Journal of Mathematical Psychology 129 (2026) 102984



Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp



Examining the feasibility of quadratic vector autoregressions for psychological intensive longitudinal data

Jingmeng Cui [a,b,*], Anna Lichtwarck-Aschoff [a], Fred Hasselman [c]

[a] Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, the Netherlands
[b] Research Group of Quantitative Psychology and Individual Differences, KU Leuven, Leuven, Belgium
[c] Behavioural Science Institute, Radboud University, Nijmegen, the Netherlands

ARTICLE INFO

Keywords:
Nonlinear dynamics
Vector autoregressive model
Quadratic regression
Network analysis

ABSTRACT

Nonlinear relationships among variables play an important role in psychological modeling and understanding changes over time from intensive longitudinal data (ILD). Most methods focus on linear relationships, with a few exceptions developed for specific nonlinear interactions or general system dynamics. Methods considering multiple possible nonlinear relationships among all variables remain rare, hindered by challenges like overfitting and interpretation difficulties. This article examines the feasibility of applying a quadratic vector autoregression method to psychological ILD, using the Regularization Algorithm under Marginality Principle (RAMP) alongside a local linearization method for interpretation. We evaluated its performance with simulated and empirical datasets using classification metrics, information criteria, and cross-validation. Results show this method often requires a long time series for satisfactory performance. While information criteria favor the quadratic model in empirical datasets, cross-validation favors simpler AR models. Nonetheless, these two challenges are comparable to those in idiographic linear VAR models. Clear evidence of nonlinear relationships among variables supports the value of this method for exploratory studies. We developed an R package, quadVAR, as an implementation of this method.

1. Introduction

Dynamics, how variables change over time, is at the heart of many psychological inquiries, such as development, affective, and treatment processes (Bogartz, 1994; Boker & Moulder, 2023; Granic, 2005; Hamaker et al., 2015; Myin-Germeys & Kuppens, 2021). Empirical studies have demonstrated various complex behaviors of systems, such as multistability, sensitive dependence on initial conditions, and sudden changes (Granic et al., 2018; Hayes et al., 2015; Loossens et al., 2020; Moulder et al., 2022; Olthof et al., 2020, 2023; Richardson et al., 2017; Tang & DeRubeis, 1999). Linear dynamics cannot generate those behaviors, as linear interactions among variables can only create one equilibrium point (Cui et al., 2023; Haslbeck et al., 2022; Haslbeck & Ryan, 2022; Hoekstra et al., 2024; Sayama, 2015). Therefore, to gain a better understanding of psychological dynamics, it is essential to move beyond linear statistics and start investigating the nonlinear part of dynamic interactions.

In the cross-sectional context, it is not a new topic to go beyond linear models and investigate nonlinear relationships among variables (Cohen et al., 2003; Failenschmid et al., 2025; Guastello, 1982; Kroc & Olvera Astivia, 2023; Rigdon et al., 1998). Linear methods (e.g., Pearson correlation, linear regression) are commonly used and provide straightforward interpretations, but they assume a stable relationship of variables throughout the variables' range. This assumption is often unrealistic, as many relationships in psychology are known to be nonlinear, for example, floor and ceiling effects (Liu & Wang, 2021; L. Wang et al., 2008), or the inverted-U curve for arousal and performance (Yerkes & Dodson, 1908). If those nonlinear effects are ignored, this can lead to false conclusions. Take the inverted-U curve as an example: there is an optimal arousal level for the best performance. When arousal is too high or too low, performance deteriorates. However, if a linear method is used, the researcher may draw one of the following three conclusions depending on the sample: the higher the arousal, the better the performance; the lower the arousal, the better the performance; arousal does not influence performance. In either case, the conclusion is misleading and fails to capture the most important feature of the relationship

This study was not preregistered.

* Corresponding author at: Tiensestraat 102, 3000 Leuven, Belgium
E-mail address: jingmeng.cui@kuleuven.be (J. Cui).

https://doi.org/10.1016/j.jmp.2026.102984
Received 4 July 2025; Received in revised form 2 March 2026; Accepted 12 March 2026
Available online 25 March 2026
0022-2496/© 2026 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

between arousal and performance. Thus, in cases where nonlinearity is involved, it is better to use a method that can actually capture nonlinear relationships to arrive at more realistic inferences (Kroc & Olvera Astivia, 2023). Examples may include polynomial regression, which adds nonlinear terms to linear regression, and nonparametric techniques, such as splines and local regression.

A challenge in dynamic modeling, however, is the balance between the number of variables, the large number of nonlinear forms they may constitute, and the relatively small number of data points usually available in psychological time series. The dynamics of psychological variables are often investigated using intensive longitudinal data (ILD), assessed with experience sampling methods (ESM, also known as ecological momentary assessment, EMA, or ambulatory assessment, AA; Kalokerinos et al., 2026; Myin-Germeys & Kuppens, 2021; Robinaugh et al., 2020; A. A. Stone et al., 2023). Those data rely on the self-reports of experiences, which can be evaluations of internal mental or emotional states, or evaluations of the experience of external events. The number of assessments per participant in ESM studies is mostly around a hundred (Vachon et al., 2019). Estimating complex models from limited data points can lead to severe overfitting issues. Previous studies have shown that even linear VAR models may be overfitting and do not provide more accurate predictions than single-variate AR models in typical psychological settings (Bulteel et al., 2018; Dablander et al., 2020; Mansueto et al., 2023). Adding a large number of nonlinear terms without proper regularization will only make the overfitting problem more severe.

Whereas the majority of ILD modeling work focuses on linear models (e.g., Bringmann et al., 2013, 2015; Epskamp et al., 2018; Hamaker et al., 2018; Jordan et al., 2020; Robinaugh et al., 2020), there are several methods focusing on nonlinear associations. Several methods focus on nonlinear trends over time (Bringmann et al., 2017; Failenschmid et al., 2025), which is an important pathway to tackle nonlinear phenomena in time series data. These methods are, however, not directly relevant to our aim, as they do not explicitly model the relationships among variables. Some methods aim to describe more complex forms of nonlinear interactions from data, focusing on the attractors in the original or reconstructed state space. The Hankel alternative view of Koopman analysis (HAVOK), for example, can mimic the behavior of dynamic systems with complex mathematical forms (Moulder et al., 2022); the fitlandr method can describe the nonlinear dynamics of a system with vector fields and stability landscapes (Cui et al., 2023); recurrence quantification analysis (RQA) and recurrence networks can describe when and how a system revisits similar states (Hasselman & Bosman, 2020; Marwan et al., 2007; Zou et al., 2019). However, those methods do not provide a direct interpretation of the *effects* among variables (i.e., how a certain value in X can predict a value of Y), which is often of interest in psychological research. Some models are closer to the traditional linear vector autoregression framework, but add a few nonlinear terms to it. For example, the moderated VAR (M-VAR) model (Adolf et al., 2017; Bringmann et al., 2024) allows adding an external moderator that can influence the VAR parameters. Adding more variables is challenging because the number of possible terms can increase exponentially as the possible nonlinear forms and the possible variables involved expand. The sparse identification of nonlinear dynamics (SINDy) algorithm, for example, applies a penalized regression for a large number of nonlinear dynamic forms and aims to recover the dynamic equations of a system from time series (Brunton et al., 2016). However, this method has a high requirement on the sampling frequency, length, and precision, making it unsuitable for psychological (self-report) ILD.

Hence, a method that can estimate the coefficients of some nonlinear terms in multivariate dynamic models from empirical time series is needed. Specifically, this method should allow for more balanced characteristics in interpretability, feasibility, and ability to represent nonlinear dynamics than currently available methods. Therefore, in this article, we introduce a new technique, the quadratic vector autoregressive model (quadVAR, accompanied by an R package of the same name), which allows fitting quadratic relationships among multiple variables in psychological ILD. By *quadratic* we mean a polynomial that includes both the linear main effect terms (e.g., x_1, x_2) and the nonlinear interaction terms, that is, both the squared terms of single variables (e.g., x_1^2) as well as all the cross-product terms between different variables (e.g., x_1x_2). We chose quadratic models that are similar to Haslbeck, Borsboom, et al. (2021) and Bringmann et al. (2024) instead of more flexible curves, such as the generalized additive model (GAM) and the kernel smoothing method (Haslbeck, Bringmann, et al., 2021; Wood, 2017), because the number of possible nonlinear terms is already very large in the case of interactions of quadratic polynomials. Consider an assessment of 10 items, the main effect terms ($x_1, x_2, \ldots, x_{10}$) and the interaction terms ($x_1^2, x_1x_2, x_1x_3, \ldots, x_2^2, x_2x_3, \ldots, x_{10}^2$) already have 65 degrees of freedom. Fitting this model for each outcome variable would increase the total degrees of freedom to 650. Considering the amount of data available in psychological studies, the possible terms in the model are already numerous. To compare with, Haslbeck, Bringmann, et al. (2021) only used time as the moderator, which reduces the number of effects to be estimated, and more degrees of freedom can be used to refine the shape of the curves. For the same reason, we did not include higher-order terms and non-polynomial terms, unlike in Brunton et al. (2016). We will revisit the pros and cons of this choice in the Discussion section.

We also start with an *idiographic* method, which means that the time series of each person is analyzed separately, and conclusions are directly drawn at the level of the individual (Molenaar, 2004; Piccirillo & Rodebaugh, 2019; Wright & Woods, 2020). A pragmatic reason for doing so is that several algorithms we introduce later are more feasible to implement for within-person time series, while accommodating different participants will require further technical developments that go beyond the current first step we are taking here. But besides those pragmatic reasons, there is also the more conceptual one, that most psychological process research (e.g., in the developmental, clinical, or affective field, Lazarus et al., 2021; Wright & Woods, 2020) concerns within-person processes. This can be quite different for different individuals, in the sense that people differ significantly in the (nonlinear) relationships among psychological variables (Molenaar, 2004). Under those conditions, when similarity across individuals is relatively limited, modelling them in a single model becomes questionable, and idiographic methods are preferred. A downside of idiographic models is that they require a lot of data per individual, which is often unfeasible, leading to issues of power in statistical inferences (e.g., Mansueto et al., 2023). We will revisit the consequences of using this idiographic approach and the potential for multi-level extensions in the General Discussion.

After choosing the general strategy of idiographic quadratic regression, two further issues remain to be addressed. First, how can the structure of the whole model be constrained? For models with interaction terms, it is desirable that the model structure has a hierarchy guided by the *marginality principle*, which means corresponding linear terms (e.g., x_1) enter the model before interaction terms (e.g., x_1x_2, Hao & Zhang, 2017; Nelder, 1977). Otherwise, an arbitrary change of the zero point of the scale can lead to a change in the number of terms in the resulting model, thus influencing the estimation of model complexity in the penalized regression procedure. For example, if we obtain a model where $y = x_1x_2$, there is only one term chosen for the predictor side, but if we arbitrarily choose another scale, $x_1' = x_1 + 10$, then we have, $y = (x_1' - 10)x_2 = x_1'x_2 - 10x_2$, where the term x_2 appears again. Because of this, the marginality principle states that when higher-order terms are included in a model, all lower terms of their components should also be included, which prevents including or excluding lower-order terms solely based on the scale of the variables. As most psychological scales are interval scales, we do not have reason to believe one particular zero point is superior to another. This makes it important to keep the aforementioned hierarchy in the regression structure. In quadVAR, we deal

with this issue by using the Regularization Algorithm under Marginality Principle (RAMP, Hao et al., 2018) algorithm, which we will introduce later in more detail.

Second, even if a model with nonlinear terms can be estimated, it can be difficult to interpret all the coefficients and draw meaningful conclusions. For linear systems, a widely used approach is the network representation, where the width of edges represents the relationships among variables. If the relationship is nonlinear, a large number of hyperedges (i.e., edges connecting three or more nodes) may be needed to describe the full dynamics, which makes inference and visualization more complicated (Bringmann et al., 2024; Haslbeck, Borsboom, et al., 2021). To deal with this issue, we adapted the proposal by Kroc and Olvera Astivia (2023) and propose using linearized networks for interpretations. We will also introduce this method in more detail in later sections.

The remainder of this article is structured as follows. First, we introduce the quadVAR method, which includes the RAMP algorithm for quadratic regressions and the local linearized networks for interpretation. After that, we use simulated datasets and empirical datasets to examine the performance of quadVAR. With simulated datasets, we examine the performance based on information criterion, cross-validation, and inference based on the true model. With the empirical datasets, we show how the method performs in real life. Finally, we discuss the pros and cons of using quadratic autoregressive models for psychological ILD and possible future developments in modeling nonlinear dynamics.

2. The quadVAR method

In this section, we introduce the quadVAR method, which includes an estimation part using RAMP and an interpretation part using local linearized networks. We will explain the motivation, specification, and advantages of the two parts separately.

2.1. RAMP: quadratic regularized regression with hierarchical structure

The core algorithm of model estimation in quadVAR is based on the RAMP algorithm (Hao et al., 2018). High-dimensional regression with nonlinear terms is a complex statistical question by itself (Hao & Zhang, 2017). Many statistical methods aimed at addressing this problem have been developed in the past decade, mainly in the context of bioinformatics. Among those methods, RAMP is a recent one that complies with the marginality principle, and previous research has shown that it performs similarly or better than other comparable models (e.g., hierNet, Bien et al., 2013, and iFOR, Hao & Zhang, 2014) with high computational efficiency and an implementation in R (Hao et al., 2018; Luo & Chen, 2021; C. Wang et al., 2021). We briefly introduce the algorithm as follows. A linear vector autoregressive model with p variables can be formulated as

$$\boldsymbol{Y}_t = \boldsymbol{c} + \mathrm{B}\boldsymbol{Y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{1}$$

where $\boldsymbol{Y}_t$ is the p-dimensional vector representing the variables at time t, $\boldsymbol{c}$ is a vector of the same length representing the constant term, B is a $p \times p$ coefficient matrix, and $\boldsymbol{\varepsilon}_t$ is a p-dimensional vector of error terms at time t. To simplify the notations, we only consider the regressive model for the i-th outcome variable (i.e., the i-th component of the p-dimensional vector). Note that here i indexes variables, not individuals (as is common in mixed-effect models). All the estimations are independent of each other, so we can always first get the regression model for each variable as the outcome variable and then combine them together to constitute the full VAR model.

$$Y_{t,i} = c_i + \boldsymbol{\beta}_i^\top \boldsymbol{Y}_{t-1} + \varepsilon_{t,i}. \tag{2}$$

The standard ordinary least squares (OLS) estimation for this regression model with T observations is then given by

$$(\widehat{c_i}, \widehat{\boldsymbol{\beta}_i}) = \underset{c_i,\boldsymbol{\beta}_i}{\operatorname{argmin}} \sum_{t=2}^{T} \left(y_{t,i} - c_i - \boldsymbol{\beta}_i^\top \boldsymbol{y}_{t-1}\right)^2, \tag{3}$$

where lowercase y denotes the observed realizations of Y. As shown above, there are $p+1$ parameters to be estimated for a single outcome variable, thus $p(p+1)$ parameters for the whole VAR model, which is often a large number compared to the number of observations. Therefore, to increase the parsimony of the model and prevent overfitting, a shrinkage method, namely the least absolute shrinkage and selection operator (LASSO), is often used (e.g., Epskamp & Fried, 2018; Haslbeck et al., 2021). This method has a penalty term for the coefficients,

$$(\widehat{c_i}, \widehat{\boldsymbol{\beta}_i}) = \operatorname{argmin}_{c_i,\beta_i} \sum_{t=2}^{T} \left(y_{t,i} - c_i - \boldsymbol{\beta}_i^\top \boldsymbol{y}_{t-1}\right)^2 + \lambda_i ||\boldsymbol{\beta}_i||_1, \tag{4}$$

which also does a variable selection, that the coefficients of unimportant terms will become zero, making the model more parsimonious. If we add interaction terms, the regression model becomes:

$$Y_{t,i} = c_i + \boldsymbol{\beta}_{\boldsymbol{Mi}}^\top \boldsymbol{Y}_{t-1,i} + \boldsymbol{\beta}_{\boldsymbol{Ii}}^\top \boldsymbol{Y}_{t-1} \circ \boldsymbol{Y}_{t-1} + \varepsilon_{t,i}, \tag{5}$$

where we follow the notation by Hao et al. (2018) and define $\boldsymbol{Y}_t \circ \boldsymbol{Y}_t$ as all the pairwise products of $\boldsymbol{Y}_t$, namely $\left(Y_{t,1}Y_{t,1}, Y_{t,1}Y_{t,2}, \ldots, Y_{t,p}Y_{t,p}\right)^T$, which is a vector with $p(p+1)/2$ dimensions, $\boldsymbol{\beta}_{\boldsymbol{Mi}}$ is a vector of length p containing the coefficients of main effect terms, and $\boldsymbol{\beta}_{\boldsymbol{Ii}}$ is a vector of length $p(p+1)/2$ containing the coefficients of all interaction terms. Taken together, there are $p(p+1) + p^2(p+1)/2$ coefficients to be estimated for the whole quadVAR model, and the possible coefficients for interaction terms are much more than the coefficients of the main effects. If we use LASSO for all the main effects and interaction terms altogether, two issues will arise. First, as the number of interaction terms is much larger than the number of main effect terms, interaction terms are more likely to be selected in the final model, assuming other factors (such as effect sizes) are comparable. This negatively affects model parsimony because we would only want to add interaction terms and complicate the model when interaction terms are necessary. Second, if we add interaction terms without corresponding main effects, the structure of the model is subject to change with a simple coding transformation, as the marginality principle we explained above.

The RAMP algorithm solves those problems with special treatment for interaction terms and hierarchical structure between the main effects and interactions. We summarize the detailed steps of RAMP in Box 1.

Specifically, the model is built with a sequence of penalty parameters, in descending order ($\lambda_{i,1} > \lambda_{i,2} > \ldots > \lambda_{i,K}$). In each step k, all the main effects that have been involved in the interactions in the step $k-1$ will not be penalized, and the other main effects that have not been involved in the interactions in the step $k-1$, together with the permitted interactions based on the main effects that are already in the equation in the step $k-1$, may enter the equation subject to the penalty parameter of the current step. Written in statistical representation:

$$\begin{aligned}&\left(\widehat{c_{i,k}}, \widehat{\boldsymbol{\beta}_{Mi,k}}, \widehat{\boldsymbol{\beta}_{Ii,k}}\right)\\ &= \operatorname{argmin}_{c_{i,k},\boldsymbol{\beta}_{Mi,k},\boldsymbol{\beta}_{Ii,k}} \sum_{t=2}^{T} \left(Y_{t,i} - c_i - \boldsymbol{\beta}_{Mi,k}^\top \boldsymbol{Y}_{t-1,i} - \boldsymbol{PI}_{\boldsymbol{\beta}_{Mi,k-1}}\left(\boldsymbol{\beta}_{Ii,k}\right)^\top \boldsymbol{Y}_{t-1} \circ \boldsymbol{Y}_{t-1}\right)^2\\ &\quad + \lambda_{i,k}\left(\left|\left|\boldsymbol{OM}_{\hat{\boldsymbol{\beta}}_{Ii,k-1}}\left(\boldsymbol{\beta}_{Mi,k}\right)\right|\right|_1 + \left|\left|\boldsymbol{\beta}_{Ii,k}\right|\right|_1\right),\end{aligned} \tag{6}$$

where $\boldsymbol{PI}_{\hat{\boldsymbol{\beta}}_{Mi,k-1}}\left(\boldsymbol{\beta}_{Ii,k}\right)$ represents choosing permitted interaction terms from $\boldsymbol{\beta}_{Ii,k}$ based on $\widehat{\boldsymbol{\beta}_{Mi,k-1}}$, and $\boldsymbol{OM}_{\hat{\boldsymbol{\beta}}_{Ii,k-1}}\left(\boldsymbol{\beta}_{Mi,k}\right)$ represents choosing other main effects from $\boldsymbol{\beta}_{Mi,k}$ that did not participate in $\widehat{\boldsymbol{\beta}_{Ii,k-1}}$. Note that $\boldsymbol{PI}$ and $\boldsymbol{OM}$ are selection operators. They do not rescale the coefficients, but determine which terms are included in the coefficients and penalty

Box 1
The RAMP procedure (Hao et al., 2018) illustrated in the context of the quadVAR package.

1. For each outcome variable i in 1 to p:

1.1. Create a list of penalty parameters $\lambda_{i,1} > \lambda_{i,2} > \ldots > \lambda_{i,K}$. By default, the λ values are chosen following the procedure implemented in the widely used R package glmnet (Friedman et al., 2010): $\lambda_{i,1}$ is the theoretically smallest value that shrinks all the coefficients to zero, $\lambda_{i,K} = \varepsilon\lambda_{i,1}$ where $\varepsilon = 0.01$ when there are more predictors than observations, and $\varepsilon = 1 \times 10^{-4}$ otherwise. The default value of K is 100, and the sequence is spaced logarithmically.

1.2. For each k in 1 to K:

1.2.1. Get the main effect terms selected in the previous $(k-1)$ step, defined as non-zero terms in $\widehat{\boldsymbol{\beta}_{Mi,k-1}}$, and the interaction terms selected in the previous step, $\widehat{\boldsymbol{\beta}_{Ii,k-1}}$. (If this is the first step, $k = 1$, then run as if no main effects or interaction terms were selected in the previous step.)

1.2.2. Get the operators to select the permitted interaction terms based on $\widehat{\boldsymbol{\beta}_{Mi,k-1}}$: $\boldsymbol{PI}_{\hat{\boldsymbol{\beta}}_{Mi,k-1}}$ Get the other main effect terms that do not participate in $\widehat{\boldsymbol{\beta}_{Ii,k-1}}$: $\boldsymbol{OM}_{\hat{\boldsymbol{\beta}}_{Ii,k-1}}$.

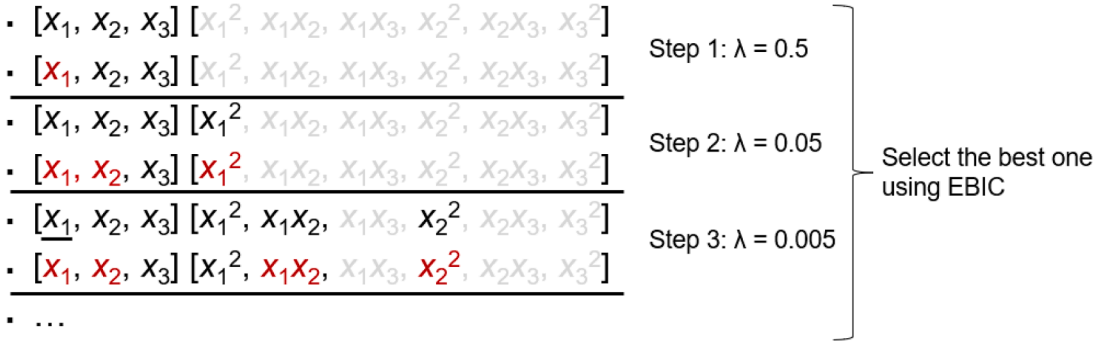
1.2.3. Estimate the coefficients of the step k, $\widehat{c_{i,k}}, \widehat{\boldsymbol{\beta}_{Mi,k}}, \widehat{\boldsymbol{\beta}_{Ii,k}}$, using Equation 6. Compared to regular LASSO, only the permitted interaction terms selected by $\boldsymbol{PI}_{\hat{\boldsymbol{\beta}}_{Mi,k-1}}$ can have non-zero coefficients, and only the main effect terms that do not participate in the previous step, selected by $\boldsymbol{OM}_{\hat{\boldsymbol{\beta}}_{Ii,k-1}}$ are regularized.

1.2.4. Calculate the EBIC value ($\text{EBIC}_{i,k}$).

1.3. Find the best model by finding the minimum of $\text{EBIC}_{i,1}$, $\text{EBIC}_{i,2}$, …, $\text{EBIC}_{i,K}$.

2. Combine the best models for all outcome variables from 1 to p as the final output.

An illustrative example is shown below, where the terms in black are eligible to be selected in each step, the terms in red are actually selected, and the terms with an underline are not regularized.



terms.

With different $\lambda_{i,k}$ values, a set of candidate models is generated for the i-th outcome variable. All the models are then evaluated using extended Bayesian information criteria (EBIC; Chen & Chen, 2008) to find the model that performs the best. EBIC is a variation of BIC with the model searching space considered. The original BIC only considers the number of parameters estimated in the penalty term. This is problematic when a large number of models are estimated and compared because when more parameters are estimated, there are often a large number of possible models. It is likely that some of them, by chance, have a better fit. This multiplicity of models is not accounted for by BIC, but it is explicitly addressed by EBIC. Hao et al. (2018) have shown that for the RAMP algorithm, using AIC and BIC would lead to overly complex models with similar out-of-sample prediction performance. By penalizing the model searching space, EBIC can better ensure the model parsimony and provides a better estimate of the model performance in new datasets. The EBIC metric has also been used in most of the common linear, binary, and mixed network estimation methods, as the standard approach (Epskamp et al., 2012; Epskamp & Fried, 2018; Haslbeck & Waldorp, 2020; van Borkulo et al., 2014).

It is worth noting that, because the RAMP algorithm follows the marginality principle, it is not able to identify the variables that have a purely quadratic relationship (i.e., the linear coefficient is zero). This may happen, for example, when the sample is distributed symmetrically about the axis of the parabola. We will further elaborate on the implications of this issue in the General Discussion.

2.2. Local linearized networks

After obtaining the estimation for the quadratic regression model, we still need to know what inferences we can draw from the estimation result. Drawing inferences from a model with nonlinear terms can be tricky because, as in analyzing interactions for ANOVA, the size of an effect may be dependent on the value of another variable or the variable itself. Here, we take the proposal by Kroc and Olvera Astivia (2023) to use a linearization method. Technically, it takes the partial derivative of the regression as the size of an effect. This means that given all other variables held constant, how much the expected change of the outcome variable is given a unit increase of the input variable. All the pairwise effects can be visualized as a network, with the weight of edges representing the effects, the same as linear longitudinal networks (Bringmann et al., 2013). For linear VAR models, the size of the effect is always the same, independent of the variable values, and equal to the corresponding coefficient. In contrast, for quadVAR models, the effects may

change with the variable values, which means the variable values will change the structure of the autoregressive networks. We provide an example of the linearization procedure in Box 2.

3. Performance of quadVAR: a simulated system

3.1. Simulation specifications

To give a general illustration and examine the performance of quadVAR, we first show its performance based on a simulated system. We use a nonlinear, bistable emotional model adapted from van de Leemput et al. (2014), which is an extension of the Lotka–Volterra equation. The basic assumption of the model is that there are two positive emotions and two negative emotions in an emotional system. The positive emotions strengthen each other, weaken the negative emotions, and vice versa. Many previous studies have used this model to illustrate nonlinear dynamics and bimodality in psychological ILD (e.g., Cui et al., 2023; Haslbeck & Ryan, 2022). The original model is described in continuous time (i.e., using differential equations), yet in the current work, we rewrite it as a discrete time version (i.e., using difference equations) because the quadVAR model we introduce is a discrete time model. If we estimate a discrete-time model from a continuous-time system, the coefficients do not have a one-on-one relationship, and it is difficult to obtain the true coefficients (Ryan & Hamaker, 2022).

The model we use is specified as follows:

$$
\begin{aligned}
x_{1,t+1} &= 0.8 + 1.5x_{1,t} - 0.1x_{1,t}^2 + 0.02x_{1,t}x_{2,t} - 0.1x_{1,t}x_{3,t} - 0.1x_{1,t}x_{4,t} + \varepsilon_{1,t},\\
x_{2,t+1} &= 0.8 + 1.5x_{2,t} - 0.1x_{2,t}^2 + 0.02x_{1,t}x_{2,t} - 0.1x_{2,t}x_{3,t} - 0.1x_{2,t}x_{4,t} + \varepsilon_{2,t},\\
x_{3,t+1} &= 0.8 + 1.5x_{3,t} - 0.1x_{3,t}^2 + 0.02x_{3,t}x_{4,t} - 0.1x_{1,t}x_{3,t} - 0.1x_{2,t}x_{3,t} + \varepsilon_{3,t},\\
x_{4,t+1} &= 0.8 + 1.5x_{4,t} - 0.1x_{4,t}^2 + 0.02x_{3,t}x_{4,t} - 0.1x_{1,t}x_{4,t} - 0.1x_{2,t}x_{4,t} + \varepsilon_{4,t}.
\end{aligned}
\tag{7}
$$

In this system, x_1 and x_2 represent two positive emotions, for example, cheerful and content, and x_3 and x_4 represent two negative emotions, for example, sad and anxious (van de Leemput et al., 2014), and ε_i $(i=1,2,3,4)$ are Gaussian noise with $SD = 1$. We first show the networks for the true model in Fig. 1. Since the model is nonlinear, the network structure differs across variable values. We can see that in the neutral situation (Fig. 1a), the influence of positive emotions on negative emotions is as strong as the effect of negative emotions on positive emotions, which is expected because the model has a symmetric structure. Yet, when the system is dominated by positive emotions (Fig. 1b), the strength of the edges becomes asymmetric. The effect of positive emotions on negative emotions is not strong anymore, but negative emotions have a stronger influence on positive emotions. The opposite pattern appears when the system is dominated by negative emotions (Fig. 1c). Such nonlinear effects are key characteristics of nonlinear models and will serve as primary indicators for model evaluation (described later).

Box 2
An example of the linearization procedure (Kroc & Olvera Astivia, 2023) illustrated in the context of the quadVAR package.

Here, we show an illustrative example of the linearization procedure.

Suppose there is a network with three variables x_1, x_2, x_3, their relationships are estimated as follows:

$x_{1,t} = 0.3x_{2,t-1} + 0.1 + \varepsilon_{1,t}$,

$x_{2,t} = x_{2,t-1}^2 + 0.1 + \varepsilon_{2,t}$,

$x_{3,t} = x_{1,t-1}x_{2,t-1} + 0.1 + \varepsilon_{3,t}$.

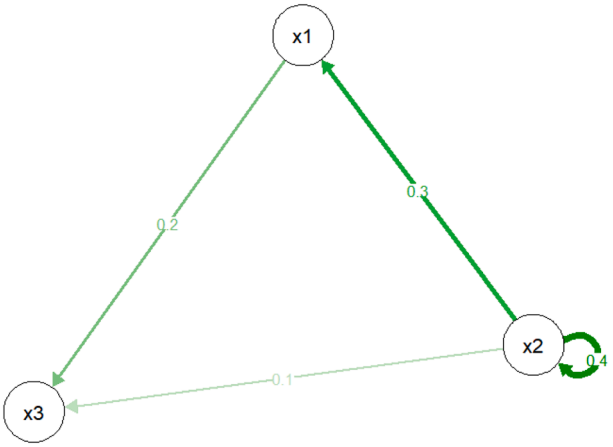
To draw the network for the system's state $x_1 = 0.1$, $x_2 = 0.2$, $x_3 = 0.3$, we have the following calculations:

1. For inbound arrows of x_1: x_1 is only influenced by x_2, and $\mathrm{d}x_{1,t}/\mathrm{d}x_{2,t-1} = 0.3$, so an arrow points from x_2 to x_1 is drawn with the weight of 0.3. (This is the same as in linear networks.)

2. For inbound arrows of x_2: x_2 is influenced by itself, and $\mathrm{d}x_{2,t}/\mathrm{d}x_{2,t-1} = 2x_{2,t-1}$. If we substitute the state of the system ($x_2 = 0.2$), this derivative becomes 0.4, so an arrow points from x_2 to x_2 is drawn with the weight of 0.4.

3. For inbound arrows of x_3: x_3 is influenced by x_1 and x_2, and $\mathrm{d}x_{3,t}/\mathrm{d}x_{1,t-1} = x_{2,t-1} = 0.2$, $\mathrm{d}x_{3,t}/\mathrm{d}x_{2,t-1} = x_{1,t-1} = 0.1$. So, an arrow points from x_1 to x_3 is drawn with the weight of 0.2, and an arrow points from x_2 to x_3 is drawn with the weight of 0.1.

Taken together, the network is drawn as follows:



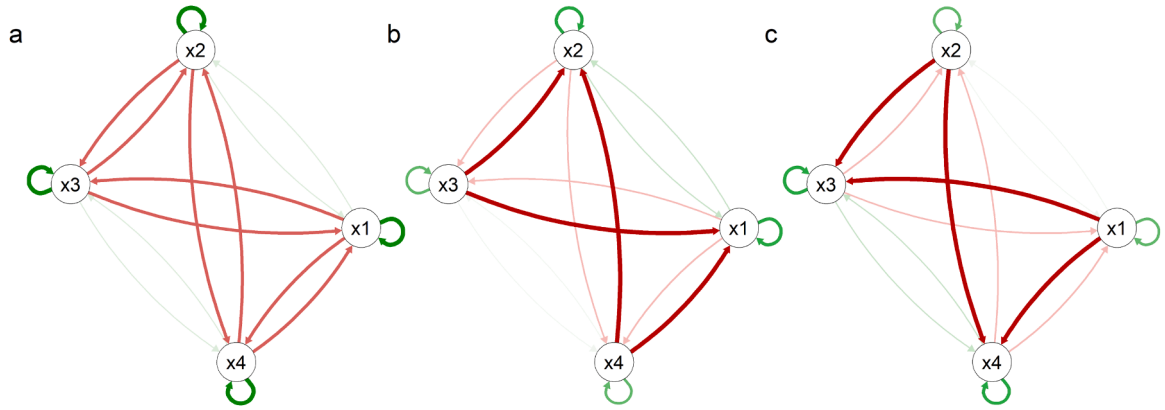


Fig. 1. Linearized networks for the true model for different system states: (a) neutral state, which is the saddle point between two phases: $\boldsymbol{x}_1 = \boldsymbol{x}_2 = \boldsymbol{x}_3 = \boldsymbol{x}_4 = 2.80$; (b) positive state, which is the local minimum of the phase dominated by positive affects, $\boldsymbol{x}_1 = \boldsymbol{x}_2 = 4.89$, $\boldsymbol{x}_3 = \boldsymbol{x}_4 = 1.36$; (c) negative state, which is the local minimum of the phase dominated by negative affects, $\boldsymbol{x}_1 = \boldsymbol{x}_2 = 1.36$, $\boldsymbol{x}_3 = \boldsymbol{x}_4 = 4.89$.

We then simulate the models for different lengths, namely 50, 100, 200, 500, and 1,000 timesteps, and estimate quadVAR models from the simulated datasets. Those data lengths correspond to different scenarios in psychological ILD collection: collecting 50 data points is an entry-level for ILD studies, and collecting 1,000 data points corresponds to the maximum possibility in ILD collection in previous studies (e.g., Delignières et al., 2004; Kossakowski et al., 2017). We first show an example for each simulation length, and the networks estimated thereafter are shown in Fig. 2. From the simulation sample, we can observe that for 50 data points, the bistability of the system is only vaguely observable, whereas for 200 or more data points, the bistability of the system is clearer. Similarly, the estimated networks also become close to the true networks for longer time series. During the simulation trials, we also found an interesting phenomenon that is specific to nonlinear multistable systems: decreasing noise does not always lead to better model estimates; on the contrary, it may make the system stay longer in one phase and reduce information on alternative phases, therefore decreasing model performance. We show an example of this finding in Appendix A.

We then run simulations for each length 1,000 times and evaluate the performance of the quadVAR models in two parts. First, we compare the performance of quadVAR with the NULL models (i.e., the intercept-only models that simply use the mean as the prediction), the AR models, the unregularized VAR models, the regularized VAR models (with the LASSO method implemented in the ncvreg package in R, Breheny & Huang, 2011), and the unregularized quadVAR models, based on the EBIC metrics and cross-validation. These conditions are adapted from the work by Bulteel et al. (2018) and are incremental in model complexity, thus can help to clarify which part of the model is crucial for the overall model performance (e.g., the AR model adds autocorrelations on the NULL models, and the regularized quadVAR model adds regularization on the unregularized quadVAR models). To consider the scenario that the true system is linear, we also simulate a linear system based on a regularized VAR estimation (coefficients reported in Appendix B) of the nonlinear model described above. This makes sure the linear model is comparable with the nonlinear model in variable ranges and correlations. To evaluate the model's performance in predicting unseen new data points, we perform 10-fold block cross-validation based on the suggestions by Bergmeir and Benítez (2012) and Bulteel et al. (2018) and take the mean squared error (MSE) as the performance metric. Although theoretical work has shown that the performance measures based on information criteria will be consistent with the measures based on cross-validation asymptotically (Shao, 1997; M. Stone, 1977), for real-life situations with limited observations, cross-validation may still give different results than information criteria. Most previous studies only used the cross-validation method for evaluating model performances (e.g., Bulteel et al., 2018; Dablander et al., 2020). In the current study, we use both to compare their differences. Second, we calculate the correlation of the estimated nonlinear coefficients from quadVAR and the true nonlinear coefficients, as well as several classification metrics, including precision, recall, specificity, sensitivity, and accuracy (see Appendix C for the definition of those metrics; Mansueto et al., 2023). The correlation between the estimated and true coefficients represents whether stronger true effects are recovered as correspondingly stronger estimates, while the classification metrics assess whether true effects are successfully selected by the model. Compared with information criteria and cross-validation, which focus more on the overall predictive power, the correlation between the estimated and true coefficients and classification metrics focus more on statistical inferences of individual relationships among variables. As quadVAR performs variable selection via LASSO, terms with nonzero coefficients are directly used for calculating classification metrics. No additional statistical significance criteria are used. In addition, we also examine the classification metrics for the key nonlinear effect described above, namely, the effects of positive emotions on negative emotions diminish when positive emotions are dominant (see Fig. 1; note that this is a negative effect, so that a higher value represents a weaker effect). Specifically, we define a true positive when the effect of *cheerful* on *sad* in a positive state ($x_1 = x_2 = 4.89$, $x_3 = x_4 = 1.36$, true effect: -0.136) is higher than the neutral state ($x_1 = x_2 = x_3 = x_4 = 2.80$, true effect: -0.280), and is a false negative otherwise. We also compare the same effect at the positive state ($x_1 = x_2 = 4.89$, $x_3 = x_4 = 1.36$, true effect: -0.136) and the positive-neutral state ($x_1 = x_2 = 4.89$, $x_3 = x_4 = 2.80$, true effect: -0.136), which should be the same for the true system. Therefore, we define a true negative when the effect of *cheerful* on *sad* at the positive state and the positive-neutral state is the same, and a false positive otherwise.

3.2. Results of the simulation study

Assessing model performance with EBIC and cross-validation. As the raw EBIC values are not comparable for different datasets, we report ΔEBIC values instead, which are the raw EBIC values subtracted by the EBIC values of the quadVAR model. The results for the original nonlinear system are shown in Fig. 3a, and the results for the comparable linear system are shown in Fig. 3c. According to the guidance by Burnham et al. (2002), an EBIC difference of 10 or larger is considered essential. We draw this criterion as a red dashed line in Figs. 3a and 4c. From the results, it is clear that quadVAR performs best for the nonlinear simulation model when the input dataset contains 100 or more data points, whereas the VAR model performs best when the actual system is linear. For short time series with a length of 50, the VAR model

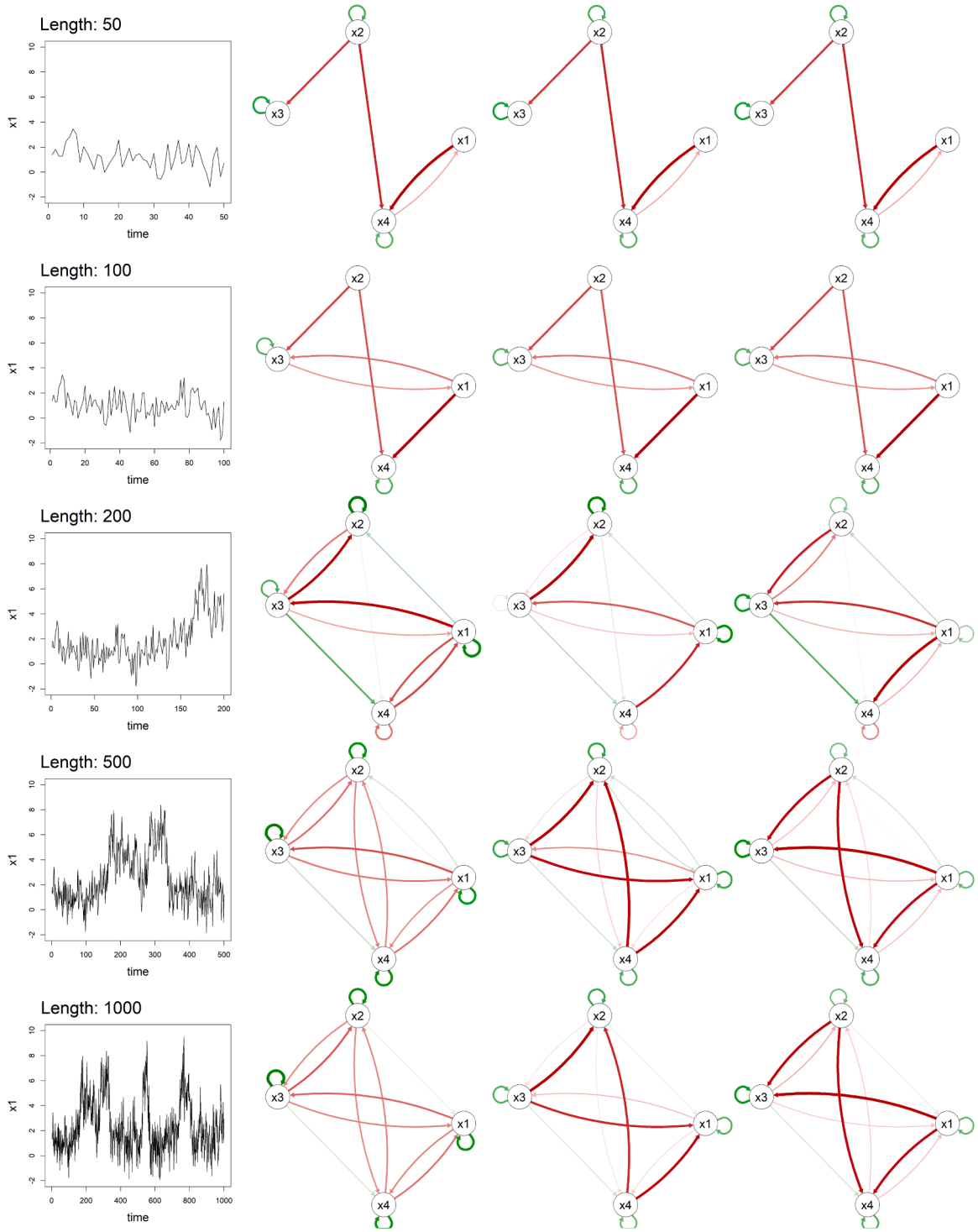


Fig. 2. Linearized networks for the simulation examples with different lengths. In each row, the first plot shows the raw time series of $\boldsymbol{x}_1$, the second to the fourth plots show the linearized networks estimated from simulation data for the neutral state ($\boldsymbol{x}_1 = \boldsymbol{x}_2 = \boldsymbol{x}_3 = \boldsymbol{x}_4 = 2.80$), the positive state ($\boldsymbol{x}_1 = \boldsymbol{x}_2 = 4.89$, $\boldsymbol{x}_3 = \boldsymbol{x}_4 = 1.36$), and the negative state ($\boldsymbol{x}_1 = \boldsymbol{x}_2 = 1.36$, $\boldsymbol{x}_3 = \boldsymbol{x}_4 = 4.89$).

outperforms the quadVAR model. This is in agreement with the results by Bulteel et al. (2018) because more complex models need more data points to be precisely estimated. If the available data points are not enough, the estimation results of the model may be inaccurate, making the performance worse than a simpler, yet mis-specified model. The AR model never performs well, which is different from previous studies (e.g., Bulteel et al., 2018; Dablander et al., 2020), and the unregularized full VAR model performs almost as well as regularized VAR models. This may be because the simulation model we used contains dense mutual influences among variables, and this is also the case for the comparable linear system, which only contains one absent effect. We will revisit how realistic the simulated systems are in later discussions. We show the MSE values from cross-validation with datasets of different lengths in Figs. 3b and 4d. In general, cross-validation MSE shows a similar pattern to the information criteria (Figs. 3a and 4c). One difference is that, when the true system is nonlinear, the unregularized quadVAR model yields

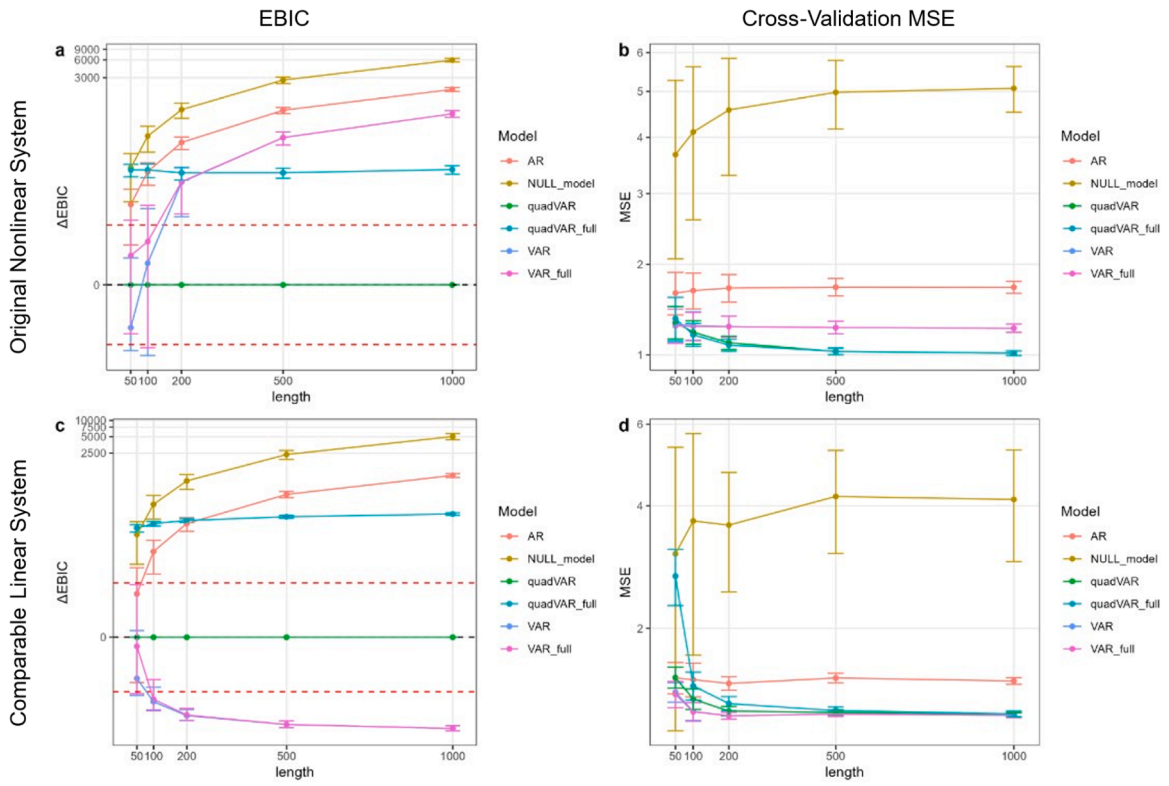


Fig. 3. EBIC and cross-validation MSE comparison for repeated simulations. The first row is the results from the original nonlinear system, and the second row is the results from the comparable linear system. The suffix "_full" represents the unregularized models (i.e., none of the parameters shrink to zero). The error bars represent the standard deviation. Note that to represent large ΔEBIC values, we used pseudo-logarithm scales for the y-axis, which means the scale is logarithmic for large values but smoothly transforms to a linear scale around 0.

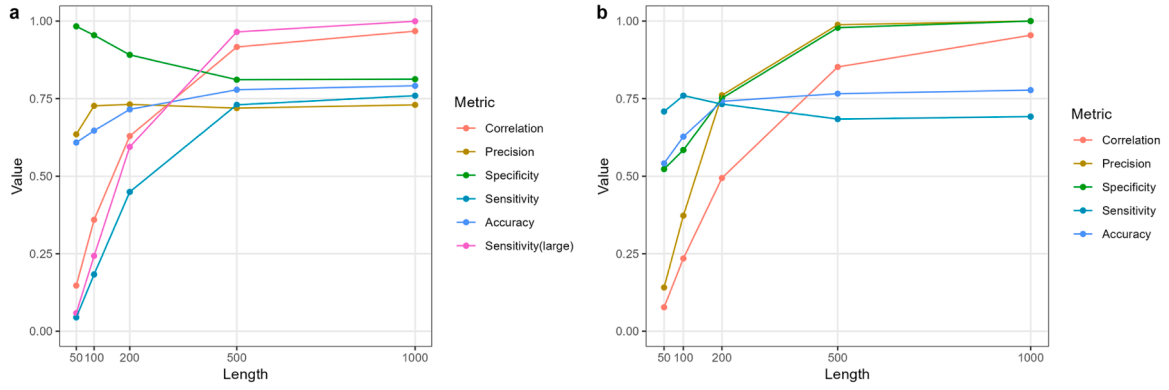


Fig. 4. Classification metrics for the models estimated from simulation data.

similar or even better predictions than the regularized quadVAR model. This indicates that the regularization procedure in quadVAR is more conservative and tends to produce models with higher parsimony, but this does not necessarily lead to a better prediction performance. Overall, both EBIC and cross-validation results indicate that the model that is closer to the true dynamics of the system tends to perform better as the dataset is long enough, and the general performance of quadVAR is satisfactory when modeling sufficient data from nonlinear systems.

3.2.1. Assessing model performance with classification metrics

The classification metrics for the coefficient estimations and the nonlinear effects are summarized in Fig. 4. We first look at Fig. 4a, which shows the classification metrics for nonlinear parameters (i.e., the coefficients for the nonlinear terms in regression). In general, many indicators seem suboptimal for short time series, especially the sensitivity. Because large effects are more important than small effects in real-life scenarios, we also checked the sensitivity for coefficients with an absolute true value larger than or the same as 0.10. As expected, the sensitivity for those coefficients is higher than that for all coefficients, but still rather low for small sample sizes. Precision and specificity are better for small sample sizes, yet the specificity declines as the sample size grows. The accuracy and correlation also increase steadily as the sample size grows. Taking together, although the overall performance of the method increases as the sample size increases, it seems to be more on

the liberal side for larger sample sizes at the coefficient level.

In contrast, the classification metrics for key nonlinear effects (i.e., whether the estimated relationships between variables change in the theoretically correct direction, Fig. 4b) show a different pattern. The sensitivity is rather high for smaller sample sizes and slightly declines as the sample size grows, yet the precision starts low and quickly grows as the sample size increases. The reason for this difference could be that multiple coefficients can lead to similar nonlinear effects. In the true system, all the variables have symmetric forms of dynamic functions. Because the positive-emotion variables (x_1 and x_2) tend to take similar values (as well as the negative-emotion variables x_3 and x_4), misattribution of effects among variables within the same valence (e.g., from $x_1 \rightarrow x_3$ to $x_2 \rightarrow x_3$) may still yield the correct nonlinear effects. Therefore, as the sample size grows, the estimation of the nonlinear effects also improves but leans toward the conservative side.

3.3. Discussion of the simulation study

In this section, we tested quadVAR on a simulated nonlinear system. We found that, overall, quadVAR has an adequate performance. For nonlinear systems, it performs better than linear models both in EBIC and cross-validation when the sample size is not too small. For linear systems, using quadVAR can result in lower EBIC than linear models, as expected. However, the cross-validation result of quadVAR is not much worse than that of the linear models, which indicates that the higher flexibility of quadVAR did not lead to excessive overfitting. For statistical inferences, the overall performance of quadVAR is not entirely satisfactory but is comparable to the individual linear models reported in previous studies (Mansueto et al., 2023), which also indicates that the overfitting issue of quadVAR is well controlled. Although some classification metrics are not very high for usual sample sizes (especially sensitivity for coefficients and precision for nonlinear effects), we would like to emphasize that, as far as we know, quadVAR is the only method to date that allows estimating nonlinear dynamics from individual ILD that provides direct interpretation of the effects among variables. If we also draw the classification metrics of other methods in Fig. 4, they will all be a straight line of zero because they will assume all nonlinear coefficients to be zero. In this sense, we suggest that quadVAR is a suitable method to estimate nonlinear dynamics for exploration purposes, yet the inferences should be drawn with caution. We also note that all the results we reported are only for one simulated system, which, as we show later, is quite different from the dynamic structure estimated from real-life data. Therefore, it is better to consider the simulation study as a proof-of-concept that confirms the order of magnitude of the data needed and illustrates some potential problems that we should take into account for modeling nonlinear models, but not as a direct guideline for data collection. We will revisit this issue after showing the results from empirical data.

4. Performance of quadVAR: empirical datasets

4.1. Description of the empirical datasets

In this section, we assess the performance of quadVAR using empirical datasets. As the true system is unknown, we cannot use classification metrics to evaluate the model's performance. Therefore, we focus on cross-validation and EBIC results. Apart from the general performance of the models, we also show several examples to draw nonlinear inferences from quadVAR estimations. We note that the following results are mainly used for examining the method instead of drawing empirical conclusions, and the analysis procedures were not preregistered.

We used two datasets from previous studies. The datasets were chosen with three criteria. First, the datasets should contain enough data points for each individual, preferably around or more than 200 data points. This is based on our simulation results, which show that the performance of quadVAR requires relatively long time series, and that time series below 100 data points per person are likely to result in poor performance. Second, the assessment should be done on a visual analog scale (VAS) instead of a Likert scale, because we consider a continuous scale is needed to adequately estimate nonlinear influences, yet a Likert scale would be too discrete and restricted to allow nonlinear features to be presented. Third, the datasets should be openly available (or easily available upon request) so that the results can be easily reproduced.

We found two datasets that suffice for the above criteria. The first one is the dataset from Rowland and Wenzel (2020). This dataset was especially interesting because Haslbeck et al. (2023) reported that this dataset contains strong bimodality that cannot be recovered using the VAR model. Therefore, by using this dataset, we can examine whether involving nonlinear terms would help to describe datasets with nonlinear features. The dataset contains 125 participants, and each of them answered 8 questions about their current affective states[1] 6 times per day in a 0–100 VAS for 40 consecutive days. On average, the participants had 172.56 valid responses (ranging from 76 to 226, SD = 35.13). To avoid the influence of different time intervals, all the overnight lags and lags with missing data in between were removed before the modeling procedure. The effective data length of the participants then had a mean of 114.06 (ranging from 30 to 180, SD = 36.83).

The second one is the dataset from Kuppens et al. (2010), requested from the EMOTE database (Kalokerinos et al., 2026) under request number 3UUCEGX6EU. In this study, each participant answers 8 questions about their affective state and emotional regulation strategies. On average, the participants had 173.6 valid responses (ranging from 46 to 211, SD = 26.54). As the two datasets showed similar patterns, we focus on the results from the first dataset in the main text and report the results from the second dataset in Appendix D.

For cross-validation, we winsorized the prediction within the possible range of answering to account for the restriction of the scale range (e.g., if a model prediction is 150 but the original data was collected using a scale from 0 to 100, we changed the prediction to 100). All other settings for the cross-validation are the same as in the simulation study reported in the previous section.

4.2. Results from the empirical datasets

General model performance by EBIC and cross-validation. The EBIC and cross-validation results for the dataset from Rowland and Wenzel (2020) are summarized in Fig. 5. For empirical data, the model selection results by EBIC and cross-validation have very different patterns. For the dataset from Rowland and Wenzel (2020), if we use EBIC for model selection, the quadVAR model is the best one for 69 out of 125 participants (55.20 %), followed by the AR model (48 participants, 38.40 %), the VAR model (5 participants, 4.00 %), the NULL model (2 participants, 1.60 %), and the unregularized quadVAR model (1 participant, 0.80 %)[2]. In general, the AR model and the quadVAR models perform the best, and their relative model performances are related to the effective data length. The correlation between the effective data length and the relative EBIC of the AR model is .19 (p = .03, 95 % CI [0.02, 0.36]), indicating that quadVAR performs better with longer datasets. However, if we use cross-validation for model selection, the AR model is the best for 105 participants (84.00 %), followed by the

[1] Those affective states include happy, excited, relaxed, satisfied, angry, anxious, depressed, and sad.

[2] Those numbers do not sum up to 125 participants because for some participants multiple procedures had the same model selected and they are tied for the best.

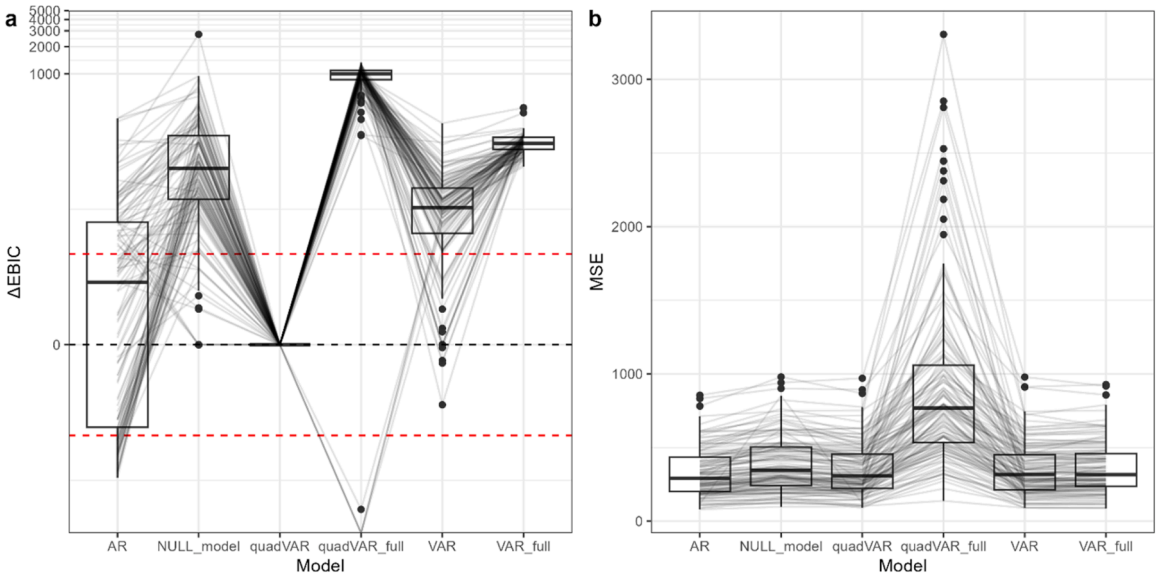


Fig. 5. EBIC and cross-validation MSE for participants in Rowland and Wenzel (2020). Gray lines represent individual participants. In Fig. 6a, there are 4 participants for whom the unregularized quadVAR models were overspecified. The EBIC values for the unregularized quadVAR models were thus negative infinity and not shown in the figure.

unregularized VAR model (9 participants, 7.20 %), the VAR model (5 participants, 4.00 %), the NULL model (5 participants, 4.00 %), and the quadVAR model (2 participants, 1.60 %).[3] This strong discrepancy between EBIC and cross-validation is not expected from either the theory (Shao, 1997; M. Stone, 1977) or the simulation study we performed in the previous section. The quadVAR model performs well according to the EBIC, but less so according to cross-validation. The performance of the AR model is much better with cross-validation than with EBIC. The NULL model also performs better in cross-validation than EBIC. Interestingly, even the best-performing AR model is not much superior to the NULL model for this dataset, which indicates that time series modeling in general does not provide substantial improvement over using the mean for prediction. It is also worth noting that the quadVAR models do not have more degrees of freedom compared to the VAR models. The quadVAR models, on average, have 10.07 degrees of freedom (SD = 8.40), and the VAR models have 11.78 degrees of freedom (SD = 9.43) for the dataset from Rowland and Wenzel (2020). To compare with, the AR models have 8 degrees of freedom, and there are 576 candidate terms in total for the quadVAR models to select from. Therefore, using quadVAR models does not necessarily lead to an increase in the complexity of the outcome model.

4.2.1. *Examples of nonlinear features in empirical data*

To illustrate the usage of quadVAR in discovering nonlinear relationships for idiographic data, we show several examples from the two datasets. The examples are arbitrarily chosen from the participants for whom the quadVAR model is the best model selected with EBIC and has at least one nonzero nonlinear term estimated. The first one is from the 6th participant from Rowland and Wenzel (2020). The nonzero parameters estimated using quadVAR for this participant are shown in Appendix E. Here, we focus on the nonlinear effect from *excited:angry* to *angry*. This indicates that different values of *excited* can result in different effects of *angry* on itself for the next time point. This can be seen more clearly from the linearized networks shown in Fig. 6, where with a higher level of *excited*, the self-reinforcing effect of *angry* becomes stronger. However, after a closer look at the data by plotting the residuals of *angry* controlled on all other effects of *angry* at the previous time point, we found that this effect is likely driven by several influential cases (Fig. 7). There was one time point when the participant had an exceptionally high level of *angry*. The level of *excited* is low at this point, but high at the previous time point. This single time point contributed strongly to a significant estimate of the interactions between *excited* and *angry*.

The second one is the 8th participant from Rowland and Wenzel (2020). The nonzero parameters estimated from quadVAR are shown in Appendix E. There is one nonlinear parameter, namely $satisfied^2$ to *happy*. This means that changing the value of *satisfied* can lead to a change in the effect from *satisfied* to *happy*. We can see this in the networks shown in Fig. 8, that with a higher level of *satisfied*, the influence of *satisfied* to *happy* becomes more positive.[4] We also checked the results with a partial plot (Fig. 9). This time, in contrast to the previous example for the 6th participant, the nonlinearity of the result seems more robust. When the level of *satisfied* is low, it does not have a strong influence on the level of *happy* at the next time point, but when the level of *satisfied* is higher, its influence on *happy* becomes more evident.

4.3. *Discussion of the results from empirical datasets*

For empirical datasets, unlike the simulation datasets, model performance was judged by information criteria (EBIC in our study). Cross-validation results, however, showed strong discrepancies. Examining the cross-validation results leads to a conclusion similar to Bulteel et al.

[3] As the scales used in the ESM measures are bounded, we applied winsorization on the prediction values of empirical datasets, as explained above. As a sensitivity test, we also examined the results when winsorization is not used. Without winsorization, the AR model is the best for 106 participants (instead of 105), and the quadVAR model is the best for 1 participant (instead of 2). Thus, the influence of winsorization on these results is minimal. A more salient difference when not using winsorization is that several extreme MSE values appear for the quadVAR_full model, which would make the comparisons of other models (Figure 5a) less clear. Therefore, we only reported the results with winsorization in the main text, while the results without winsorization are available together with the reproducible code online.

[4] Note that the comparisons of linear plots are meant to represent the influence strength between variables, instead of representing a significant change in the direction of the influence. For example, the arrow from *satisfied* to *happy* appears negative in Figure 8a. This does not mean that the influence of *satisfied* on *happy* is significantly negative when the level of *satisfied* is at *Mean* – *SD*.

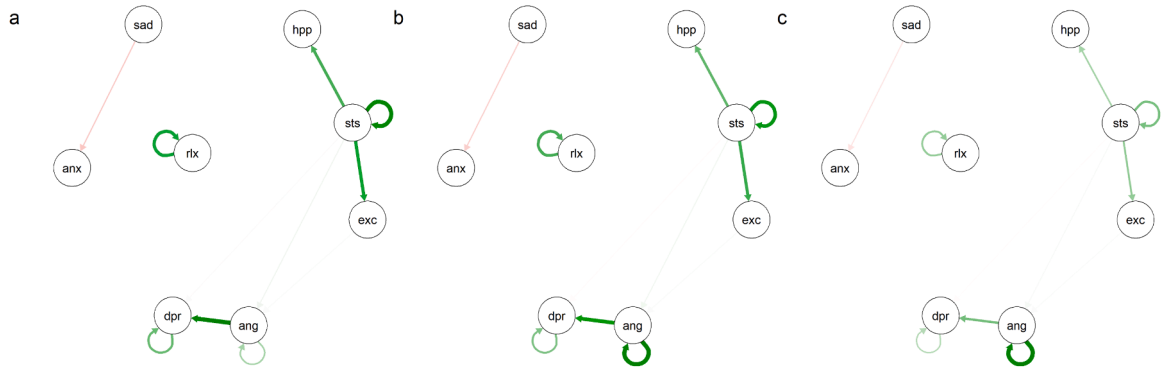


Fig. 6. Linearized networks for the 6th participant in Rowland and Wenzel (2020) with different levels of *excited*: (a) *Mean* − *SD*; (b) *Mean*; (c) *Mean* + *SD*. The following abbreviations are used: "hpp" represents happy, "exc" represents excited, "rlx" represents relaxed, "sts" represents satisfied, "ang" represents angry, "anx" represents anxious, "dpr" represents depressed, and "sad" represents sad.

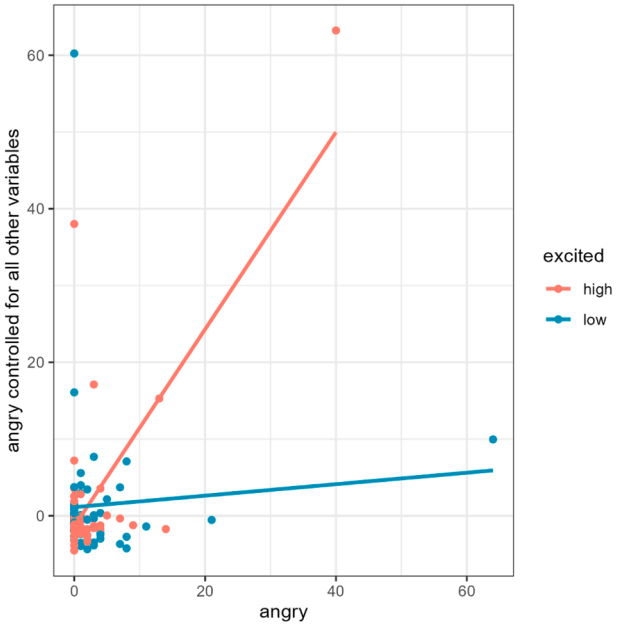


Fig. 7. Partial plot of the residual of *angry*, controlled for all other variables excluding *angry* at the previous time point, on *angry* at the previous time point. Different colors represent high or low values of the moderator, *excited*, at the previous time point, divided by the median. The straight lines represent linear regression lines based on the data points shown in the plot (i.e., data split by the median of the focal variable and controlled for all other variables).

(2018), that the AR model performs best in predicting unseen data. However, if we look at EBIC, which is also intended to be an estimate of the predictive power for unseen data (Burnham et al., 2002; Chen & Chen, 2008; Wagenmakers & Farrell, 2004), quadVAR performs better for more than half of the participants. The most likely explanation for the better performance of the AR model compared to the simulation data is the differences in the data generation process. Compared to the networks for simulation data (Fig. 1), the networks estimated with quadVAR from empirical data contain more self-reinforced loops and far fewer edges, which indicates that the empirical data generation process is intrinsically more similar to an AR process than the simulated system. However, this does not explain the difference between EBIC results and cross-validation results. To our knowledge, the studies regarding the difference between the two performance indicators are sparse, especially in the context of psychology. We suspect that the influential cases and multistability in the datasets may contribute to this, as some cross-validation folds may not include those influential cases, thus having very different parameter estimates than using the whole dataset. Future studies are still needed to fully resolve this issue. At the same time, we would not suggest removing those influential data points before analysis, as those observations are from the same individual, and the influential data points may represent rare yet important states of the system. Therefore, we recommend always checking the influential cases before drawing conclusions and treating them with caution.

The example we showed exhibited nonlinear patterns that are theoretically interesting. From the 8th participant from Rowland and Wenzel (2020), we found a nonlinear relationship between *satisfied* and *happy* (Fig. 9). This relationship is understandable because when satisfaction is low, it may not significantly influence happiness, but when satisfaction is high, it contributes notably to happiness. However, we also observed few overlaps of nonlinear effects between the two

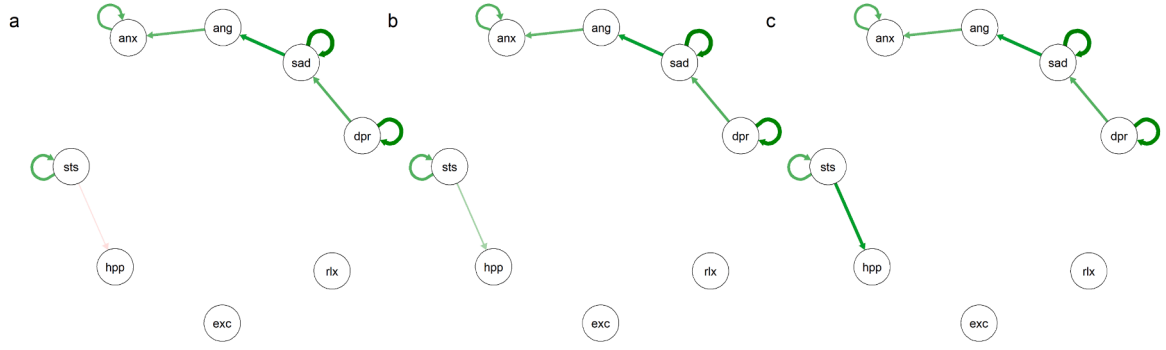


Fig. 8. Linearized networks for the 8th participant in Rowland and Wenzel (2020) with different levels of *satisfied*: (a) *Mean* − *SD*; (b) *Mean*; (c) *Mean* + *SD*. The following abbreviations are used: "hpp" represents happy, "exc" represents excited, "rlx" represents relaxed, "sts" represents satisfied, "ang" represents angry, "anx" represents anxious, "dpr" represents depressed, and "sad" represents sad.

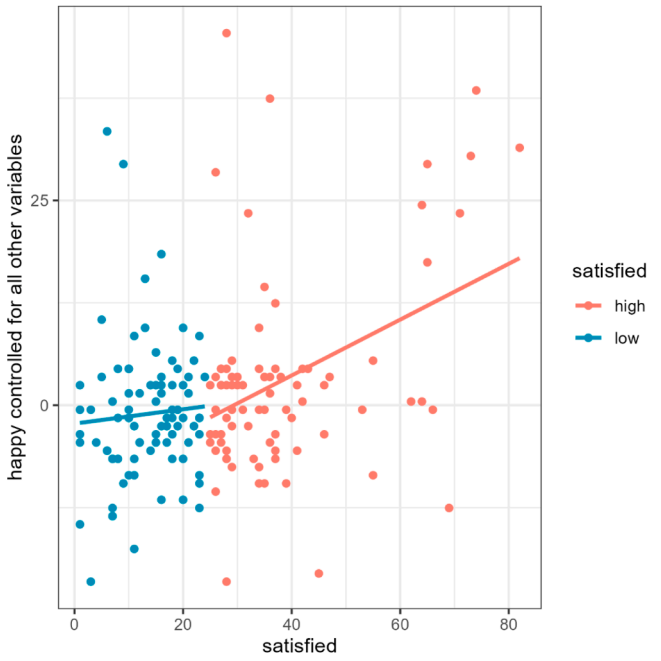


Fig. 9. Partial plot of the residual of *happy*, controlled for all other variables, excluding *satisfied* at the previous time point, on *satisfied* at the previous time point. Different colors represent high or low values of the moderator, *satisfied*, at the previous time point, divided by the median. The straight lines represent linear regression lines based on the data points shown in the plot (i.e., data split by the median of the focal variable and controlled for all other variables).

participants. As the quadVAR model is currently an idiographic model, it does not consider any possible between-person similarities. Therefore, whether the nonlinear relationships we found are only applicable to a single participant or can be generalized to the group level is still to be investigated.

5. General discussion

In this article, we introduced the quadVAR method with the accompanying R package, which enables testing nonlinear effects in vector autoregressive models for psychological ILD. Our method implements a state-of-the-art quadratic regression algorithm, RAMP (Hao et al., 2018), that takes the hierarchy of regression structure and a large amount of possible nonlinear terms into account. The linearization method was used to facilitate the interpretation of the model output (Kroc & Olvera Astivia, 2023). The performance of the method with both simulated and empirical datasets was examined. In simulation datasets, we found that quadVAR outperforms linear methods for a nonlinear system, as long as the sample size is large enough. When we look at the performance for individual parameters and nonlinear effects, quadVAR faces a similar challenge as linear models, that the general accuracy is not very high and requires a large sample size to reach satisfactory performance (Mansueto et al., 2023). This indicates that quadVAR is more suitable for exploratory purposes and requires cautious interpretation and confirmatory repetitions before reaching a solid conclusion regarding specific nonlinear characteristics of a process. With empirical datasets, we found that quadVAR performs the best for more than half of the participants judged by EBIC, yet cross-validation results favor the simpler AR model. Further detailed investigations into individual participants showed that, although sometimes the nonlinear effects were driven by several influential data points, there were cases where a robust nonlinear relationship between psychological variables seemed to exist. This evidences the usefulness of investigating nonlinear relationships for psychological ILD but also requires future investigations.

To our knowledge, the current work is the first in the field to explicitly model continuous, intrinsic, nonlinear dynamics from psychological ILD. Including nonlinear terms in autoregression models clearly induced more challenges. For the method development part, we need special treatments for both the estimation and interpretation of the results. A challenge in the simulation part is that nonlinear systems can easily become unstable (i.e., grow into infinity), which makes it difficult to systematically vary the parameters to test their influences on model performance. It also prohibits us from using a more realistic simulation example using parameters estimated with empirical datasets. The noise level is another challenge because it does not have a monotonic relationship with the model's performance due to the double role noise plays. On the one hand, stronger noise decreases the signal-to-noise ratio, making it more difficult to retrieve dynamic information. On the other hand, driven by noise, the system travels through the state space more easily, adding more information to the data. Therefore, it is difficult to systematically evaluate the influence of noise on the model's performance. Challenges also arise for the performance benchmarks. In a model with nonlinear terms, the linear coefficients are not independent of the nonlinear coefficients, making it difficult to tell how well the linear part is estimated within the nonlinear framework[5]. We also found that the EBIC and the cross-validation results favor different models, making it more difficult for model selection. This may also be related to the nonlinearity of the quadVAR method.

There are a few limitations of the quadVAR method that are worth noting. First, as the RAMP algorithm can only be applied to different outcome variables separately, we assumed that the error term for each outcome variable is independent. This assumption may sometimes be unrealistic. Second, the quadratic form assumed in quadVAR may be unrealistic, either. It does not accommodate more complex forms of dynamic relationships, such as S-shaped curves. Moreover, the estimated relationships near the boundaries, where fewer data points are available, may also be distorted. Third, as a within-person model, quadVAR requires long ESM data from each participant, and it cannot summarize the similarities across people and draw group-level inferences. And fourth, although the application of the marginality principle helps to facilitate interpretation and prevent overfitting, it may fail to pick up quadratic relationships when the linear regression coefficient is close to zero. We still consider this trade-off to be beneficial, as ignoring the marginality principle often leads to a large number of spurious relationships (Hao et al., 2018). However, it should be noted that the RAMP algorithm may lead to false negatives in this situation.

Nevertheless, we also see promising signs of introducing such a method into the world of idiographic dynamic analysis and network modeling. First, counterintuitively, we found that introducing nonlinear terms does not necessarily lead to a stronger overfitting issue. The degree of freedom of the quadVAR models is even slightly lower than that of the VAR models for the empirical datasets we used. This indicates that the RAMP estimation algorithm we implemented performed well in constraining the overall model complexity and choosing a small number of nonlinear terms from a large reservoir. Theoretically, if some dynamic characteristics of a system can be modeled with one nonlinear term, forcing the system to be linear only makes it necessary to have more terms, which actually increases model complexity. Therefore, using nonlinear terms does not necessarily lead to more severe overfitting problems. We also see from the partial plots that quadVAR has the potential to discover nonlinear dynamical relationships from multidimensional time series. Taking all the evidence together, we suggest that exploring nonlinear dynamical relationships for psychological ILD is a

[5] For example, if the true model is $y = x_1x_2 + 0.2x_1$, but the estimated model omits the interaction term and yields $\hat{y} = 0.4x_1$, then the linear coefficient in the estimated model (0.4) is not directly comparable with the true linear coefficient (0.2). This is because the true model can be written as $y = (x_2 + 0.2)x_1$, which means that the effective coefficient for x_1 depends on x_2 and varies across observations.

promising direction. At the same time, when a linear method is used, the existence of nonlinearity and outliers in the datasets should be checked.

Several directions of development are worth future consideration. First, the order of the polynomial terms may be extended further. As shown by Kroc and Olvera Astivia (2023), including cubic terms is beneficial to model many nonlinear relationships in psychology, including the ceiling and flooring effects. Using higher-order terms inevitably makes the model selection process more difficult. Although it may not be feasible to estimate all possible cubic and even higher-order interaction terms for psychological ILD, adding cubic terms for single variables may be possible. Future studies may also consider using robust estimations in quadVAR and idiographic modeling in general, as we saw influential data points may be present in individual ILD that may change the parameter estimations significantly. Yet no matter such robust method can be developed or not, we would emphasize again the importance of checking influential cases in dynamic modeling studies in general, which is hardly done in current empirical studies, as most of the shrinkage methods and multilevel methods based on least squares estimates are actually sensitive to outliers (Dedrick et al., 2009; H. Wang et al., 2007).

Second, it may be possible to develop a multilevel extension for the quadVAR method. Previous research has shown that multilevel models perform better in cross-validation because they can also use information from other participants for estimation, therefore partly overcoming the constraints of limited data points per participant (Bulteel et al., 2018). However, multilevel models assume the coefficients of each participant vary around the group mean, which means that the within-person dynamics should be similar enough across people. This may not hold, as we see great variability in network structures among individuals. Alternatively, more idiographic frameworks, like the group iterative multiple model estimation (GIMME, Lane & Gates, 2017; Wright et al., 2019), may also be considered to estimate the generalizability of the relationship at the group level. A major difficulty in developing such group-level methods is that the RAMP algorithm and some alternative quadratic regression algorithms (e.g., hierNet, Bien et al., 2013) are all based on the LASSO penalized regression framework, which shrinks the parameter estimates to zero instead of the group mean. Recent developments have shown the possibility of combining multilevel modeling and LASSO in a single estimation procedure (e.g., glmmLasso, Schelldorfer et al., 2014). Combining this idea with the quadVAR method may be a productive approach.

Third, data collection and study design may also be improved to provide datasets that are more suitable for discovering nonlinear dynamics. In the empirical examples, we found that the estimated networks are rather sparse, with very few connections across variables. This may partly be because the variables assessed were aimed at providing a general description of the affective state of participants, instead of examining a specific relationship between some important variables. With the estimation method for nonlinear relationships available, future research may be specifically designed to recover nonlinear dynamics that are theoretically predicted. A well-designed algorithm may help to discover potential effects and examine certain hypotheses, but it can never do the conceptual job for researchers. As written by Burnham et al. (2002, p. 147), " 'Let the computer find out' is a poor strategy for researchers who do not bother to think clearly about the problem of interest and its scientific setting." Therefore, what should be fed into the method should always be carefully considered.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used ChatGPT (OpenAI) for language refinement and stylistic suggestions. The authors critically reviewed and revised all outputs and take full responsibility for the final content of the published article.

Funding

JC was partly supported by a KU Leuven postdoctoral mandate, Grant No. PDMT2/25/036. ALA was supported by an NWO VIDI grant, Grant No. VI.Vidi.191.178.

CRediT authorship contribution statement

Jingmeng Cui: Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Conceptualization, Funding acquisition. **Anna Lichtwarck-Aschoff:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Fred Hasselman:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare no competing interests.

Acknowledgements

We thank the Center for Information Technology of the University of Groningen, particularly Pedro Santos Neves, for their support and for providing access to the Hábrók high-performance computing cluster. We thank Laura Bringmann and Yong Zhang for their helpful feedback and suggestions on the earlier version of this manuscript.

Appendix A. The Influence of Noise and Transitions on quadVAR Model Estimation

For linear models, the influence of noise on model estimates is rather straightforward, in that larger noise reduces the signal-to-noise ratio, thus making the parameter estimations less accurate (e.g., Mansueto et al., 2022). However, for quadratic models, the relationship between noise and model estimation is more complicated. We show an example to illustrate this idea, in which we reduce the noise level from $SD = 1$ to $SD = 0.2$. The results are shown in Fig. A1. The networks estimated from this example are similar to the linearized networks for the negative state shown in Fig. 3. However, no nonlinear effects were estimated in this case, making the networks for different states the same as each other. This is because the noise in the system is too weak, making the system always stay in the negative phase. The simulated data do not contain enough information about the system in the alternative phase, making it difficult to estimate the nonlinear effects.

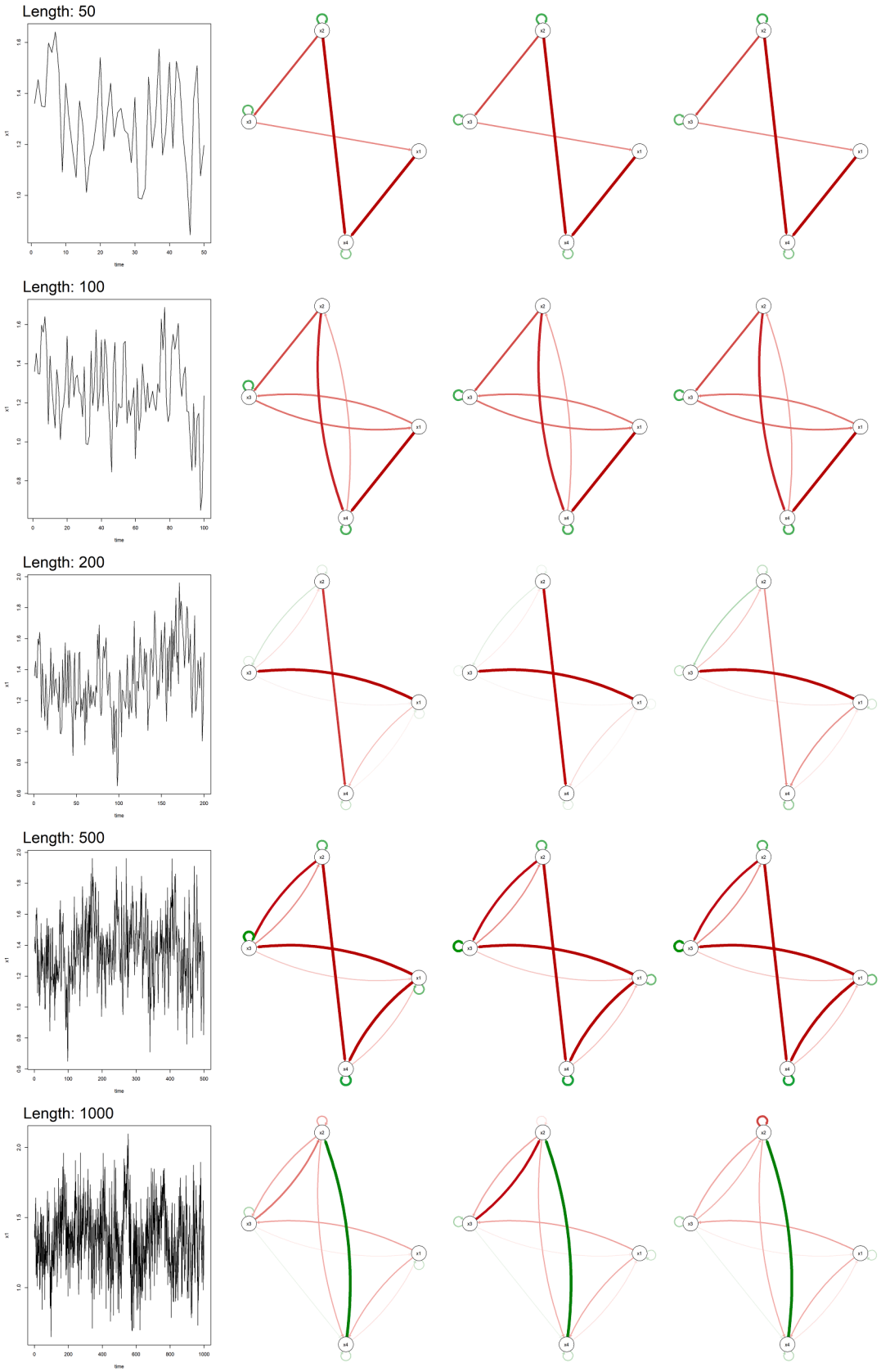


Fig. A1. Linearized networks for the simulation examples with different lengths, with noise $SD = 0.2$. In each row, the first plot shows the raw time series of x_1, the second to the fourth plots show the linearized networks estimated from simulation data for the neutral state ($x_1 = x_2 = x_3 = x_4 = 2.80$), the positive state ($x_1 = x_2 = 4.89$, $x_3 = x_4 = 1.36$), and the negative state ($x_1 = x_2 = 1.36$, $x_3 = x_4 = 4.89$).

Appendix B. The Comparable Linear System

The comparable linear system is specified using parameters of a VAR model estimated from the simulated data from the nonlinear, bistable system. The simulated dataset contains 1000 data points. The coefficients for the VAR model are as follows. The precise value without rounding can also be reproduced with the code online.

$$x_{t+1} = Ax_t + c$$

$$A = \begin{bmatrix} 0.44 & 0.08 & -0.18 & -0.26 \\ 0.14 & 0.41 & -0.18 & -0.23 \\ -0.17 & -0.25 & 0.46 & 0.08 \\ -0.2 & -0.18 & 0.13 & 0.44 \end{bmatrix}$$

$$c = \begin{bmatrix} 2.76 \\ 2.61 \\ 2.68 \\ 2.46 \end{bmatrix}$$

Appendix C. Definitions of Classification Metrics

For the classification metrics, we do not only consider whether a value is zero but also consider whether the sign of the value is correct according to the true value. Therefore, we are interested in the correctness of the directed inferences of the coefficients. Specifically, we define the following:

(1) True positive, when the true coefficient is nonzero, and the estimated coefficient is also nonzero and has the same sign as the true coefficient.
(2) True negative, when the true coefficient is zero, and the estimated coefficient is also zero.
(3) False positive, when the estimated coefficient is nonzero, but the true coefficient is zero or is nonzero but has the opposite sign as the estimated coefficient.
(4) False negative, when the estimated coefficient is zero, but the true coefficient is nonzero.

Based on the definitions above, we calculate the following metrics:

(1) Accuracy = (true positive + true negative)/(true positive + true negative + false positive + false negative).
(2) Precision = true positive/(true positive + false positive).
(3) Sensitivity = true negative/(true negative + false positive).
(4) Specificity = true positive/(true positive + false negative).

The correlation is defined as the Pearson correlation between the estimated coefficients and the true coefficients.

Appendix D. Results from the Second Empirical Dataset

The results from the dataset from Kuppens et al. (2010) were similar to the results of the dataset Rowland and Wenzel (2020; reported in the main text). The EBIC and cross-validation results for the dataset from Kuppens et al. (2010) are shown in Fig. D1.

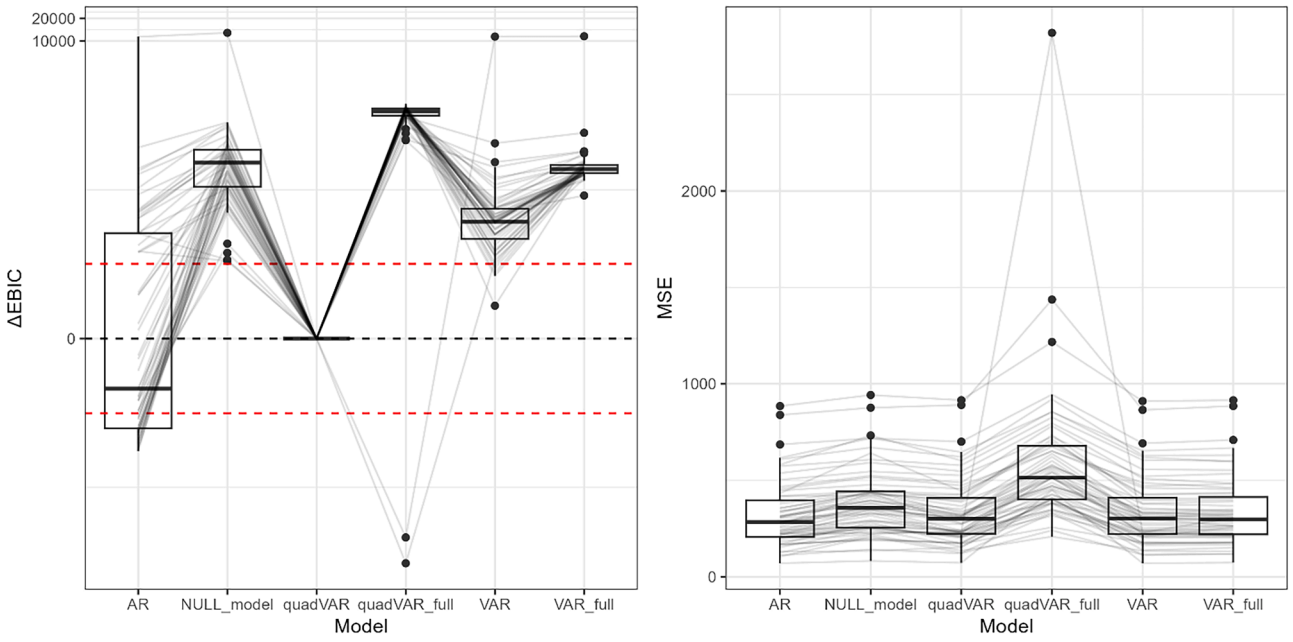


Fig. D1. EBIC and cross-validation MSE for participants in Kuppens et al. (2010). Gray lines represent individual participants.

Appendix E. quadVAR Model Estimations for Selected Participants

Table E1
quadVAR model estimations for the 6th participant in Rowland and Wenzel (2020).

Model	Effect	Estimate
happy	satisfied	0.3551
excited	satisfied	0.4287
relaxed	relaxed	0.4105
satisfied	satisfied	0.5321
angry	excited	-0.0783
angry	satisfied	0.0322
angry	angry	-0.1278
angry	excited:angry	0.0231
anxious	sad	-0.3702
anxious	sad:sad	0.0375
depressed	satisfied	-0.1500
depressed	angry	-1.4642
depressed	depressed	0.2909
depressed	satisfied:angry	0.0391

Table E2
quadVAR model estimations for the 8th participant in Rowland and Wenzel (2020).

Model	Effect	Estimate
happy	satisfied	-0.1669
happy	satisfied:satisfied	0.0056
satisfied	satisfied	0.2585
angry	sad	0.3214
anxious	angry	0.2355
anxious	anxious	0.2537
depressed	depressed	0.4029
sad	depressed	0.2574
sad	sad	0.4248

Data availability

The necessary code for replicating the findings presented in this article can be obtained from the following URL: https://osf.io/xzf3n/.

References

Adolf, J. K., Voelkle, M. C., Brose, A., & Schmiedek, F. (2017). Capturing context-related change in emotional dynamics via fixed moderated time series analysis. *Multivariate Behavioral Research, 52*(4), 499–531. https://doi.org/10.1080/00273171.2017.1321978

Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences, 191*, 192–213. https://doi.org/10.1016/j.ins.2011.12.028

Bien, J., Taylor, J., & Tibshirani, R. (2013). A lasso for hierarchical interactions. *The Annals of Statistics, 41*(3), 1111–1141. https://doi.org/10.1214/13-AOS1096

Bogartz, R. S. (1994). The future of dynamic systems models in developmental psychology in the light of the past. *Journal of Experimental Child Psychology, 58*(2), 289–319. https://doi.org/10.1006/jecp.1994.1036

Boker, S. M., & Moulder, R. G. (2023). Dynamical systems and differential equation models of change. In *APA handbook of research methods in psychology: Data analysis and research publication* (2nd ed,, *3* pp. 337–350). American Psychological Association. https://doi.org/10.1037/0000320-015

Breheny, P., & Huang, J. (2011). Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Annals of Applied Statistics, 5*(1), 232–253. https://doi.org/10.1214/10-AOAS388

Bringmann, L. F., Ariens, S., Ernst, A. F., Snippe, E., & Ceulemans, E. (2024). Changing networks: Moderated idiographic psychological networks. *Advances.in/Psychology, 2*, Article e658296. https://doi.org/10.56296/aip00014

Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2017). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods, 22*(3), 409–425. https://doi.org/10.1037/met0000085

Bringmann, L. F., Lemmens, L. H. J. M., Huibers, M. J. H., Borsboom, D., & Tuerlinckx, F. (2015). Revealing the dynamic network structure of the Beck Depression Inventory-II. *Psychological Medicine, 45*(4), 747–757. https://doi.org/10.1017/S0033291714001809

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., Borsboom, D., & Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLOS ONE, 8*(4), Article e60188. https://doi.org/10.1371/journal.pone.0060188

Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences, 113*(15), 3932–3937. https://doi.org/10.1073/pnas.1517384113

Bulteel, K., Mestdagh, M., Tuerlinckx, F., & Ceulemans, E. (2018). VAR(1) based models do not always outpredict AR(1) models in typical psychological applications. *Psychological Methods, 23*(4), 740–756. https://doi.org/10.1037/met0000178

Burnham, K. P., Anderson, D. R., & Burnham, K. P. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed). Springer.

Chen, J., & Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika, 95*(3), 759–771. https://doi.org/10.1093/biomet/asn034

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression correlation analysis for the behavioral sciences* (Third edition). Routledge Taylor & Francis Group.

Cui, J., Hasselman, F., & Lichtwarck-Aschoff, A. (2023). Unlocking nonlinear dynamics and multistability from intensive longitudinal data: A novel method. *Psychological Methods*. https://doi.org/10.1037/met0000623

Dablander, F., Ryan, O., & Haslbeck, J. M. B. (2020). Choosing between AR(1) and VAR (1) models in typical psychological applications. *PLOS ONE, 15*(10), Article e0240730. https://doi.org/10.1371/journal.pone.0240730

Dedrick, R. F., Ferron, J. M., Hess, M. R., Hogarty, K. Y., Kromrey, J. D., Lang, T. R., Niles, J. D., & Lee, R. S. (2009). Multilevel modeling: A review of methodological issues and applications. *Review of Educational Research, 79*(1), 69–102. https://doi.org/10.3102/0034654308325581

Delignières, D., Fortes, M., & Ninot, G. (2004). The fractal dynamics of self-esteem and physical self. *Nonlinear Dynamics, Psychology, and Life Sciences, 8*(4), 479–510.

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software, 48*(1), 1. https://doi.org/10.18637/jss.v048.i04. Article.

Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods, 23*(4), 617–634. https://doi.org/10.1037/met0000167

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*(4), 453–480. https://doi.org/10.1080/00273171.2018.1454823

Failenschmid, J. I., Vogelsmeier, L. V. D. E., Mulder, J., & Jongerling, J. (2025). Modelling non-linear psychological processes: Reviewing and evaluating non-parametric approaches and their applicability to intensive longitudinal data. *British Journal of Mathematical and Statistical Psychology*. https://doi.org/10.1111/bmsp.12397

Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*, 1–22. https://doi.org/10.18637/jss.v033.i01

Granic, I. (2005). Timing is everything: Developmental psychopathology from a dynamic systems perspective. *Developmental Review, 25*(3), 386–407. https://doi.org/10.1016/j.dr.2005.10.005

Granic, I., Lewis, M. D., & Lichwarck-Aschoff, A. (2018). Complex pathways to psychopathology. *Psychological Inquiry, 29*(3), 145–150. https://doi.org/10.1080/1047840X.2018.1513682

Guastello, S. J. (1982). Moderator regression and the cusp catastrophe: Application of two-stage personnel selection, training, therapy, and policy evaluation. *Behavioral Science, 27*(3), 259–272. https://doi.org/10.1002/bs.3830270305

Hamaker, E. L., Asparouhov, T., Brose, A., Schmiedek, F., & Muthén, B. (2018). At the frontiers of modeling intensive longitudinal data: Dynamic structural equation models for the affective measurements from the COGITO study. *Multivariate Behavioral Research, 53*(6), 820–841. https://doi.org/10.1080/00273171.2018.1446819

Hamaker, E. L., Ceulemans, E., Grasman, R. P. P. P., & Tuerlinckx, F. (2015). Modeling affect dynamics: State of the art and future challenges. *Emotion Review, 7*(4), 316–322. https://doi.org/10.1177/1754073915590619

Hao, N., Feng, Y., & Zhang, H. H. (2018). Model selection for high-dimensional quadratic regression via regularization. *Journal of the American Statistical Association, 113*(522), 615–625. https://doi.org/10.1080/01621459.2016.1264956

Hao, N., & Zhang, H. H. (2014). Interaction screening for ultrahigh-dimensional data. *Journal of the American Statistical Association, 109*(507), 1285–1301. https://doi.org/10.1080/01621459.2014.881741

Hao, N., & Zhang, H. H. (2017). A note on high-dimensional linear regression with interactions. *The American Statistician, 71*(4), 291–297. https://doi.org/10.1080/00031305.2016.1264311

Haslbeck, J. M. B., Borsboom, D., & Waldorp, L. J. (2021). Moderated network models. *Multivariate Behavioral Research, 56*(2), 256–287. https://doi.org/10.1080/00273171.2019.1677207

Haslbeck, J. M. B., Bringmann, L. F., & Waldorp, L. J. (2021). A tutorial on estimating time-varying vector autoregressive models. *Multivariate Behavioral Research, 56*(1), 120–149. https://doi.org/10.1080/00273171.2020.1743630

Haslbeck, J. M. B., & Ryan, O. (2022). Recovering within-person dynamics from psychological time series. *Multivariate Behavioral Research, 57*(5), 735–766. https://doi.org/10.1080/00273171.2021.1896353

Haslbeck, J. M. B., Ryan, O., & Dablander, F. (2023). Multimodality and skewness in emotion time series. *Emotion (Washington, D.C.), 23*(8), 2117–2141. https://doi.org/10.1037/emo0001218

Haslbeck, J. M. B., Ryan, O., Robinaugh, D., Waldorp, L., & Borsboom, D. (2022). Modeling psychopathology: From data models to formal theories. *Psychological Methods, 27*(6), 930–957. https://doi.org/10.1037/met0000303

Haslbeck, J. M. B., & Waldorp, L. J. (2020). mgm: Estimating time-varying mixed graphical models in high-dimensional data. *Journal of Statistical Software, 93*(8), 1–46. https://doi.org/10.18637/jss.v093.i08

Hasselman, F., & Bosman, A. M. T. (2020). Studying complex adaptive systems with internal states: A recurrence network approach to the analysis of multivariate time-series data representing self-reports of human experience. *Frontiers in Applied Mathematics and Statistics, 6*. https://doi.org/10.3389/fams.2020.00009

Hayes, A. M., Yasinski, C., Ben Barnes, J., & Bockting, C. L. H. (2015). Network destabilization and transition in depression: New methods for studying the dynamics of therapeutic change. *Clinical Psychology Review, 41*, 27–39. https://doi.org/10.1016/j.cpr.2015.06.007

Hoekstra, R. H. A., De Ron, J., Epskamp, S., Robinaugh, D., & Borsboom, D. (2024). Mapping the dynamics of idiographic network models to the network theory of psychopathology using stability landscapes. *PsyArXiv*. https://doi.org/10.31234/osf.io/9sguw

Jordan, D. G., Winer, E. S., & Salem, T. (2020). The current status of temporal network analysis for clinical science: Considerations as the paradigm shifts? *Journal of Clinical Psychology, 76*(9), 1591–1612. https://doi.org/10.1002/jclp.22957

Kalokerinos, E. K., Russo-Batterham, D., Koval, P., Moeck, E. K., Grewal, K. K., Greenaway, K. H., Shrestha, K. M., Garrett, P., Michalewicz, A., Garber, J., & Kuppens, P. (in preparation). 2026. *The EMOTE Database: An open, searchable database of experience sampling data mapping everyday life.* [Unpublished manuscript].

Kossakowski, J., Groot, P., Haslbeck, J., Borsboom, D., & Wichers, M. (2017). Data from 'critical slowing down as a personalized early warning signal for depression. *Journal of Open Psychology Data, 5*, 1. https://doi.org/10.5334/jopd.29. Article.

Kroc, E., & Olvera Astivia, O. L. (2023). The case for the curve: Parametric regression with second- and third-order polynomial functions of predictors should be routine. *Psychological Methods*. https://doi.org/10.1037/met0000629

Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology, 99*(6), 1042–1060. https://doi.org/10.1037/a0020962

Lane, S. T., & Gates, K. M. (2017). Automated selection of robust individual-level structural equation models for time series data. *Structural Equation Modeling: A Multidisciplinary Journal, 24*(5), 768–782. https://doi.org/10.1080/10705511.2017.1309978

Lazarus, G., Song, J., Crawford, C. M., & Fisher, A. J. (2021). A close look at the role of time in affect dynamics research. In C. E. Waugh, & P. Kuppens (Eds.), *Affect dynamics* (pp. 95–116). Springer International Publishing. https://doi.org/10.1007/978-3-030-82965-0%255C_5.

Liu, Q., & Wang, L. (2021). T-test and ANOVA for data with ceiling and/or floor effects. *Behavior Research Methods, 53*(1), 264–277. https://doi.org/10.3758/s13428-020-01407-2

Loossens, T., Mestdagh, M., Dejonckheere, E., Kuppens, P., Tuerlinckx, F., & Verdonck, S. (2020). The Affective Ising Model: A computational account of human affect dynamics. *PLOS Computational Biology, 16*(5), Article e1007860. https://doi.org/10.1371/journal.pcbi.1007860

Luo, S., & Chen, Z. (2021). Sequential interaction group selection by the principle of correlation search for high-dimensional interaction models. *Statistica Sinica, 31*(1), 197–221.

Mansueto, A. C., Wiers, R. W., van Weert, J. C. M., Schouten, B. C., & Epskamp, S. (2023). Investigating the feasibility of idiographic network models. *Psychological Methods, 28*(5), 1052–1068. https://doi.org/10.1037/met0000466

Marwan, N., Carmen Romano, M., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports, 438*(5), 237–329. https://doi.org/10.1016/j.physrep.2006.11.001

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research and Perspectives, 2*(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1

Moulder, R. G., Martynova, E., & Boker, S. M. (2022). Extracting nonlinear dynamics from psychological and behavioral time series through HAVOK analysis. *Multivariate Behavioral Research, 0*(0), 1–25. https://doi.org/10.1080/00273171.2021.1994848

Myin-Germeys, I., & Kuppens, P. (Eds.). (2021). *The open handbook of experience sampling methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies*. Center for Research on Experience Sampling and Ambulatory Methods Leuven.

Nelder, J. A. (1977). A reformulation of linear models. *Journal of the Royal Statistical Society. Series A (General), 140*(1), 48–77. https://doi.org/10.2307/2344517

Olthof, M., Hasselman, F., & Lichtwarck-Aschoff, A. (2020). Complexity in psychological self-ratings: Implications for research and practice. *BMC Medicine, 18*(1), 317. https://doi.org/10.1186/s12916-020-01727-2

Olthof, M., Hasselman, F., Oude Maatman, F., Bosman, A. M. T., & Lichtwarck-Aschoff, A. (2023). Complexity theory of psychopathology. *Journal of Psychopathology and Clinical Science, 132*(3), 314–323. https://doi.org/10.1037/abn0000740

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

Richardson, M., Paxton, A., & Kuznetsov, N. (2017). Nonlinear methods for understanding complex dynamical phenomena in psychological science. *APA Psychological Science Agenda*.

Rigdon, E. E., Schumacker, R. E., & Wothke, W. (1998). A comparative review of interaction and nonlinear modeling. *Interaction and nonlinear effects in structural equation modeling*. Routledge.

Robinaugh, D. J., Hoekstra, R. H. A., Toner, E. R., & Borsboom, D. (2020). The network approach to psychopathology: A review of the literature 2008–2018 and an agenda for future research. *Psychological Medicine, 50*(3), 353–366. https://doi.org/10.1017/S0033291719003404

Rowland, Z., & Wenzel, M. (2020). Mindfulness and affect-network density: Does mindfulness facilitate disengagement from affective experiences in daily life? *Mindfulness, 11*(5), 1253–1266. https://doi.org/10.1007/s12671-020-01335-4

Ryan, O., & Hamaker, E. L. (2022). Time to intervene: A continuous-time approach to network analysis and centrality. *Psychometrika, 87*(1), 214–252. https://doi.org/10.1007/s11336-021-09767-0

Sayama, H. (2015). *Introduction to the modeling and analysis of complex systems*. Open SUNY Textbooks.

Schelldorfer, J., Meier, L., & Bühlmann, P. (2014). GLMMLasso: An algorithm for high-dimensional generalized linear mixed models using ℓ_1-penalization. *Journal of Computational and Graphical Statistics, 23*(2), 460–477.

Shao, J. (1997). An asymptotic theory for linear model selection. *Statistica Sinica, 7*(2), 221–242.

Stone, A. A., Schneider, S., & Smyth, J. M. (2023). Evaluation of pressing issues in ecological momentary assessment. *Annual Review of Clinical Psychology, 19*, 107–131. https://doi.org/10.1146/annurev-clinpsy-080921-083128

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society. Series B (Methodological), 39*(1), 44–47.

Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology, 67*(6), 894–904. https://doi.org/10.1037/0022-006X.67.6.894

Vachon, H., Viechtbauer, W., Rintala, A., & Myin-Germeys, I. (2019). Compliance and retention with the experience sampling method over the continuum of severe mental disorders: Meta-analysis and recommendations. *Journal of Medical Internet Research, 21*(12), Article e14475. https://doi.org/10.2196/14475

van Borkulo, C. D., Borsboom, D., Epskamp, S., Blanken, T. F., Boschloo, L., Schoevers, R. A., & Waldorp, L. J. (2014). A new method for constructing networks from binary data. *Scientific Reports, 4*(1), 1. https://doi.org/10.1038/srep05918. Article.

van de Leemput, I. A., Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., van Nes, E. H., Viechtbauer, W., Giltay, E. J., Aggen, S. H., Derom, C., Jacobs, N., Kendler, K. S., van der Maas, H. L. J., Neale, M. C., Peeters, F., Thiery, E., Zachar, P., & Scheffer, M. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences, 111*(1), 87–92. https://doi.org/10.1073/pnas.1312114110

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11*(1), 192–196. https://doi.org/10.3758/BF03206482

Wang, C., Jiang, B., & Zhu, L. (2021). Penalized interaction estimation for ultrahigh dimensional quadratic regression. *Statistica Sinica, 31*(3), 1549–1570.

Wang, H., Li, G., & Jiang, G. (2007). Robust regression shrinkage and consistent variable selection through the LAD-lasso. *Journal of Business & Economic Statistics, 25*(3), 347–355. https://doi.org/10.1198/073500106000000251

Wang, L., Zhang, Z., McArdle, J. J., & Salthouse, T. A. (2008). Investigating ceiling effects in longitudinal data analysis. *Multivariate Behavioral Research, 43*(3), 476–496. https://doi.org/10.1080/00273170802285941

Wood, S. N. (2017). *Generalized additive models: An introduction with R, Second Edition* (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279

Wright, A. G. C., 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*(4), 502–515. https://doi.org/10.1037/pas0000617

Wright, A. G. C., & Woods, W. C. (2020). Personalized models of psychopathology. *Annual Review of Clinical Psychology, 16*(1), 49–74. https://doi.org/10.1146/annurev-clinpsy-102419-125032

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18*(5), 459–482. https://doi.org/10.1002/cne.920180503

Zou, Y., Donner, R. V., Marwan, N., Donges, J. F., & Kurths, J. (2019). Complex network approaches to nonlinear time series analysis. *Physics Reports, 787*, 1–97. https://doi.org/10.1016/j.physrep.2018.10.005