STRUCTURAL EQUATION MODELING
2025, VOL. 32, NO. 3
ISSN: 1070-5511 | E-ISSN: 1532-8007

Unsupervised Model Construction in Continuous-Time

Jonathan J. Park[a], Zachary F. Fisher[b], Michael D. Hunter[b], Chad Shenk[c], Michael Russell[b], Peter C. M. Molenaar[b], and Sy-Miin Chow[b]

[a]University of California, Davis; [b]The Pennsylvania State University; [c]University of Rochester

ABSTRACT

Many of the advancements reconciling individual- and group-level results have occurred in the context of a discrete-time modeling framework. Discrete-time models are intuitive and offer relatively simple interpretations for the resulting dynamic structures; however, they do not possess the flexibility of models fitted in the continuous-time framework. We introduce `ct-gimme`, a continuous-time extension of the group iterative multiple model estimation (GIMME) procedure which enables researchers to fit complex, high-dimensional dynamic networks in continuous time. Our results indicate that `ct-gimme` outperforms $N = 1$ model fitting in continuous time by pooling information across multiple subjects. Likewise, `ct-gimme` outperforms group-level model fitting in the presence of within-sample heterogeneity. We conclude with an empirical illustration and highlight the limitations of the approach relating to the identification of meaningful starting values.

KEYWORDS

Continuous-time; dynamic network analysis; network analysis; state-space modeling; structural equation modeling

The social and behavioral sciences have witnessed a surge in the popularity of person-specific models for describing individual processes overtime over the last decade. This is due—in part—to the convergence of myriad developments such as frequent calls for increased attention to individuals (Fisher et al., 2018; Molenaar, 2004) alongside technological developments for gathering large streams of data at little to no burden to participants via smartphones, mobile applications, and wearable devices. Indeed, publications concerning "person-specific" modeling on Google Scholar have risen from 15,600 articles published between 1910 and 2010 to over 21,800 published only between 2010 and 2023. This certainly indicates an increased awareness and interest in modeling person-specific dynamics. However, this area of statistical modeling in the social and behavioral sciences is still in development.

Several issues limit our ability to describe and model individual dynamics such as the (mis)alignment of our data, theories, and models, as well as how to best identify group-level signals from a sample of noisy person-specific processes. Further, the popularity of person-specific modeling has brought along with it criticisms on how to best derive group-level inference from a sample of $N = 1$ results (Hamaker, 2004; Lundh, 2015; Runyan, 1983). These arguments are cogent when many person-specific analyses find heterogeneity to be a rule more often than an exception (e.g., De Vos et al., 2017; Gates & Molenaar, 2012; Kim et al., 2007). However, while individuals may vary greatly from each other in their prototypical dynamics over time, some similarities have been consistently identified within the literature. For instance, the concept of "inertia" has been replicated consistently (Koval et al., 2012; Kuppens et al., 2010). Methods over the last decade have been developed with this purpose in mind and have ranged from fully constrained models fitted to the "chained" time series of a sample of individuals (e.g., Epskamp et al., 2018) to more relaxed methods that characterize groups based on features common to a majority of individuals in the sample (e.g., GIMME; Gates & Molenaar, 2012).

In the current work, we adapt a popular algorithmic procedure developed in the discrete-time framework: the Group Iterative Multiple Model Estimation (GIMME; Gates & Molenaar, 2012) procedure, to operate in continuous time. In so doing, we introduce a novel contribution to the literature that identifies systematic covariations in change processes across a sample of individuals with person- and (possibly) time-specific time intervals between successive occasions, and assist in the construction of group- and individual-level dynamic networks henceforth referred to as continuous-time GIMME or `ct-gimme`. Using a continuous-time model as the computational backbone, the `ct-gimme` algorithm readily handles methodological concerns that may affect its discrete-time counterparts, such as unequal intervals between successive measurement occasions, and other identification challenges, particularly in determining the directionality of contemporaneous cross-process influences.

We provide a review of the original, discrete-time GIMME algorithm. Then, we introduce the continuous-time extension: `ct-gimme`. We describe how it is fitted in the continuous-time framework, and discuss differences between

the original GIMME algorithm and the continuous-time form. The proposed changes and alterations naturally motivate the questions which are evaluated in a simulation study followed by an empirical illustration that highlights the strengths and weaknesses of the proposed algorithm.

Gimme

GIMME is a powerful methodological tool that allows researchers to identify group-level commonalities in person-specific processes (Gates & Molenaar, 2012). We detail the modeling framework from which GIMME is based and provide a brief description of the algorithm to provide context for `ct-gimme` in the following sections.

Modeling in Discrete-Time

The vector autoregression (VAR) is the most utilized statistical model in person-specific modeling (Piccirillo & Rodebaugh, 2019). VAR models are intuitive as variables at one moment in time relate to others at a given lag order (Lütkepohl, 2005). For the remainder of this work, we will only be concerned with lag-order 1 models. VAR models have seen extensive application in the social and behavioral sciences ranging from econometrics (Lütkepohl, 2005), affective modeling (Chow et al., 2010; De Vos et al., 2017), and neuroimaging (Gates & Molenaar, 2012; Wang et al., 2023). Work with these models has led to key insights relating model parameters VAR-type models to real-world phenomena. For instance, "emotional inertia" relates to the auto-regressive parameters of a VAR model and describes the rate at which a variable returns to its baseline levels (Kuppens et al., 2010). Likewise, cross-regression coefficients have been used to describe how and why individuals exhibit broader forms of "inertia" by means of strongly connected dynamics that prevent substantial change (i.e., network density; Lydon-Staley et al., 2019; Pe et al., 2015). The standard VAR(1) model may be described as follows:

$$\boldsymbol{\eta}_t = \boldsymbol{\mu} + \Phi_1^* \boldsymbol{\eta}_{t-1} + \boldsymbol{\zeta}_t^* \tag{1}$$

where $\boldsymbol{\eta}_t$ is a p-variate vector of latent variables measured at t, $\boldsymbol{\mu}$ is a p-variate vector of latent intercepts, Φ_1^* is the lag-1 regression coefficients matrix or transition matrix which relates variables at the previous time-point to the current time point, and $\boldsymbol{\zeta}_t$ is a p-variate residual vector that is assumed to be multivariate normally distributed with zero means and covariance matrix, Ψ^*.[1] Notably, the $\{^*\}$ superscript serves as an indicator to differentiate the standard, discrete-time VAR from alternative representations introduced in the following sections.

The discrete-time GIMME procedure uses an extension of the standard VAR introduced in Equation (1) known as the structural VAR (SVAR; Lütkepohl, 2005). Formally, we may express the SVAR as:

$$\boldsymbol{\eta}_t = \boldsymbol{\mu} + \Phi_0 \boldsymbol{\eta}_t + \Phi_1 \boldsymbol{\eta}_{t-1} + \boldsymbol{\zeta}_t \tag{2}$$

where $\boldsymbol{\eta}$ is a p-variate vector of latent variables measured at t, $\boldsymbol{\mu}$ is a p-variate vector of latent intercepts, Φ_0 is the lag-0 regression coefficients matrix whose diagonal elements are fixed to 0, Φ_1 is the lag-1 regression coefficients matrix or transition matrix which relates variables at the previous time-point to the current time point, and $\boldsymbol{\zeta}_t$ is a p-variate residual vector that is assumed to be multivariate normally distributed with zero means and a diagonal covariance matrix, Ψ.

The SVAR differs from the standard VAR model via the inclusion of the contemporaneous effects captured by, Φ_0. These contemporaneous relations capture effects that occurred in the time between two successive measurement occasions when all higher-order lags have been accounted for (Lütkepohl, 2005). For example, momentary fluctuations in emotions at the daily level may indicate that a substantial number of associations exist within the Φ_0 matrix. This indicates that relations between emotions fluctuate at faster time scales than those captured by day-to-day assessments. The modeling of contemporaneous relations has been beneficial in allowing researchers to model processes that unfold faster than the observed time differences between successive occasions present in their data (e.g., Δ_t; Gates & Molenaar, 2012; Wright et al., 2019).

The GIMME Algorithm

The procedure for GIMME defines a null model for all individuals in a sample where the diagonal elements of the Φ_1^* matrix (i.e., the AR coefficients) is freed for estimation alongside the diagonal elements of the residual covariance matrix, Ψ^* (Gates et al., 2010; Gates & Molenaar, 2012). Following this, the modification indices (MIs; Sörbom, 1989) for all individuals are assessed and the parameter which would improve model fit for a user-specified proportion of individuals is added to all individuals. This procedure is repeated iteratively until no parameters can be added to improve the model fit for all individuals in the sample. The models then proceed with individual model fitting via the same iterative procedure where their baseline model is the group model derived in the previous step. The procedure stops when no paths improve model fit or when model fit criteria are satisfied (e.g., the Root Mean Squared Error of Approximation or RMSEA; Gates & Molenaar, 2012). Once paths can no longer be added, paths that are no longer statistically significant are pruned from the models.

GIMME has been extensively applied to the modeling of functional connectivity in neuroimaging research (Gates & Molenaar, 2012; Henry et al., 2019) where processes may unfold many times faster than our ability to sample. Further, GIMME's pooling of the modification indices of many subjects has been shown to improve its ability to recover individual signal from noise relative to person-specific modeling of individual time series (e.g., Gates & Molenaar, 2012; Lane et al., 2019; Park et al., 2023). These benefits withstanding, some limitations may still affect the performance of GIMME. Being formulated in the discrete-

[1] An alternative, arguably more popular form of the VAR model with non-zero intercepts is structured such that the current deviation from the mean relates to previous deviation from the mean as an AR process of lag 1 (i.e., $\boldsymbol{\eta}_t - \boldsymbol{\mu} = \boldsymbol{\Phi}_1^*(\boldsymbol{\eta}_{t-1} - \boldsymbol{\mu}) + \zeta_t^*$ (e.g., Li et al., 2022; Lütkepohl, 2005)

time framework and based on the VAR framework, GIMME requires that data be equidistant for parameter estimates to be accurate and the SVAR parameterization may fail to capture group- and subgroup-level dynamic patterns that exist within the data (Park et al., 2024). Further, changes in the sampling frequency of the data may impact the coefficients and their comparability across studies; that is, the parameters obtained from a VAR model are temporally dependent (Lütkepohl, 2005; Ryan & Hamaker, 2022). The assumption of equally-spaced data-points can be controlled for by careful planning of the study design. However, failure to satisfy this condition can result in the VAR coefficients being a blend of the multiple lags that were contained in the sampling intervals (Ryan & Hamaker, 2022). Further, the comparability of VAR estimates may be compromised if studies use wildly different sampling intervals for the same underlying processes (Gollob & Reichardt, 1987; Voelkle et al., 2012).

The following section introduces the novel contribution of the current work: `ct-gimme`. We provide a description of the model fitting in continuous time and relate it to fitting in discrete-time as described above. We then describe in detail the algorithm of `ct-gimme` and additional considerations that separate it from its discrete-time counterpart and provide motivation for our simulation and empirical illustrations.

Continuous-Time GIMME

Modeling in Continuous-Time

An alternative approach for describing processes through time is by explicitly modeling their change via continuous-time modeling. This framework is powerful in relating changes among variables explicitly to changes in their predefined time scales (Arminger, 1986; Boker & Graham, 1998) and is ideal for our implementation of `ct-gimme`. These models can tell us how our current state of depression may be associated with quick or slow changes in our resulting anxiety. Continuous-time models have already been used in the literature on psychological disorders such as depression and antisocial behavior (e.g., Delsing & Oud, 2008). Continuous processes such as these may be described as a set of differential equations that describe how changes in one variable may affect changes in another variable. These relations can be expressed in the following general form:

$$\mathrm{d}\boldsymbol{\eta}(t) = [\mathbf{b} + \mathbf{A}\boldsymbol{\eta}(t)]\mathrm{d}t + \mathbf{G}\mathrm{d}\mathbf{W}(t) \tag{3}$$

where $\mathrm{d}\boldsymbol{\eta}$ is a p-variate set of differentials in the latent variables, $\boldsymbol{\eta}(t)$, $\mathbf{A}$ is a $p \times p$ drift matrix which describes how changes in the values of η at time relate to itself and other variables, $\mathbf{b}$ is a vector of intercepts or home-bases, $\mathbf{W}(t)$ is a vector of process noises (specifically, standard Wiener processes) whose changes between any two-time points, $\mathrm{d}\mathbf{W}(t)$, are assumed to be normally distributed with zero means and variance-covariances that depend on $\mathbf{G}$ and the time interval between two time points, Δt and $\mathbf{G}\mathbf{G}' = \mathbf{Q}$ which represents the diffusion matrix (Arnold, 1974; Voelkle et al., 2012).

Analytic solutions have been proposed for Equation (3) to delineate $\boldsymbol{\eta}$ for any real or integer values of time, including discrete-time solutions that relate $\boldsymbol{\eta}_t$ to $\boldsymbol{\eta}_{t-\Delta t}$ for which the time difference, Δt, can only assume integer values (e.g., Δt=1, 2, and so on; Voelkle et al., 2012; Chow et al., 2022; Hecht & Zitzmann, 2021a). These exact transformations are provided below and allow one to transform elements of the general SDE to the equivalent VAR forms presented in Equation (1). Given a specified Δt, the mapping between the discrete- and continuous-time formulations can be expressed as:

$$\boldsymbol{\Phi}^*(\Delta t) = e^{\mathbf{A}\Delta t} \tag{4}$$

$$\boldsymbol{\Psi}^*(\Delta t) = \mathrm{irow}\{\mathbf{A}_{\#}^{-1}[e^{\mathbf{A}_{\#}\Delta t} - \mathbf{I}]\mathrm{row}(\mathbf{Q})\} \tag{5}$$

where $\mathbf{A}_{\#} = \mathbf{A} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{A}$; row(.) is a row operator which transforms a matrix row-wise into a column vector and irow(.) is the reverse operation of turning a column vector into a matrix, $\mathbf{Q}$ is the continuous-time error covariance matrix also known as the diffusion matrix (Arnold, 1974; Voelkle et al., 2012). Of note is that the sampling interval becomes encoded into both the Φ^* and Ψ^* matrices at their given lag-orders. Thus showing how the parameters of the standard VAR are dependent on the sampling interval of the data. We provide an illustrative figure in Figure 1, which highlights how parameters from the continuous-time model transform to those in the discrete-time VAR model at $\Delta t = 1$ or 10.

The limitations of the discrete-time VAR models are addressed by fitting models in the continuous-time framework (Chow et al., 2022; Ryan & Hamaker, 2022). Continuous-time models can handle irregularly spaced data and are not sensitive to the sampling rate of the data because of their approximation of the derivatives of the underlying processes (Gollob & Reichardt, 1987; Ryan & Hamaker, 2022). While a growing body of work has been established to ease researchers into interpreting continuous-time models (Van Montfort et al., 2018), relatively less has work has been done to simplify entry into continuous-time modeling. Likewise, beyond a handful of methods, relatively few avenues exist for identifying group-level structures in individual-level processes in the continuous-time framework. Those that do typically fall within multilevel frameworks that impose strict assumptions on the structure of individual-level models (e.g., Boker & Graham, 1998; Nestler & Humberg, 2021).

The `ct-Gimme` Algorithm

The general procedure for fitting `ct-gimme` is presented in Algorithm 1. Here, we discuss the algorithm of `ct-gimme` for unsupervised model construction in continuous time via modification indices. The use of modification indices has a rich history in both Structural Equation Modeling as well as the discrete-time GIMME framework (Gates & Molenaar, 2012; Whittaker, 2012). `ct-gimme` adopts the use of modification indices to stay in line with the precedent

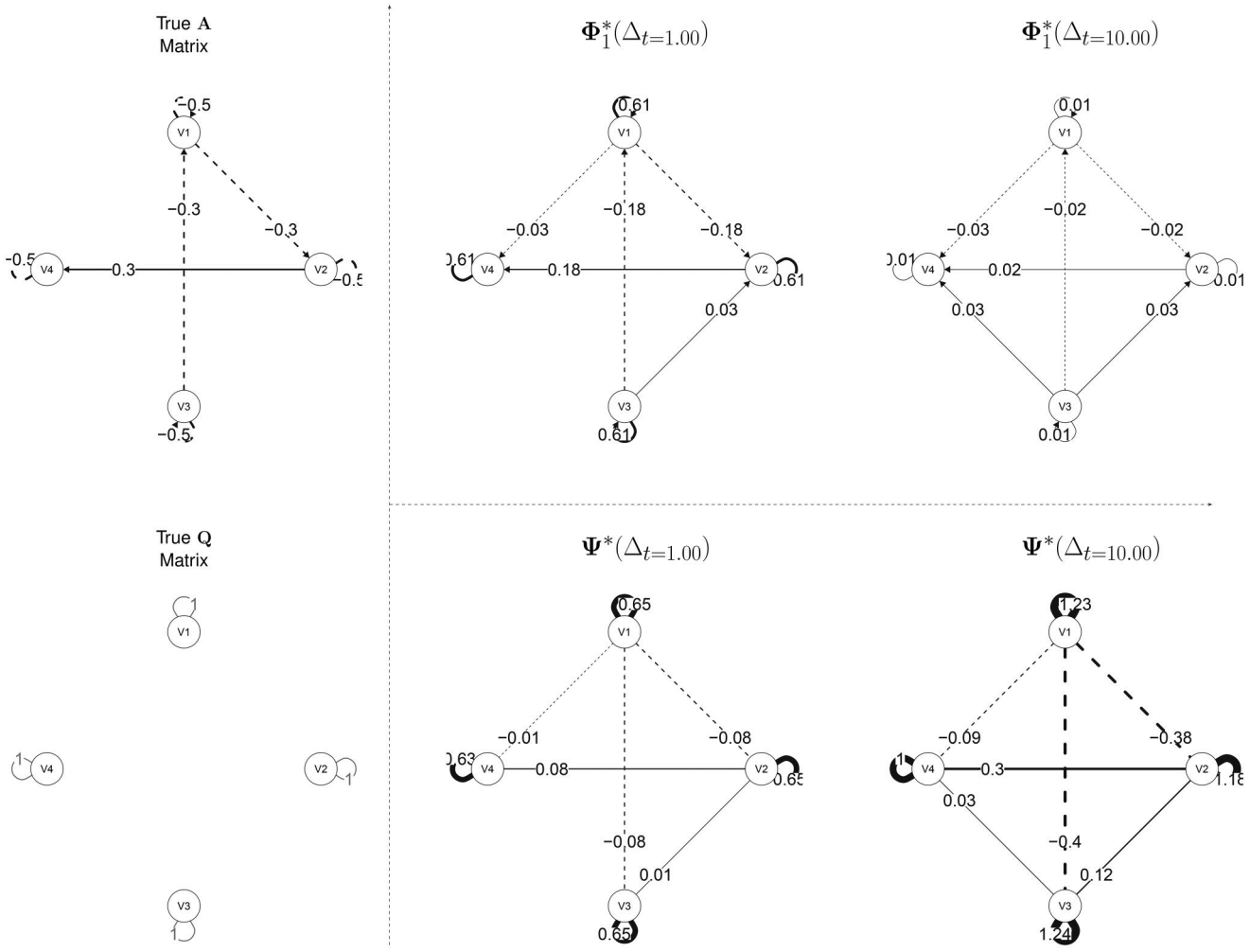

Figure 1. Relating the true CT VAR (**A** and **Q**) to the DT VAR at a $\Delta_t = 1$ ($\mathbf{\Phi}_1^*(\Delta_{t=1})$ and $\mathbf{\Psi}^*(\Delta_{t=1})$) and $\Delta_t = 10$ ($\mathbf{\Phi}_1^*(\Delta_{t=10})$ and $\mathbf{\Psi}^*(\Delta_{t=10})$). Solid lines indicate positive coefficients while dashed lines indicate negative coefficients. Notably, paths in **A** become weaker as Δ_t increases but manifest as covariances in the $\mathbf{\Psi}^*$ matrices.

set by the discrete-time GIMME which has shown reliable performance in detecting effects and controlling for error inflation (Lane et al., 2019). Here, we focus on the specifics of `ct-gimme` which draws heavily from its discrete-time counterpart: GIMME (Gates & Molenaar, 2012). `ct-gimme` is fit in the state-space framework which can be described as a set of state and space equations. The state equation, as shown in Equation (3), describes the dynamics of the system of interest in continuous time. In addition to the state equations are the measurement equations which are described by:

$$\mathbf{y}_{t,k} = \mathbf{C}\boldsymbol{\eta}_{t,k} + \mathbf{r}_{t,k} \tag{6}$$

where $\mathbf{y}_{t,k}$ is q-variate vector of observed data of the k^{th} measurement occasion at time, t, $\mathbf{C}$ is a $q \times p$ factor loadings matrix which relates latent variables to observed variables and is analogous to a factor loadings matrix in the structural equation modeling framework, $\boldsymbol{\eta}_{t,k}$ is the p-variate vector of latent variables, and $\mathbf{r}_{t,k}$ is a q-variate measurement errors with covariance matrix, $\mathbf{R}$. In the implementation for `ct-gimme`, we assume by default that single variables directly correspond to single factors. In such a case, $\mathbf{C}$ reduces to an identity matrix and may be omitted for parsimony.

A user inputs their pre-processed N, p-variate time-series and `ct-gimme` defines a null state-space model to begin optimization. By default, the diagonal elements of the drift, **A**, and process noise, **Q** matrices are freed for estimation whereas all off-diagonal elements are fixed to 0.00. The factor loadings, **C**, are assumed to be diagonal with one item loading onto each factor; though, this may be changed at the user's preference to form multi-indicator latent factors. Users may specify that the measurement error variances in **R** be freely estimated by using the argument `ME = TRUE`, with starting values specified via the argument `rvals` to be a vector of p-values, or a single value to be set for all variables. By default, this is disabled and measurement error variances are assumed to be near 0.00 (i.e., $1e^{-5}$) but can be enabled to capture measurement error variances. Finally, non-informative initial values are set for the latent variables, with zero means and an identity matrix as their covariance matrix, but users are encouraged to input their own values whenever possible.

When executed, `ct-gimme` fits the null model specified above to all N subjects (see Algorithm 1 for full details). Then, modification indices are extracted for all subjects to identify the parameter which would improve the model fit for a majority of individuals. The discrete-time GIMME algorithm used a pre-defined value of 75.0% (Gates & Molenaar, 2012). Thus, if a model parameter would improve

the model fit for at least 75.0% of the sample, then the model parameter would be added for all individuals. In the case of `ct-gimme`, the program default is set to 51%. This lower setting is to err on the side of discovery of group-level structure but can be changed by user specification. Of all parameters that pass this "majority criterion" within an iteration, the parameter with the largest average modification index across all subjects will first be freed up for all subjects in the sample, and model fitting will begin again. This procedure repeats iteratively at the group-level until one of the stopping criterion is met: 1. The dynamic matrix, **A**, is full or 2. The largest modification index is no longer statistically significant for a majority of individuals (suggesting e.g., that 51% of the sample does not share a path).

Error Inflation

The first stopping criterion for `ct-gimme` is straight forward. A full drift matrix is saturated and no further parameters can be added. However, such iterative model adaptations through repeated use can lead to inflation in Type-I error rates. The discrete-time GIMME has implemented corrections for this exact issue by penalizing the α-level by the sample size and implementing stopping criteria based on model fit (e.g., NNFI, CFI, and RMSEA; Lane et al., 2024). However, certain metrics such as RMSEA are not straightforward to incorporate in the continuous-time framework due to the need to define a saturated model for comparison (Oh et al., 2024). To this end, we propose and evaluate two avenues for controlling α-inflation in addition to serving as feasible stopping criteria for the `ct-gimme` algorithm: 1. An adjusted form of the Benjamini-Hochberg correction and 2. Model selection via sample-size adjusted Bayesian information criterion (BIC) to reflect the original corrections of discrete-time GIMME via direct α correction as well as model-based selection via information criterion. As these implementations for error inflation are new to the GIMME framework in both discrete- and continuous-time, these conditions are likewise tested and evaluated in our simulation studies in the sections that follow.

Benjamini–Hochberg Correction. The Benjamini–Hochberg correction for multiple comparisons is a well-established and evaluated method for adjusting the family-wise error rate (Benjamini & Hochberg, 1995, 1997) and have been shown to be relatively easy to implement and more powerful than alternatives such as the Bonferroni method (Thissen et al., 2002). Simply, the adjusted Benjamini-Hochberg correction is applied to the α-levels of each test of the modification indices by becoming progressively more conservative with each additional parameter (Benjamini & Hochberg, 1997, 1995). Formally, the correction can be calculated as:

$$\frac{i}{m}\alpha \tag{7}$$

where α is the pre-selected α-level, i is the ranking of a given p-value in descending order, and m is the total number of tests being executed. Since `ct-gimme` operates in a model construction framework, the total number of tests cannot be known ahead of time. Thus, we constrain `ct-gimme`'s value of m to be the number of freeable parameters in the dynamic matrix. Thus, a 5-variate model with the diagonal elements fixed to be estimated would result in $m = 20$ total possible tests. Likewise, the ranking of each test is taken in sequence. In `ct-gimme`, the Benjamini-Hochberg correction simply replaces the p-value of each successive MI until statistical significance is not achieved and the resulting model is retained. Once the individual-model fitting state begins, each subjects' models are estimated iteratively but pick up from the last Benjamini-Hochberg adjusted p-value from the group fitting stage. In theory, this would result in lower Type-I error rates with a possible decrease in power as well (i.e., higher Type-II error rates). However, when compared to the discrete-time GIMME algorithm's penalty by sample size, the Benjamini-Hochberg correction may be less conservative.

Sample-Size Adjusted BIC. In the absence of fit measures for continuous-time models such as RMSEA, we chose to make use of the sample-size adjusted BIC (BIC_n; Sclove, 1987). Information criterion (IC) based model selection is a large field of study and several potential IC measures could have served as potential candidates for evaluating model performance in `ct-gimme`. The decision to make use of IC-based measures was influenced by work indicating their strength over similar model fit indices such as the CFI and TLI in the structural equation modeling framework (e.g., Bollen et al., 2014). Likewise, the BIC_n has been shown to outperform other IC-based measures that were readily available such as the AIC (Tein et al., 2013). The sample size adjusted BIC may be expressed as:

$$BIC_n = k \ln\left(\frac{n+2}{24}\right) - 2\ln(\hat{L}) \tag{8}$$

where k indicates the number of parameters in the model, n is the sample-size, and $\hat{L}$ represents the maximized value of the likelihood function at the estimated parameters. In the case of `ct-gimme`'s implementation in `OpenMx`, this is obtained via prediction error decomposition by means of a hybrid Kalman filter approach (Boker et al., 2011; Schweppe, 1965). Models with lower BIC_n's are preferred and–intuitively–this selects models which exhibit the best fit to the data scaled by the number of parameters in the model and the sample size. The implementation of BIC_n in `ct-gimme` tracks the BIC_n for each subject at each parameter. Once the best candidate MI is not statistically significant at the $\alpha = 0.05$-level, 3-additional parameters are added based on the highest average MI value across the sample. Once these additional 3-parameters are added, the model-state that corresponded to the lowest average BIC_n across the sample is selected based on a Convex Hull (CHULL; Ceulemans & Kiers, 2006) where:

$$st_k = \frac{BIC_{n,(K-1)} - BIC_{n,(K)}}{BIC_{n,(K)} - BIC_{n,(K+1)}} \tag{9}$$

where st_k indicates the model stage at the k^{th} additional parameter which is calculated by the ratio of the differences in

BIC_n between the current $BIC_{n,(K)}$ and prior, $BIC_{n,(K-1)}$, and current versus future, $BIC_{n,(K+1)}$, model states. This procedure was chosen over simply selecting the lowest possible BIC_n to reduce the possibility of over-fitting due to sampling variability, and sensitivity to individual outliers. Thus, in theory, this procedure would select the earliest model that maximized the improvement to BIC_n. Once this procedure has been performed at the group level, it is repeated at the individual level using the person-specific BIC_n, instead of the average BIC_n. Both the Benjamini-Hochberg and BIC_n corrections can be enabled via standard arguments in `ct-gimme`.

Algorithm 1 `CT-GIMME`

Input: $\mathbf{y}_{t,k}$, for $i = 1, ..., N$.
Output: $\widehat{\mathbf{A}}_i$, $\widehat{\mathbf{Q}}_i$, for $i = 1, ..., N$.

Stopping criterion satisfied by meeting or or or :
Dynamics matrix, $\mathbf{A}_i$ is full
Largest Modification Index is not statistically significant for a majority of individuals in the sample
If `check.BIC = TRUE`, run 3-extra iterations following largest MI and select model corresponding to best BIC relative to model complexity during model search procedure
If `Benjamini.Hochberg = TRUE`, penalize p-value scaled by number of parameters added using Equation (7)

for $i = 1, ..., N$ **do**
Initialize continuous-time model with diag($\mathbf{A}_i$) and diag ($\mathbf{Q}_i$) free
while *stopping criterion not met* **do**
- Fit continuous-time model to all, N individuals
- Obtain MIs
- If in group-stage, check parameter corresponding to maximum *average* MI in raw value is statistically significant at $\chi^2(1)$ for a majority of individuals
- If in individual-stage, check parameter corresponding to maximum MI in raw value is statistically significant at $\chi^2(1)$
- If so, free element of the drift matrix, $\mathbf{A}_i$

end
Prune non-significant paths
Begin do loop for each individual with `NULL` model set as the group-level model
end

Methods

Simulation Study

Throughout the design of `ct-gimme`, several aspects of the algorithm had to be adjusted from the original discrete-time GIMME. This led to research questions regarding the performance of `ct-gimme` as well as how it may perform against alternate approaches for fitting models in either discrete- or continuous-time. We begin this section via an explicit description of the research questions which motivate our simulation study.

Motivating Research Questions

Our motivating research questions (RQs) focused on evaluation of the performance of `ct-gimme` under ideal and misspecified conditions, and comparisons of `ct-gimme` against benchmarks in discrete and continuous time. The first research question (RQ1) was to compare the different model-fitting options for `ct-gimme` when modeling assumptions are met: standard MI, Benjamini-Hochberg corrected MI, or BIC_n selected models across conditions of sample size, effect size, and within-sample heterogeneity conditions. Upon determination of the best performing version of `ct-gimme`, all subsequent RQs only relied on that selected criterion. The expectation of RQ1 was that both the Benjamini-Hochberg and BIC_n corrections would outperform the raw modification index; however, their performance against one another in deriving acceptable models was less clear and of interest for the current investigation.

The second research question (RQ2) centered around the strengths and weaknesses of continuous-time modeling with `ct-gimme` under model misspecification. We were interested in the performance of `ct-gimme` in conditions where the continuous-time model is either 1. provided with incorrectly labeled time-intervals which are assumed equidistant and 2. incorrectly neglect modeling of measurement errors which may impact the quality of resulting parameter estimates. Broadly, RQs 1 and 2—broadly—assessed the general performance of `ct-gimme` and validate its characteristics in a vacuum.

The third research question (RQ3) introduced comparisons of `ct-gimme` to alternative approaches in continuous-time that empirical researchers may consider when fitting models to real-world data. Namely, 1. strictly $N = 1$ modeling and 2. strictly group-based modeling. We evaluated the performance of `ct-gimme` against these two alternative approaches for model fitting in continuous time across conditions of sample size, effect size, and within-sample heterogeneity conditions. It was expected that the strictly group-based models would outperform `ct-gimme` when samples were homogeneous but perform worse when individual differences in dynamic network structures were present but unaccounted for. Likewise, we hypothesized that `ct-gimme` would outperform strictly $N = 1$ model fitting when some groups structures were present due to `ct-gimme`'s ability to pool information at the sample-level prior to individual model fitting.

Finally, RQ4 evaluated `ct-gimme` to discrete-time GIMME. One may expect that `ct-gimme` would outperform discrete-time GIMME in instances where the measurements were not equally spaced. However, a unique characteristic of the original GIMME algorithm is that it is formulated in the structural VAR which enables it to model "faster" dynamics via the contemporaneous effects. Thus, the degree to which `ct-gimme` might outperform GIMME–if at all–was of interest for the current investigation.

Simulation Design

Data were simulated from the Ornstein-Uhlenbeck (OU) model which is a special case of Equation (3) (Oravecz & Tuerlinckx, 2011). The OU model can be thought of as the continuous-time analogue for a discrete-time VAR model with some differences in parameterization. Where $\mathbf{A} = -\mathbf{B}$ indicating the drift matrix and $\mathbf{b} = \mathbf{B}\boldsymbol{\mu}$ where $\boldsymbol{\mu}$ is a mean-vector. For the current work, the means are assumed to be a vector of 0's. Manipulations of the within-sample heterogeneity in dynamics will alter $\mathbf{B}$. Specifically, all subjects shared a common dynamic model when heterogeneity was not present. When heterogeneity was present, all subjects shared a common set of connections in the off-diagonal elements of $\mathbf{B}$ comprised of 15% of the total available off-diagonal elements. In conditions where within-sample heterogeneity was enabled, an additional 10% of off-diagonal paths in $\mathbf{B}$ were randomly added to each participant. Depending on the effect size condition, these cross-process relations may take values of either 0.30 or 0.90.

Data were simulated at a temporal resolution of $t = 0.1$. When Δt was manipulated, the raw time-series were subsampled at the corresponding time intervals (e.g., $\Delta t = 0.50$ or 1.0). When irregularly sampling time-points, time-points were sampled from the following scheme: $t_i = \sum_{j=1}^{i}(1 + \Delta t_k)$ where Δt_k was assumed to follow a gamma distribution with a shape parameter of 2, and a scale parameter of 0.5. That is, each subsequent measurement occasion was drawn such that it was at least 1-unit in time later plus a randomly drawn value from a gamma distribution that was heavily concentrated around the value of 1.0. This yielded an irregularly spaced sampling structure that was both shifted in time as well as random in nature with an average $\Delta t = 2.0$ but highly skewed. The resulting N-subject time series was then fitted to the different modeling frameworks.

For `ct-gimme`, models were fit in `R`. Results from using different model selection measures were compared for RQ1 (i.e., under raw MI, Benjamini-Hochberg correction, and BIC_n). Following this, the approach that yielded the best overall performance were used to test all subsequent RQs. Unless otherwise specified, time-indices were labeled correctly with diagonal elements of the drift and process noise variances freed for initial estimation with measurement errors freed when the conditions dictated they be freed.

For the strict $N = 1$ models, continuous-time state-space models were fitted to each subject with all drift parameters freed for estimation alongside the diagonal elements of the process noises. Once the full model was fitted, non-significant parameters were removed from the models and refitted to reflect a naive approach to fitting these models. For the strict group-level models, a multi-group model was specified in `OpenMx` where a common, dynamic model was fitted to all subjects and parameters in the full drift matrices and diagonal elements of the process noise matrices were constrained to equality. Once a group-level model was fitted to the data, the group-level model was pruned and refitted. For all other arguments, model configurations were identical to `ct-gimme`. Finally, when discrete-time GIMME implemented using the `gimme` package in `R` using program defaults (Lane et al., 2024) assuming that the time intervals were equally spaced. All models across all conditions were tested across 500 Monte Carlo replications per condition.

Simulation Conditions

All simulation metrics and their broad rationale are depicted in Table 2 by their corresponding research question. Below, we detail the specifics of these conditions and relate them to our three motivating questions.

The first two parameters in the simulation study which were not manipulated or adjusted: sample sizes and network sizes. The rationale for fixing these two were because the effects of manipulating these variables are well established in the literature (e.g., Park et al., 2020, 2023). Notably, manipulations of sample size would be more apparent when taking into account subgroups of individuals within the broader sample (e.g., Park et al., 2024). However the current investigation does not study subgrouping thus the impacts were deemed minimal. Likewise, the network size was not manipulated for similar reasons. In particular, as the number of estimable parameters increases, the sample sizes required to sufficiently estimate them generally increases as well; however, given our manipulation of proportions of

Table 2. Research questions crossed with simulation factors and rationales.

	Model	Time	Effect Size	Sampling Rate	Heterogeneity	Sampling Regularity	Measurement Error
RQ1	`ct-gimme` only with: raw MI, Benjamini-Hochberg, and BIC_n	$T = 50$ & 100	$EF = 0.30$ & 0.90	$\Delta t = 0.50$ & 1.0	No & Yes	Regular	None
RQ2	`ct-gimme` with BIC_n	$T = 100$	$EF = 0.30$	$\Delta t = 1.0$	No	$t_i = \sum_{j=1}^{i}(1 + X_j)$ where $X_j \sim \gamma(2, 0.5)$	$\kappa \approx 0.80$
RQ3	`ct-gimme` with BIC_n model selection, $N = 1$, and group-level	$T = 50$ & 100	$EF = 0.30$ & 0.90	$\Delta t = 0.50$ & 1.0	No & Yes	Regular	None
RQ4	`ct-gimme` with BIC_n and GIMME	$T = 100$	$EF = 0.30$	$\Delta t = 0.50$ & 1.0	No	$t_i = \sum_{j=1}^{i}(1 + X_j)$ where $X_j \sim \gamma(2, 0.5)$	None

Note: Conditions are enabled during each research question. T = number of time-points per individual, EF = magnitude of cross-process dynamics, Δt = sampling rate of true time-series, Heterogeneity can be No/Yes indicating whether individual differences were simulated, Sampling Regularity denotes whether data were sampled regularly or irregularly and by what mechanism, Measurement Error denotes if measurement error variances were simulated and the average item-level reliability of the simulated data.

effects, we expected our design to generalize to larger networks with similar proportions of zero- to non-zero paths.

For RQ1, *Time*, *Effect Size*, *Sampling Rate*, and *Heterogeneity* conditions were manipulated and crossed with one another to broadly assess the performance of the different corrections for `ct-gimme` when undergoing model modification via raw modification indices, the Benjamini-Hochberg correction, and BIC_n. By crossing these factors, we were able to assess the performance of the three different approaches for handling α-inflation in the model modification search in different configurations of sample-size, effect size, and heterogeneity across sampling rates. In particular, sample sizes ranged from 50 to 100 time points, reflecting the sample sizes typically seen in empirical applications involving time series data (e.g., Fisher et al., 2017; Lane et al., 2019; Wright et al., 2019). Effect sizes were selected to span from nigh undetectable to nearly explosive by means of varying the weights of the cross-process dynamics while keeping the centralizing tendencies fixed at −0.50. Sampling rates differed from $\Delta t = 0.50$ to 1.00 to mirror conditions with faster sampling faster than a unit time interval. Finally, heterogeneity was simulated at 0% to 10% of cross-process paths being randomly added to individual models to simulate cases where individual variation was prominent in the dataset.

RQ2 took the best performing argument from RQ1 and defined targeted simulations to test how well `ct-gimme` performed under conditions of model misspecification via irregularly spaced time-series with [in]correctly labeled time-series as well as in conditions where measurement errors are present in the data and modeled or ignored. The additional factor of regular spacing was sampled from a γ distribution as noted in the prior section as well as in Table 2. This was to simulate a condition where simulated participants regularly responded both late (i.e., $\overline{\Delta t} = 2.0$) as well as with a non-normally distributed range. Likewise, the inclusion of the measurement error variances were meant to simulate cases where the true, data-generating process exhibits measurement noise that is left unmodeled to assess the effects of those measurement error variances on the estimated dynamics of the model.

For RQ3, `ct-gimme` was compared to $N = 1$ and group-level modeling across conditions of *Time*, *Effect Size*, Δt, and *Heterogeneity* with regular sampling and no measurement errors. These conditions largely matched the broad tests of `ct-gimme` in RQ1 but compared the performance of `ct-gimme` with BIC_n model selection to the $N = 1$ and group-level conditions to evaluate the performance of `ct-gimme` relative to reasonable alternative approaches for fitting models in continuous-time.

Finally, for RQ4 `ct-gimme` with Benjamini-Hochberg correction was compared to the discrete-time GIMME algorithm in a targeted set of conditions in the presence of irregularly spaced data and different sampling rates to highlight the strengths of `ct-gimme`.

With the goals of essentially comparing various model search procedures in continuous- and discrete-time, there are some metrics that should be considered when concluding whether one approach performs "better" or worse than others. In what follows, we describe how the data were generated for the simulations then, describe the performance measures of the simulation study and how they relate to my broad research questions and goals.

Performance Measures

The performance measures for our simulations are presented below and are related to specific research questions when appropriate. When comparing the discrete-time GIMME algorithm to `ct-gimme`, we were most interested in the performance and recovery of the point-estimates. Thus, we assessed measures of Type-I and Type-II error rates, bias, and the root mean squared errors (RMSEs). Type-I error rates were calculated as:

$$\alpha_\theta = \frac{FP_\theta}{TN_\theta + FP_\theta} \tag{10}$$

where θ is the parameter estimate, FP_θ is the number of false positives and TN_θ is the number of true negatives. Type-II error rates were calculated as:

$$\beta_\theta = \frac{FN_\theta}{FN_\theta + TP_\theta} \tag{11}$$

where FN_θ is the number of false negatives TP_θ is the number of true positives.

Absolute biases were be used to measure the degree to which parameter estimates deviated from their true values and were derived as:

$$\text{ABias}_\theta = \frac{1}{H}\sum_{h=1}^{H}\left|\hat{\theta}_h - \theta\right| \tag{12}$$

where H is the total number of Monte Carlo runs, $\hat{\theta}_h$ is the estimated parameter estimate in the h^{th} Monte Carlo run, and θ is the true parameter estimate. Absolute biases were selected over standard biases to avoid relatively minor biases that averaged to zero during the simulations.

Similarly, the variance of the parameter estimates needs to be accounted for. Thus, RMSEs of the parameter estimates were calculated and are given by:

$$RMSE_\theta = \sqrt{\frac{1}{H}\sum_{h=1}^{H}(\hat{\theta}_h - \theta)^2} \tag{13}$$

Taken together, these performance measures provided insight as to how discrete-time GIMME performed relative to `ct-gimme`. To evaluate the quality of the standard error estimates, we focused on the `ct-gimme`, $N = 1$ modeling, and the multigroup-method, all of which returned standard error estimates associated with the continuous-time parameters. Standard error estimates returned by other discrete-time approaches required additional transformations to the continuous-time metrics and were not considered in this article. We examined the mean of the standard error estimates for each parameter as:

$$\overline{SE}_\theta = \frac{1}{H}\sum_{h=1}^{H}\widehat{SE}_\theta \tag{14}$$

where $\overline{SE}_\theta$ is the average of the standard error estimates across the Monte Carlo runs. We also computed the standard deviation of the standard error estimates as:

$$SD(SE)_\theta = \sqrt{\frac{1}{H-1}\sum_{h=1}^{H}\left(\widehat{SE}_\theta^{(h)} - \overline{SE}_\theta\right)^2} \quad (15)$$

where $\widehat{SE}_\theta^{(h)}$ is the estimated standard error for a given parameter, θ, in the h^{th} Monte Carlo run. Together, these $\overline{SE}_\theta$ and the $SD(SE)_\theta$ summarized the quality of the standard errors in terms of their size and spread across Monte Carlo runs.

Results

The Results of the Four Research Questions Are Discussed in Detail Using Results from the Monte Carlo Simulations

RQ1: Which Model Selection Method for `ct-Gimme` Performs Best?

We focused on elaborating results pertaining to Type-I and Type-II error rates because the choice of correction method had very little effects on other performance measures when the correctly specified continuous-time model was within the search spaces of `ct-gimme`.

Figure 2's left column displays the Type-I and II error rates from the low- and high-effect size conditions at $\Delta t = 0.50$ for $T = 50$ and $T = 100$. The right column highlights the same conditions only when $\Delta t = 1.0$. Generally, Type-I and II error rates were lower with the BIC_n correction in homogeneous samples with the Benjamini-Hochberg correction and raw MI performing comparably to one another (both ≈ 0.08 vs ≈ 0.065 for BIC_n). The Benjamini-Hochberg correction and raw MI performed comparably to one another in Type-I and II error rates and better than the BIC_n correction indicating that the BIC_n correction exhibited a bias towards model parsimony.

When samples contained heterogeneous dynamics and large effect sizes, the Benjamini-Hochberg outperformed the raw MI in error rates. These results were expected as the adjustments for α under the Benjamini-Hochberg approach were more salient toward the end of the search, especially in evaluating the statistical significance of parameters freed up in the individual models. This is highlighted in Figure 3 which displays the Type-I and II error rates in heterogeneous conditions when $T = 100$ and $\Delta t = 1.0$ across effect size conditions.

These results make sense given the BIC's tendency to prefer model parsimony as well as our use of a CHULL procedure which prefers the configuration of the model which best describes the data with the fewest parameters possible. Thus, the expectation is that we would have lower power due to the conservative bias induced by this procedure. This is similar to findings using applications of the CHULL procedure in other dynamic network modeling contexts to find subgroups where CHULL methods tended to be more conservative than alternative approaches (see, Park et al., 2024).

Due to these results, the conclusions for RQ1 are as follows: the BIC_n-based correction led to the lowest Type-I (lower even than the nominal rate of α) and Type-II error

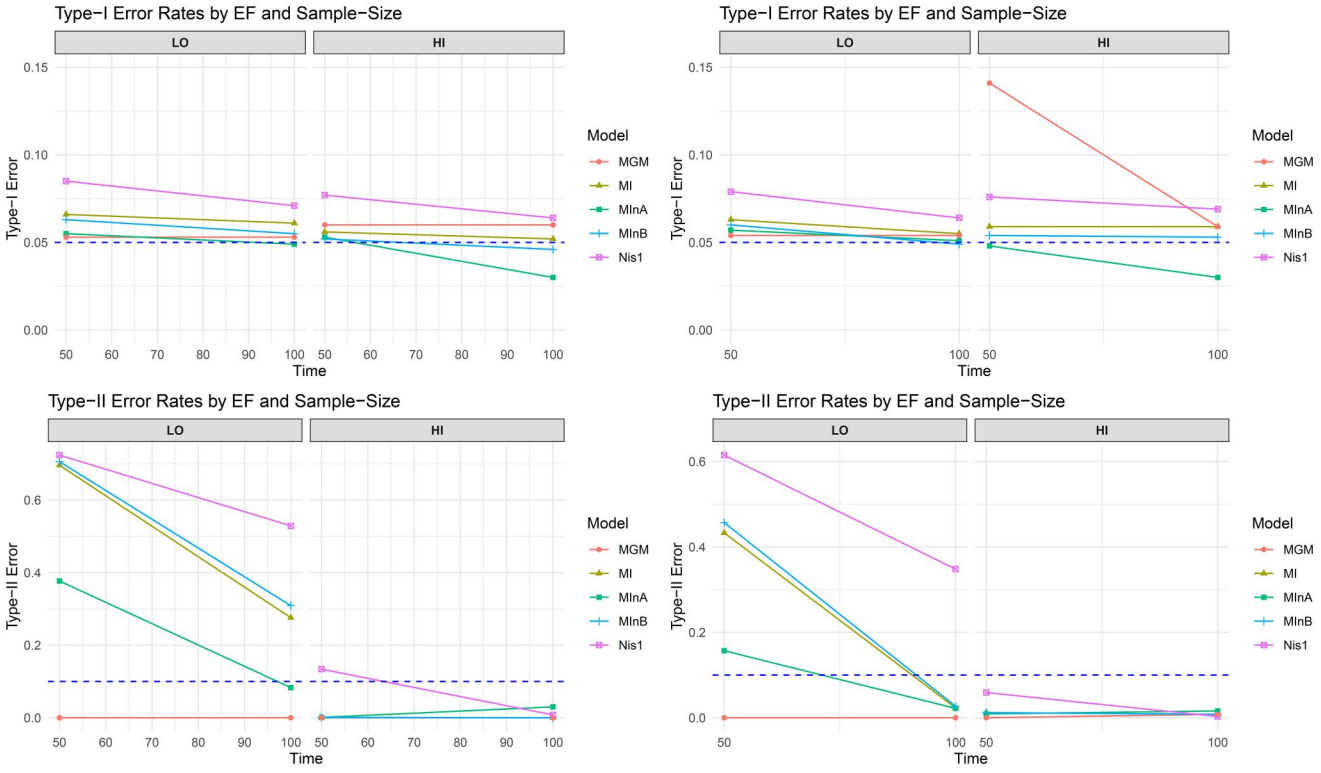

Figure 2. Type-I and II error rates across sample size ($T = 50$ and $T = 100$) when $\Delta t = 0.50$ (left column) and 1.00 (right column). MGM: Group-level model, Nis1: $N = 1$ modeling, MI: Raw MI `ct-gimme`, MInA: `ct-gimme` with BIC_n model selection, and MInB: `ct-gimme` with Benjamini-Hochberg correction. Generally, Type-I and II error rates tended to be stable across Δt when the sample-size was held constant with some minor differences across the columns. Notably, the multigroup method exhibited much higher Type-I error rates when $T = 50$ at a $\Delta t = 1.00$ compared to $\Delta t = 0.50$.

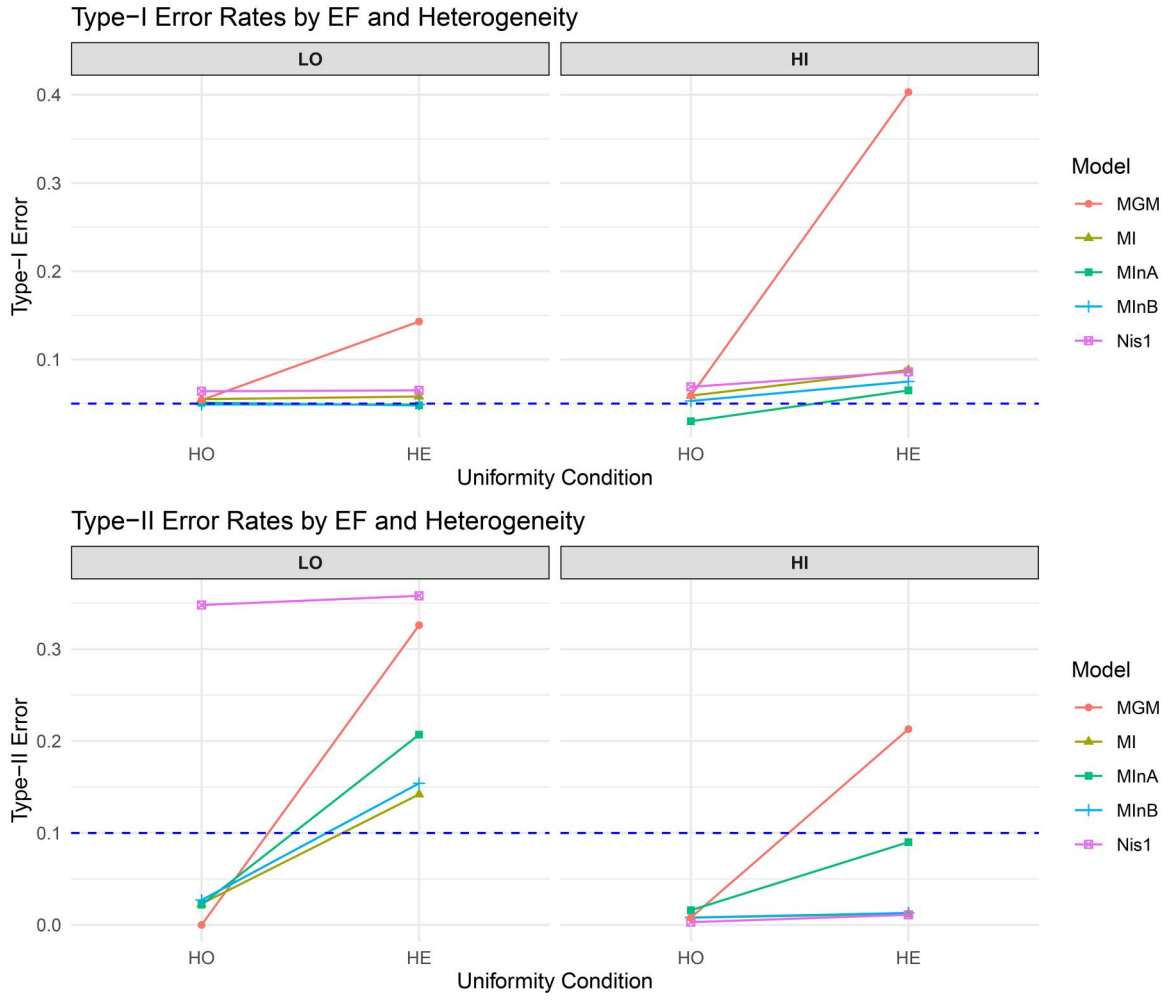

Figure 3. Type-I and II error rates across homogeneity (HO/HE) conditions crossed with effect size (EF = LO/HI) when $T = 100$ and $\Delta t = 1.0$. dashed line indicates the expected nominal rate for Type-I error and an acceptable threshold for Type-II error at 0.10 or 90% power. MGM: Group-level model, Nis1: $N = 1$ modeling, MI: Raw MI `ct-gimme`, MInA: `ct-gimme` with BIC_n model selection, and MInB: `ct-gimme` with Benjamini-Hochberg correction.

rates when samples were entirely homogeneous. In the presence of person-specific variations in dynamic network structure, it still exhibited remarkable Type-I error rates but tended to exhibit less statistical power for detecting person-specific dynamics. The Benjamini-Hochberg correction yielded Type-I error rates that were closer to the nominal level compared to raw MIs and showed higher power than the BIC_n correction in the presence of person-specific variations. Thus, the general recommendation would be to use BIC_n if parsimony is the key goal, or when groups are expected to be relatively homogeneous. In situations with person-specific variations, the Benjamini-Hochberg correction might be preferred. Both of these model selection approaches showed improvements over raw MIs. To address other remaining research questions (i.e., RQs 2-4) we used only the BIC_n as the model selection procedure in light of its lower Type-II error rates (i.e., higher power or sensitivity), especially under smaller T and effect size conditions.

RQ2: How Does `ct-Gimme` Perform in Misspecified Conditions of Time and Measurement?

`ct-gimme` handles these issues well when correctly specified with shortcomings when either measurement errors or sampling intervals are neglected or incorrect. Table 3 highlights conditions where `ct-gimme` was tested against itself under these two special conditions. First, in the presence of measurement error variances and second when sampling intervals were irregular and skewed.

RQ2 utilized $T = 100$ time points per subject with small effect sizes ($EF = 0.30$) for the cross-process dynamics and a fixed $\Delta t = 1.0$ under the measurement error condition. These configurations were identical in the case of the irregularly spaced condition, except that Δt was no longer fixed. Instead, Δt was either set at the correct person- and time-varying values, or ignored and labeled as equidistant with a $\Delta t = 1.0$. We present the results of RQ2 in two broad sections now.

`ct-gimme` had the option for measurement errors either enabled or disabled to effectively model the measurement errors variances simulated in the true data. Our simulations indicated that accounting for measurement errors led to substantial reductions in the absolute biases for the diagonal elements of the dynamics (e.g., $ABias_{\text{ME}} = 0.09$ vs $ABias_{\text{No ME}} = 0.23$). This reduction for the diagonal elements of the dynamics was reflected as well in the RMSEs (e.g., $RMSE_{\text{ME}} = 0.12$ vs $RMSE_{\text{No ME}} = 0.25$) and the standard errors (e.g., $SE_{\text{ME}} = 0.11$ vs $SE_{\text{No ME}} = 0.17$). While not

Table 3. Simulation results comparing measurement error and irregular sampling through time.

	Measurement Error		Irregularly Sampled		
	Modeled ME	Ignored ME	Correct Δt	Wrong Δt	GIMME
AR Bias	0.090	0.234	0.110 (0.063)	0.535 (0.235)	- (0.240)
CR Bias	0.014	0.015	0.025 (0.015)	0.065 (0.020)	- (0.027)
AR RMSE	0.116	0.247	0.154 (0.082)	0.914 (0.253)	- (0.266)
CR RMSE	0.041	0.043	0.073 (0.041)	0.206 (0.048)	- (0.066)
AR SE	0.111	0.165	0.236	0.686	–
CR SE	0.012	0.013	0.021	0.044	–
AR SD.SE	0.029	0.012	0.096	12.549	–
CR SD.SE	0.026	0.027	0.044	0.528	–
Type-I	0.048	0.045	0.049 (0.137)	0.049 (0.145)	- (0.058)
Type-II	0.001	0.000	0.009 (0.013)	0.007 (0.008)	- (0.344)

Note: AR - indicates diagonal elements of the dynamics while CR indicates off-diagonal values. RMSE is the root mean squared error, SE is the standard error, SD.SE is the standard deviation of SEs, and Type-I and Type-II indicate error rates. Values in parentheses indicate values for performance metrics when transformed to the VAR(1) metric. Hyphens indicate values that were not calculated. For instance, standard errors for GIMME would need to be transformed from the SVAR to VAR metric and were beyond the scope of this simulation study.

reported explicitly, failing to account for the measurement errors also led to pronounced biases in the process noise variances. The off-diagonal elements of the drift matrices were not as affected by failure to account for the measurement error variances. Finally, the Type-I and II error rates did not substantially differ between `ct-gimme` with and without estimated measurement error variances ($Error_{Type-I} = 0.048; Error_{Type-II} = 0.001$ vs $Error_{Type-I} = 0.045; Error_{Type-II} = 0.000$ for `ct-gimme` without estimated measurement errors). These results agreed with the extant literature regarding neglect of measurement errors and their corresponding effects. In our simplistic example with only measurement error variances, the most prominent effects occurred along these diagonal elements of the dynamics and process noise variances but in more complex scenarios, neglecting measurement errors may results in more substantial down-stream effects on the cross-processes as well (e.g., Schuurman & Hamaker, 2019).

`ct-gimme` performed well when the data-generating OU models were sampled at irregular intervals and fitted to the data at correctly specified time intervals (labeled henceforth as information restrictions or IR). Compared to results with misspecified time intervals (abbreviated henceforth as no IR), correct specification of the time intervals yielded lower absolute biases for the diagonal elements of the dynamics (i.e., $ABias_{\text{IR}} = 0.11$ vs $ABias_{\text{No IR}} = 0.54$) as well as the off-diagonal elements (i.e., $ABias_{\text{IR}} = 0.03$ vs $ABias_{\text{No IR}} = 0.07$). This pattern was reflected in the RMSEs for the diagonal (i.e., $RMSE_{\text{IR}} = 0.15$ vs $RMSE_{\text{No IR}} = 0.91$) and off-diagonal (i.e., $RMSE_{\text{IR}} = 0.07$ vs $RMSE_{\text{No IR}} = 0.21$) elements of the drift matrices. Similarly, the standard errors were better when `ct-gimme` modeled the correct time rather than assuming equally-spaced intervals between successive measurement occasions for diagonal ($SE_{\text{IR}} = 0.24$ vs $SE_{\text{No IR}} = 0.69$) and off-diagonal ($SE_{\text{IR}} = 0.02$ vs $SE_{\text{No IR}} = 0.04$) elements of the drift matrices. Consistent with results from the continuous-time literature (Chow et al., 2016), misspecification of the time intervals lead to higher biases and lower efficiency in the point estimates. The Type-I and II error rates were not affected strongly by this misspecification. This lack of differences suggested that the presence/absence of signals relative to noise could still be recovered reasonably under the misspecification magnitudes considered although the signs (e.g., polarity) and quality (e.g., bias) of those estimates tended to deviate further from their true values. Likewise, when Δt was manipulated from 0.50 to 1.00, minor differences were observed in the performance of many of the continuous-time models (see Figure 2). Still, the various options of `ct-gimme` performed consistently across different effect size, Δt, and T conditions. For the same T and Δt, all approaches tended to perform similarly with respect to Type-I and II error rates regardless of whether they were in the low or high effect size conditions.

Overall, the results indicate the following conclusions: First, `ct-gimme`'s ability to model measurement errors can lead to substantial reductions in biases of the point estimates when measurement errors are present. Our simulations only considered measurement error *variances* in the true, data-generating models; however, when measurement error covariances are more complex, these can lead to substantial differences in the modeling outcomes and dynamics (e.g., Schuurman & Hamaker, 2019). Thus, the ability to model and account for these errors in `ct-gimme` is a significant benefit. Second, when `ct-gimme` was fitted with correct temporal labeling we found improvements in the quality of all dynamic parameters in terms of both bias, RMSE (e.g., variation), as well as the standard errors of these estimates. Broadly, the conclusions of these simulations highlight and reinforce the strengths of modeling in continuoustime as well as with `ct-gimme`. Despite requiring the estimation of 5-additional parameters (i.e., the measurement error variances), `ct-gimme` with measurement errors outperformed its misspecified counterpart assuming no measurement errors. Similarly, by accounting for the true temporal sequencing of events, `ct-gimme` could derive the true, data-generating process better than if data were assumed equidistant when they were not. Thus, our recommendation would be to enable measurement errors in `ct-gimme` when feasible as allowed by sample size as well as collecting explicit information reflecting the time elapsed between successive occasions.

RQ3: How Does `ct-Gimme` *Perform Against Strictly* $N = 1$ *and Group-Level Modeling?*

Generally, quite well; although this is context dependent. The full extent of comparisons between `ct-gimme`, person-specific modeling (i.e., $N = 1$), and group-level modeling may be found in Table 4. Here, we provide a direct discussion of selected results and comparisons.

In homogeneous conditions `ct-gimme` did not outperform the group-level method on nearly any performance metric but consistently outperformed the person-specific approach. In the *HO/HI/T100/DT1* condition in Table 4, the group-level model exhibited the smallest biases ($ABias_{MGM} = 0.04$) when compared to either `ct-gimme`

Table 4. Simulation results across all major conditions.

	HO/LO/T100/DT1			HO/HI/T100/DT1			HE/LO/T100/DT1			HE/HI/T100/DT1			HO/LO/T100/DT0.5		
	`ct-gimme`	$N = 1$	MGM	`ct-gimme`	$N = 1$	MGM	`ct-gimme`	$N = 1$	MGM	`ct-gimme`	$N = 1$	MGM	`ct-gimme`	$N = 1$	MGM
AR ABias	0.115	0.116	0.024	0.102	0.099	0.038	0.115	0.116	0.035	0.108	0.097	0.196	0.147	0.150	0.030
CR ABias	0.028	0.041	0.006	0.032	0.033	0.028	0.046	0.056	0.040	0.063	0.052	0.147	0.039	0.058	0.008
AR RMSE	0.152	0.152	0.031	0.141	0.135	0.051	0.151	0.152	0.043	0.152	0.136	0.227	0.192	0.193	0.037
CR RMSE	0.080	0.104	0.018	0.092	0.090	0.059	0.106	0.122	0.097	0.174	0.123	0.305	0.112	0.140	0.022
AR SE	0.186	0.185	0.030	0.181	0.179	0.031	0.186	0.184	0.029	0.189	0.181	0.021	0.164	0.163	0.037
CR SE	0.024	0.019	0.005	0.024	0.026	0.009	0.029	0.026	0.007	0.033	0.038	0.010	0.030	0.022	0.007
AR SD.SE	0.043	0.044	0.002	0.045	0.041	0.010	0.043	0.043	0.002	0.063	0.045	0.009	0.030	0.030	0.003
CR SD.SE	0.050	0.045	0.011	0.052	0.052	0.014	0.053	0.051	0.012	0.057	0.059	0.011	0.065	0.057	0.014
Type-I	0.051	0.064	0.054	0.030	0.069	0.059	0.048	0.065	0.143	0.065	0.086	0.403	0.049	0.071	0.053
Type-II	0.022	0.348	0.000	0.016	0.003	0.008	0.207	0.358	0.326	0.090	0.011	0.213	0.083	0.529	0.000
	HO/HI/T100/DT0.5			HO/LO/T50/DT1			HO/HI/T50/DT1			HO/LO/T50/DT0.5			HO/HI/T50/DT0.5		
	`ct-gimme`	$N = 1$	MGM	`ct-gimme`	$N = 1$	MGM	`ct-gimme`	$N = 1$	MGM	`ct-gimme`	$N = 1$	MGM	`ct-gimme`	$N = 1$	MGM
AR ABias	0.118	0.118	0.030	0.170	0.169	0.035	0.138	0.147	0.046	0.211	0.209	0.043	0.168	0.174	0.043
CR ABias	0.028	0.033	0.010	0.050	0.066	0.009	0.039	0.049	0.028	0.068	0.086	0.012	0.044	0.063	0.013
AR RMSE	0.160	0.160	0.039	0.218	0.216	0.044	0.186	0.200	0.061	0.263	0.259	0.054	0.221	0.227	0.055
CR RMSE	0.097	0.099	0.026	0.135	0.156	0.025	0.118	0.144	0.061	0.180	0.202	0.033	0.139	0.192	0.036
AR SE	0.158	0.158	0.037	0.271	0.266	0.043	0.258	0.254	0.044	0.234	0.233	0.052	0.225	0.225	0.053
CR SE	0.029	0.034	0.008	0.033	0.023	0.008	0.035	0.036	0.011	0.035	0.028	0.010	0.049	0.048	0.011
AR SD.SE	0.028	0.028	0.010	0.089	0.089	0.004	0.081	0.084	0.015	0.061	0.060	0.004	0.056	0.057	0.015
CR SD.SE	0.066	0.070	0.016	0.073	0.062	0.016	0.074	0.074	0.019	0.090	0.082	0.020	0.104	0.101	0.023
Type-I	0.030	0.064	0.060	0.057	0.079	0.054	0.048	0.076	0.141	0.055	0.085	0.053	0.053	0.077	0.060
Type-II	0.030	0.008	0.000	0.157	0.615	0.000	0.009	0.059	0.000	0.377	0.724	0.000	0.001	0.134	0.000

Note: Simulation headings indicate Heterogeneity condition (HO/HE), Effect-size (LO/HI), Sample size (T50/T100), and Δt (DT0.5/DT1). Thus HO/LO/T100/DT1 indicates homogeneous condition with low effect sizes with 100 time-points at $\Delta t = 1.0$. AR indicates diagonal elements of the dynamics while CR indicates off-diagonal values. RMSE is the root mean squared error, SE is the standard error, SD.SE is the standard deviation of SEs, and Type-I and Type-II indicate error rates.

($ABias_{ctg} = 0.10$) and $N = 1$ modeling ($ABias_{N=1} = 0.10$). These results were reflected for all other metrics. However, `ct-gimme` performed well with Type-I and Type-II error rates with ($Type_I = 0.03; Type_{II} = 0.02$) which were comparable to the group-level approach ($Type_I = 0.06; Type_{II} = 0.01$) than $N = 1$ approaches ($Type_I = 0.07; Type_{II} = 0.003$). This pattern was consistent across T as well. In the low-effect size conditions (e.g., *HO/LO/T100/DT1*), error rates in `ct-gimme` exceeded that of $N = 1$ approaches ($Type_I = 0.05; Type_{II} = 0.02$ compared to ($Type_I = 0.06; Type_{II} = 0.035$). These improvements over the $N = 1$ approach manifested in reductions in biases and RMSEs for the cross-process effects with $ABias = 0.028, RMSE = 0.080$ for `ct-gimme` and $ABias = 0.041, RMSE = 0.104$ for $N = 1$ (see condition HO/LO/T100/DT1 in Table 4). Logically, when all subjects entirely share their dynamic structure, modeling them as a homogeneous group will leverage all of their collective information to improve parameter estimation. Similarly, `ct-gimme` outperforming the $N = 1$ approaches highlighted the unique strengths of the GIMME-like approach. By pooling the modification indices, `ct-gimme` could delineate signal from noise better than single-subject modeling could. In particular during simulation conditions with smaller effect sizes; a finding well established with discrete-time GIMME (Gates & Molenaar, 2012).

`ct-gimme` highlighted its strengths when samples were comprised of a common, group-level structure with individual differences. `ct-gimme` exhibited excellent Type-I and II error rates ($Type_I = 0.07; Type_{II} = 0.09$) compared to both $N = 1$ modeling ($Type_I = 0.09; Type_{II} = 0.01$) and the group-level methods ($Type_I = 0.40; Type_{II} = 0.21$; see Figure 3). The expectation was that $N = 1$ models should exhibit consistent performance across conditions of homogeneity and heterogeneity since group-level information was not being accounted for. Despite minor deviations in Type-I and Type-II error rates, this expectation was met with fluctuations likely being to the additional parameters being estimated in heterogeneous conditions. The presence of any degree of heterogeneity in the dynamic structures resulted in inflated absolute biases for the group-level method on the diagonal ($ABias_{MGM} = 0.20$) and off-diagonal dynamics ($ABias_{MGM} = 0.15$) when compared to `ct-gimme` ($ABias_{ctg} = 0.11$; $ABias_{ctg} = 0.06$) or the $N = 1$ approach ($ABias_{N=1} = 0.10$; $ABias_{N=1} = 0.05$). While the homogeneous conditions favored the group-level model, all heterogeneous conditions across Δt, *Effect Size*, and *Time* highlighted the flaws in assuming common structures throughout the sample. These results highlighted the strengths of the $N = 1$ approaches; however, these advantages only held when effect sizes were relatively large (i.e., $EF = 0.90$). When $EF = 0.30$, the $N = 1$ modeling approaches became far less powerful (i.e., $Type_{II} = 0.36$) and in-line with the group-level modeling approaches ($Type_{II} = 0.33$). Overall, `ct-gimme` performed the best in terms of statistical power ($Type_{II} = 0.21$) while still achieving a near-nominal Type-I error rates of 4.8%. Logically, the group-level method was expected to perform poorly as the presence of individual differences in dynamics leads to "forcing" dynamic parameters onto individuals that did not need them because they are present for others in the sample.

`ct-gimme` exhibited poorer performance with smaller sample sizes (i.e., $T = 50$ vs $T = 100$). Notably, the biases of the point estimates for the homogeneous condition with low effect sizes only changed from 0.170 to 0.115 when transitioning from $T = 50$ to $T = 100$, respectively. However, `ct-gimme` maintained high power even with as few at 50-measurement occasions ($T50_{Type-II} = 0.16$ vs $T100_{Type-II} = 0.02$). While a substantial increase compared to the 100-measurement occasion condition, `ct-gimme` performed substantially better than the person-specific approach, ($T50_{Type-II} = 0.62$).

One condition did not align with expectations; that is, the results for the Type-I error rates for the group-level model in condition *HO/HI/T50/DT1*. In this condition–illustrated in Figure 3–the Type-I error rates for the group-level model were substantially higher than the nominal level (e.g., $\approx 14\%$). Closer inspection of modeling results revealed a small number of spurious paths that emerged in the group-level models at relatively small magnitudes (e.g., ≈ 0.09). These spurious paths tended to emerge when the processes were close to the boundary of being unstable (showing increasing deviations from the baseline of 0). As an example, Figure 4 highlights two time series of the same variable for the same subject but in different configurations of effect size (i.e., low versus high). Of note, the higher effect size condition–while mathematically stationary–still exhibits a greater degree of variation over time. When these time series are then compared across multiple subjects, additional patterns could potentially emerge that appear to be caused by "phantom" paths or dynamics.

Ultimately, the results of RQ3 indicated that `ct-gimme`'s performance is situated between that of $N = 1$ modeling and group-level modeling. When researchers can assume dynamic structures are homogeneous across samples, the group-level modeling approach will provide the best results across all performance metrics applied in our simulation studies across any configuration of sample size, effect size, and Δt. However, instances where complete homogeneity exists in dynamic modeling are not commonly seen in empirical applications of dynamic networks (e.g., De Vos et al., 2017; Ebrahimi et al., 2024; Hamaker et al., 2005; Park et al., 2023; Wright et al., 2019). Likewise, we saw that $N = 1$ modeling approaches did well when samples were heterogeneous and effect sizes were large; however, the performance of the $N = 1$ procedures did not significantly outperform `ct-gimme`. In contrast, when effect sizes were smaller, the $N = 1$ approaches were eclipsed by `ct-gimme` in terms of statistical power via its ability to leverage information across the samples.

RQ4: How Does `ct-Gimme` Perform Against Discrete-Time GIMME?

Generally, quite well. Comparisons between the parameter estimates of `ct-gimme` and GIMME were compared on the metric of the standard VAR(1) by transforming both the

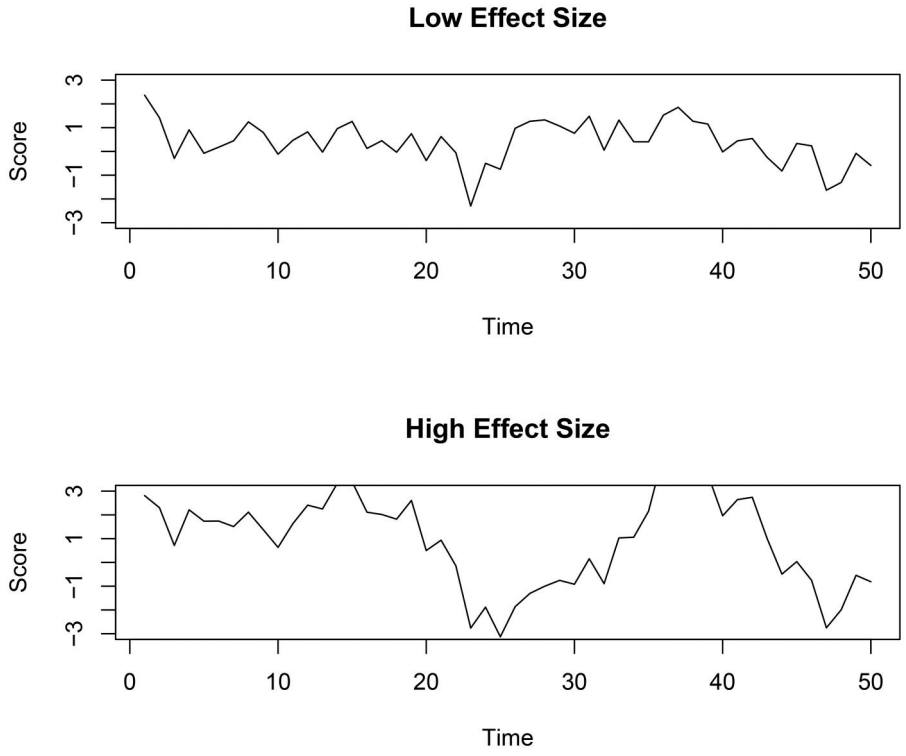

Figure 4. Visualization of single trajectory in low- and high-effect size conditions. While both processes are stationary the "wandering" in the high effect size conditions may have resulted in pronounced Type-I errors in the group-level models.

OU models and SVAR models to the VAR(1) metric using transformations described in Equation (4). Note that such transformations were not necessary for inferential purposes within the ct-gimme framework per se, but were done only to facilitate comparisons of estimates from the two GIMME approaches under a common model.

When data were measured irregularly, `ct-gimme` ($ABias = 0.06, RMSE = 0.08$) outperformed the discrete-time GIMME in terms of absolute biases and RMSEs when placed upon the common metric of the VAR(1) ($ABias = 0.24, RMSE = 0.03$; see parenthetical values in Table 3). The cross-process effects exhibited similar patterns for `ct-gimme` ($ABias = 0.02, RMSE = 0.04$) compared to GIMME ($ABias = 0.27, RMSE = 0.07$). These results indicated that parameter estimates for discrete-time GIMME tended to be more biased and varied than those from `ct-gimme`. Interestingly, discrete-time GIMME's performance was similar to `ct-gimme` with incorrect coding for the time-intervals (i.e., Δ_t) in terms of ABiases and RMSEs. These results highlight the unique strengths of modeling processes in the continuous-time framework where accounting for irregular spacing between data-points may yield significant biases (e.g., $\sim 3.81\times$ greater) when spacing between measurement occasions is neglected.

GIMME outperformed the `ct-gimme` algorithm in terms of Type-I error rates (5.8% compared to 13.7% for `ct-gimme`), however. As a trade-off GIMME exhibited less power compared to `ct-gimme` with Type-II error rates approaching 34% compared to 1.3% for `ct-gimme`. This is in line with prior work that suggests that the false-positive rate for GIMME tends to be good (Gates & Molenaar, 2012) while exhibiting low power in the presence of weak dynamics (Nestler & Humberg, 2021).

To elucidate the higher Type-I error rates of ct-gimme relative to GIMME in the VAR metrics, we show below the raw and transformed values for the drift matrix in `ct-gimme`. The inflated Type-I error rates for `ct-gimme` can be explained by nature of the transformations required to transform the continuous-time VAR to the discrete-time form. Take for instance a true matrix, $\mathbf{A}$, and its discrete-time transform at $\Delta t = 1.0$:

$$\mathbf{A} = \begin{bmatrix} -0.50 & 0.00 & 0.00 & 0.00 & -0.30 \\ 0.00 & -0.50 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & -0.50 & 0.00 & 0.30 \\ -0.30 & 0.00 & 0.00 & -0.50 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & -0.50 \end{bmatrix},$$

$$\Phi^*(\Delta t_{1.0}) = \begin{bmatrix} 0.61 & 0.00 & 0.00 & 0.00 & -0.18 \\ 0.00 & 0.61 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.61 & 0.00 & 0.18 \\ -0.18 & 0.00 & 0.00 & 0.61 & 0.03 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.61 \end{bmatrix} \quad (16)$$

Then, see a similar transformation for an estimated dynamics matrix, $\hat{\mathbf{A}}$, and its discrete-time counterpart:

$$\hat{\mathbf{A}} = \begin{bmatrix} -0.66 & 0.00 & 0.00 & \mathbf{-0.38} & -0.44 \\ 0.00 & -0.69 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & -0.49 & \mathbf{-0.32} & 0.30 \\ -0.18 & 0.00 & 0.00 & -0.85 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & -0.54 \end{bmatrix},$$

$$\Phi^*(\Delta t_{1.0}) = \begin{bmatrix} 0.53 & 0.00 & 0.00 & \mathbf{-0.18} & -0.25 \\ 0.00 & 0.50 & 0.00 & 0.00 & 0.00 \\ \mathbf{0.01} & 0.00 & 0.61 & \mathbf{-0.17} & 0.17 \\ -0.08 & 0.00 & 0.00 & 0.44 & \mathbf{0.02} \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.58 \end{bmatrix} \quad (17)$$

In this illustrative case, a misspecification of 2-parameters yields 4-incorrect, non-zero parameters when transformed to the discrete-time case which are highlighted in bold.

These results indicated that the `ct-gimme` algorithm performed well when compared to GIMME in discrete-time; however, GIMME does exhibit strong performance with respect to its false discovery rate by maintaining a near nominal Type-I error rate (i.e., 5.8%) even when modeling abnormally sequenced time-series.

Broad-Level Simulation Conclusions

Our simulations highlighted the strengths and weaknesses of the new `ct-gimme` algorithm in a vacuum as well as more broadly against alternative methods in continuous- and discrete-time. Notably, of the adjustment options integrated into `ct-gimme`, the BIC_n model selection criterion outperformed the raw MI approach and the Benjamini-Hochberg corrections in homogeneous conditions. With heterogeneity in dynamics, the Benjamini-Hochberg performed better in terms of both Type-I and Type-II error rates; thus, generally, `ct-gimme`'s BIC_n procedure may be preferable when groups are considered more homogeneous and the Benjamini-Hochberg correction when samples are suspected to be more heterogeneous in favor of discovery of dynamic structures.

Our validations of `ct-gimme` indicated strong performance. By developing `ct-gimme` in the state-space framework, options for integrating and estimating measurement errors are readily allowed and our simulations indicated that neglecting to incorporate them into the continuous-time models biased our centralizing tendencies. Likewise, when data were sampled from a skewed distribution to simulate unequally spaced data, we found that explicitly modeling the time differences resulted in notable improvements to all dynamic parameters in terms of their accuracy and confidence.

Likewise, `ct-gimme` possessed the strengths of both $N = 1$ and group-level modeling procedures with few of their drawbacks. Across homogeneous conditions, the group-level fitting outperformed `ct-gimme` and $N = 1$ modeling. However, in heterogeneous conditions, `ct-gimme` outperformed both $N = 1$ and group-level modeling when effect sizes were relatively weak; a situation more likely to be encountered in the social and behavioral sciences. By leveraging information from the whole sample to identify common features prior to estimating person-specific dynamics, `ct-gimme` almost unilaterally outperformed the $N = 1$ approach on metrics of bias, variance (RMSE), quality of the parameter estimates (SEs), and Type-I/II error rates. This advantage held even in the presence of individual differences in dynamics. Likewise, the presence of individual differences in dynamics posed a unique challenge for the group-level approach which assumes a common dynamic pattern for the entire sample. This often led to the group-level approach performing quite poorly in configurations involving any degree of within-sample heterogeneity. By contrast, `ct-gimme`'s tended to remain much more stable between homogeneous and heterogeneous conditions.

Finally, when compared to the discrete-time GIMME algorithm, `ct-gimme` performed well–particularly in conditions when data were irregularly spaced as is to be expected by modeling in continuous-time. However, `ct-gimme` did exhibit higher Type-I error rates in relation to GIMME possibly as a result of the nature of how the continuous-to-discrete transformation operates.

Empirical Illustration

Modeling Dynamics

For our empirical illustration, we leveraged data taken from Fisher et al. (2017). These data were selected due to past work fitting discrete-time GIMME-methods to this dataset providing avenues for comparisons between `ct-gimme` and past results. We fitted `ct-gimme` to symptom state-level data and evaluated the continuous-time dynamic networks. We expected that the `ct-gimme` would identify group-level structures in the continuous-time dynamics should they be present. We also compare the results obtained from the discrete-time literature on the same data to highlight any differences from fitting the models in this new framework.

Sample

Data were comprised of $N = 40$ participants with clinical-levels of generalized anxiety disorder (GAD), major depressive disorder (MDD), or comorbid GAD and MDD. The sample was majority female ($n_{female} = 26$) and White ($n_{White} = 19$). A full description of the sample and its characteristics may be found in Fisher et al. (2017).

Measures

Participants in the study were assessed over ≈ 30 days and responded to 21-items of mood and anxiety symptoms 4-times a day. Symptoms were drawn from the Diagnostic and Statistical Manual of Mental Disorders, Fifth edition for GAD and MDD. Items participants responded to included: *down and depressed*, *hopeless*, *loss of interest or pleasure*,

worthless or guilty, *worried*, and *restless*. Participants were asked to rate from 0 to 100 their experience of each item since the previous measurement occasion. For instance, a participant may rate the degree to which they felt "*hopeless*" from 0 to 100.

Measurement occasions were random and exact time-codes were provided in the dataset on an hourly-basis. Hourly times were transformed to indices denoting the hourly-time difference from the initial measurement occasion and divided by 24. This implied that a $\Delta t = 1.0$ is associated with a daily measurement occasion. This allowed for sufficient density in measured time-points for each participant. When the gap between any two subsequent measurement occasions was larger than one third of a day, empty rows were inserted into the dataset for each participant.

For the illustration, 5-symptoms were selected to match the network size of our simulations. These symptoms were: *Irritable*, *Fatigue*, *Concentration Difficulties*, *Rumination*, and *Avoiding Activities*.

On average, participants completed 130.43 ($SD = 19.27$) reports with a range of 87 to 212 completed reports. Data were preprocessed by methods detailed by Fisher et al. (2017) and gathered second-hand. Additional preprocessing was done in the current study by within-person and within-variable standardizing each variable's scores.

Model Fitting

A group-level continuous-time model was fitted to the N p-variate time-series in `R` using the `OpenMx` package. All subjects were constrained to equality on all dynamic parameters, process noise variances, and measurement error variances. These parameters were also were freed for estimation at the group-level. Once completed, the process noise and measurement error variances were extracted from the model output and used as the rational starts from which `ct-gimme` would begin estimation.

`ct-gimme` was fitted to the N p-variate time-series in `R`. Program settings were set to default with the BIC_n adjustment selected based on its performance in our earlier simulation results. If the performance based on BIC_n was not sufficient (i.e., not finding a common, group-level structure), the Benjamini-Hochberg correction for family-wise α would be used instead as it was found to exhibit more power in heterogeneous samples. Process noise and measurement error variances were fixed to values obtained during the group-level modeling stage to serve as rational starting values. For this illustration, measurement error variances were fixed but process noise variances were freed for estimation for each subject. Group-level paths were determined using a threshold of 51.00% to err on the side of discovery.

The results of `ct-gimme` yielded a set of participants with estimable models which represented $\approx 45\%$ of the sample. The remaining $\approx 55\%$ of the sample encountered issues relating to model convergence, specifically relating to the recovery of the standard errors for their dynamic parameters and/or process noise variances. We detail the dynamics and characteristics of participants with converged results. Following this, we provide a brief summary of, and some speculations of the reasons for non-convergence to facilitate future developments of `ct-gimme`.

A common, group-level structure was not found by `ct-gimme` when using the BIC_n model selection procedure. To err on the side of discovery, we implemented the Benjamini-Hochberg correction which exhibited higher power when samples were heterogeneous and a common, group-level structure still did not emerge. This indicated a high degree of heterogeneity in the person-specific networks. Figure 5 displays the continuous-time dynamic network structures of 3-randomly selected participants whose dynamics pose interesting theoretical possibilities. All three participants exhibited strong, excitatory relations from *Concentration Difficulties* to *Avoiding Activities*. This indicated that greater difficulty concentrating was associated with increased avoidance of activities. Likewise, all centralizing tendencies were strongly negative indicating that each symptom tended to decay in severity with time. Interestingly, Participant No. 12 exhibited *Fatigue* as a core symptom which would cause several downstream effects (i.e., greater fatigue led to greater irritability but less concentration difficulties leading to less avoidance of activities). In contrast, *Concentration Difficulties* was more influential for

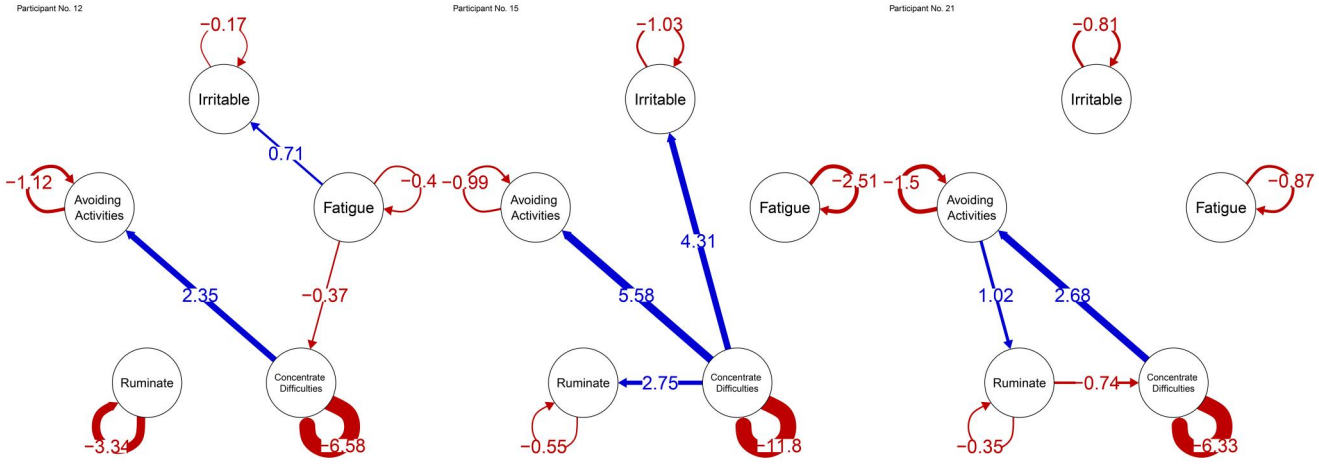

Figure 5. Continuous-time dynamic networks for 3 randomly selected participants. Blue edges indicate excitatory dynamics and red edges indicate inhibitory dynamics. Centralizing tendencies (self-looping) were forced to be freed for estimation for all participants.

Participant No. 15 with greater levels of difficulty being associated with greater irritability, rumination, and activity avoidance. Finally, Participant No. 21 exhibited a feedback loop where high levels of *Concentration Difficulties* were associated with avoidance of activities which would then lead to greater rumination which would then suppress concentration difficulties.

One of the strengths of modeling in continuous-time is the ability to translate the continuous-time dynamic structure to different intervals of time to see what the relations between variables looks like at various practical intervals (Driver & Voelkle, 2018; Ryan et al., 2018). Likewise, these transformations can be used to inform future work for identifying optimal time-scales to maximize the likelihood of detecting effects (Hecht & Zitzmann, 2021a). Figure 6 highlights the discretization of Participant No. 12's continuous-time dynamic network into hourly, bi-daily, daily, and weekly dynamic networks. The transformations indicate that–on an hourly basis–Participant No. 12's dynamics are largely inertial (Kuppens et al., 2010). That is, symptoms did not tend to influence other symptoms on an hour-by-hour basis and–instead–were regulated by themselves in the prior hour. For instance, feeling fatigue an hour ago would highly relate to fatigue at the current moment but fatigue an hour ago wouldn't be strongly predictive of irritability now. The cross-process dynamics begin to show prominence when transformed to a bi-daily (12-hour) schedule. The model indicates that fatigue for Participant No. 12 was related to greater irritability and lower activity avoidance and concentration difficulties 12-hours later. On a daily basis, the connection between fatigue and irritability becomes even stronger indicating that fatigue today would be strongly associated with irritability the following day in addition to other downstream effects. Finally, at a weekly scale, the strongest influence that remains is between fatigue and irritability suggesting that fatigue exhibits fairly long-lasting impacts on Participant No. 12's irritability from week-to-week.

These results highlight an important connection to discrete-time applications conducted on this same dataset by Fisher et al. (2017). Notably, both applications highlighted significant heterogeneity in the temporal dynamics across the participants. In our applications, no dynamic patterns were present for more than 51% of the sample. Likewise, in their larger network application, Fisher et al. (2017) noted that the heterogeneity in the data indicated that the disorders under investigation may be "too great" to limit to a purely nomothetic nosology of clinical psychopathology. While symptom states in both applications were meaningful across both `ct-gimme` and in work by Fisher et al. (2017), the key difference lied in how those symptoms affected other symptoms in the network. As noted above, *Concentration Difficulties* was influential in all three of our randomly selected participants; however, the down-stream effects of *Concentration Difficulties* were largely heterogeneous even across our own sample.

As noted, we encountered convergence issues in a subset of the participants. Closer inspection suggested that one participant exhibited a dramatic shift in their dynamic

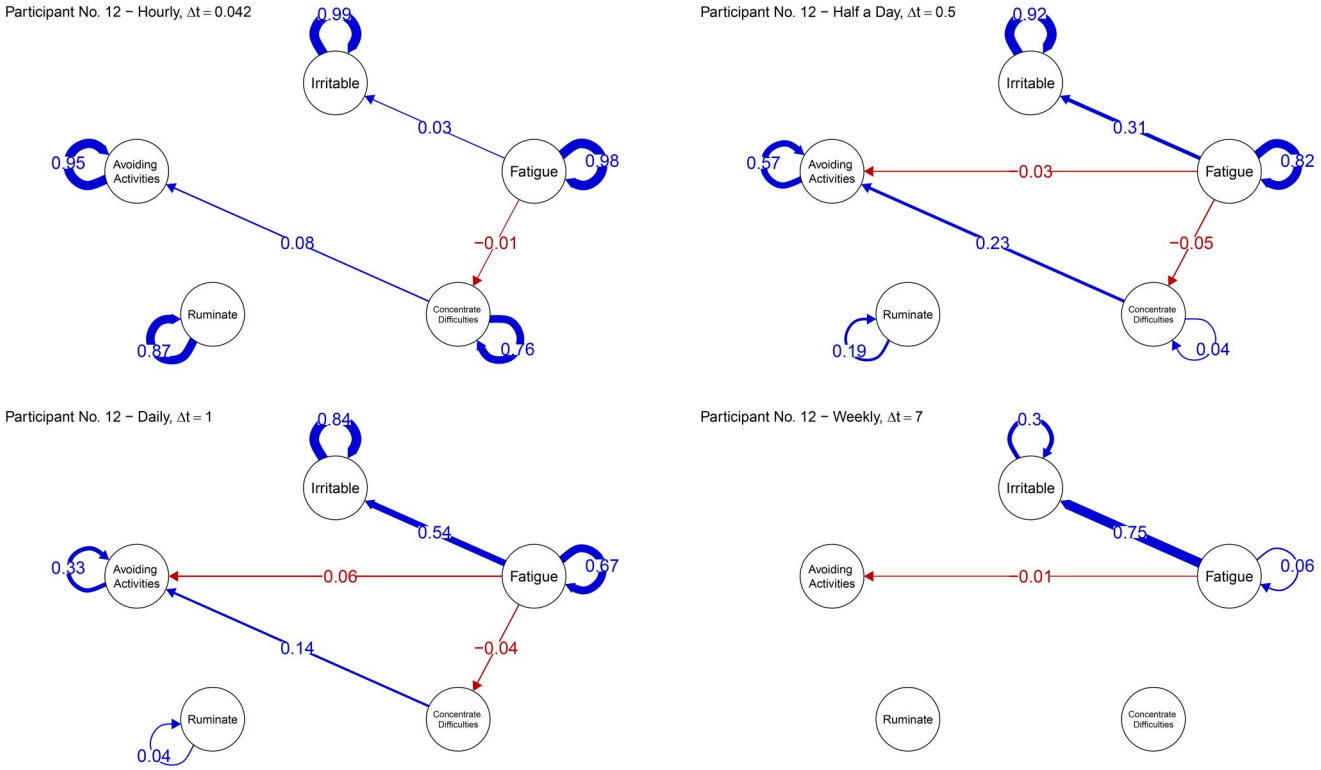

Figure 6. Discretized dynamic network of participant No. 12 at different Δt configurations of 0.042, 0.50, 1.00, and 7.00 indicating hourly, bi-daily, daily, and weekly relations among the five symptoms, respectively. Blue edges indicate positive associations between variables at one moment in time and the next. Red edges indicate negative associations. Self-loops indicate the effect a variable has on itself at a subsequent time.

processes roughly halfway through the study period (Figure 7). Specifically, Participant No. 5 exhibited dynamics with periodic rises and falls in their symptom levels but following the halfway mark, the participant switched to exhibiting a floor effect with significantly less variation on all 5 symptoms used in this investigation. This type of change or alteration in dynamics may be characterized as a regime change. These regime changes can be modeled explicitly in the continuous-time framework (e.g., Chow et al., 2018); however, this was beyond the scope of the application.

Likewise, we found that person-specific estimation of the measurement error variances could differ dramatically from the measurement error variances derived from the group-level model. Given the observation that individuals were largely heterogeneous in their dynamic network structures, it should come as no surprise that the measurement error and process noise variances would be person-specific as well. When we attempted to model both the measurement error and process noise variances, we ran into additional convergence issues likely relating to the relatively small sample-size in the study and the number of parameters being estimated. Thus, the compromise was to keep the fixed measurement error variances in favor of process noise variances for the purposes of this illustrative example. Further work and development needs to be conducted in order to allow for `ct-gimme` to dynamically evaluate when process noise, measurement error variances, or both can be freed for individuals and prioritize the quality of those estimates in addition to recovery of the dynamic parameters.

The empirical illustration highlighted several substantively interesting features across a select group of participants as well as the technical challenges that need to be addressed for widespread adoption and proliferation of these complex continuous-time models. Results of the three

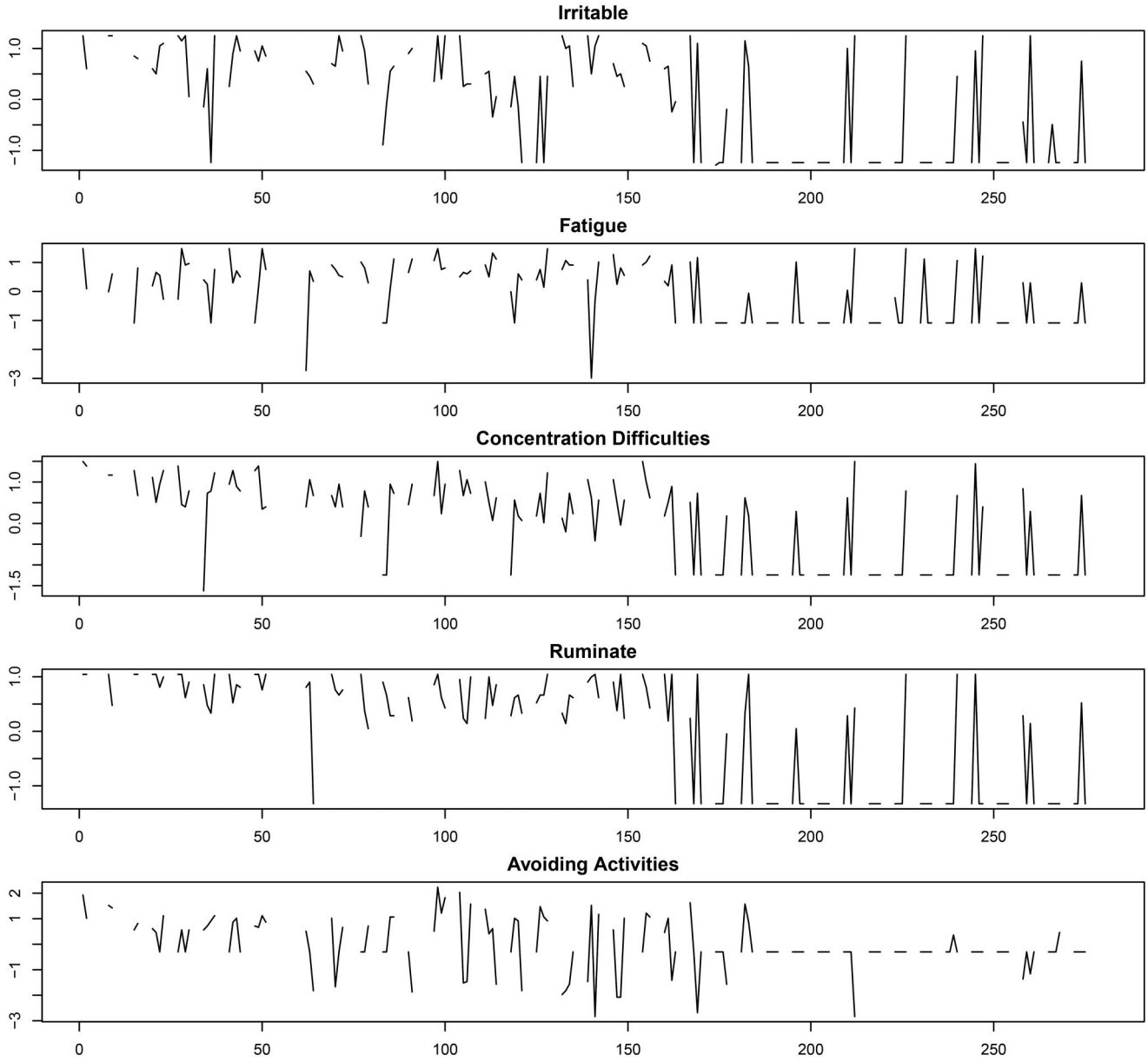

Figure 7. Time series of participant No. 5 for all variables. Notably, participant seems to exhibit a distinct change or alteration to their dynamics around halfway through the study.

selected participants indicated high centralizing tendencies for variables such as *Concentration Difficulties* and *Rumination* with relatively weaker centralizing tendencies for variables such as *Fatigue* and *Irritability*. When translated to a discrete-time analogue, this implies that feelings of *Fatigue* and *Irritability* tend to exhibit high "inertia" (i.e., high autoregressive terms; see Figure 6). Indeed, for Participant No. 12, this implies that feelings of *Fatigue* and *Irritability* tend to persist for long periods of time, take longer to recover from, and exhibit extensive connections to other variables in their dynamic networks affecting their *Concentration* and *Avoidance of activities*. These results aligned with prior findings conducted in the discrete-time framework using the same dataset (see Fisher et al., 2017). The results of `ct-gimme` also aligned with other results from Fisher et al. (2017) as noted above such as the presence of large degrees of heterogeneity in dynamic patterns underscoring the relative consistency between the two studies in both the discrete- and continuous-time frameworks.

Participant No. 15 exhibited significant dependence of *Irritability*, *Activity Avoidance*, and *Rumination* on *Concentration Difficulties*. This implies that this participant–when faced with difficulties concentrating–tends to become more irritable, avoids activities more, and becomes more ruminatory. Recent literature has attempted to identify *bridge symptoms* (Jones et al., 2021). These symptoms may be indicative of symptoms or variables that are highly influential to the state of the overall network; particularly in the case of symptoms networks of multiple disorders. In the case of Participant No. 15, it becomes evident that the state of their dynamics relies heavily on *Concentration Difficulties* and may serve as a valuable target for intervention in future investigations.

Additionally, the failure for several subjects to successfully converge highlighted key weaknesses and areas to improve upon the `ct-gimme` algorithm and how it initializes person-specific models in particular when estimating both measurement error variances and process noise variances. In addition, the process of fitting `ct-gimme` could be further streamlined by adding checks for when single subjects encounter difficulties during optimization and model fitting to flag users.

Discussion

Despite the myriad advantages of modeling dynamic processes in continuous-time, the application of these complex models is still critically underutilized (Ryan & Hamaker, 2022). Furthermore, tools for addressing within-sample heterogeneity in the continuous-time framework are still few in number (Hunter, 2014, 2024; Liu et al., 2021). The current work contributes multiple innovations to the literature. First and foremost, a novel extension of the GIMME algorithm for the continuous-time framework. This extension allows researchers to effectively model dynamic processes in the continuous-time framework and draw upon many of its strengths such as the identification of optimal time-scales via transformation of the continuous-time drift matrix among others (e.g., Hecht et al., 2019; Hecht & Zitzmann, 2021a). Additionally, `ct-gimme` allows users to identify common, sample-level dynamics while drawing upon sample-level information to delineate signal from noise. Furthermore, the GIMME framework differs from other popular frameworks such as multilevel modeling in that group-level structures do not constrain individuals to a particular final model structure. This is due to the two-stage estimation which identifies a common, group-level structure followed by individual model fitting (see Gates & Molenaar, 2012, for more information regarding GIMME). Second, the development of `ct-gimme` contributes a user-friendly means for empirical researchers to explore continuous-time modeling. While many complexities belie continuous-time modeling, this serves as one–of many other–first step towards disseminating these complex models.

Our results indicated that `ct-gimme` performed well when compared to benchmark measures. Specifically, $N = 1$ or person-specific model fitting and group-level modeling. Both comparisons represent two extreme ends of how researchers may view dynamic processes unfolding as either entirely idiosyncratic or uniform across all individuals. Across our simulation studies, we found that `ct-gimme` reliably performed better than the $N = 1$ procedures due to its ability to draw information across the sample to identify key, group-level dynamics. These paths would then further strengthen `ct-gimme`'s ability to identify person-specific dynamics later on. This ability to leverage sample-level information for person-specific parameter recovery allowed `ct-gimme` to compensate for relatively short person-specific time-series. These results are contingent on the degree of homogeneity within a given sample and prior research has provided guidelines on these relations in continuous-time modeling when balancing temporal sampling, T and participants N (see, Hecht & Zitzmann, 2021b). Likewise, `ct-gimme` tended to outperform the group-level procedure in the presence of individual differences in dynamic structures by identifying common group features but enabling person-specific expression via the two-stage approach.

Our illustration highlighted the steps researchers may take when implementing `ct-gimme` as well as the benefits and pitfalls of modeling in continuous-time. As noted, the illustration failed to recover sound models for a large number of participants due to myriad convergence issues relating to shifts in dynamics, boundary limits that–when freed–exploded, and difficulty in determining whether measurement error or process noise variances should be freed in lieu of the other. On individual probing, these decisions were ultimately person-specific with some participant's converging when some process noises were fixed with measurement error variances were freed and vice versa. The `ct-gimme` approach attempts to automate many of the decisions to construct continuous-time models from the ground up; however, many decisions are still–ultimately–left to the user. That being said, the `ct-gimme` approach of iteratively adding a single parameter tended to yield fewer convergence issues when compared to fitting person-specific

continuous-time models with all dynamic parameters freed for estimation.

Limitations

Several limitations exist within the context of the current work. First, the OU models used in our simulations are not entirely reflective of typical continuous-time dynamic models. The relatively clean nature of our OU models was designed to give a clear idea on how manipulations of specific parameters and factors could result in changes to the results. As a result–however–the models themselves are somewhat artificial. Future research could benefit from a more extensive simulation study that begins from a model derived from empirical results. Likewise, the estimation of `ct-gimme` imposes a diagonal process noise covariance structure which may not be practical for real-world continuous-time processes which can exhibit notable covariances in their parameterization (e.g., Oravecz & Tuerlinckx, 2011). `ct-gimme` could–in theory–be extended to search the process noise covariances in addition to the dynamics matrices when undergoing model construction; however, more extensive testing would be required. Extending upon this further, due to the unique flexibility presented by `OpenMx`, almost all model parameters could be tested via the modification indices. This could include building the measurement model in addition to the dynamic model in a purely data-driven fashion. This would allow for specification of multi-indicator dynamic factor models with latent variables. Our simulations also assumed that no means were present in the data of either our simulation or empirical illustration (due to centering). As a result, we exclude a vital component of psychological processes relating to trends. Past work has illustrated how neglecting trend-level information may bias dynamic parameters but can also be explicitly accounted for in continuous-time models (Lohmann et al., 2022, 2024). These developments highlight the importance of explicitly modeling trend-level information and serves as an avenue for future development for `ct-gimme`.

Our comparison models (i.e., $N = 1$ and Group-level models) represented two extreme ends of the "idio-thetic" spectrum of models available to researchers. While continuous-time models certainly lag behind discrete-time methods, our comparisons were not exhaustive by any means. Future research should compare the performance of `ct-gimme` against approaches which assume continuous- rather than discrete differences between participants in a sample such as the multilevel or hierarchical frameworks. These assume a common fixed structure about all model parameters with differences between participants being associated with person-specific variation about those fixed estimates. These differences between `ct-gimme` and multilevel models would further elucidate long-standing considerations on whether groups differ by continuous or discrete differences in dynamic structures (Hunter, 2024).

The empirical illustration highlighted some flaws in `ct-gimme`. Notably, the specification of initial conditions and starting values can significantly impact the resulting models and whether or not they converge. Entire dissertations could be written on the specification of initial conditions and starting values. The avenue we took was to fit a group-level model and use the resulting parameter estimates for the process noise and measurement error variances as starting values for `ct-gimme`. However, it could be the case that process noises and/or measurement error variances are person-specific and thus the starting values we used may have been inappropriate for many in the sample if heterogeneity were present as was the case in our application. Due to the relatively short time-series relative to the number of estimated parameters, we were not able to free up both the process noise variances and the measurement error variances for estimation during the fitting of `ct-gimme`. More investigation needs to be done regarding these issues to better understand the limitations of `ct-gimme` in relation to data quality, sample size, and network size. Likewise, some participants exhibited clearly non-stationary dynamics (e.g., Participant 5; Figure 7).

Regime-switching methods have already been extended into the continuous-time framework and have been applied readily to various contexts (Chow et al., 2018) and software packages (Ou et al., 2019). `ct-gimme` could be extended to account for non-stationarity in dynamics within or between individuals. One such approach could be to segment or subgroup individuals through time and seeing whether individuals coalesce with themselves at other time periods via subgrouping methods in a manner similar to how subgroups are derived for heterogeneous time-series of multiple subjects (e.g., Gates et al., 2017; Park et al., 2020) or by inclusion of time-varying parameters (e.g., Chen et al., 2018; Chow et al., 2011; Fisher et al., 2022). In the former approach, a rolling window could be applied to individual time-series and test whether individuals get "clustered" with themselves based on their dynamics. In such cases, individuals would roughly exhibit similar dynamic patterns. In the latter case, the non-stationarity could be explicitly integrated into the modeling framework by accounting for how dynamic parameters change over time. Notably, recent work has demonstrated how time-varying continuous-time models may be implemented (Hecht et al., 2024). Moreover, the current application of `ct-gimme` investigated scenarios where samples are comprised of a single common dynamic structure with individual differences in dynamics. Alternatively, samples could be comprised of multiple constituent subgroups of individuals who each share more in common with a select group of others than they do with other members of the overall sample. The GIMME-framework has already been extended to this subgrouping case (e.g., S-GIMME; Gates et al., 2017) but it–as well as other subgrouping methods–are solely developed in the discrete-time framework (e.g., Gates et al., 2017; Park et al., 2024). Future work should extend `ct-gimme` to include subgrouping routines to further parse out within-sample heterogeneity in dynamic structures. Further, experiments could address how well subgroups derived in the discrete-time framework align with those derived from the

continuous-time framework and whether subgroup solutions are dependent on Δt.

Despite these limitations, `ct-gimme` exhibits promising performance in relation to alternative methods for fitting continuous-time models (e.g., group-level and person-specific approaches) and tends to outperform its discrete-time form in select scenarios. Further, by extending the GIMME framework to the continuous-time framework, researchers may draw upon the many benefits of modeling in continuous-time whilst also benefiting from the strength of the GIMME approach. Theoretically, the GIMME framework is attractive as it implies a common, group-level structure without imposing any real-valued constraint on those parameters. In contrast to other popular approaches that "pull" participants towards a specific value, individuals in the GIMME framework may share common structural paths with entirely opposite signs (i.e., $X \rightarrow Y = +$ for one participant vs $X \rightarrow Y = -$ for another). This flexibility encourages theoretical conceptualizations on the relevance of variables in addition to their magnitude and polarity.

References

Arminger, G. (1986). Linear stochastic differential equation models for panel data with unobserved variables. *Sociological Methodology*, *16*, 187–212. https://doi.org/10.2307/270923

Arnold, L. (1974). *Stochastic differential equations: Theory and application* (1st ed.). Wiley Interscience.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Benjamini, Y., & Hochberg, Y. (1997). Multiple hypotheses testing with weights. *Scandinavian Journal of Statistics*, *24*, 407–418. https://doi.org/10.1111/1467-9469.00072

Boker, S. M., & Graham, J. (1998). A dynamical systems analysis of adolescent substance abuse. *Multivariate Behavioral Research*, *33*, 479–507. https://doi.org/10.1207/s15327906mbr3304_3

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., Spies, J., Estabrook, R., Kenny, S., Bates, T., Mehta, P., & Fox, J. (2011). Openmx: An open source extended structural equation modeling framework. *Psychometrika*, *76*, 306–317. https://doi.org/10.1007/s11336-010-9200-6

Bollen, K. A., Harden, J. J., Ray, S., & Zavisca, J. (2014). Bic and alternative Bayesian information criteria in the selection of structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*, 1–19. https://doi.org/10.1080/10705511.2014.856691

Ceulemans, E., & Kiers, H. A. (2006). Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method. *The British Journal of Mathematical and Statistical Psychology*, *59*, 133–150. https://doi.org/10.1348/000711005X64817

Chen, M., Chow, S., Hunter, M. D., & van Montford, K. (2018). Stochastic differential equation models with time-varying parameters. In. J. H. L. Oud, & M. C. Voelkle (Eds.), *Continuous-time modeling in the behavioral and related sciences* (pp. 205–238). Springer International Publishing. https://doi.org/10.1007/978-3-319-77219-6_9

Chow, S.-M., Ho, M.-H. R., Hamaker, E. J., & Dolan, C. V. (2010). Equivalences and differences between structural equation and state-space modeling frameworks. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*, 303–332. https://doi.org/10.1080/10705511003661553

Chow, S.-M., Losardo, D., Park, J., & Molenaar, P. (2022). *Continuous-time dynamic models: Connections to structural equation models and other discrete-time models. Structural equation modeling: Concepts, issues, and applications*. Guilford.

Chow, S.-M., Lu, Z., Sherwood, A., & Zhu, H. (2016). Fitting nonlinear ordinary differential equation models with random effects and unknown initial conditions using the Stochastic Approximation Expectation Maximization (SAEM) algorithm. *Psychometrika*, *81*, 102–134. https://doi.org/10.1007/s11336-014-9431-z

Chow, S.-M., Ou, L., Ciptadi, A., Prince, E. B., You, D., Hunter, M. D., Rehg, J. M., Rozga, A., & Messinger, D. S. (2018). Representing sudden shifts in intensive dyadic interaction data using differential equation models with regime switching. *Psychometrika*, *83*, 476–510. https://doi.org/10.1007/s11336-018-9605-1

Chow, S.-M., Zu, J., Shifren, K., & Zhang, G. (2011). Dynamic factor analysis models with time-varying parameters. *Multivariate Behavioral Research*, *46*, 303–339. https://doi.org/10.1080/00273171.2011.563697

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

Delsing, M. J., & Oud, J. H. (2008). Analyzing reciprocal relationships by means of the continuous-time autoregressive latent trajectory model. *Statistica Neerlandica*, *62*, 58–82. https://doi.org/10.1111/j.1467-9574.2007.00386.x

Driver, C. C., & Voelkle, M. C. (2018). Hierarchical Bayesian continuous time dynamic modeling. *Psychological Methods*, *23*, 774–799. https://doi.org/10.1037/met0000168

Ebrahimi, O. V., Borsboom, D., Hoekstra, R. H., Epskamp, S., Ostinelli, E. G., Bastiaansen, J. A., & Cipriani, A. (2024). Towards precision in the diagnostic profiling of patients: Leveraging symptom dynamics as a clinical characterisation dimension in the assessment of major depressive disorder. *The British Journal of Psychiatry: The Journal of Mental Science*, *224*, 157–163. https://doi.org/10.1192/bjp.2024.19

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018). The Gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research*, *53*, 453–480. https://doi.org/10.1080/00273171.2018.1454823

Fisher, Z. F., Chow, S.-M., Molenaar, P. C. M., Fredrickson, B. L., Pipiras, V., & Gates, K. M. (2022). A square-root second-order extended Kalman filtering approach for estimating smoothly time-varying parameters [PMID]. *Multivariate Behavioral Research*, *57*, 134–152. https://doi.org/10.1080/00273171.2020.1815513

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, E6106–E6115. https://doi.org/10.1073/pnas.1711978115

Fisher, A. J., Reeves, J. W., Lawyer, G., Medaglia, J. D., & Rubel, J. A. (2017). Exploring the idiographic dynamics of mood and anxiety via network analysis. *Journal of Abnormal Psychology*, *126*, 1044–1056. https://doi.org/10.1037/abn0000311

Gates, K. M., Lane, S. T., Varangis, E., Giovanello, K., & Guskiewicz, K. (2017). Unsupervised classification during time-series model building. *Multivariate Behavioral Research*, *52*, 129–148. https://doi.org/10.1080/00273171.2016.1256187

Gates, K. M., & Molenaar, P. C. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage*, *63*, 310–319. https://doi.org/10.1016/j.neuroimage.2012.06.026

Gates, K. M., Molenaar, P. C., Hillary, F. G., Ram, N., & Rovine, M. J. (2010). Automatic search for fmri connectivity mapping: An alternative to granger causality testing using formal equivalences among sem path modeling, var, and unified sem. *NeuroImage*, *50*, 1118–1125. https://doi.org/10.1016/j.neuroimage.2009.12.117

Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development*, *58*, 80–92.

Hamaker, E. L. (2004). *Time series analysis and the individual as the unit of psychological research*. Universiteit van Amsterdam.

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. (2005). Statistical modeling of the individual: Rationale and application of multivariate stationary time series analysis. *Multivariate Behavioral Research*, *40*, 207–233. https://doi.org/10.1207/s15327906mbr4002_3

Hecht, M., Hardt, K., Driver, C. C., & Voelkle, M. C. (2019). Bayesian continuous-time Rasch models. *Psychological Methods*, *24*, 516–537. https://doi.org/10.1037/met0000205

Hecht, M., Lohmann, J., & Zitzmann, S. (2024). Time-varying continuous-time models. *Manuscript Submitted for Publication*.

Hecht, M., & Zitzmann, S. (2021a). Exploring the unfolding of dynamic effects with continuous-time models: Recommendations concerning statistical power to detect peak cross-lagged effects. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*, 894–902. https://doi.org/10.1080/10705511.2021.1914627

Hecht, M., & Zitzmann, S. (2021b). Sample size recommendations for continuous-time models: Compensating shorter time series with larger numbers of persons and vice versa. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*, 229–236. https://doi.org/10.1080/10705511.2020.1779069

Henry, T. R., Feczko, E., Cordova, M., Earl, E., Williams, S., Nigg, J. T., Fair, D. A., & Gates, K. M. (2019). Comparing directed functional connectivity between groups with confirmatory subgrouping gimme. *NeuroImage*, *188*, 642–653. https://doi.org/10.1016/j.neuroimage.2018.12.040

Hunter, M. D. (2014). *State space dynamic mixture modeling: Finding people with similar patterns of change* [Phd thesis]. University of Oklahoma.

Hunter, M. D. (2024). State space mixture modeling: Finding people with similar patterns of change. *Multivariate Behavioral Research*, *59*, 1253–1269. https://doi.org/10.1080/00273171.2023.2261224

Jones, P. J., Ma, R., & McNally, R. J. (2021). Bridge centrality: A network approach to understanding comorbidity. *Multivariate Behavioral Research*, *56*, 353–367. https://doi.org/10.1080/00273171.2019.1614898

Kim, J., Zhu, W., Chang, L., Bentler, P. M., & Ernst, T. (2007). Unified structural equation modeling approach for the analysis of multisubject, multivariate functional mri data. *Human Brain Mapping*, *28*, 85–93. https://doi.org/10.1002/hbm.20259

Koval, P., Kuppens, P., Allen, N. B., & Sheeber, L. (2012). Getting stuck in depression: The roles of rumination and emotional inertia. *Cognition & Emotion*, *26*, 1412–1427. https://doi.org/10.1080/02699931.2012.667392

Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional inertia and psychological maladjustment. *Psychological Science*, *21*, 984–991. https://doi.org/10.1177/0956797610372634

Lane, S., Gates, K., Fisher, Z., Arizmendi, C., Molenaar, P., Merkle, E., Hallquist, M., Pike, H., Henry, T., Duffy, K., et al. (2024). *gimme: Group iterative multiple model estimation* (R package version 0.7-18) [Computer software]. CRAN.

Lane, S. T., Gates, K. M., Pike, H. K., Beltz, A. M., & Wright, A. G. (2019). Uncovering general, shared, and unique temporal patterns in ambulatory assessment data. *Psychological Methods*, *24*, 54–69. https://doi.org/10.1037/met0000192

Li, Y., Wood, J., Ji, L., Chow, S.-M., & Oravecz, Z. (2022). Fitting multilevel vector autoregressive models in stan, jags, and mplus. *Structural Equation Modeling: a Multidisciplinary Journal*, *29*, 452–475. https://doi.org/10.1080/10705511.2021.1911657

Liu, S., Ou, L., & Ferrer, E. (2021). Dynamic mixture modeling with dynr. *Multivariate Behavioral Research*, *56*, 941–955. https://doi.org/10.1080/00273171.2020.1794775

Lohmann, J. F., Zitzmann, S., & Hecht, M. (2024). Studying between-subject differences in trends and dynamics: Introducing the random coefficients continuous-time latent curve model with structured residuals. *Structural Equation Modeling: A Multidisciplinary Journal*, *31*, 151–164. https://doi.org/10.1080/10705511.2023.2192889

Lohmann, J. F., Zitzmann, S., Voelkle, M. C., & Hecht, M. (2022). A primer on continuous-time modeling in educational research: An exemplary application of a continuous-time latent curve model with structured residuals (ct-lcm-sr) to pisa data. *Large-Scale Assessments in Education*, *10*, 5. https://doi.org/10.1186/s40536-022-00126-8

Lundh, L.-G. (2015). The person as a focus for research—The contributions of Windelband, Stern, Allport, Lamiell, and Magnusson. *Journal for Person-Oriented Research*, *1*, 15–33. https://doi.org/10.17505/jpor.2015.03

utkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Lydon-Staley, D. M., Xia, M., Mak, H. W., & Fosco, G. (2019). Adolescent emotion network dynamics in daily life and implications for depression. *Journal of Abnormal Child Psychology*, *47*, 717–729. https://doi.org/10.1007/s10802-018-0474-y

Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research & Perspective*, *2*, 201–218. https://doi.org/10.1207/s15366359mea0204_1

Nestler, S., & Humberg, S. (2021). Gimme's ability to recover group-level path coefficients and individual-level path coefficients. *Methodology*, *17*, 58–91. https://doi.org/10.5964/meth.2863

Oh, H., Hunter, M. D., & Chow, S.-M. (2024). Extending fit indices L¨from structural equation models to state space models [Presentation]. *International Meeting of the Psychometric Society (IMPS)*.

Oravecz, Z., & Tuerlinckx, F. (2011). The linear mixed model and the hierarchical Ornstein–Uhlenbeck model: Some equivalences and differences. *The British Journal of Mathematical and Statistical Psychology*, *64*, 134–160. https://doi.org/10.1348/000711010X498621

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*, 91–111. https://doi.org/10.32614/rj-2019-012

Park, J. J., Chow, S. M., Epskamp, S., & Molenaar, P. C. M. (2024). Subgrouping with chain graphical VAR models. *Multivariate Behavioral Research, 59*, 543–565. https://doi.org/10.1080/00273171.2023.2289058

Park, J. J., Chow, S.-M., Fisher, Z. F., & Molenaar, P. C. (2020). Affect and personality: Ramifications of modeling (non-)directionality in dynamic network models. *European Journal of Psychological Assessment*, *36*, 1009–1023. https://doi.org/10.1027/1015-5759/a000612

Park, J. J., Fisher, Z. F., Chow, S.-M., & Molenaar, P. C. M. (2023). Evaluating discrete time methods for subgrouping continuous processes. *Multivariate Behavioral Research*, *59*, 1240–1252. https://doi.org/10.1080/00273171.2023.2235685

Pe, M. L., Kircanski, K., Thompson, R. J., Bringmann, L. F., Tuerlinckx, F., Mestdagh, M., Mata, J., Jaeggi, S. M., Buschkuehl, M., Jonides, J., Kuppens, P., & Gotlib, I. H. (2015). Emotion-network density in major depressive disorder. *Clinical Psychological Science*, *3*, 292–300. https://doi.org/10.1177/2167702614540645

Piccirillo, M. L., & Rodebaugh, T. L. (2019). Foundations of idiographic methods in psychology and applications for psychotherapy. *Clinical Psychology Review*, *71*, 90–100. https://doi.org/10.1016/j.cpr.2019.01.002

Wang, Y., Yan, G., Wang, X., Li, S., Peng, L., Tudorascu, D. L., & Zhang, T. (2023). A variational Bayesian approach to identifying whole-brain directed networks with FMRI data. *The Annals of Applied Statistics*, *17*, 518–538. https://doi.org/10.1214/22-AOAS1640

Whittaker, T. A. (2012). Using the modification index and standardized expected parameter change for model modification. *The Journal of Experimental Education*, *80*, 26–44. https://doi.org/10.1080/00220973.2010.531299

Wright, A. G., Gates, K. M., Arizmendi, C., Lane, S. T., Woods, W. C., & Edershile, E. A. (2019). Focusing personality assessment on the person: Modeling general, shared, and person specific processes in personality and psychopathology. *Psychological Assessment*, *31*, 502–515. https://doi.org/10.1037/pas0000617

Runyan, W. M. (1983). Idiographic goals and methods in the study of lives. *Journal of Personality*, *51*, 413–437. https://doi.org/10.1111/j.1467-6494.1983.tb00339.x

Ryan, O., & Hamaker, E. L. (2022). Time to intervene: A continuous-time approach to network analysis and centrality. *Psychometrika*, *87*, 214–252. https://doi.org/10.1007/s11336-021-09767-0

Ryan, O., Kuiper, R. M., & Hamaker, E. L. (2018). A continuous-time approach to intensive longitudinal data: What, why, and how? In *Continuous time modeling in the behavioral and related sciences*. (pp. 27–54). Springer.

Schuurman, N. K., & Hamaker, E. L. (2019). Measurement error and person-specific reliability in multilevel autoregressive modeling. *Psychological Methods*, *24*, 70–91. https://doi.org/10.1037/met0000188

Schweppe, F. (1965). Evaluation of likelihood functions for Gaussian signals. *IEEE Transactions on Information Theory*, *11*, 61–70. https://doi.org/10.1109/TIT.1965.1053737

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*, 333–343. https://doi.org/10.1007/BF02294360

Sörbom, D. (1989). Model modification. *Psychometrika*, *54*, 371–384. https://doi.org/10.1007/BF02294623

Tein, J.-Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural Equation Modeling: a Multidisciplinary Journal*, *20*, 640–657. https://doi.org/10.1080/10705511.2013.824781

Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics*, *27*, 77–83. https://doi.org/10.3102/10769986027001077

Van Montfort, K., Oud, J. H., & Voelkle, M. C. (2018). *Continuous time modeling in the behavioral and related sciences*. Springer.

Voelkle, M. C., Oud, J. H., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological Methods*, *17*, 176–192. https://doi.org/10.1037/a0027543