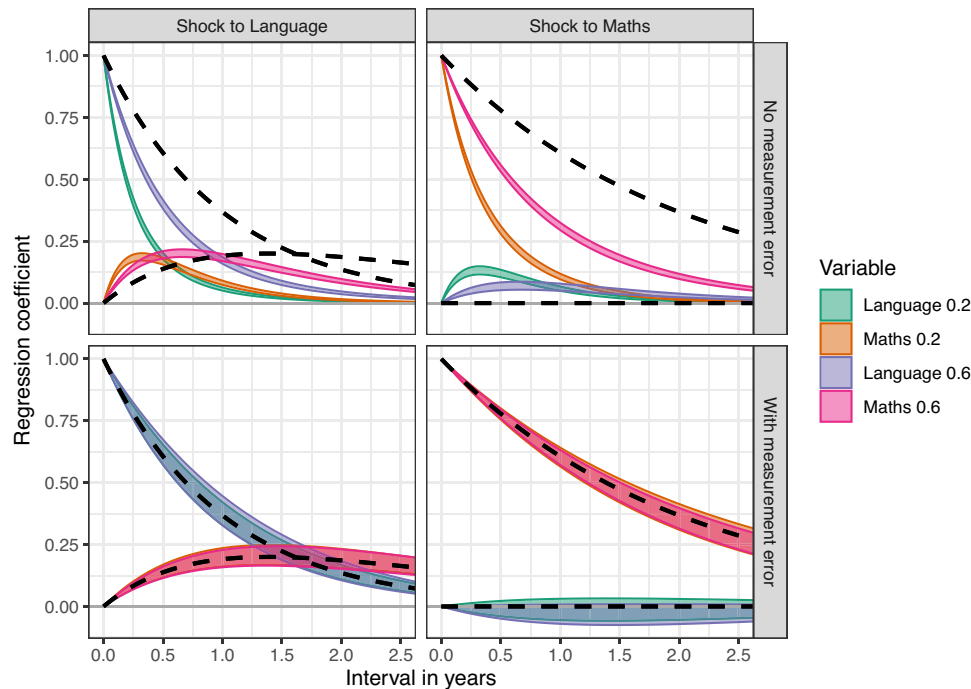
where $\mathbf{x}_u$ now represents the observations at measurement u and $\mathbf{y}_u$ the latent processes, measured with error $\boldsymbol{\epsilon}_u$. The errors are random with variance(s) given by $\boldsymbol{\Theta}$. This can be handled using structural equation or state-space modeling approaches, as used in the already mentioned software options. This model is often extended to a linear factor model with a factor loading matrix and intercepts (e.g., [Driver & Voelkle, 2018](#)).

To demonstrate the problematic phenomena discussed regarding measurement error and sampling rate, I simulated data for 500 subjects with 30 measurement occasions each, based on a hypothetical (first order) model where language abilities facilitate mathematics ability development, but not the reverse, and both experience correlated random shocks due to external sources (e.g., motivation or brain development). Latent variables are observed via a single indicator each, with Gaussian measurement errors. The measurement error variance is approximately half that of the latent constructs. All subjects share the same model parameters, and were observed either every 0.2 or 0.6 years. These sampling rates were crossed with model specification. For one model, measurement error variance was assumed nonexistent, while for the other it was freely estimated, and otherwise the models are correctly specified. [Figure 6](#) shows the results of this, with data fit using the ctsem software ([Driver & Voelkle, 2018](#)). These plots show the core points mentioned, with the top row (wrongly assuming no measurement error) showing: spurious (not genuinely causal) cross effects from maths ability to language (green and purple curves in top right); precise but incorrect (too fast) estimates of temporal relations generally; and differential results that depend on the sampling interval—while the orange/pink and green/purple curves represent the same underlying phenomenon, the curves from the more frequent observation regime show faster effects. The lower half of the figure, where measurement error is appropriately included in the estimation model, shows approximately correct estimates (the true values are included in the confidence regions shown), and the sampling frequency does not impact the time scale of estimated processes. [Appendix A](#) shows very similar results but for a second-order case, wherein both language and maths processes are damped linear oscillators. Simulation scripts are in the [online supplemental materials](#).

Two caveats are important to emphasize here. While accounting for measurement error can avoid the dependency of estimates on the sampling rate, as already mentioned the information gained from each measurement for various parameter estimates is a complex function depending in part on the amount of true change and measurement error in the observation. One will struggle to estimate slow yearly dynamics with microsecond-based estimates over a minute, just as one will not capture microsecond dynamics with yearly sampling. The second caveat is that although we do not need to know the precise “true lag” of the system, “lag” and “order” are terms with important differences that are sometimes confused. The lag of a system refers to how many steps in time before a

Figure 6

Discrete-Time Auto and Cross-Regression 95% Confidence Regions Estimated Based on Either a Correctly Specified First-Order Continuous-Time Model (Bottom Row) or the Same Model but Without Accounting for Measurement Error (Top Row)



Note. The variable types are suffixed with the sampling time intervals of either 0.2 or 0.6, but given no difference in the underlying process both the maths coefficients (orange/pink), and language coefficients (green/purple), regions should overlap the true coefficients (in black), as seen in the correctly specified lower row. The top row where measurement error is not accounted for shows spurious cross-effects, inaccurate estimates, and changes in estimated time-scale (i.e., slower/faster) depending on sampling rate. These problems do not occur once measurement error is accounted for, as in the lower row. See the online article for the color version of this figure.

change in one variable predicts a change in another, while the order of a system refers to the number of interrelated processes needed to represent the dynamics of one observed variable. In many discrete-time estimation approaches lag and order can be treated interchangeably, but this is not the case when it comes to continuous-time. Estimating a model where the order of the system is too low is a problem—the model is not sufficiently complex to capture the dynamics and may lead to invalid inferences and poor predictions. Substantial work has occurred in this domain, I will point to only a few examples, and demonstrate a general approach to model checking that can highlight misspecifications.

Checking Model Order and Structure

While the majority of psychology-oriented research with VAR-type models relies on first-order linear approaches, more complex models have been explored from time to time, particularly oscillatory models (for some examples see Boker et al., 2016; Chow et al., 2005; Hu et al., 2014; Oud et al., 2018). Non, or semi, parametric approaches are often a useful tool to check and or explore system structure, as they are less dependent on specific assumptions, and Chow (2019) discusses the use of such approaches for model development. Parametric approaches are also possible, Belcher et al. (1994) provided an alternative

parameterization of continuous-time models that allows for simplified model-order assessment, and Ha and Welsh (2015) demonstrated model-order selection using regularization approaches. Whichever route is chosen, there is substantial complexity to accurate model development. To convey some understanding of the issue in regards to continuous-time, I will demonstrate the basics of both residual checks, and comparisons between the observed data and new data generated based on the fitted model. For these purposes, I have simulated multiple subjects data from a second-order linear continuous-time system, that also exhibits a trend over time. Time-intervals between observations were generated with substantial randomness. For simplicity, no stable individual differences were simulated. Precise details about the simulation are not needed to understand the demonstration, but can be found in the [online supplemental materials](#).

As a first step in this example workflow, I fitted a relatively typical first-order continuous-time model to the simulated data, using the ctsem software. The model accounted for both measurement error and random changes, but at this stage was misspecified due to both the unaccounted for trend, and because the dynamic system was unable to account for the second-order dynamics (i.e., nonrandom oscillations). The resulting first-order process was estimated with auto-effect parameters very close to the default upper limit of zero. While it would have been possible to relax the boundary, in

general, a nonnegative auto-effect (or one very close to a parameter boundary) can be a good initial clue that the process is misspecified—unless there are reasons to suspect genuine random-walk behavior of the underlying process, where variance continually increases. Note that a negative auto-effect provides no guarantee the process is well specified, it is simply no evidence of a problem.

Residual based checks are a classic approach to detecting misspecification, with many forms of check possible (for more discussion, see Conn et al., 2018; Lewis-Beck & Lewis-Beck, 2015). The general idea is simply that if the model is well specified, the residuals should not contain information that could have been used by the model to improve predictions. Put differently, residuals should be independent and identically distributed with a mean of zero. A fundamental residual check for time-series modeling is whether past residuals can be used to predict future residuals, and one way this is examined is via the autocorrelation function (ACF). The ACF of the residuals should be approximately zero for all time-intervals or lags (depending on whether a continuous or discrete-time ACF is used). There may be cases where it is sensible to check in both a continuous and discrete-time sense, as for instance, if we assume the underlying process operates in continuous-time, but worry that there may be an effect of the measurement occasion itself—an example of such could be a course of psychotherapy, in which the patients anxiety is in continual flux, but the occasions at which they are observed (therapy sessions) exert a unique influence.

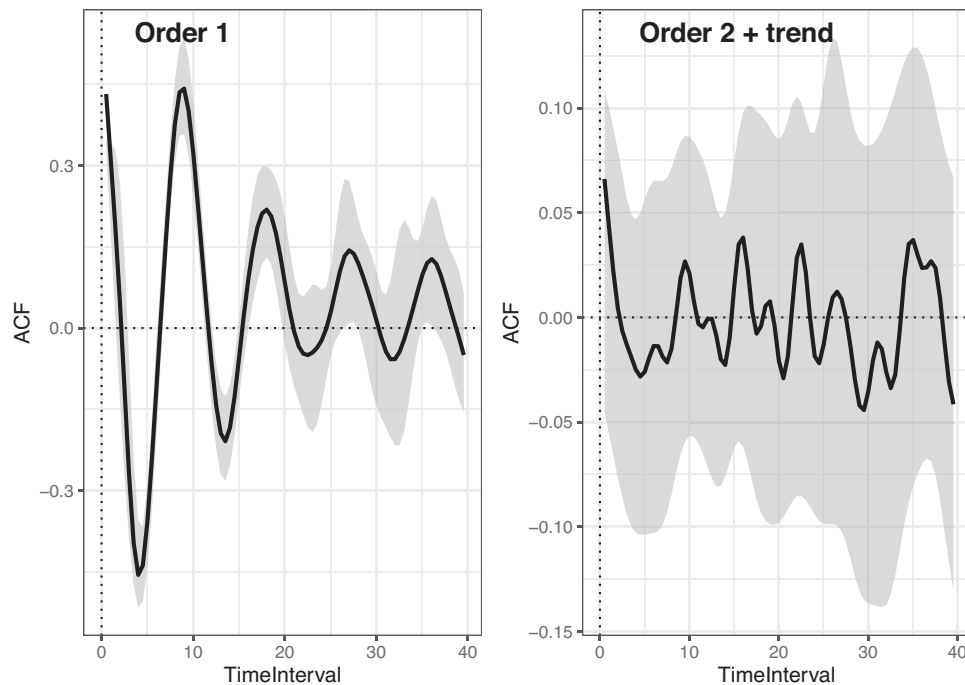
The left of Figure 7 shows an approximated continuous-time ACF with bootstrapped confidence intervals, for the first-order model fit. This was computed (within ctsem) by discretizing the residual

process to a small time-step (i.e., inserting missing data and rounding where necessary to achieve equal time intervals between observations), then computing the discrete-time ACF as normal (Shumway & Stoffer, 2017) and aggregating over subjects. The expct package for R (Ryan, 2023) offers an alternative approach to this. The figure shows a clear oscillation pattern in the residuals, pointing us to the need for a higher-order model.

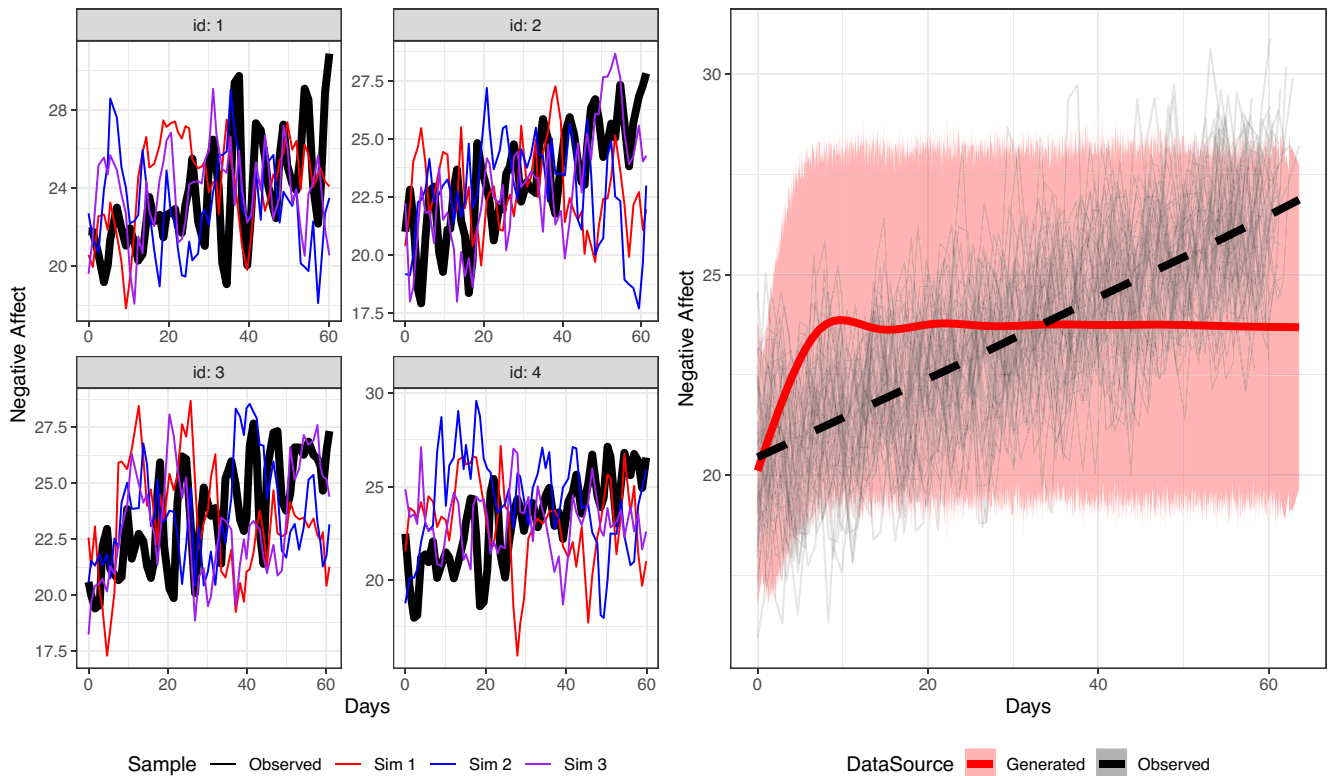
However, if we extend the first-order model to a second-order model, we obtain a very similar pattern of results. This is because the trend is still not accounted for, and as such the temporal dynamics parameters that result in best fit, are those that also best accommodate the trend in the data. In this case, the result was again autoregression strengths of ~ 1.00 , independent of the time interval—a random-walk. So, although we have seen evidence of higher-order dynamics and then accounted for them in our model, we need further model improvements before the higher-order dynamics will actually capture the observed oscillations.

Another approach to examine misspecification is to compare multiple realizations of data generated from the fitted values, to the original data. Such approaches are common with Bayesian modeling and known as posterior-predictive checks, but are just as applicable in non-Bayesian contexts also Gelman and Shalizi (2013). If the fitted model represents the underlying process that generated the original data, we would expect to observe similar patterns in data generated based on the fitted model, as we find in the original data. Such comparisons can be done both at the level of individual subjects, as seen in the left side of Figure 8, as well as at the level of summary statistics, as seen on the right side. Considering the

Figure 7
Continuous-Time ACF Plots of Residuals for Order 1, and Order 2 With Trend, Models



Note. Showing median with 95% quantiles shaded. Note the clear predictability of residuals in the order 1 model, indicating unused information that could improve the model. ACF = autocorrelation function.

Figure 8*Comparing Data Generated Based on the Fitted Second-Order Model to the Empirical Data*

Note. The left shows exemplar individuals (id) observed data to random samples from the fitted model (Sim 1, 2, 3). The right plot shows summary statistics over time, with empirical data and overall trend in black, and 95% quantiles and means of generated data in red. Note the discrepancies between generated and empirical data, on both summary statistics and individual-level plots. See the online article for the color version of this figure.

subject-level plots, if we only considered them individually, it may not be obvious there is any misspecification in the model, though when looking at many such plots, we would probably notice that the actual data, in black, tend to follow a similar upward trend, while samples from the fitted model (in red, green, and blue) show far more variation. This additional variation is most noticeable, however, when we look at the summary statistics plot on the right, where we see that while the actual data in black is relatively tightly clustered around an upward trajectory, the 95% quantiles of the generated data are far outside any of the original data points, and the overall trend is missing. When we look to Figure 9, examining a model with higher-order dynamics as well as a trend, both the subject-level plots and the summary statistics over time look sensible. For further details on specifying higher-order dynamics in continuous-time see Oud et al. (2018), and for the inclusion of trends see Driver and Tomasik (2023).

With this example, I have demonstrated some basic model-checking concepts, though with real-world empirical data more checks are necessary, because the underlying data-generating processes are generally only very roughly understood. Such checks could include examinations of individual differences or more thorough trend modeling (Ryan et al., 2023). We will now turn to a multivariate empirical example, where individual differences, nonlinear measurement models, and more complex model forms are also considered.

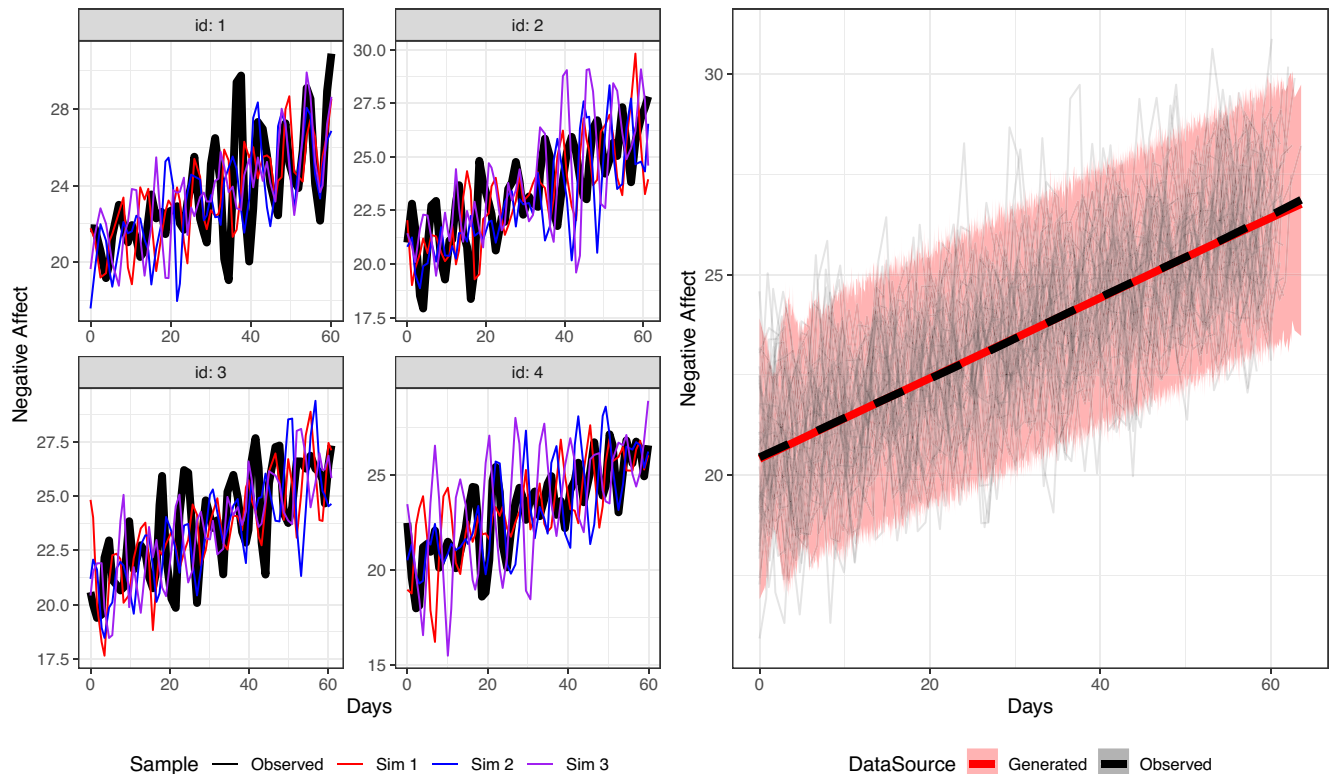
Empirical Example

To demonstrate the use of continuous-time approaches, I re-analyzed a subset of variables from an ecological momentary assessment study conducted in the Netherlands during the initial stages of the COVID-19 pandemic (for full details, see Fried et al., 2022). The study monitored 80 undergraduate students, who answered a questionnaire 4 times per day over a 14-day period. The observations occurred at ~ 12 p.m., 3 p.m., 6 p.m., and 9 p.m. each day, so there is a longer time-interval between the last of the day and the first of the next day. The data collection focused on assessing mental health, social contact, and various COVID-19-related variables, but for my purposes, I am looking at the variables “anhedonia,” “angry,” “hungry,” “tired,” and their relationships over time. At each measurement occasion, participants were asked how much, over the previous 3 hr, they endorsed a certain feeling or behavior (1 = *not at all*, 2 = *slightly*, 3 = *moderately*, 4 = *very*, 5 = *extremely*).

As a starting point, I fit a first-order continuous-time model containing a process for each of the observed variables, relying largely on the ctsem defaults which include measurement error for each variable type, correlated system noise, and temporal dynamics parameters. Individual differences in average level of each process are accounted for by random-effects on the continuous intercepts (i.e., the value that is constantly added to the process to ensure that it does not revert to zero). Expanded details,

Figure 9

Comparing Data Generated Based on the Fitted Second-Order Plus Trend Model to the Empirical Data



Note. The left shows exemplar individuals (id) observed data to random samples from the fitted model (Sim 1, 2, 3). The right plot shows summary statistics over time, with empirical data and overall trend in black, and 95% quantiles and means of generated data shown in red. In this case, the fitted model accurately matches the underlying generative model for the data, and data generated based on the fit both qualitatively look similar for individual cases (left), and exhibit similar summary characteristics over time (right). See the online article for the color version of this figure.

including the full system matrix representation of the model and fit, are available in the [online supplemental materials](#). While not strictly necessary at this stage, because more complex models will also be tested I used the ctsem default priors, providing some moderate level of regularization on the system dynamics. Full details of the priors are available by [Driver and Voelke \(2018, 2021\)](#), but they are not of substantial importance to the points made here, optimization was still used (in contrast to the far slower Bayesian sampling) and interpretation is similar to typical maximum likelihood fits.

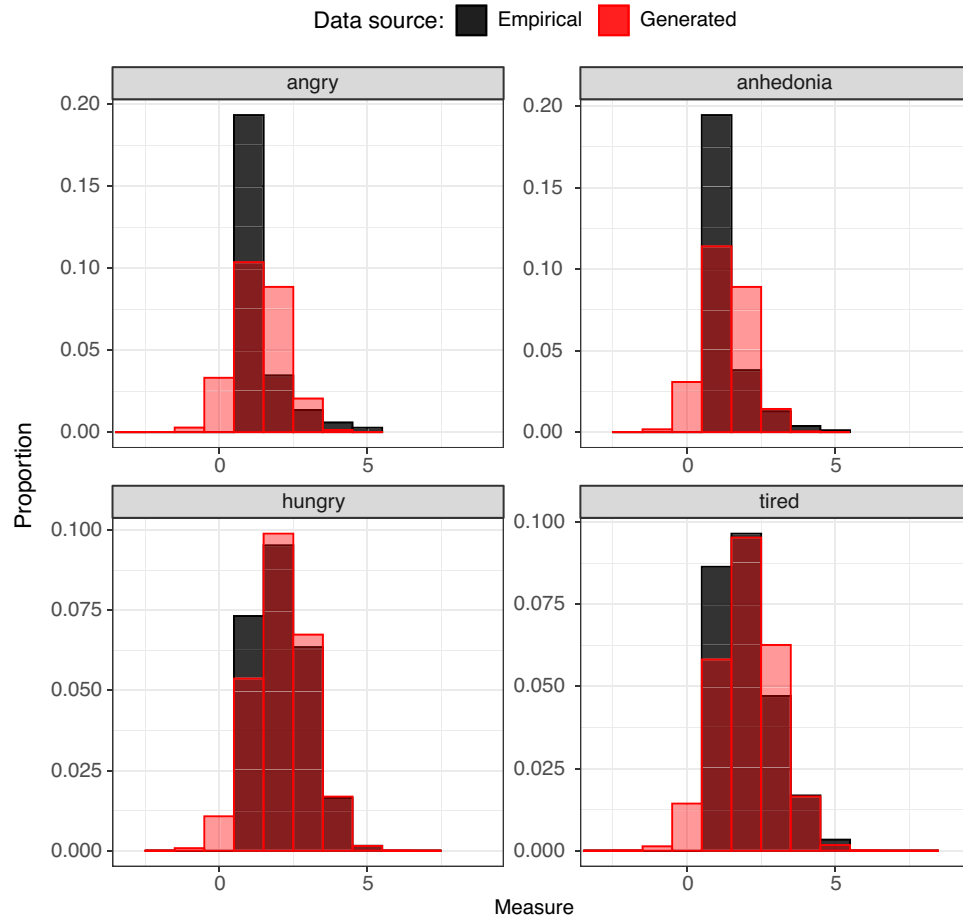
When comparing data generated based on the fitted model to the actual data, as shown in [Figure 10](#), substantial differences in the overall distribution of the variables were visible. The empirical data were more compressed toward the lower values (i.e., more positively skewed), and limited to a minimum of 1 (“not at all”). Such nonnormal distributions are quite common, and a wide variety of approaches, ranging from ignoring to elaborate item response theory models, are possible. For this example, as in the initial work from [Fried et al. \(2022\)](#), I will treat the data as approximately continuous, though will include two approaches to account for at least some of the potential nonlinearity—a nonlinear link between the latent process and observations, and individual differences in system dynamics and measurement properties as a linear function of the individuals’ mean on each of the observed variables.

The nonlinear link between the latent process and observations essentially stretches the underlying latent process (with Gaussian system noise) so that the resulting observation predictions are positively skewed. The more a process is “stretched,” the faster it will tend to change when it takes on higher values, and vice versa, thereby resulting in more time spent in the regions that correspond to lower values on observations. This formulation does not require the estimation of many additional parameters, and does not alter the interpretation of estimated dynamics substantially.

While the stretch factor may account for some of the nonlinearity, it may also be the case that individuals who tend to score highly on a certain variable also show different dynamics and measurement properties—if, for example, an individual high in anhedonia is also more likely to show greater fluctuations in anhedonia, this would also lead to a positively skewed distribution. The simplest approach in such cases is often to estimate the system and measurement properties separately for each individual, however, unless large amounts of data are available this is likely to result in substantial overfitting. Random-effects or hierarchical models are in some sense ideal, but can be computationally very costly once there are many parameters ([Driver & Voelke, 2018](#)). Here, I opt for a middle-ground approach that treats intercept parameters as random, and allows all other model parameters to be moderated by the subject-level mean of each observed variable. A conservative Laplace

Figure 10

Distributions of Variables From the Empirical Data (Black) and Data Generated Based on the Initial First-Order Model Fit



Note. See the online article for the color version of this figure.

prior (interpretable as an L1/Lasso penalty as described earlier) is used for the moderation effects, to reduce overfitting tendencies.

The overall distributions of variables can point to possible issues as shown, however, a more nuanced approach to comparing the data simulated from the fitted model to the empirical data is to consider the auto and cross correlations of both, across a range of time intervals. When analyzing these plots, two interesting discrepancies were visible. The first is visible when examining the correlations between all variables and later tiredness, as shown in Figure 11. Here, the autocorrelation in the top-left corner shows that the model has captured the primary features reasonably, in that the starting point of the correlations and the general pattern of decline over time to some stable level (often thought of as the between-subjects correlation) match reasonably. However, the autocorrelation plot for the empirical data shows an oscillating pattern, which may not be very surprising for a variable like tiredness. To attempt to capture such an oscillation, the model can be expanded to a second-order model as described earlier.

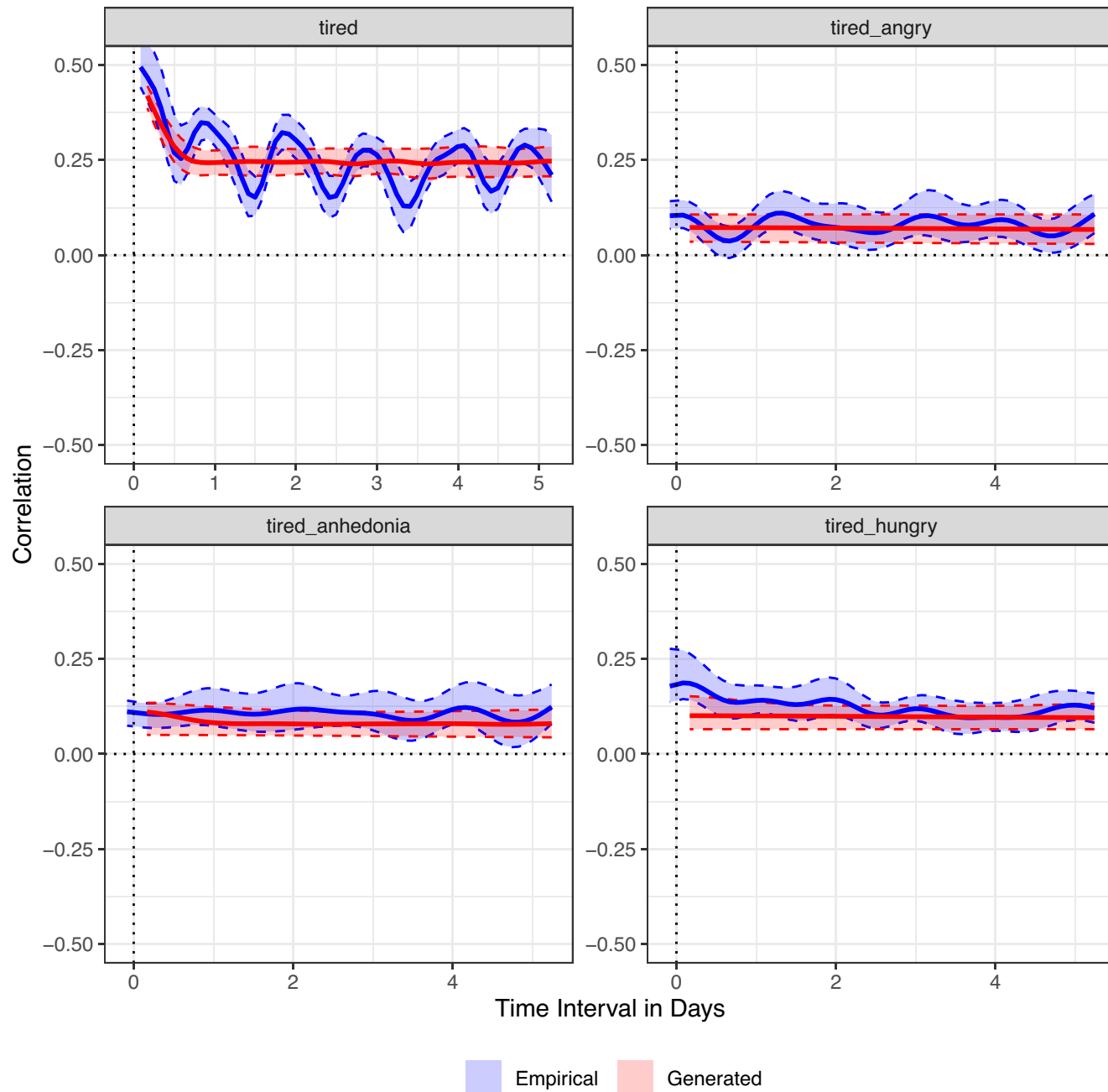
The other interesting feature that becomes apparent from the autocorrelation and cross-correlation plots is subject-level heterogeneity,

as already discussed. When computing the auto and cross-correlation plots conditional on a median split for each variable, differences in some of the patterns as a function of whether a subject is typically high or low on the variable become apparent. The most obvious such difference is shown in Figure 12, which shows the autocorrelation and cross-correlation between all variables and later anger, split at the median of the subject-level mean for anger. Here we see that individuals in the empirical data who are typically high anger show a higher autocorrelation, while those who are typically low show relatively little autocorrelation. The fitted model had no scope to account for this, so although the simulated data were also split at the anger median, both high and low anger simulated individuals show an autocorrelation somewhere in between the empirical plots.

At this point, there are a number of model extensions to incorporate and test. To facilitate this and also to test a broader range of models, I fit all combinations of the noted extension possibilities (nonlinear, individual differences, and higher order), generated summary and comparison plots (see the [online supplemental materials](#)), and computed out-of-sample likelihoods based on repeated 10-fold cross-validation. I also included discrete-time variants of each model, as there are

Figure 11

Autocorrelation and Cross-Correlation Effects With Later Tiredness, on Empirical and Simulated Data Sets



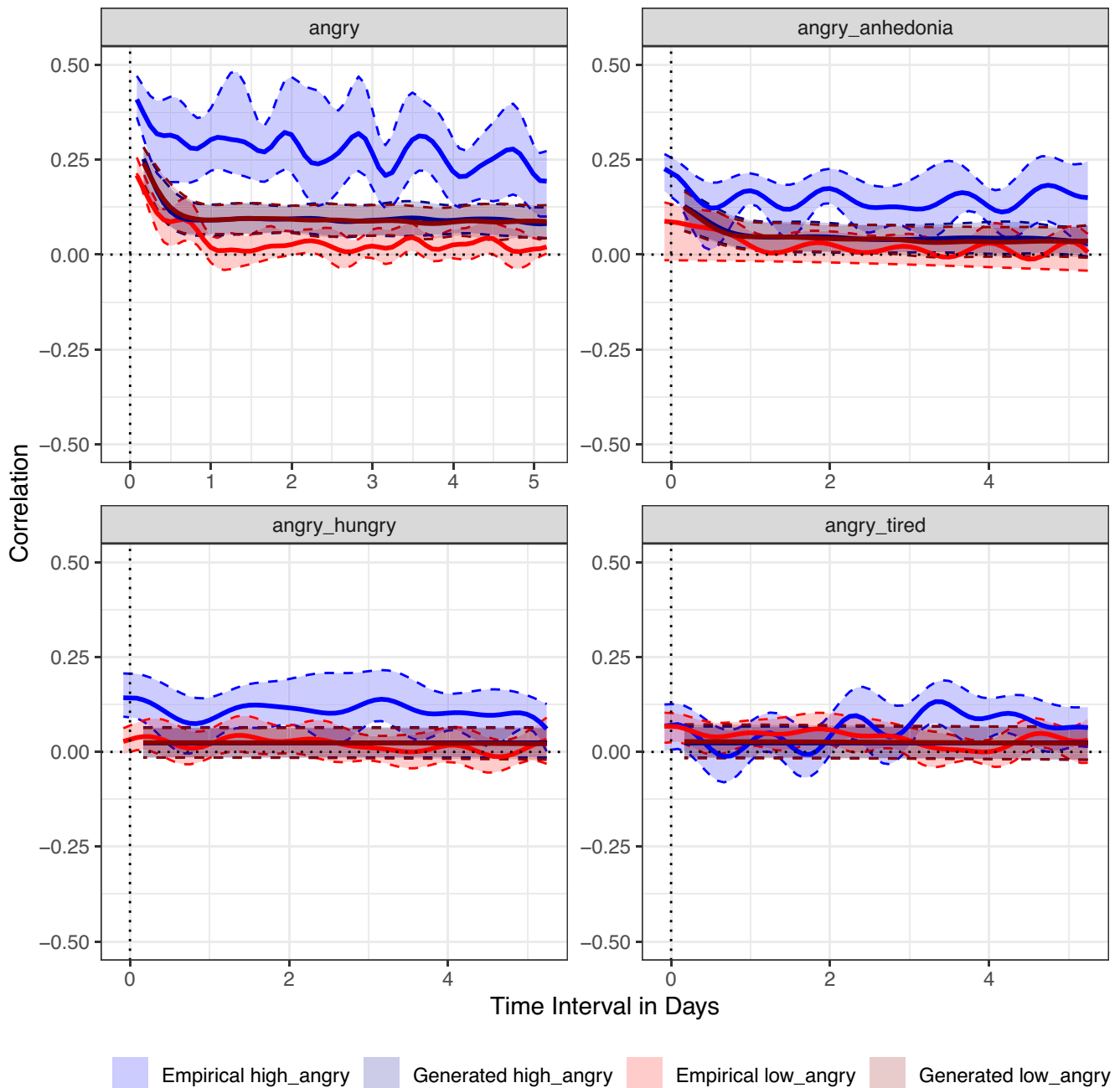
Note. Median and 50% probability intervals are shown. See the online article for the color version of this figure.

some circumstances where discrete-time models, because they are not restricted to being interpretable at all time intervals, can fit data better for the same order of model (e.g., a first-order discrete-time model can sometimes approximately fit oscillatory data). A final model possibility included was a variant of each model without cross-effects, as the presence or absence of these are often of substantial interest. This resulted in 32 models to fit and compare. The out-of-sample likelihood comparison in [Appendix B](#) shows that: both the nonlinear measurement model (NL) and individual-level heterogeneity (hetero) were valuable

extensions; continuous-time (CT) forms generally outperformed discrete-time (DT); removing cross-effects (NoCross) worsened model performance; and second-order models (O2) performed worse than first order. Note that during initial testing, models with specific time of day effects as well as general trends were considered, but these did not improve performance and have been left out for simplicity. Also note that these comparisons are not highly specific, in that they only consider whether cross-effects for all processes versus no processes are helpful—if a specific individual

Figure 12

Autocorrelation and Cross-Correlations With Later Anger, on Empirical and Simulated Data Sets, Split on the Median of Individuals Mean Anger



Note. Median and 50% probability intervals are shown. See the online article for the color version of this figure.

effect (e.g., the cross-effect from anger to tiredness) is of substantial interest then comparisons focusing only on that difference would be important.

The model comparisons are of course conditional on the available data—given substantially more data (e.g., sampled more often, for longer, with more subjects) the results would likely favor higher-order possibilities, and the various model checks would likely show further details to consider in the modeling. Given this, and

the generally high level of uncertainty around the modeling of psychological phenomena, ensuring that key inferences are consistent across an array of plausible specifications is often wise. For this example, I will provide some output from the best fitting model, but also consider patterns of effects found in the second best model, which is the same but also includes second-order dynamics rather than only first order. More extensive output can be found in the [online supplemental materials](#), Appendix C contains

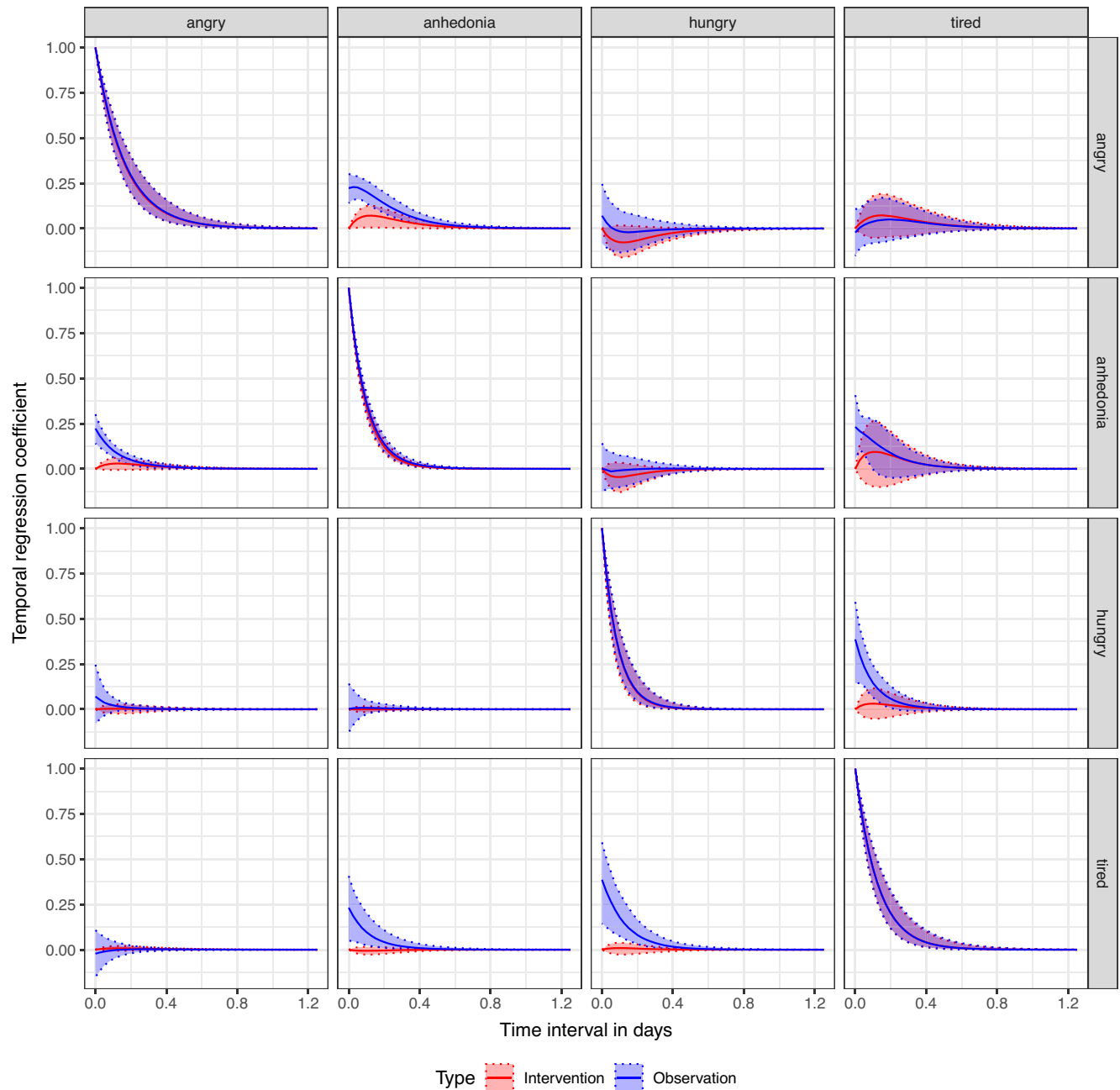
the expanded system matrices, and [Appendix D](#) contains the individual difference distributions for the best-fitting model.

To understand the estimated dynamics, while parameter estimates are important and can be found in [Table B1](#) in [Appendix B](#), plots of the model implied temporal regressions over time are often more intuitive. [Figure 13](#) shows the estimated dynamics as they are

typically considered, where the temporal regressions (based on the drift matrix A) are the prime feature of interest. These, shown in red, represent the discrete-time regressions between one of the processes at an earlier time, and one of the processes at a later time. The only significant cross-effect (given typical 95% confidence intervals) is the influence of anhedonia on anger (found in the top

Figure 13

Temporal Regression Estimates and 95% Quantiles for the Best-Fitting Model, a First-Order Continuous-Time Model With Cross Effects, Allowing for Nonlinear Measurement Properties, and Individual Differences in Dynamics



Note. The column represents the “causal” or predicting variable, and the row the affected or predicted variable. Conditional expectations based on two different hypothetical scenarios are shown. In the intervention type, the plot shows the effect of a change in the causal variable only, while the observational type shows the expected effect conditional on observing a change in the causal variable, and accounting for correlated random changes in the system. See the online article for the color version of this figure.

row, second column), given by the parameter “drift_angry_anhedonia.” For this, we see a positive effect, such that if anhedonia were to increase we would expect anger to also increase over the next hours. Overlaid in blue we see a similar plot, but one in which the correlated system-noise is also considered. This can be important because in many cases the system-noise correlation may contribute substantially to the overall dynamics. Such a plot shows how the system is expected to change, conditional on observing a change in the predictor (column) variable. As such, when there is correlation in the system-noise between the two variables, the blue cross-effect plot will start above or below zero, according to the extent of correlation. This is in contrast to the more common approach (in red), in which we see the expected change conditional on a hypothetical intervention on the predictor variable only, so the cross-effect plot will always start at zero. More description of this distinction is available by Driver (2023). Looking at the same anhedonia-to-anger relationship, we see that indeed there is a positive correlation in the random changes between the two, such that when anhedonia increases due to sources which are not explicitly included in the model, anger is also expected to increase, and to a greater degree than given based only on the cross-effect. Interpretation of the significant cross-effect as any indicator of causality under such circumstances may be overly ambitious, given this evidence for a time-varying common-cause between the two. Considering the results from the model with second-order dynamics as a sensitivity check here also doesn’t clarify matters—the cross-effect parameter is still significantly different from zero, the positive correlation in system-noise is still there, but the expected change in anger after observing a positive change in anhedonia actually becomes negative quite quickly (before leveling off). The second-order plots and estimates are in Appendix B.

Considering the system-noise alone (given by the prefix “diff,” for diffusion, in Table B1 in Appendix B), unpredictable fluctuations were significantly positively correlated between anhedonia and tiredness, anhedonia and anger, and hunger and tiredness. While the general pattern remained the same, in the second-order model only the correlation between anhedonia and anger remained significant.

Regarding individual differences, significant moderation effects are shown in Table B3 in Appendix B, with full results for all models found in the online supplemental materials. For all variables except anhedonia the temporal auto-effect was moderated according to the subject-level mean for that same variable, though not always in the same direction. For individuals with high average tiredness and hunger, these processes had a less negative auto-effect, suggesting a slower return to baseline, while individuals with high average anger had more negative coefficient and thus faster return to baseline. Individuals high in anhedonia tended to also have more variability in anhedonia. The most interesting feature in the individual differences however is that cross-effects from hunger to each of the other variables is also moderated based on the variable in question—the effect of hunger on anhedonia or anger rises when anhedonia or anger (respectively) tends to be higher, while the effect of hunger on tiredness lowers when overall tiredness is higher. This pattern is largely maintained when considering the second-order model. So, although there was no substantial evidence found for an overall cross-effect of hunger, it appears likely that there are effects for at least some people.

The nonlinear measurement “stretch” factor, denoted by the prefix “expbalance” in the parameter tables, exhibited high values for both anhedonia and anger across diverse models. This aligns with

anticipated outcomes upon observing the distributions presented in Figure 10. The implication is that variations in the upper regions of the latent process distributions yield more pronounced changes in observed values compared to analogous movements in the lower regions. The persistence of the nonlinear measurement extension’s importance, even after accommodating individual differences, hints at the inherent nonlinearity of the measurement scales themselves—at least with respect to the sample in question. In this case, it appears that accounting for the scales’ nonlinearity allowed for more dynamics to become relevant, as the same model but without measurement nonlinearity had most cross-effects much closer to zero (and performed worse).

While some clear patterns emerge from the analysis, the results are perhaps not as clear or confident with respect to lead-lag relations as are sometimes generated from analysis of such datasets. Unfortunately, this may be the grim reality we have to contend with when considering relations across time between psychological variables—there are simply many unknowns when faced with measurement error, nonlinearity, individual differences, highly uncertain time-scales at which processes operate, variable time-scales at which they are observed, unobserved confounding variables, and likely many complex interactions. As such, inferences regarding mechanism or causality from straightforward solutions such as the classical first-order discrete-time VAR without measurement error may be rather optimistic. I would love to offer an alternative straightforward solution, but at present must settle for highlighting some of the complexities regarding the treatment of time, and pointing a path to taking these more seriously.

Discussion

In this work, I have focused on the problems that using discrete-time representations of continuous-time systems can cause, when inference regarding the underlying processes is of interest. When parameters are freely estimated, discrete-time estimates can provide a perfectly valid representation of expected change in a system over a specific length of time, and can be used to estimate the effects of certain hypothetical interventions accurately. In the motivation → exercise → fitness example, for instance, a freely estimated discrete-time temporal matrix could provide an accurate estimate of the influence of changing motivation on fitness 2 years later. The discrete-time approach in general cannot be relied upon for typical usages wherein nonzero or zero cross-effect parameters are used to infer the presence or absence of a direct causal relation. Turning again to the motivation example, even though the relation between motivation and fitness was completely mediated by exercise, the cross-regression from motivation to fitness after one year was not zero. These problems also affect efforts to regularize temporal relation parameters—for a given level of sparsity in a continuous-time system (i.e., the number of variables with no direct relation), discrete-time systems tend to be less sparse, and the zeroes that are found may not represent zeroes in the underlying system relationships.

Parameterizing the system as a stochastic-differential equation allows for a truly continuous-time representation, and offers a solution to the problems associated with relying on discrete-time representations for inference about continuous-time systems. Confronted with such problems, however, an alternate approach might be to decide the task of accurately specifying and estimating such a system is simply too difficult—there are certainly instances where this may

be the most effective option. The hope contained in this article is not to offer any complete package of causal inference from observational data. Rather, the aim is to point to the substantial confusion related to discrete-time representations, and with the remaining sections on regularization, observational time lag, model order, and model checking, to point out a general path forwards for those aiming for more realistic representations of the underlying systems.

In the section on true time lag and the importance of measurement error, I discussed the known lag assumption described by [Shojaie and Fox \(2022\)](#), in which the system needs to be observed at the same frequency as it changes, and that this may be difficult to reconcile with continuous-time systems. Then I demonstrated that once measurement error is accounted for, the estimated dynamics of a continuous-time system are not strongly dependent on the sampling rate. Alternatively put, I demonstrated how a strong dependency on the sampling time-interval is induced when measurement error is not accounted for.

Time lag and model order are often the same in discrete-time systems, but this is not the case for continuous-time. While a full treatment of potential forms of model specification and how to determine such is a substantial task and well beyond the scope here, both alternative parametric and non or semiparametric approaches can serve as valuable checks and or model development guides. [Chow \(2019\)](#) offers many good suggestions, and I have demonstrated the use of both residual and posterior-predictive checks, auto and cross-correlation plots, and individual and summary statistic level analysis.

Beyond the issues discussed so far, unmeasured confounders are recognized as a major threat to any attempt at system identification and or causal inference. There are a range of approaches to handle the problems posed by unmeasured time invariant, or at least “slow” changing, variables, [Balestra and Nerlove \(1966\)](#) and these have recently received more attention in the psychological domain ([Hamaker et al., 2015](#); [Lüdtke & Robitzsch, 2022](#)). The ctsem software in use throughout this work allows for both measured and unmeasured time invariant effects across all system and measurement parameters, originally only via Bayesian estimation ([Driver & Voelkle, 2018](#)) but for some time available also via maximum likelihood or maximum a posteriori estimation. Unmeasured time-varying variables are much more problematic, and substantial consideration should be given to such possibilities whenever causal relations between variables in longitudinal studies is of interest. In the empirical example, I briefly discussed the relation between system-noise correlation and unmeasured common-causes, but this is a limited approach and will likely miss slower changing common-causes. For a review of approaches to measured time-varying confounding, see [Clare et al. \(2019\)](#), while [Lubin et al. \(2018\)](#), [Luiz and Cabral \(2010\)](#), and [Streeter et al. \(2017\)](#) considered possibilities when the unmeasured case is to be taken seriously.

A variety of software is available for fitting data to continuous-time systems with stochastic elements, as described here. Of those targeting social science fields, the R software packages ctsem ([Driver et al., 2017](#)) and dynr ([Ou et al., 2019](#)) are purpose built for addressing such systems as described here as well as nonlinear variants, and OpenMx ([Neale et al., 2016](#)) offers functionality for linear dynamic models within its broader scope. Between these three, ctsem includes more direct support for multilevel and or Bayesian approaches, dynr includes regime-switching support, and OpenMx is based on an extended structural equation modeling framework allowing a range of model choices.

Beyond the more conceptual and theoretical concerns discussed in this work, many practical questions arise when a switch from discrete to continuous-time representations is considered. A full treatment is well beyond scope, but two very common concerns include computational demands and data requirements. Continuous-time models are computationally more demanding, particularly when for example calculations cannot be reused across subjects due to the inclusion of individual differences or variable time intervals ([Driver, 2021](#)). When such calculations can be reused, time points are very few and number of subjects very high, structural equation model-based approaches as offered by ctsemOMX (the original form of ctsem presented by [Driver et al., 2017](#)) tend to be fastest—though such an approach quickly reaches its limit as variability is introduced or time points increase. State-space approaches (as offered by ctsem, dynr, and OpenMx) chunk computations into more, smaller blocks, and are thus more tractable as complexity or time points rise. Important to consider is that common discrete-time usages are often optimized for simpler cases such as assuming zero measurement error, which can make computational demand differences seem even greater than they are.

The question of “how much data do I need” is an extremely common one, and there is rarely a good answer—however, at least the answer is not very different to discrete-time approaches. If similar model structures are used to answer similar questions, a continuous-time approach may even need somewhat less data as the system is a little more constrained, and time interval variability may help to distinguish between aspects such as stable individual differences, genuine process change, and measurement error. In general, simulations are useful both as a means to understand the different representations, and to determine power under various circumstances.

Conclusion

Advances in data collection technologies have led to an explosion in longitudinal studies and interest in modeling approaches. Cross-lagged effect parameters have provided both substantial interest and substantial controversy for many years. In this work, I have described issues related to problematic interpretation of discrete-time cross-effects as directly causal, when the underlying system is in fact continually changing. This issue can be resolved and leveraged by applying more appropriate models such as offered by the ctsem ([Driver & Voelkle, 2018](#)), dynr ([Ou et al., 2019](#)), and OpenMx ([Neale et al., 2016](#)) software packages, among others. The critique and approaches shown here are but one small piece of the puzzle when it comes to accurate representations of psychological systems, but one may hope that the vast increases in data, coupled with improvements in computational performance and many such incremental works as this, do slowly lead to improved representations—the more accurate our representations of psychological systems, the better we can generalize to new circumstances, times, and interventions.

References

- Aalen, O. O. (1987). Dynamic modelling and causality. *Scandinavian Actuarial Journal*, 1987(3–4), 177–190. <https://doi.org/10.1080/03461238.1987.10413826>
- Aalen, O. O., Røysland, K., Gran, J. M., Kouyos, R., & Lange, T. (2016). Can we believe the DAGs? A comment on the relationship between causal

- DAGs and mechanisms. *Statistical Methods in Medical Research*, 25(5), 2294–2314. <https://doi.org/10.1177/0962280213520436>
- Adolf, J. K., Loossens, T., Tuerlinckx, F., & Ceulemans, E. (2021). Optimal sampling rates for reliable continuous-time first-order autoregressive and vector autoregressive modeling. *Psychological Methods*, 26(6), 701–718. <https://doi.org/10.1037/met0000398>
- Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(3), 359–388. <https://doi.org/10.1080/10705511.2017.1406803>
- Balestra, P., & Nerlove, M. (1966). Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas. *Econometrica*, 34(3), 585–612. <https://doi.org/10.2307/1909771>
- Belcher, J., Hampton, J. S., & Wilson, G. T. (1994). Parameterization of continuous time autoregressive models for irregularly sampled time series data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(1), 141–155. <https://doi.org/10.1111/j.2517-6161.1994.tb01966.x>
- Boker, S. M., Staples, A. D., & Hu, Y. (2016). Dynamics of change and change in dynamics. *Journal of Person-Oriented Research*, 2(1–2), 34–55. <https://doi.org/10.17505/jpor>
- Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2017). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods*, 22(3), 409–425. <https://doi.org/10.1037/met0000085>
- Bulteel, K., Tuerlinckx, F., Brose, A., & Ceulemans, E. (2018). Improved insight into and prediction of network dynamics by combining VAR and dimension reduction. *Multivariate Behavioral Research*, 53(6), 853–875. <https://doi.org/10.1080/00273171.2018.1516540>
- Campbell, D. T. (1963). From description to experimentation: Interpreting trends as quasi-experiments. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 212–242). University of Wisconsin Press.
- Chow, S.-M. (2019). Practical tools and guidelines for exploring and fitting linear and nonlinear dynamical systems models. *Multivariate Behavioral Research*, 54(5), 690–718. <https://doi.org/10.1080/00273171.2019.1566050>
- Chow, S.-M., Losardo, D., Park, J., & Molenaar, P. C. M. (2023). Chapter 32: Continuous-time dynamic models: Connections to structural equation models and other discrete-time models. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications*. Guilford Press.
- Chow, S.-M., Ram, N., Boker, S. M., Fujita, F., & Clore, G. (2005). Emotion as a thermostat: Representing emotion regulation using a damped oscillator model. *Emotion (Washington, D.C.)*, 5(2), 208–225. <https://doi.org/10.1037/1528-3542.5.2.208>
- Chow, S.-M., & Zhang, G. (2008). Continuous-time modelling of irregularly spaced panel data using a cubic spline model. *Statistica Neerlandica*, 62(1), 131–154. <https://doi.org/10.1111/stan.2008.62.issue-1>
- Clare, P. J., Dobbins, T. A., & Mattick, R. P. (2019). Causal models adjusting for time-varying confounding—A systematic review of the literature. *International Journal of Epidemiology*, 48(1), 254–265. <https://doi.org/10.1093/ije/dyy218>
- Conn, P. B., Johnson, D. S., Williams, P. J., Melin, S. R., & Hooten, M. B. (2018). A guide to Bayesian model checking for ecologists. *Ecological Monographs*, 88(4), 526–542. <https://doi.org/10.1002/ecm.2018.88.issue-4>
- Deboeck, P. R., & Preacher, K. J. (2016). No need to be discrete: A method for continuous time mediation analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(1), 61–75. <https://doi.org/10.1080/10705511.2014.973960>
- Deboeck, P. R., Preacher, K. J., & Cole, D. A. (2018). Mediation modeling: Differing perspectives on time alter mediation inferences. In K. van Montfort, J. H. L. Oud, & M. C. Voelkle (Eds.), *Continuous time modeling in the behavioral and related sciences* (pp. 179–203). Springer International Publishing.
- Degtiar, I., & Rose, S. (2023). A review of generalizability and transportability. *Annual Review of Statistics and Its Application*, 10(1), 501–524. <https://doi.org/10.1146/statistics.2023.10.issue-1>
- Dekkers, O. M., Romijn, J. A., & Vandenbroucke, J. P. (2010). How to assess the external validity of therapeutic trials: A conceptual approach. *International Journal of Epidemiology*, 39(1), 89–94. <https://doi.org/10.1093/ije/dyp174>
- Dickens, C. (2020). *Modeling sleep and psychological well-being dynamics in daily life using group iterative multiple model estimation* [Doctoral dissertation]. The Pennsylvania State University. <https://etda.libraries.psu.edu/catalog/17633czd171>
- Dormann, C., & Griffin, M. A. (2015). Optimal time lags in panel studies. *Psychological Methods*, 20(4), 489–505. <https://doi.org/10.1037/met0000041>
- Driver, C. C. (2021). Computational efficiency in continuous (and discrete!) time models—Comment on Hecht and Zitzmann. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(5), 791–793. <https://doi.org/10.1080/10705511.2021.1877547>
- Driver, C. C. (2023). *Inference with cross-lagged effects—Common causes in dynamic systems*. PsyArXiv. <https://doi.org/10.31234/osf.io/y3e9d>
- 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. <https://doi.org/10.18637/jss.v077.i05>
- Driver, C. C., & Tomasik, M. J. (2023). Formalizing developmental phenomena as continuous-time systems: Relations between mathematics and language development. *Child Development*, 94(6), 1454–1471. <https://doi.org/10.1111/cdev.v94.6>
- Driver, C. C., & Voelkle, M. C. (2018). Hierarchical Bayesian continuous time dynamic modeling. *Psychological Methods*, 23(4), 774–799. <https://doi.org/10.1037/met0000168>
- Driver, C. C., & Voelkle, M. C. (2021). Hierarchical continuous time modeling. In J. F. Rauthmann (Ed.), *The handbook of personality dynamics and processes* (pp. 887–908). Academic Press. <https://doi.org/10.1016/B978-0-12-813995-0.00034-0>
- Ellison, W. D., Levy, K. N., Newman, M. G., Pincus, A. L., Wilson, S. J., & Molenaar, P. C. M. (2020). Dynamics among borderline personality and anxiety features in psychotherapy outpatients: An exploration of nomothetic and idiographic patterns. *Personality Disorders: Theory, Research, and Treatment*, 11(2), 131–140. <https://doi.org/10.1037/per0000363>
- Epskamp, S. (2020). Psychometric network models from time-series and panel data. *Psychometrika*, 85(1), 206–231. <https://doi.org/10.1007/s11336-020-09697-3>
- Epskamp, S., Waldorp, L. J., Möttus, R., & Borsboom, D. (2018). The Gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research*, 53(4), 453–480. <https://doi.org/10.1080/00273171.2018.1454823>
- Fried, E. I., Papanikolaou, F., & Epskamp, S. (2022). Mental health and social contact during the COVID-19 pandemic: An ecological momentary assessment study. *Clinical Psychological Science*, 10(2), 340–354. <https://doi.org/10.1177/21677026211017839>
- Gates, K. M., Fisher, Z. F., & Bollen, K. A. (2020). Latent variable GIMME using model implied instrumental variables (MIIVs). *Psychological Methods*, 25(2), 227–242. <https://doi.org/10.1037/met0000229>
- Gates, K. M., & Molenaar, P. C. M. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage*, 63(1), 310–319. <https://doi.org/10.1016/j.neuroimage.2012.06.026>
- Gelman, A., & Shalizi, C. R. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, 66(1), 8–38. <https://doi.org/10.1111/bmsp.2013.66.issue-1>
- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3), 424–438. <https://doi.org/10.2307/1912791>
- Granger, C. W. J. (1980). Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control*, 2, 329–352. [https://doi.org/10.1016/0165-1889\(80\)90069-X](https://doi.org/10.1016/0165-1889(80)90069-X)

- Ha, H., & Welsh, J. S. (2015). *Model order selection for continuous time instrumental variable methods using regularization*. 2015 54th IEEE Conference on Decision and Control (CDC) (pp. 771–776). <https://doi.org/10.1109/CDC.2015.7402323>
- Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, 20(1), 102–116. <https://doi.org/10.1037/a0038889>
- Hansen, N., & Sokol, A. (2014). Causal interpretation of stochastic differential equations. *Electronic Journal of Probability*, 19, 1–24. <https://doi.org/10.1214/EJP.v19-2891>
- Hlaváčková-Schindler, K., Paluš, M., Vejmelka, M., & Bhattacharya, J. (2007). Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports*, 441(1), 1–46. <https://doi.org/10.1016/j.physrep.2006.12.004>
- Hopwood, C. J., Bleidorn, W., & Wright, A. G. (2022). Connecting theory to methods in longitudinal research. *Perspectives on Psychological Science*, 17(3), 884–894. <https://doi.org/10.1177/17456916211008407>
- Hu, Y., Boker, S., Neale, M., & Klump, K. L. (2014). Coupled latent differential equation with moderators: Simulation and application. *Psychological Methods*, 19(1), 56–71. <https://doi.org/10.1037/a0032476>
- Jacobs, J. P. A. M., & Otter, P. W. (2008). Determining the number of factors and lag order in dynamic factor models: A minimum entropy approach. *Econometric Reviews*, 27(4–6), 385–397. <https://doi.org/10.1080/07474930801960196>
- Jacobson, N. C., Chow, S.-M., & Newman, M. G. (2019). The differential time-varying effect model (DTVEM): A tool for diagnosing and modeling time lags in intensive longitudinal data. *Behavior Research Methods*, 51(1), 295–315. <https://doi.org/10.3758/s13428-018-1101-0>
- Jacobucci, R., & Grimm, K. J. (2018a). Comparison of frequentist and Bayesian regularization in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(4), 639–649. <https://doi.org/10.1080/10705511.2017.1410822>
- Jacobucci, R., & Grimm, K. J. (2018b). Regularized estimation of multivariate latent change score models. In E. Ferrer, S. M. Boker, & K. J. Grimm (Eds.), *Longitudinal multivariate psychology* (pp. 109–125). Routledge.
- Kuiper, R. M., & Ryan, O. (2018). Drawing conclusions from cross-lagged relationships: Re-considering the role of the time-interval. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(5), 809–823. <https://doi.org/10.1080/10705511.2018.1431046>
- Lewis, D. (1973). Causation. *Journal of Philosophy*, 70(17), 556–567. <https://doi.org/10.2307/2025310>
- Lewis-Beck, C., & Lewis-Beck, M. (2015, June 18). *Applied regression: An introduction* (Vol. 22). SAGE Publications.
- Liang, M., Koslovsky, M. D., Hébert, E. T., Kendzor, D. E., Businelle, M. S., & Vannucci, M. (2023). Bayesian continuous-time hidden Markov models with covariate selection for intensive longitudinal data with measurement error. *Psychological Methods*, 28(4), 880–894. <https://doi.org/10.1037/met0000433>
- Lubin, J. H., Hauptmann, M., & Blair, A. (2018). Indirect adjustment of relative risks of an exposure with multiple categories for an unmeasured confounder. *Annals of Epidemiology*, 28(11), 801–807. <https://doi.org/10.1016/j.annepidem.2018.09.003>
- Lüdtke, O., & Robitzsch, A. (2022). A comparison of different approaches for estimating cross-lagged effects from a causal inference perspective. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(6), 888–907. <https://doi.org/10.1080/10705511.2022.2065278>
- Luiz, R. R., & Cabral, M. D. B. (2010). Sensitivity analysis for an unmeasured confounder: A review of two independent methods. *Revista Brasileira de Epidemiologia*, 13(2), 188–198. <https://doi.org/10.1590/S1415-790X201000200002>
- Maclaren, O. J., & Nicholson, R. (2021). *Models, identifiability, and estimability in causal inference*. 38th International Conference on Machine Learning. Workshop on the Neglected Assumptions in Causal Inference. ICML. <https://sites.google.com/view/naci2021/home>.
- McArdle, J. J. (2001). A latent difference score approach to longitudinal dynamic structural analysis. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 342–380). Scientific Software International.
- Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., Estabrook, R., Bates, T. C., Maes, H. H., & Boker, S. M. (2016). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika*, 81(2), 535–549. <https://doi.org/10.1007/s11336-014-9435-8>
- Orcutt, G. H., & Irwin, J. O. (1948). A study of the autoregressive nature of the time series used for Tinbergen's model of the economic system of the United States, 1919–1932. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(1), 1–53. <https://doi.org/10.1111/rssb.1948.10.issue-1>
- Orzek, J. H., Arnold, M., & Voelkle, M. C. (2023). Striving for sparsity: On exact and approximate solutions in regularized structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 30(6), 956–973. <https://doi.org/10.1080/10705511.2023.2189070>
- Orzek, J. H., & Voelkle, M. C. (2023). Regularized continuous time structural equation models: A network perspective. *Psychological Methods*, 28(6), 1286–1320. <https://doi.org/10.1037/met0000550>
- Ou, L., Hunter, M. D., & Chow, S.-M. (2019). What's for dynr: A package for linear and nonlinear dynamic modeling in R. *The R Journal*, 11(1), 91–111. <https://doi.org/10.32614/RJ-2019-012>
- Oud, J. H. L., Voelkle, M. C., & Driver, C. C. (2018). First- and higher-order continuous time models for arbitrary n using sem. In K. van Montfort, J. H. Oud, & M. C. Voelkle (Eds.), *Continuous time modeling in the behavioral and related sciences* (pp. 1–26). Springer International Publishing. https://doi.org/10.1007/978-3-319-77219-6_1
- Park, J. J., Chow, S.-M., Fisher, Z. F., & Molenaar, P. C. M. (2020). Affect and personality: Ramifications of modeling (non-)directionality in dynamic network models. *European Journal of Psychological Assessment*, 36(6), 1009–1023. <https://doi.org/10.1027/1015-5759/a000612>
- Park, T., & Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, 103(482), 681–686. <https://doi.org/10.1198/016214508000000337>
- Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*, 3, 96–146. <https://doi.org/10.1214/09-SS057>
- Rothwell, P. M. (2006). Factors that can affect the external validity of randomised controlled trials. *PLoS Clinical Trials*, 1(1), Article e9. <https://doi.org/10.1371/journal.pctr.0010009>
- Rubin, D. B. (2005). Causal inference using potential outcomes. *Journal of the American Statistical Association*, 100(469), 322–331. <https://doi.org/10.1198/016214504000001880>
- Ruissen, G. R., Zumbo, B. D., Rhodes, R. E., Puterman, E., & Beauchamp, M. R. (2022). Analysis of dynamic psychological processes to understand and promote physical activity behaviour using intensive longitudinal methods: A primer. *Health Psychology Review*, 16(4), 492–525. <https://doi.org/10.1080/17437199.2021.1987953>
- Ryan, O. (2023, June 5). *Expct* (Version 0.1.0) [Software]. Retrieved November 27, 2023, from <https://github.com/ryanoisin/expct>
- Ryan, O., & Hamaker, E. L. (2022). Time to intervene: A continuous-time approach to network analysis and centrality. *Psychometrika*, 87(1), 214–252. <https://doi.org/10.1007/s11336-021-09767-0>
- Ryan, O., Haslbeck, J., & Waldorp, L. (2023). *Non-stationarity in time-series analysis: Modeling stochastic and deterministic trends*. PsyArXiv. <https://doi.org/10.31234/osf.io/z7ja2>
- Schuurman, N. K., & Hamaker, E. L. (2019). Measurement error and person-specific reliability in multilevel autoregressive modeling. *Psychological Methods*, 24(1), 70–91. <https://doi.org/10.1037/met0000188>
- Schuurman, N. K., Houtveen, J. H., & Hamaker, E. L. (2015). Incorporating measurement error in $n=1$ psychological autoregressive modeling. *Frontiers in Psychology*, 6, Article 1038. <https://doi.org/10.3389/fpsyg.2015.01038>

- Shojaie, A., & Fox, E. B. (2022). Granger causality: A review and recent advances. *Annual Review of Statistics and Its Application*, 9(1), 289–319. <https://doi.org/10.1146/statistics.2022.9.issue-1>
- Shumway, R. H., & Stoffer, D. S. (2017). Time series regression and exploratory data analysis. In R. H. Shumway & D. S. Stoffer (Eds.), *Time series analysis and its applications: With R examples*, Springer Texts in Statistics (pp. 45–74). Springer International Publishing. https://doi.org/10.1007/978-3-319-52452-8_2
- Spirtes, P., Glymour, C. N., & Scheines, R. (2000). *Causation, prediction, and search*. MIT Press.
- Streeter, A. J., Lin, N. X., Crathorne, L., Haasova, M., Hyde, C., Melzer, D., & Henley, W. E. (2017). Adjusting for unmeasured confounding in non-randomized longitudinal studies: A methodological review. *Journal of Clinical Epidemiology*, 87, 23–34. <https://doi.org/10.1016/j.jclinepi.2017.04.022>
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Timmons, A. C., & Preacher, K. J. (2015). The importance of temporal design: How do measurement intervals affect the accuracy and efficiency of parameter estimates in longitudinal research? *Multivariate Behavioral Research*, 50(1), 41–55. <https://doi.org/10.1080/00273171.2014.961056>
- Voelkle, M. C., Gische, C., Driver, C. C., & Lindenberg, U. (2019). The role of time in the quest for understanding psychological mechanisms. *Multivariate Behavioral Research*, 53(6), 782–805. <https://doi.org/10.1080/00273171.2018.1496813>
- Voelkle, M. C., & Oud, J. H. (2015). Relating latent change score and continuous time models. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(3), 366–381. <https://doi.org/10.1080/10705511.2014.935918>
- Voelkle, M. C., Oud, J. H. L., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological Methods*, 17(2), 176–192. <https://doi.org/10.1037/a0027543>
- Wiener, N. (1956). The theory of prediction. In E. Beckenbach (Ed.), *Modern mathematics for engineers* (pp. 165–190). McGraw-Hill.
- Yuen, T. P., Wong, H., & Yiu, K. F. C. (2018). On constrained estimation of graphical time series models. *Computational Statistics & Data Analysis*, 124, 27–52. <https://doi.org/10.1016/j.csda.2018.01.019>
- Zyphur, M. J., Allison, P. D., Tay, L., Voelkle, M. C., Preacher, K. J., Zhang, Z., Hamaker, E. L., Shamsollahi, A., Pierides, D. C., Koval, P., & Diener, E. (2020). From data to causes I: Building a general cross-lagged panel model (GCLM). *Organizational Research Methods*, 23(4), 651–687. <https://doi.org/10.1177/1094428119847278>

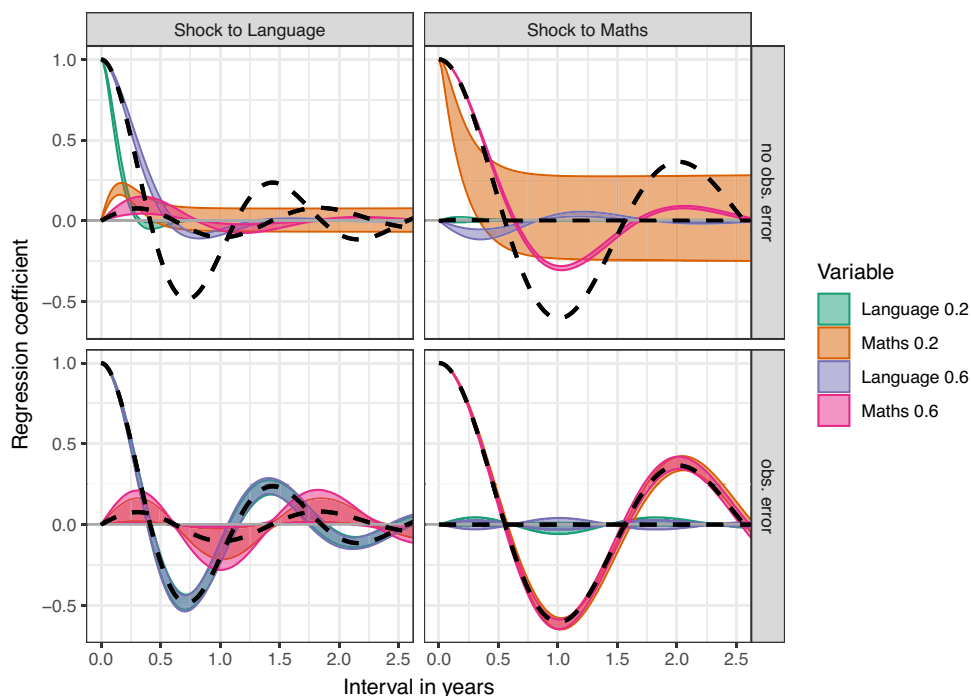
(Appendices follow)

Appendix A

Implications of Unaccounted for Measurement Error in a Second-Order Model

Figure A1

Discrete-Time Auto and Cross Regression 95% Confidence Regions Estimated Based on Either a Correctly Specified Second-Order Continuous-Time Model (Bottom Row) or the Same Model but Without Accounting for Measurement Error (Top Row)



Note. The variable types are suffixed with the sampling time intervals of either 0.2 or 0.6, but as there is no difference in the underlying process both the maths (orange/pink), and language (green/purple), regions should overlap the true coefficients (in dashed black), as in the correctly specified lower row. The top row where measurement error is not accounted for shows spurious cross-effects, inaccurate estimates, and changes in estimated time-scale (i.e., slower/faster) depending on sampling rate. These problems do not occur once measurement error is accounted for, as in the lower row. obs. = observations. See the online article for the color version of this figure.

Appendix B

Extended Empirical Analysis Details

This section contains a range of points and figures that are relevant to the specific analyses, but not highly relevant to understanding the application of continuous-time models.

Additional Measurement Model Details

Because we generally assume the underlying phenomena (e.g., anger) are approximately continuous, but they are measured on a 1–5 integer scale, a random value between -0.5 and 0.5 was added to each observation and a minimum level of measurement error was enforced for the model. This avoids the possible issue where a very high or very low autocorrelation is estimated due

only to the measurement discretization, and not the underlying process. This issue became evident in this case when plotting sample autocorrelation versus model autocorrelation functions, but is not specific to continuous-time approaches or the ctsem software.

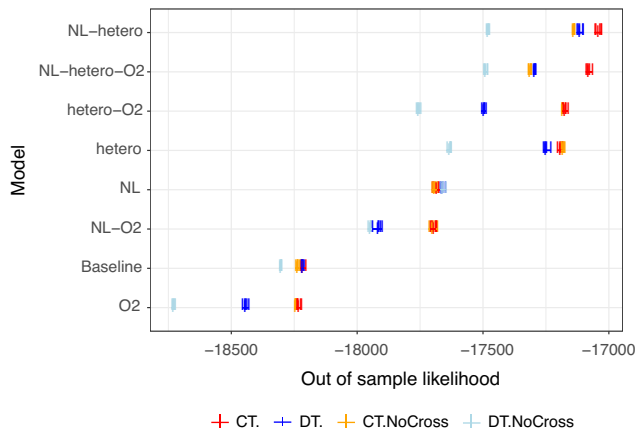
The nonlinear measurement model incorporating a stretch factor to account for compression to lower scale values is:

$$\mathbf{x}_{uj} = (1 - \mathbf{s}_j)\mathbf{y}_{uj} + \mathbf{s}_j e^{\mathbf{y}_{uj}} + \boldsymbol{\epsilon}_{uj}, \quad \text{where } \boldsymbol{\epsilon}_{uj} \sim N(0, \boldsymbol{\Theta}_j). \quad (\text{B1})$$

Using j to index the latent variable, when for instance $\mathbf{s}_j$ is zero, the measurement model is equivalent to the earlier shown linear model. When $\mathbf{s}_j$ is exactly 1, the measurement model is simply

(Appendices continue)

Figure B1
Out-of-Sample Likelihoods for 32 Different Fitted Models



Note. The models are either CT or DT, may have an NL measurement model, may account for heterogeneity in individual dynamics and measurement (hetero), and may have cross-effects fixed to zero (NoCross). The smaller lines represent individual runs of 10-fold cross-validation, the larger line is the mean over the five runs. NL = nonlinear; hetero = heterogeneity; O2 = second-order model; CT = continuous-time; DT = discrete-time; NoCross = removing cross-effects. See the online article for the color version of this figure.

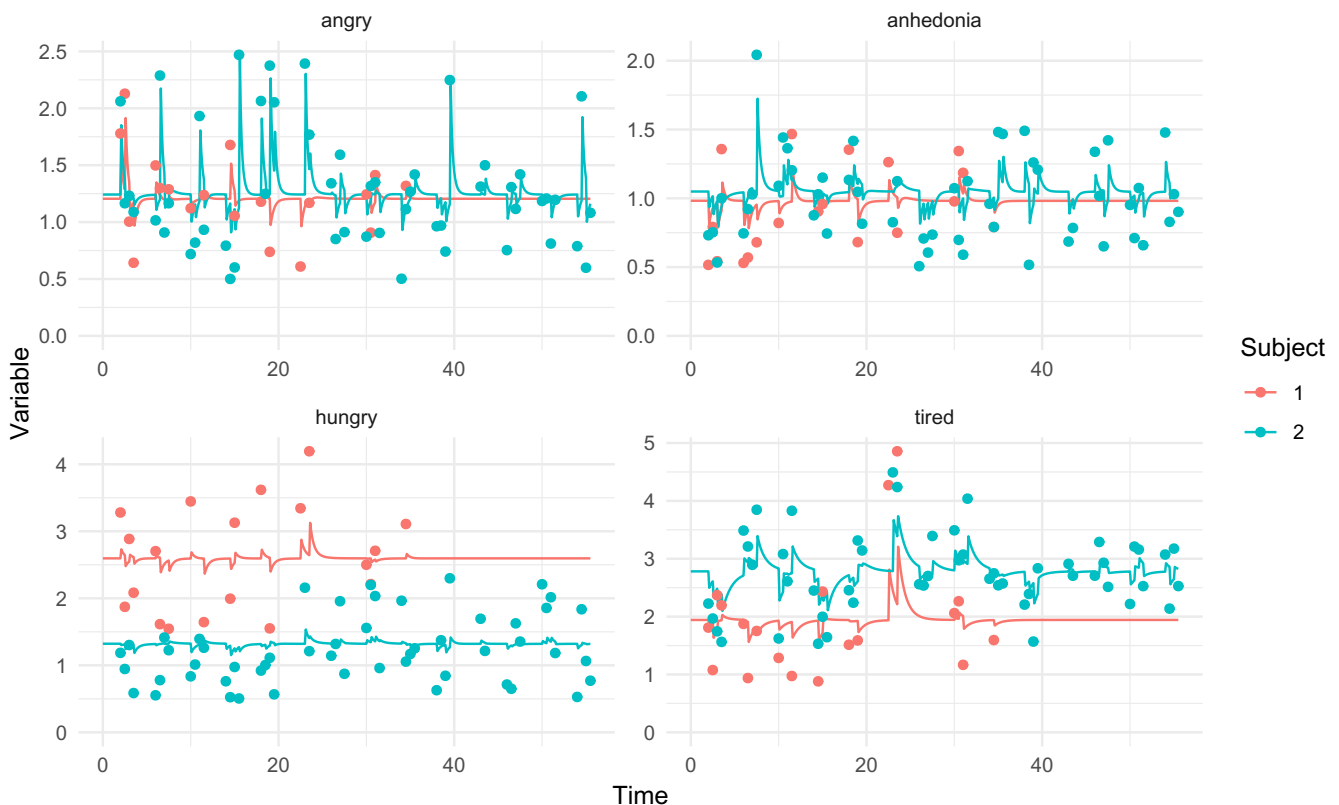
that the observation $\mathbf{x}_{uj}$ is the exponential of the latent process $\mathbf{y}_{uj}$, with the usual intercept and residual terms. When s_j is between 0 and 1, the observation is a mixture of the latent value $\mathbf{y}_{uj}$ and its exponential.

Model Comparisons Plot

Interpreting the Estimated Model

We can get some sense of how the best performing model functions by looking at subject-level prediction plots, as shown in Figure B2. From this, we can see, for instance, that changes in observed hunger are only weakly informative compared to changes in observed anger (i.e., the forecast is not as influenced when a new data point arrives). This is because anger is estimated as having substantially lower measurement error, so the updated prediction once new data arrives relies is weighted more heavily toward the pre-existing prediction. Whether or not this is truly measurement error, or simply that the model and data combination performs best when more information about hunger is discarded, is of course not possible to say without further data collection (such as multiple anger and hunger measurement types).

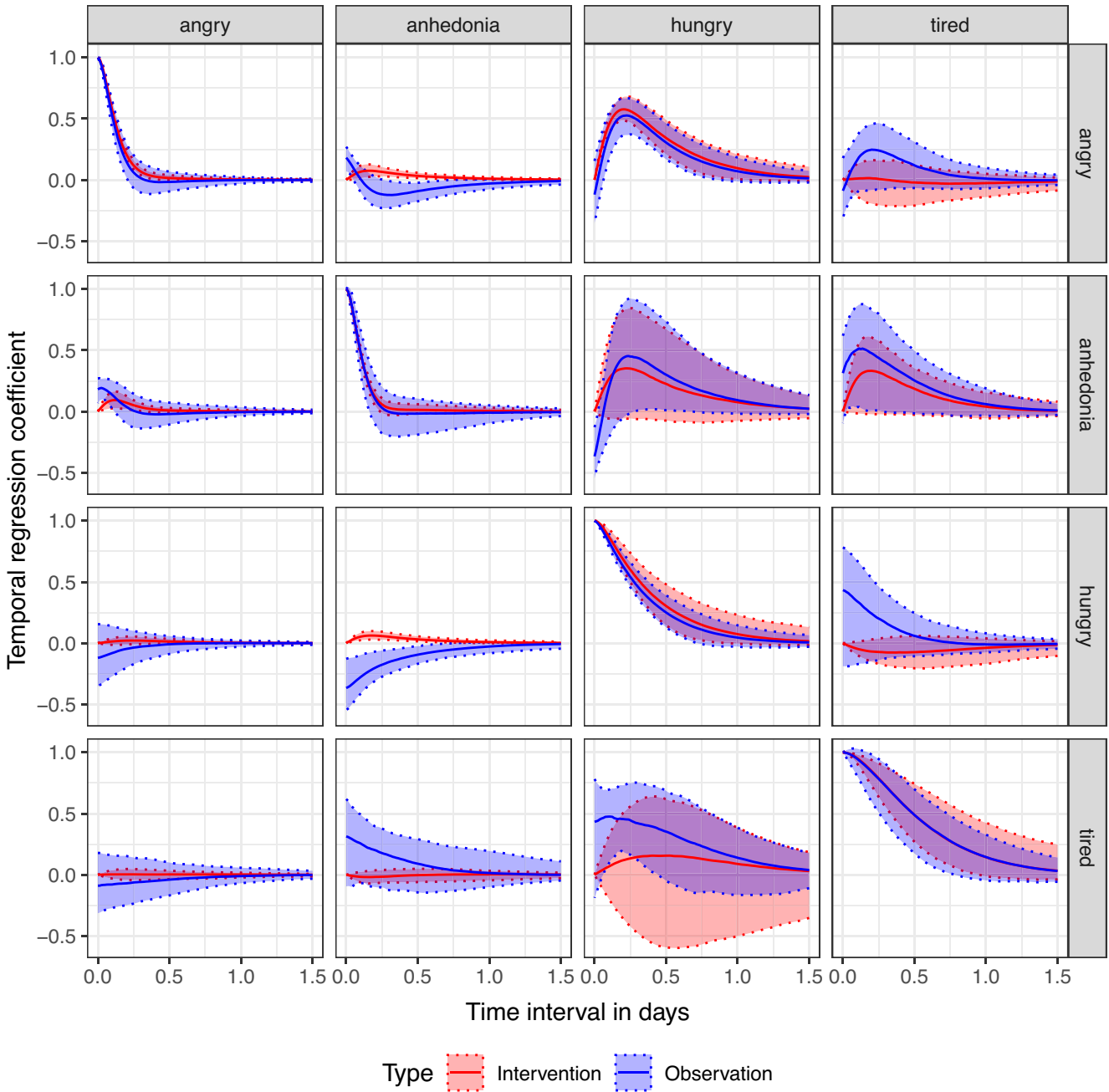
Figure B2
One-Step Ahead Forward Predictions (Lines) From the Model With Nonlinear Measurement and Heterogeneity, and Empirical Data (Dots), for Two Subjects



Note. See the online article for the color version of this figure.

(Appendices continue)

Figure B3
Temporal Regression Estimates and 95% Quantiles for the Second Best-Fitting Model, a Second-Order Continuous-Time Model With Cross Effects, Allowing for Nonlinear Measurement Properties, Individual Differences in Dynamics



Note. The column represents the “causal” or predicting variable, and the row the affected or predicted variable. Conditional expectations based on two different hypothetical scenarios are shown. In the intervention type, the plot shows the effect of a change in the causal variable only, while the observational type shows the expected effect conditional on observing a change in the causal variable, and accounting for correlated random changes in the system. See the online article for the color version of this figure.

Table B1

Parameter Estimates for the Best-Fitting Model, a First-Order Continuous-Time Model With Cross Effects, Allowing for Nonlinear Measurement Properties, and Individual Differences in Dynamics

Parameter	Est.	SD	2.5%	97.5%
drift_anhedonia	-2.66	0.21	-3.06	-2.28
drift_anhedonia_tired	0.59	0.59	-0.59	1.65
drift_anhedonia_hungry	-0.36	0.31	-0.96	0.27
drift_anhedonia_angry	0.17	0.09	-0.03	0.35
drift_tired_anhedonia	-0.05	0.06	-0.17	0.09
drift_tired	-2.02	0.33	-2.72	-1.39
drift_tired_hungry	0.05	0.11	-0.17	0.27
drift_tired_angry	0.04	0.04	-0.03	0.11
drift_hungry_anhedonia	0.00	0.07	-0.14	0.13
drift_hungry_tired	0.21	0.28	-0.37	0.75
drift_hungry	-3.03	0.54	-4.13	-2.08
drift_hungry_angry	0.02	0.08	-0.14	0.16
drift_angry_anhedonia	0.40	0.18	0.05	0.74
drift_angry_tired	0.30	0.31	-0.31	0.90
drift_angry_hungry	-0.45	0.30	-1.05	0.15
drift_angry	-1.59	0.24	-2.06	-1.18
diff_anhedonia	0.81	0.23	0.45	1.33
diff_tired_anhedonia	0.12	0.05	0.03	0.23
diff_tired	0.25	0.04	0.18	0.33
diff_hungry_anhedonia	-0.01	0.04	-0.08	0.06
diff_hungry_tired	0.21	0.07	0.07	0.34
diff_hungry	0.37	0.10	0.21	0.61
diff_angry_anhedonia	0.12	0.02	0.08	0.17
diff_angry_tired	-0.02	0.03	-0.09	0.04
diff_angry_hungry	0.04	0.05	-0.04	0.14
diff_angry	2.06	0.40	1.38	2.91
mvaranhedonia	0.27	0.01	0.25	0.30
mvartired	0.58	0.04	0.50	0.68
mvarhungry	0.70	0.04	0.64	0.78
mvarangry	0.29	0.03	0.26	0.37
expbalance_anhedonia	0.92	0.02	0.87	0.96
expbalance_tired	0.25	0.04	0.18	0.33
expbalance_hungry	0.00	0.00	0.00	0.00
expbalance_angry	0.96	0.02	0.92	0.98
offset_anhedonia	0.29	0.16	-0.03	0.60
offset_tired	-2.21	0.27	-2.73	-1.68
offset_hungry	-5.85	0.63	-7.11	-4.61
offset_angry	0.66	0.10	0.48	0.86
cint_anhedonia	-1.25	1.96	-5.04	2.63
cint_tired	3.13	0.71	1.68	4.49
cint_hungry	17.41	2.81	11.94	23.19
cint_angry	-3.52	2.59	-8.84	1.47

Note. Individual difference parameters not included. Est. = estimate.

Table B2

Parameter Estimates for the Second Best-Fitting Model, a Second-Order Continuous-Time Model With Cross Effects, Allowing for Nonlinear Measurement Properties, Individual Differences in Dynamics

Parameter	Est.	SD	2.5%	97.5%
drift_anhedonia_tired	0.95	0.51	-0.11	1.94
drift_anhedonia_hungry	0.79	0.50	-0.21	1.77
drift_anhedonia_angry	0.37	0.15	0.07	0.65
drift_tired_anhedonia	-0.06	0.07	-0.21	0.08
drift_tired_hungry	0.20	0.38	-0.53	0.95
drift_tired_angry	0.01	0.06	-0.10	0.13
drift_hungry_anhedonia	0.18	0.06	0.07	0.28
drift_hungry_tired	-0.16	0.09	-0.32	0.02
drift_hungry_angry	0.03	0.05	-0.06	0.12
drift_angry_anhedonia	0.20	0.11	0.00	0.41
drift_angry_tired	0.03	0.21	-0.38	0.42
drift_angry_hungry	1.42	0.11	1.20	1.62
drift_danhedonia_anhedonia	-16.70	1.44	-19.38	-13.82
drift_anhedonia	-7.12	0.69	-8.43	-5.78
drift_tired_tired	-0.70	0.26	-1.28	-0.29
drift_tired	-1.81	0.59	-3.04	-0.82
drift_dhungry_hungry	-3.12	0.95	-4.99	-1.32
drift_hungry	-4.60	1.49	-7.66	-2.04
drift_dangry_angry	-12.85	1.10	-14.94	-10.73
drift_angry	-6.71	0.52	-7.73	-5.67
diff_anhedonia	1.12	0.50	0.46	2.33
diff_tired_anhedonia	0.21	0.10	-0.01	0.40
diff_tired	0.14	0.05	0.07	0.26
diff_hungry_anhedonia	-0.23	0.06	-0.34	-0.10
diff_hungry_tired	0.27	0.16	-0.07	0.55
diff_hungry	0.12	0.03	0.08	0.18
diff_angry_anhedonia	0.10	0.03	0.03	0.16
diff_angry_tired	-0.05	0.07	-0.19	0.10
diff_angry_hungry	-0.05	0.07	-0.19	0.09
diff_angry	1.12	0.15	0.86	1.43
mvaranhedonia	0.26	0.01	0.25	0.29
mvartired	0.63	0.06	0.53	0.77
mvarhungry	0.73	0.04	0.67	0.81
mvarangry	0.27	0.02	0.25	0.32
expbalance_anhedonia	0.96	0.01	0.93	0.98
expbalance_tired	0.22	0.11	0.06	0.48
expbalance_hungry	0.01	0.00	0.00	0.01
expbalance_angry	0.95	0.02	0.91	0.98
offset_anhedonia	0.48	0.15	0.20	0.77
offset_tired	-1.13	0.64	-2.33	0.14
offset_hungry	-4.91	0.61	-6.13	-3.75
offset_angry	0.46	0.08	0.30	0.62
cint_anhedonia	-8.10	3.76	-14.73	-0.54
cint_tired	-0.54	1.91	-4.22	3.05
cint_hungry	3.65	0.60	2.50	4.81
cint_angry	-10.15	1.52	-13.08	-7.06

Note. Individual difference parameters not included. Est. = estimate.

Table B3

Significant Covariate Effects for the Best-Fitting Model, a First-Order Continuous-Time Model With Cross Effects, Allowing for Nonlinear Measurement Properties, and Individual Differences in Dynamics

Parameter	Est.	SD	2.5%	97.5%
anhedoniaMu_drift_anhedonia_hungry	0.33	0.04	0.24	0.41
anhedoniaMu_diff_anhedonia	0.21	0.04	0.10	0.27
tiredMu_drift_tired	0.46	0.10	0.29	0.67
tiredMu_drift_tired_hungry	-0.06	0.02	-0.09	-0.03
hungryMu_drift_hungry	0.19	0.05	0.10	0.28
angryMu_drift_angry_hungry	0.52	0.04	0.44	0.61
angryMu_drift_angry	-0.94	0.13	-1.19	-0.69

Note. Est. = estimate.

(Appendices continue)

Appendix C

Best-Fitting Model Specification—Subject Level

$$\begin{aligned}
 &\text{Deterministic change :} \quad d \begin{bmatrix} \text{anhedonia} \\ \text{tired} \\ \text{hungry} \\ \text{angry} \end{bmatrix} \eta(t) = \left(\begin{bmatrix} \text{drift_anhedonia} & \text{drift_anhedonia_tired} & \text{drift_anhedonia_hungry} & \text{drift_anhedonia_angry} \\ \text{drift_tired_anhedonia} & \text{drift_tired} & \text{drift_tired_hungry} & \text{drift_tired_angry} \\ \text{drift_hungry_anhedonia} & \text{drift_hungry_tired} & \text{drift_hungry} & \text{drift_hungry_angry} \\ \text{drift_angry_anhedonia} & \text{drift_angry_tired} & \text{drift_angry_hungry} & \text{drift_angry} \end{bmatrix} \underbrace{\mathbf{A}}_{\text{DRIFT}} \underbrace{\begin{bmatrix} \text{anhedonia} \\ \text{tired} \\ \text{hungry} \\ \text{angry} \end{bmatrix}}_{\eta(t)} + \underbrace{\begin{bmatrix} \text{cint_anhedonia} \\ \text{cint_tired} \\ \text{cint_hungry} \\ \text{cint_angry} \end{bmatrix}}_{\mathbf{b}_{\text{CINT}}} \right) dt + \\
 &\text{Random change :} \quad UcorSDtoChol \left\{ \begin{bmatrix} \text{diff_anhedonia} & 0 & 0 & 0 \\ \text{diff_tired_anhedonia} & \text{diff_tired} & 0 & 0 \\ \text{diff_hungry_anhedonia} & \text{diff_hungry_tired} & \text{diff_hungry} & 0 \\ \text{diff_angry_anhedonia} & \text{diff_angry_tired} & \text{diff_angry_hungry} & \text{diff_angry} \end{bmatrix} \underbrace{\mathbf{G}}_{\text{DIFFUSION}} \underbrace{\begin{bmatrix} W_1 \\ W_2 \\ W_3 \\ W_4 \end{bmatrix}}_{dW(t)} \right\} \\
 &\text{Observations :} \quad \mathbf{Y}(t) = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{\underbrace{\mathbf{\Lambda}}_{\text{LAMBDA}}} \underbrace{\begin{bmatrix} \text{anhedonia} \\ \text{tired} \\ \text{hungry} \\ \text{angry} \end{bmatrix}}_{\eta(t)} + \underbrace{\begin{bmatrix} \text{anhedonia} * (1 - \text{expbalance_anhedonia}) + \text{exp(anhedonia)} * \text{expbalance_anhedonia} + \text{offset_anhedonia} \\ \text{tired} * (1 - \text{expbalance_tired}) + \text{exp(tired)} * \text{expbalance_tired} + \text{offset_tired} \\ \text{hungry} * (1 - \text{expbalance_hungry}) + \text{exp(hungry)} * \text{expbalance_hungry} + \text{offset_hungry} \\ \text{angry} * (1 - \text{expbalance_angry}) + \text{exp(angry)} * \text{expbalance_angry} + \text{offset_angry} \end{bmatrix}}_{\underbrace{\mathbf{\Gamma}}_{\text{MANIFESTMEANS}}} \\
 &\text{Observation noise :} \quad \underbrace{\begin{bmatrix} \text{mvaranhedonia} & 0 & 0 & 0 \\ 0 & \text{mvertired} & 0 & 0 \\ 0 & 0 & \text{mvarhungry} & 0 \\ 0 & 0 & 0 & \text{mvarangry} \end{bmatrix}}_{\underbrace{\mathbf{\Theta}}_{\text{MANIFESTVAR}}} \underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}}_{\epsilon(t)} \\
 &\text{System noise distribution per time step :} \quad \Delta \left[W_{j \in [1,4]} \right] (t - u) \sim N(0, t - u) \quad \text{Observationnoise distribution :} \quad \left[\epsilon_{j \in [1,4]} \right] (t) \sim N(0, 1)
 \end{aligned}$$

Note. UcorSDtoChol converts lower tri matrix of standard deviations and unconstrained correlations to Cholesky factor, UcorSDtoCov = transposed cross product of UcorSDtoChol, to give covariance, see [Driver and Voelkle \(2018, p. 11\)](#).

(Appendices continue)

Appendix D

Best-Fitting Model Specification—Individual Differences

Continuous intercept values, setting the baseline level for individuals, are modeled via random effects:

$$\underbrace{\begin{bmatrix} \text{cint_anhedonia}_i \\ \text{cint_tired}_i \\ \text{cint_hungry}_i \\ \text{cint_angry}_i \end{bmatrix}}_{\Phi(i)} \sim N \left(\begin{bmatrix} \text{raw_cint_anhedonia} \\ \text{raw_cint_tired} \\ \text{raw_cint_hungry} \\ \text{raw_cint_angry} \end{bmatrix}, \begin{bmatrix} \text{rawPCov_1_1} & \text{rawPCov_2_1} & \text{rawPCov_3_1} & \text{rawPCov_4_1} \\ \text{rawPCov_2_1} & \text{rawPCov_2_2} & \text{rawPCov_3_2} & \text{rawPCov_4_2} \\ \text{rawPCov_3_1} & \text{rawPCov_3_2} & \text{rawPCov_3_3} & \text{rawPCov_4_3} \\ \text{rawPCov_4_1} & \text{rawPCov_4_2} & \text{rawPCov_4_3} & \text{rawPCov_4_4} \end{bmatrix} \right)$$

Individual differences in temporal dynamics, random fluctuations, and measurement error are modeled via four moderators (mean levels of each variable within person), which exert a linear effect on the unconstrained parameter scale. This is then constrained (e.g., positivity in the case of standard deviation parameters) via the “tform” operator. See [Driver and Voelkle \(2018\)](#) for more details.

(Appendices continue)

drift_anhedonia _i	raw_drift_anhedonia	drift_anhedonia_anhedoniaMu	drift_anhedonia_tiredMu	drift_anhedonia_hungryMu	drift_anhedonia_angryMu
drift_anhedonia_tired _i	raw_drift_anhedonia_tired	drift_anhedonia_tired_anhedoniaMu	drift_anhedonia_tired_tiredMu	drift_anhedonia_tired_hungryMu	drift_anhedonia_tired_angryMu
drift_anhedonia_hungry _i	raw_drift_anhedonia_hungry	drift_anhedonia_hungry_anhedoniaMu	drift_anhedonia_hungry_tiredMu	drift_anhedonia_hungry_hungryMu	drift_anhedonia_hungry_angryMu
drift_anhedonia_angry _i	raw_drift_anhedonia_angry	drift_anhedonia_angry_anhedoniaMu	drift_anhedonia_angry_tiredMu	drift_anhedonia_angry_hungryMu	drift_anhedonia_angry_angryMu
drift_tired_anhedonia _i	raw_drift_tired_anhedonia	drift_tired_anhedonia_anhedoniaMu	drift_tired_anhedonia_tiredMu	drift_tired_anhedonia_hungryMu	drift_tired_anhedonia_angryMu
drift_tired _i	raw_drift_tired	drift_tired_anhedoniaMu	drift_tired_tiredMu	drift_tired_hungryMu	drift_tired_angryMu
drift_tired_hungry _i	raw_drift_tired_hungry	drift_tired_hungry_anhedoniaMu	drift_tired_hungry_tiredMu	drift_tired_hungry_hungryMu	drift_tired_hungry_angryMu
drift_tired_angry _i	raw_drift_tired_angry	drift_tired_angry_anhedoniaMu	drift_tired_angry_tiredMu	drift_tired_angry_hungryMu	drift_tired_angry_angryMu
drift_hungry_anhedonia _i	raw_drift_hungry_anhedonia	drift_hungry_anhedonia_anhedoniaMu	drift_hungry_anhedonia_tiredMu	drift_hungry_anhedonia_hungryMu	drift_hungry_anhedonia_angryMu
drift_hungry_tired _i	raw_drift_hungry_tired	drift_hungry_tired_anhedoniaMu	drift_hungry_tired_tiredMu	drift_hungry_tired_hungryMu	drift_hungry_tired_angryMu
drift_hungry _i	raw_drift_hungry	drift_hungry_anhedoniaMu	drift_hungry_tiredMu	drift_hungry_hungryMu	drift_hungry_angryMu
drift_hungry_angry _i	raw_drift_hungry_angry	drift_hungry_angry_anhedoniaMu	drift_hungry_angry_tiredMu	drift_hungry_angry_hungryMu	drift_hungry_angry_angryMu
drift_angry_anhedonia _i	raw_drift_angry_anhedonia	drift_angry_anhedonia_anhedoniaMu	drift_angry_anhedonia_tiredMu	drift_angry_anhedonia_hungryMu	drift_angry_anhedonia_angryMu
drift_angry_tired _i	raw_drift_angry_tired	drift_angry_tired_anhedoniaMu	drift_angry_tired_tiredMu	drift_angry_tired_hungryMu	drift_angry_tired_angryMu
drift_angry_hungry _i	raw_drift_angry_hungry	drift_angry_hungry_anhedoniaMu	drift_angry_hungry_tiredMu	drift_angry_hungry_hungryMu	drift_angry_hungry_angryMu
drift_angry _i	raw_drift_angry	drift_angry_anhedoniaMu	drift_angry_tiredMu	drift_angry_hungryMu	drift_angry_angryMu
mvaranhedonia _i	raw_mvaranhedonia	mvaranhedonia_anhedoniaMu	mvaranhedonia_tiredMu	mvaranhedonia_hungryMu	mvaranhedonia_angryMu
mvartired _i	raw_mvartired	mvartired_anhedoniaMu	mvartired_tiredMu	mvartired_hungryMu	mvartired_angryMu
mvahungry _i	raw_mvahungry	mvahungry_anhedoniaMu	mvahungry_tiredMu	mvahungry_hungryMu	mvahungry_angryMu
mvarangry _i	raw_mvarangry	mvarangry_anhedoniaMu	mvarangry_tiredMu	mvarangry_hungryMu	mvarangry_angryMu

$\sim$ tform

$+$

z

anhedoniaMu_i
tiredMu_i
hungryMu_i
angryMu_i