Modeling Cycles, Trends and Time-Varying Effects in Dynamic Structural Equation Models with Regression Splines

Ø. Sørensen & E. M. McCormick

View supplementary material ⧉

Published online: 02 Jun 2025.

Submit your article to this journal ⧉

View related articles ⧉

View Crossmark data ⧉

Routledge
Taylor & Francis Group

Modeling Cycles, Trends and Time-Varying Effects in Dynamic Structural Equation Models with Regression Splines

Ø. Sørensen[a] and E. M. McCormick[b]

[a]Department of Psychology, Center for Lifespan Changes in Brain and Cognition, University of Oslo, Oslo, Norway; [b]Educational Statistics and Data Science, College of Education, University of Delaware, Newark, DE, USA

ABSTRACT

Intensive longitudinal data with a large number of timepoints per individual are becoming increasingly common. Such data allow going beyond the classical growth model situation and studying population effects and individual variability not only in trends over time but also in autoregressive effects, cross-lagged effects, and the noise term. Dynamic structural equation models (DSEMs) have become very popular for analyzing intensive longitudinal data. However, when the data contain trends, cycles, or time-varying predictors which have nonlinear effects on the outcome, DSEMs require the practitioner to specify the correct parametric form of the effects, which may be challenging in practice. In this paper, we show how to alleviate this issue by introducing regression splines which are able to flexibly learn the underlying function shapes. Our main contribution is thus a building block to the DSEM modeler's toolkit, and we discuss smoothing priors and hierarchical smooth terms using the special cases of two-level lag-1 autoregressive and vector autoregressive models as examples. We illustrate in simulation studies how ignoring nonlinear trends may lead to biased parameter estimates, and then show how to use the proposed framework to model weekly cycles and long-term trends in diary data on alcohol consumption and perceived stress.

Introduction

Dynamic structural equation models (DSEMs) (Asparouhov et al., 2018) offer a general framework for Bayesian multilevel modeling of timeseries data. Applications include analysis of positive and negative affect (Hamaker et al., 2018), interindividual differences in cognitive variability (Judd et al., 2024), affective dynamics and alcohol-related outcomes (Feinn et al., 2023), and the impact of weather conditions on the use of cycling for transportation (Bjørnarå et al., 2023). Recent developments of DSEM include incorporation of measurement models (McNeish et al., 2021; Oh & Jahng, 2023), modeling of cycles using sine-cosine curves (Muthén et al., 2024), and open-source implementations (Li et al., 2022).

The focus of this paper is on dealing with the effect of time-varying covariates, trends, or cycles on the outcomes of interest. In classical time series analysis, in which a single unit of observation is followed over time, it is common to detrend the data prior to analysis. This is not as straightforward in the multilevel case, since it has to be decided whether to detrend at the individual level or at the population level. In addition, trends caused by an experimental intervention are often of substantial interest and should arguably not be removed prior to analysis (Wang & Maxwell, 2015). When the trend can be captured by a linear model, possibly including some transformation of the time-dependent covariate, it can be accounted for by adding a linear regression term to the within-level equation of the DSEM (Hamaker et al., 2018, Sec. 6.3.2). With more general nonlinear trends—particularly in the multivariate case—how to take these systematic changes into account is more of an open question. While nonlinear parametric models are one possibility, they require the practitioner to specify their complete parametric form, which may be challenging. In the rest of this paper, the terms "linear" and "linearity" refer specifically to a straight line, whereas "linear in the parameters"

explicitly describes a linear regression model that may include nonlinear transformations of the explanatory variables.

Asparouhov et al. (2018, p. 381) suggested modeling trends using splines, without pursuing it further. In this paper, we follow up on this suggestion and show how nonparametric regression splines can be used to account for trends or effects of time-varying covariates on the outcome of interest both in the univariate and in the multivariate setting. One advantage of this approach is that we only need to assume that the function to be estimated is smooth. Lack of smoothness—denoted wiggliness in the smoothing literature—is quantified by the second derivative of the function in a sense to be precisely described by Equation (11). By this definition, a straight line is perfectly smooth. This means that a priori the function is assumed more likely to be close to linear than highly nonlinear. The Bayesian estimation procedure then learns the functional shape concurrently with estimating all other model parameters. Cyclic patterns can also be accommodated using cyclic splines (Wood, 2017a), which allow a much larger range of function shapes than more-restrictive sine and cosine functions.

We note that detrending of time series using splines has been proposed in a range of fields, including astronomy (Hippke et al., 2019), econometrics (Kauermann et al., 2011), dendrochronology (Klesse, 2021), functional magnetic resonance imaging (Tanabe et al., 2002), and global warming (Wu et al., 2007). However, all of these applications consider the case in which a single unit of observation—possibly multivariate—is followed over time. Since DSEMs are multilevel models aimed at analyzing data following multiple individuals over time, several new methodological issues arise. Firstly, for each observed outcome we need to distinguish between a common trend and individual-specific deviations from this trend, both of which may be nonlinear. Secondly, in the multivariate case, there may be substantial reasons to assume that the shape and/or degree of nonlinearity is similar across multiple domains. Both of these issues call for the use of hierarchical modeling to increase statistical power, and we hence extend previous frequentist approaches for hierarchical nonlinear smooths (Brumback & Rice, 1998; Pedersen et al., 2019) to the Bayesian setting.

To summarize, our goal is to provide an additional tool in the DSEM modeler's toolkit. To keep the presentation simple, we consider lag-1 autoregressive (AR (1)) and lag-1 vector autoregressive (VAR (1)) models

with manifest variables in the main paper and show the extension to the fully general cross-classified DSEM framework in Appendix A. We have implemented the proposed models in the open source probabilistic programming language Stan (Carpenter et al., 2017). A particular advantage of Stan in the present context is that its Hamiltonian Monte Carlo algorithm (Hoffman & Gelman, 2014) scales very well with the number of parameters (Betancourt, 2018), which quickly runs into the tens of thousands for the models considered here. Also, in contrast to methods using Gibbs sampling, Stan gives the user large flexibility in extending the models to new use cases by changing the priors and data generating distributions.

The paper proceeds as follows. In "Two-level AR(1) models", we present a two-level AR(1) model with regression splines. In "Two-level VAR(1) models", we proceed to the multivariate case, and consider two-level VAR(1) models with vector-valued regression splines, and also discuss hierarchical smooth terms. In "Simulation experiment", we conduct simulation experiments comparing the proposed methodology to alternative approaches. In "Analysis of daily diary data on alcohol consumption", we present an example application in which we analyze diary data with recordings of alcohol consumption and stress levels. Finally, in "Discussion", we discuss the results and propose some future extensions. In Appendix A, we show how smooth terms can be incorporated in the within-level models of the fully general DSEM framework and derive the AR(1) and VAR(1) models as special cases.

Two-level AR(1) models

We now consider a two-level AR(1) model, in which the first level refers to the internal dynamics of single individuals and the second level refers to the population level parameters. In what follows, subscripts i and t are assumed repeated for $i = 1, ..., N$ individuals and $t = 1, ..., T$ timepoints, respectively.

In residual formulation, the level-1 model is

$$y_{it} = \alpha_i + \beta_i x_{it} + \phi_i(y_{i,t-1} - \alpha_i - \beta_i x_{i,t-1}) + \delta_{it}, \quad (1)$$

where α_i is the mean of participant i, β_i is the effect of a time-varying covariate x_{it}, ϕ_i is the autoregressive coefficient of participant i, and δ_{it} is a normally distributed residual term, $\delta_{it} \sim N(0, \psi_i^2)$, where ψ_i is the residual standard deviation of participant i. The term $\beta_i x_{it}$ may be a trend, e.g., with $x_{it} = t$, or it can describe the effect of some time-varying covariate – for instance, in McNeish and Hamaker (2020, eq. (4a)), x_{it} is a measure of depressive symptoms and y_{it}

is urge to smoke, thus β_i is the effect of depression on urge to smoke for subject i. It is straightforward to include both a trend term and multiple time-varying covariates by using vectors of regression coefficients and covariates, $\boldsymbol{\beta}_i' \boldsymbol{x}_{it}$.

One limitation of model (1) is that the trend term is restricted to be linear in the parameters. Although this admits nonlinear transformations of x_{it} and can easily be extended to nonlinear parametric models, it requires the researcher to specify the parametric form a priori. An alternative is to instead assume that the effect of x_{it} on the response for subject i takes some generic functional form $f_i(x_{it})$, and formulate the model as

$$y_{it} = \alpha_i + f_i(x_{it}) + \phi_i\{y_{i,t-1} - \alpha_i - f_i(x_{i,t-1})\} + \delta_{it}. \tag{2}$$

Before presenting the details we also state the level-2 model,

$$\alpha_i = \gamma_\alpha + u_{\alpha i} \tag{3}$$

$$\phi_i = \gamma_\phi + u_{\phi i} \tag{4}$$

$$\log \psi_i^2 = \gamma_\psi + u_{\psi i} \tag{5}$$

$$f_i(x_{it}) = s(x_{it}) + s_i(x_{it}). \tag{6}$$

In Equations (3)–(5), we assume that the individual-level mean, autoregressive effect, and log residual variance have a common form with a population component $\boldsymbol{\gamma} = (\gamma_\alpha, \gamma_\phi, \gamma_\psi)'$ and an individual component $\boldsymbol{u}_i = (u_{\alpha i}, u_{\phi i}, u_{\psi i})'$. Note here that since in Bayesian statistics all parameters have associated probability distributions, we refrain from using the terms "fixed" and "random" effects, and instead talk about common/population parameters and individual parameters. The prior distribution of $\boldsymbol{\gamma}$ will vary between applications, but is typically weakly informative and based on domain knowledge. The individual deviations are assumed normally distributed, $\boldsymbol{u}_i \sim N(\boldsymbol{0}, \boldsymbol{T})$, where the covariance matrix is

$$\boldsymbol{T} = \begin{bmatrix} \sigma_\alpha^2 & \sigma_{\alpha\phi} & \sigma_{\alpha\psi} \\ \sigma_{\alpha\phi} & \sigma_\phi^2 & \sigma_{\phi\psi} \\ \sigma_{\alpha\psi} & \sigma_{\phi\psi} & \sigma_\psi^2 \end{bmatrix}. \tag{7}$$

For easier specification of prior distributions we decompose $\boldsymbol{T}$ into a scaling matrix $\text{diag}(\boldsymbol{\tau})$ and a correlation matrix $\boldsymbol{\Omega}$,

$$\boldsymbol{T} = \text{diag}(\boldsymbol{\tau}) \boldsymbol{\Omega} \text{diag}(\boldsymbol{\tau})$$

$$= \begin{bmatrix} \tau_\alpha & & \\ & \tau_\phi & \\ & & \tau_\psi \end{bmatrix} \begin{bmatrix} 1 & \rho_{\alpha\phi} & \rho_{\alpha\psi} \\ \rho_{\alpha\phi} & 1 & \rho_{\phi\psi} \\ \rho_{\alpha\psi} & \rho_{\phi\psi} & 1 \end{bmatrix} \begin{bmatrix} \tau_\alpha & & \\ & \tau_\phi & \\ & & \tau_\psi \end{bmatrix}. \tag{8}$$

For each component of $\boldsymbol{\tau}$ we use a half Cauchy prior on the positive real line,

$$\pi(\tau_j|\xi_j) = \frac{1_{\tau_j \in \mathbb{R}_+}}{2\pi\gamma\left[1 + \left(\tau_j/\xi_j\right)^2\right]}, \quad j \in \{\alpha, \phi, \psi\}, \tag{9}$$

where ξ_j is a scale parameter. For $\boldsymbol{\Omega}$ we use an LKJ prior with shape $\zeta \geq 1$ (Lewandowski et al., 2009).

Common smooth terms

Equation (6) contains the function $f_i(x)$ from (2) decomposed into a common part $s(x)$ and an individual-specific part $s_i(x)$, where we now have dropped the subscripts on x_{it} and let x represent some generic argument to the functions $f_i(x)$, $s(x)$, and $s_i(x)$. We define each of the two component functions using regression splines, whereby the functional form is constructed as a weighted sum over a set of basis functions. For the common term we have

$$s(x) = \sum_{k=1}^{K} \beta_k b_k(x) \tag{10}$$

where β_k are weights and $b_k(x)$ are basis functions. Cubic regression splines (Wood, 2017a, Ch. 5.3.1), thin-plate regression splines (Wood, 2003) and Bayesian P-splines (Brezger & Lang, 2006; Eilers & Marx, 1996; Wood, 2017b) are directly applicable in the proposed framework.

We use the intermediate rank approach to smoothing (Wood, 2011), in which the number of basis functions K is chosen large enough to support a wide range of functional shapes yet much smaller than the number of observations. Without further penalization, simply plugging the term (10) into (2) will overfit considerably unless K is very small. We define the degree of wiggliness of $s(x)$ as the integral of its squared second derivative over the domain $\mathcal{X}$ of x,

$$\int_{\mathcal{X}} \left[s''(x)\right]^2 \mathrm{d}x. \tag{11}$$

For a perfectly smooth function with no wiggliness the integral equals zero. Our assumption that $s(x)$ is smooth can hence be formalized by assuming that (11) is a priori more likely to be close to zero than not.

We refer to Wood (2017a, Ch. 4.2.4, Ch. 5.8) for details, but note that the integral in (11) can be written in terms of the coefficients $\boldsymbol{\beta} = (\beta_1, ..., \beta_K)'$ in (10) as $\int_{\mathcal{X}}[s''(x)]^2 \mathrm{d}x = \boldsymbol{\beta}' \boldsymbol{S} \boldsymbol{\beta}$, where $\boldsymbol{S}$ is a symmetric $K \times K$ penalty matrix. The set of basis functions $\{b_1(x), ..., b_K(x)\}$ used in (10) spans a space

containing both constant, linear, and nonlinear functions. That is, although $s(x)$ in general is nonlinear it is possible to set $\boldsymbol{\beta}$ such that $s(x) = ax + b$ for some constants a and b. However, these values of $\boldsymbol{\beta}$ yield $\int_{\mathcal{X}}[s''(x)]^2\mathrm{d}x = 0$ since $s''(x) = 0$ when $s(x)$ is linear. Positive definiteness of a symmetric matrix $\boldsymbol{S}$ is defined by $\boldsymbol{\beta}'\boldsymbol{S}\boldsymbol{\beta} > 0$ for all $\boldsymbol{\beta}$. As we have seen, in this case there are two dimensions (constant and linear functions) along which $\boldsymbol{\beta}'\boldsymbol{S}\boldsymbol{\beta} = 0$, and it follows that $\boldsymbol{S}$ is only positive semidefinite, with rank $K - 2$.

Next, for the intercepts α_i in (2) to be identifiable, $s(x)$ should not have its own intercept. This is imposed by removing all constant functions from the spline basis using sum-to-zero constraints (Wood, 2017a, Ch. 4.3.1, Ch. 5.4.1). The result is a transformed basis with $K - 1$ functions $\{\bar{b}_1(x), ..., \bar{b}_{K-1}(x)\}$, and the representation

$$s(x) = \sum_{k=1}^{K-1} \bar{\beta}_k \bar{b}_k(x). \tag{12}$$

With this basis, for a given sample of size NT, we have $\sum_{i=1}^{N}\sum_{t=1}^{T} s(x_{it}) = 0$ for any choice of weights $\bar{\beta}_k$. Importantly, any function shape which can be represented by (10) can also be represented by (12), and they will only differ by a constant term. In this new basis, the integral (11) is given by $\int_{\mathcal{X}}[s''(x)]^2\mathrm{d}x = \bar{\boldsymbol{\beta}}'\bar{\boldsymbol{S}}\bar{\boldsymbol{\beta}}$, where $\bar{\boldsymbol{\beta}} = (\bar{\beta}_1, ..., \bar{\beta}_{K-1})'$ and $\bar{\boldsymbol{S}}$ is a $(K-1) \times (K-1)$ penalty matrix. It is now impossible to choose $\bar{\boldsymbol{\beta}}$ such that $s(x)$ is constant, but a linear function can still be constructed. More technically, $\bar{\boldsymbol{\beta}}'\bar{\boldsymbol{S}}\bar{\boldsymbol{\beta}} = 0$ for any $\bar{\boldsymbol{\beta}}$ such that $s(x) = kx$ for some constant $k \neq 0$, and all linear functions are said to be in the nullspace of $\bar{\boldsymbol{S}}$. On the other hand, for any $\bar{\boldsymbol{\beta}}$ such that $s(x)$ is nonlinear, $\bar{\boldsymbol{\beta}}'\bar{\boldsymbol{S}}\bar{\boldsymbol{\beta}} > 0$, and all nonlinear functions are said to be in the range space of $\bar{\boldsymbol{S}}$. It follows that the $(K-1) \times (K-1)$ symmetric matrix $\bar{\boldsymbol{S}}$ is positive semidefinite and has rank $(K-1) - 1 = K - 2$.

In a Bayesian setting, the smoothness assumption can be naturally encoded by assuming that the weights $\bar{\boldsymbol{\beta}}$ have a multivariate normal prior centered at zero, with covariance matrix proportional to the inverse of the penalty matrix $\bar{\boldsymbol{S}}$. Since $\bar{\boldsymbol{S}}$ is rank deficient (as is the original $\boldsymbol{S}$), the standard matrix inverse cannot be computed. However, the Moore-Penrose pseudoinverse (Moore, 1920; Penrose, 1955), which we denote by $\bar{\boldsymbol{S}}^-$ can be used instead. For definitions and properties of the pseudoinverse we refer to Gentle (2024, Ch. 3.7.2). For our purposes, the key is that it lets us write the smoothing prior as a normal distribution

$$\bar{\boldsymbol{\beta}} \sim N\left(\boldsymbol{0}, \tau_\beta^2 \bar{\boldsymbol{S}}^-\right), \tag{13}$$

where the scale parameter τ_β quantifies the amount of deviation from linearity. The prior in (13) is a degenerate multivariate normal distribution for which values of $\bar{\boldsymbol{\beta}}$ corresponding to a linear $s(x)$ are not penalized. In particular, if no deviation from linearity is allowed by setting $\tau_\beta = 0$, it effectively turns into an improper uniform prior on the slope k of a linear function of the form $s(x) = kx$.

For practical model fitting, the spline basis and coefficients are transformed in yet another step into a representation in matrix-vector form

$$\tilde{\boldsymbol{x}}_1 \tilde{\beta}_1 + \tilde{\boldsymbol{X}}_2 \tilde{\boldsymbol{\beta}}_2 \tag{14}$$

where $\tilde{\boldsymbol{x}}_1$ is a vector of length NT containing the linear term and $\tilde{\boldsymbol{X}}_2$ is a matrix of size $NT \times (K - 2)$ containing the part of the spline basis in the range space of the penalty matrix. In this transformed representation, the smoothing prior amounts to assuming that the penalized coefficients $\tilde{\boldsymbol{\beta}}_2$ are independent and identically distributed $\tilde{\boldsymbol{\beta}}_2 \sim N(\boldsymbol{0}, \tau_\beta^2 \boldsymbol{I}_{K-2})$, where $\boldsymbol{I}_{K-2}$ is an identity matrix of size $(K - 2) \times (K - 2)$. As for the other variance components, we use the half Cauchy prior (9) for τ_β with scale parameter ξ_β. The precision, τ_β^{-2} now is a regularization parameter; a high value of τ_β^{-2} means that all components of $\tilde{\boldsymbol{\beta}}_2$ are close to zero and thus $s(x)$ close to linear, and the opposite for large values of τ_β^{-2}. For the linear part we recommend a weakly informative prior based on prior knowledge.

For our purposes, the R (R Core Team, 2024) package mgcv (Wood, 2017a) contains functions which set up the basis $\{b_1(x), ..., b_K(x)\}$ and the penalty matrix $\boldsymbol{S}$ and transform them to the form (14). This is also how the brms package (Bürkner, 2017) sets up smooth functions for models which are subsequently estimated in Stan (Carpenter et al., 2017). However, brms does not directly support estimating the DSEMs discussed in this paper.

Individual smooth terms

The individual smooth terms $s_i(x)$ reflect how each participant deviates from $s(x)$ and are composed in the same way as (10),

$$s_i(x) = \sum_{l=1}^{L} \beta_{li} d_l(x), \tag{15}$$

where there are now L basis functions $\{d_1(x), ..., d_L(x)\}$ with weights β_{li}. The construction described in the previous paragraph also applies to $s_i(x)$, and we use the smoothing prior $\boldsymbol{\beta}_i \sim N(\boldsymbol{0}, \tau_{\beta, \mathrm{ind}}^2 \boldsymbol{S}_{\mathrm{ind}}^-)$, where we assume that the trajectories for all individuals are sampled from a distribution in which the typical deviation from linearity is quantified

by the standard deviation $\tau_{\beta,\text{ind}}$. Importantly, this construction allows individual curves to differ in their degree of nonlinearity. For $\tau_{\beta,\text{ind}}$ we again use a half Cauchy prior, with scale parameter $\xi_{\beta,\text{ind}}$.

A more restrictive model in which the individual-specific trajectories only deviate linearly from the common curve can be obtained by setting $s_i(x) = \beta_i x$. In this case the transformed model (14) does not contain a penalized part $\tilde{X}_2 \tilde{\beta}_2$ and hence no smoothing parameter $\tau_{\beta,\text{ind}}$ is needed.

Two-level VAR(1) models

In a VAR(1) model, P outcome variables are measured at each timepoint. Level-1 model (2) thus generalizes to

$$y_{it} = \alpha_i + f_i(x_{it}) + \Phi_i \{y_{i,t-1} - \alpha_i - f_i(x_{i,t-1})\} + \delta_i,$$
(16)

where boldface lowercase letters now denote vectors of size P, y_{it} contains observed outcomes for subject i at timepoint t, α_i contains intercepts for subject i, $f_i(x_{it})$ is a vector-valued function of a vector of time-varying covariates x_{it}, and δ_i is a vector of residual terms. Each component of δ_i is assumed normally distributed, so $\delta_{ip} \sim N(0, \psi_{ip}^2)$ and we define $\psi_i = (\psi_{i1}, ..., \psi_{iP})'$. The autoregressive and cross-lagged effects between the outcomes are represented by the $P \times P$ matrix Φ_i, which we write as

$$\Phi_i = \begin{bmatrix} \phi_{i,11} & \phi_{i,12} & \cdots & \phi_{i,1P} \\ \phi_{i,21} & \phi_{i,22} & \cdots & \\ \vdots & & \ddots & \\ \phi_{i,P1} & & \cdots & \phi_{i,PP} \end{bmatrix}.$$

The level-2 model (3)–(6) generalizes to

$$\alpha_i = \gamma_\alpha + u_{\alpha i}$$
(17)

$$\text{vec}(\Phi_i) = \gamma_\phi + u_{\phi i}$$
(18)

$$\log \psi_i^2 = \gamma_\psi + u_{\psi i}$$
(19)

$$f_i(x_{it}) = s(x_{it}) + s_i(x_{it}),$$
(20)

where $\text{vec}(\Phi_i)$ denotes the vector operator which turns the $P \times P$ matrix Φ into a column vector of size P^2. Again we collect all the individual deviations into a vector $u_i = (u'_{\alpha i}, u'_{\phi i}, u'_{\psi i})'$, which is now of size $2P + P^2$. The deviations are assumed normally distributed around zero with covariance matrix T, with the latter decomposed as in (8). The half Cachy prior (9) for the scale parameters and the LKJ prior for the correlation matrix still apply.

Equation (20) shows the vector valued smooth function decomposed into a common component $s(x_{it})$ and an individual-level component $s_i(x_{it})$. We assume each vector component takes the form

$$f_i(x_{it}) = \begin{bmatrix} f_{1i}(x_{1,it}) \\ \vdots \\ f_{Pi}(x_{P,it}) \end{bmatrix} = \begin{bmatrix} s_1(x_{1,it}) \\ \vdots \\ s_P(x_{P,it}) \end{bmatrix} + \begin{bmatrix} s_{1i}(x_{1,it}) \\ \vdots \\ s_{Pi}(x_{P,it}) \end{bmatrix}$$
(21)

where $x_{p,it}$ is a scalar time-varying covariate used in the pth smooth function. Both the common and individual terms can be constructed as described in the previous section.

Multivariate hierarchical smooth terms

In some applications, it may be reasonable to assume a priori that the common terms are similar in terms of wiggliness or overall shape. If the assumption is correct, it will yield increased power compared to the case of estimating the terms independently. We refer to Pedersen et al. (2019) for a thorough introduction to such hierarchical generalized additive models in the maximum likelihood setting. Here we discuss the model assumptions which seem most useful in cognitive science, and describe the prior distributions necessary to perform Bayesian estimation.

When the components of y have similar trends, e.g., if y contains measurements of closely related skills like different mathematical tasks (Judd et al., 2024), we can rewrite (21) to contain both a smooth term $s(x)$ which is common for all outcomes and a smooth term $s_p(x)$ which is unique to each outcome, yielding

$$f_{pi}(x) = s(x) + s_p(x) + s_{pi}(x), p = 1, ..., P.$$
(22)

We can further assume that the component specific deviations $s_p(x)$ have similar (but not identical) degrees of nonlinearity, by which we mean that all $s_p(x)$ for $p = 1, ..., P$ are sampled from a common distribution in which the typical amount of deviation from nonlinearity is quantified by a scale parameter. Making the same assumption for the individual deviations $s_{pi}(x)$, we have one scale parameter τ_β for $s(x)$, one scale parameter $\tau_{\beta,2}$ for $s_p(x)$ $(p = 1, ..., P)$ and finally $\tau_{\beta,\text{ind}}$ for $s_{pi}(x)$ $(p = 1, ..., P)$.

An example of such hierarchical smooth terms is shown in Figure 1, simulated for $P = 3$ outcome variables and $N = 5$ subjects. Note that both the overall trend and the overall smoothness is similar across the outcome variables.
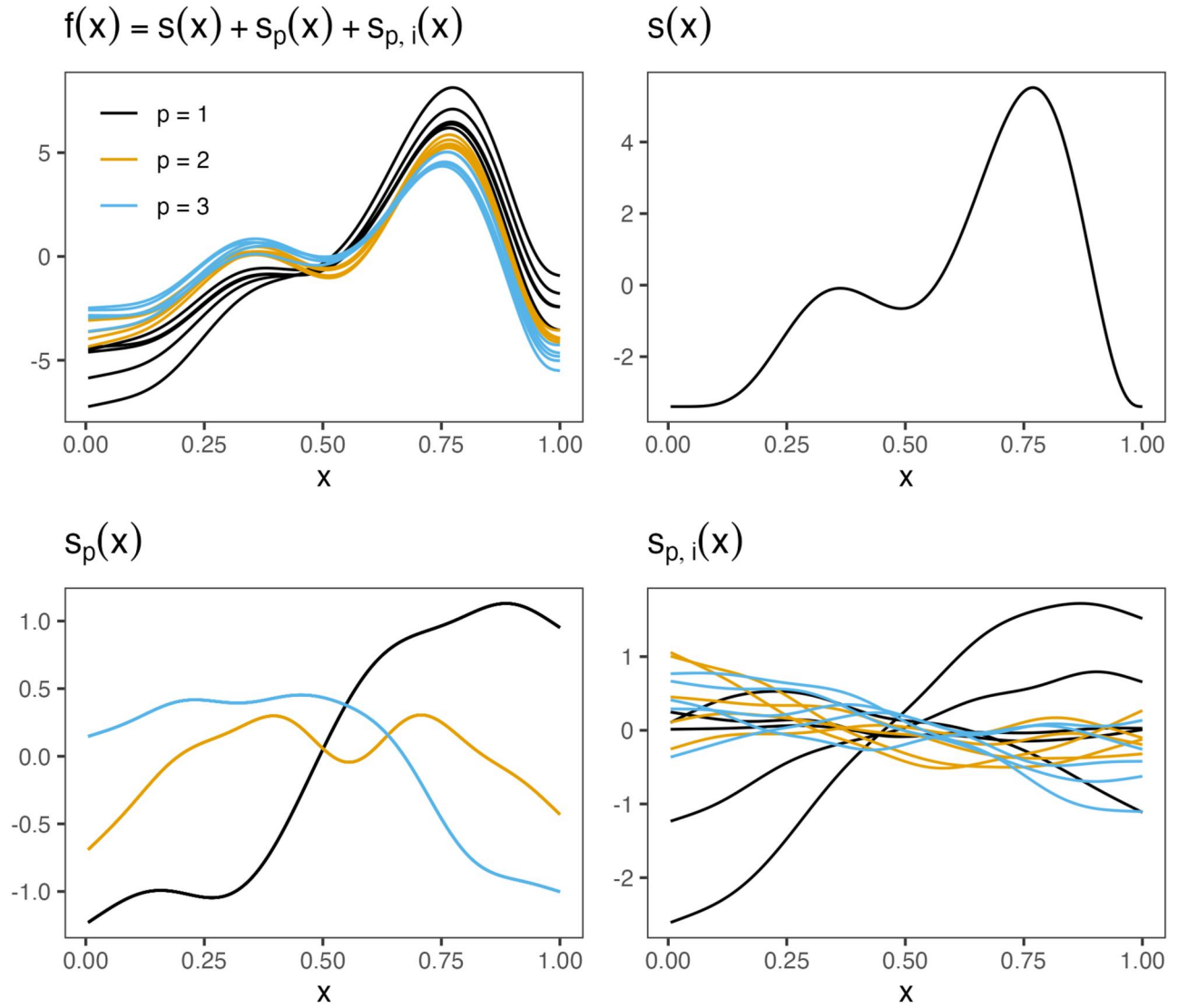
Figure 1. Example of hierarchical smooth terms with similar trends and similar degree of nonlinearity. The upper left plot shows the overall trend for three domains and five individuals. The upper right plot shows the common trend component, the lower left plot shows the domain-specific components, and the lower right plot shows the individual-specific components.

If no common trend can be assumed, as would be the case when the outcome variables are of different character, Equation (22) reduces to (21). In this case each $s_p(x)$ represents the overall trend for the outcome variable in question. If in addition no assumption about similar degree of nonlinearity can be made, we get P scale parameters $\tau_{\beta, p}$ for $p = 1, ..., P$.

Simulation experiment

We here report results from a simulation experiment aimed at understanding the consequences of ignoring nonlinear effects in the level-1 model, and also the cost in statistical power of using the more flexible regression splines when the truth is linear.

Data generation

We simulated 200 random datasets from an AR(1) model as specified by Equations (2) and (3)–(6), with $N = 50$ participants measured at $T = 200$ identically spaced timepoints and with $f_i(x_{it})$ representing a time trend. We let $x_t = \Delta t$ represent chronological time at timepoint t, with $\Delta = 1/200$, so $x_t \in [0.005, 1]$. The common trend had the functional form

$$s(x) = 0.2 \cdot (1 - x)^{11} \cdot (10 \cdot x)^6 + 10 \cdot \left\{ 10 \cdot (1 - x) \right\}^3 \cdot x^{10}$$

with its mean subtracted so it summed to zero over the range of x_t. The function can be seen in the upper right panel of Figure 1 and is a reverted version of a function introduced by Gu and Wahba (1991).

For the individual deviations $s_i(x)$ we first set up a basis consisting of 10 thin-plate regression splines (Wood, 2003), and then transformed it to the form (14). The linear part was sampled from a standard normal distribution and the penalized part from a normal distribution with variance $\tau^2_{\beta,\,\mathrm{ind}} = 4$. We set $\gamma_\alpha = 0$ and $\sigma_\alpha = 1$, indicating that the individual means are normally distributed around zero with a standard deviation of one. For the autoregressive parameter, we set the mean to $\gamma_\phi = 0.6$ and the standard deviation to $\sigma_\phi = 0.05$. The mean of the logarithm of the residual variance was set to $\gamma_\psi = -3$ with a standard deviation $\sigma_\psi = 0.3$, implying that the mean of the residual standard deviation ψ_i was $\sqrt{\exp\{\gamma_\psi + \sigma_\psi^2/2\}} \approx 0.308$, using the properties of the log-normal distribution. We assumed no correlation between the individual deviations α_i, ϕ_i, and $\log\psi_i^2$, so all off-diagonal elements of the covariance matrix T in (7) were zero.

We also generated 200 additional datasets in which both the common and individual trends were linear, by setting $\tilde{\beta}_2 = 0$ in the transformed representation (14).

Bayesian estimation

For each simulated dataset we estimated the AR(1) model (2)–(6) with the smooth term $f_i(x_t)$ consisting of a common component $s(x_t)$ and individual components $s_i(x_t)$. Four different choices of basis functions were compared. The most restrictive model had only linear terms identically to (1), and hence no smoothing parameters. The second was a close-to-linear spline model with $K = 4$ and $L = 4$ thin-plate regression splines for the common and individual component, respectively. The next two were more flexible spline models, with $K = 20$ and $L = 10$, and $K = 40$ and $L = 20$, thin-plate regression splines for the common and individual terms, respectively. The goal with estimating these models was to understand the consequences of assuming linearity or close-to-linearity when the truth is nonlinear, as well as the consequence of using a basis that is more flexible than necessary.

For both smoothing parameters we used the half Cauchy prior (9) with scale $\xi = 1$. For the linear coefficients we used a weakly informative normal distribution prior with mean zero and standard deviation 3 and for the level-2 means $\gamma = (\gamma_\alpha, \gamma_\phi, \gamma_\psi)'$ we used a normal distribution with zero correlation and standard deviation 3. For the correlation matrix Ω we used an LKJ prior (Lewandowski et al., 2009) with shape

$\zeta = 2$, thus the correlation of the level-2 parameters in γ was estimated although it was set to zero in the data generation process. Finally, for the scale parameters, $\mathrm{diag}(\tau)$ in (8), we used a half Cauchy prior with scale $\xi = 2.5$.

For all models and simulated datasets we ran four independent chains with 6000 iterations of the NUTS sampler (Hoffman & Gelman, 2014) discarding the first 3000 as burn-in, yielding 12,000 samples from the posterior distribution. The default options of version 2.32.6 of the `rstan` package were used (Carpenter et al., 2017), except that the maximum treedepth was set to 20.

Results

Table 1 shows simulation results for the population mean parameters and their standard deviations, which are typically the quantities of main interest in a DSEM analysis. When the true trend was linear, there was essentially no difference between the four models, and they all had low bias. That is, allowing a nonlinear trend by using the flexible spline models had minimal cost in terms of accuracy and power.

When the true trend was nonlinear, on the other hand, both the linear and the constrained spline model suffered from high bias in all parameters, whereas the two flexible spline models recovered all parameters quite well. In particular, both the linear model and the constrained model on average estimated the autoregressive effect γ_ϕ to be one, and at the same time erroneously estimated the standard deviation σ_ϕ to be zero, indicating no individual variability. In contrast, the two flexible spline models on average had a slight upward bias in this parameter,

Table 1. Simulation results with AR(1) models.

	Linear	Splines (4/4)	Splines (20/10)	Splines (40/20)
Linear data				
$\gamma_\alpha = 0$	−0.01 (0.14)	−0.02 (0.14)	−0.01 (0.13)	−0.02 (0.13)
$\gamma_\phi = 0.6$	0.60 (0.01)	0.60 (0.01)	0.60 (0.01)	0.60 (0.01)
$\gamma_\psi = -3$	−3.00 (0.05)	−2.99 (0.05)	−3.00 (0.05)	−3.00 (0.05)
$\sigma_\alpha = 1$	1.05 (0.10)	1.03 (0.10)	1.03 (0.11)	1.03 (0.10)
$\sigma_\phi = 0.05$	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)
$\sigma_\psi = 0.3$	0.31 (0.04)	0.31 (0.04)	0.32 (0.04)	0.32 (0.04)
Nonlinear data				
$\gamma_\alpha = 0$	15.27 (2.83)	13.79 (2.04)	−0.10 (0.31)	−0.06 (0.23)
$\gamma_\phi = 0.6$	1.00 (0.01)	0.99 (0.01)	0.67 (0.10)	0.67 (0.10)
$\gamma_\psi = -3$	−2.12 (0.03)	−2.29 (0.04)	−2.96 (0.07)	−2.97 (0.07)
$\sigma_\alpha = 1$	0.98 (0.15)	0.61 (0.07)	0.97 (0.16)	0.99 (0.16)
$\sigma_\phi = 0.05$	0.00 (0.00)	0.00 (0.00)	0.06 (0.03)	0.06 (0.03)
$\sigma_\psi = 0.3$	0.13 (0.03)	0.17 (0.03)	0.31 (0.04)	0.31 (0.04)

Notes: The terms "Splines (K/L)" indicate that the common smooth term had K basis functions and the individual smooth terms had L basis functions. True values of each parameter are shown in the leftmost column. The number shows average posterior means, with standard deviations of posterior means in parentheses.
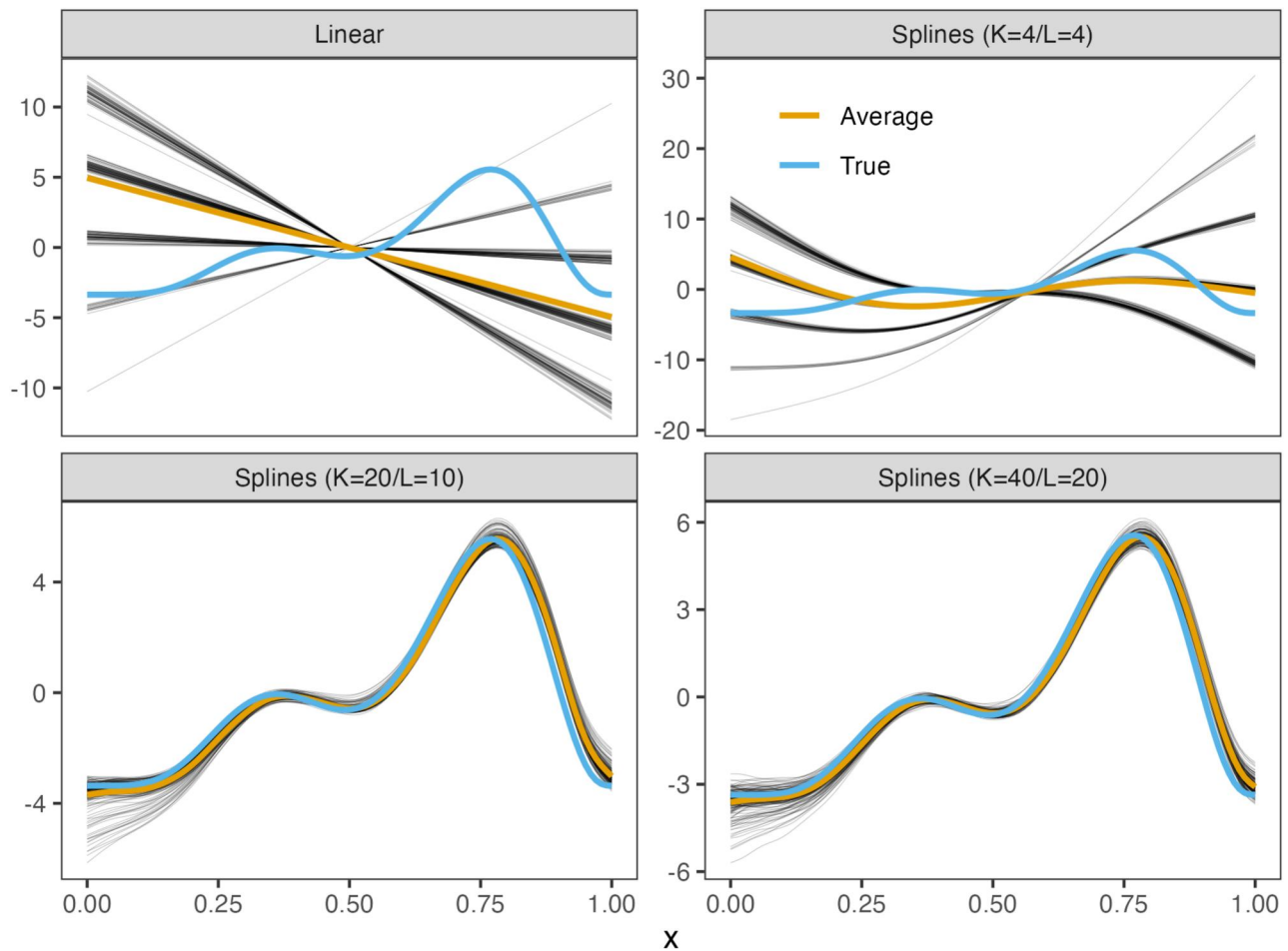
Figure 2. The black lines show the posterior mean of the common smooth function $s(x)$ for each simulated dataset when using a linear model (top left), constrained splines (top right), more flexible splines (bottom left), and very flexible splines (bottom right). The thick orange line shows the average across simulations and the thick blue line shows the true function. Note that the y-axis scale differs between figures.

but estimated the standard deviation very close to its true value.

Figure 2 shows the posterior means of $s(x)$ across all simulated datasets when the true trend was nonlinear, together with the average of the posterior means and the true function. With the two most flexible spline models, the estimates for each dataset as well as the average are close to the true value, whereas both the linear model and the constrained spline model failed spectacularly in capturing the true trend.

Table 1 and Figure 2 also illustrate a practical method of choosing the number of basis functions for a particular analysis. Firstly, we see that the smoothing prior protects against overfitting, since the most flexible model ($K = 40/L = 20$) had practically identical performance to the model with $K = 20$ and $L = 10$. Hence, the main concern is to ensure that the splines are flexible enough to allow proper estimation of the underlying function. This can be achieved by gradually increasing the number of basis functions,

while monitoring the estimates of the parameters of interest. When the parameters of interest no longer change notably when the basis size increases, we have a clear indication that the chosen values of K and L are large enough. In this simulation example it was thus clear that ($K = 4/L = 4$) was too small but ($K = 20/L = 10$) was sufficiently large.

Analysis of daily diary data on alcohol consumption

We here present an application of the proposed methods using a dataset from Carney et al. (2000). The data consist of daily reports of alcohol consumption, desire to drink, and perceived stress from 93 individuals followed over a period of 60 days. Carney et al. (2000), which we refer to for further details, analyzed a subset of 83 individuals, removing 10 who were unemployed at the time of data collection. Figure 3 shows descriptive summaries of the three main
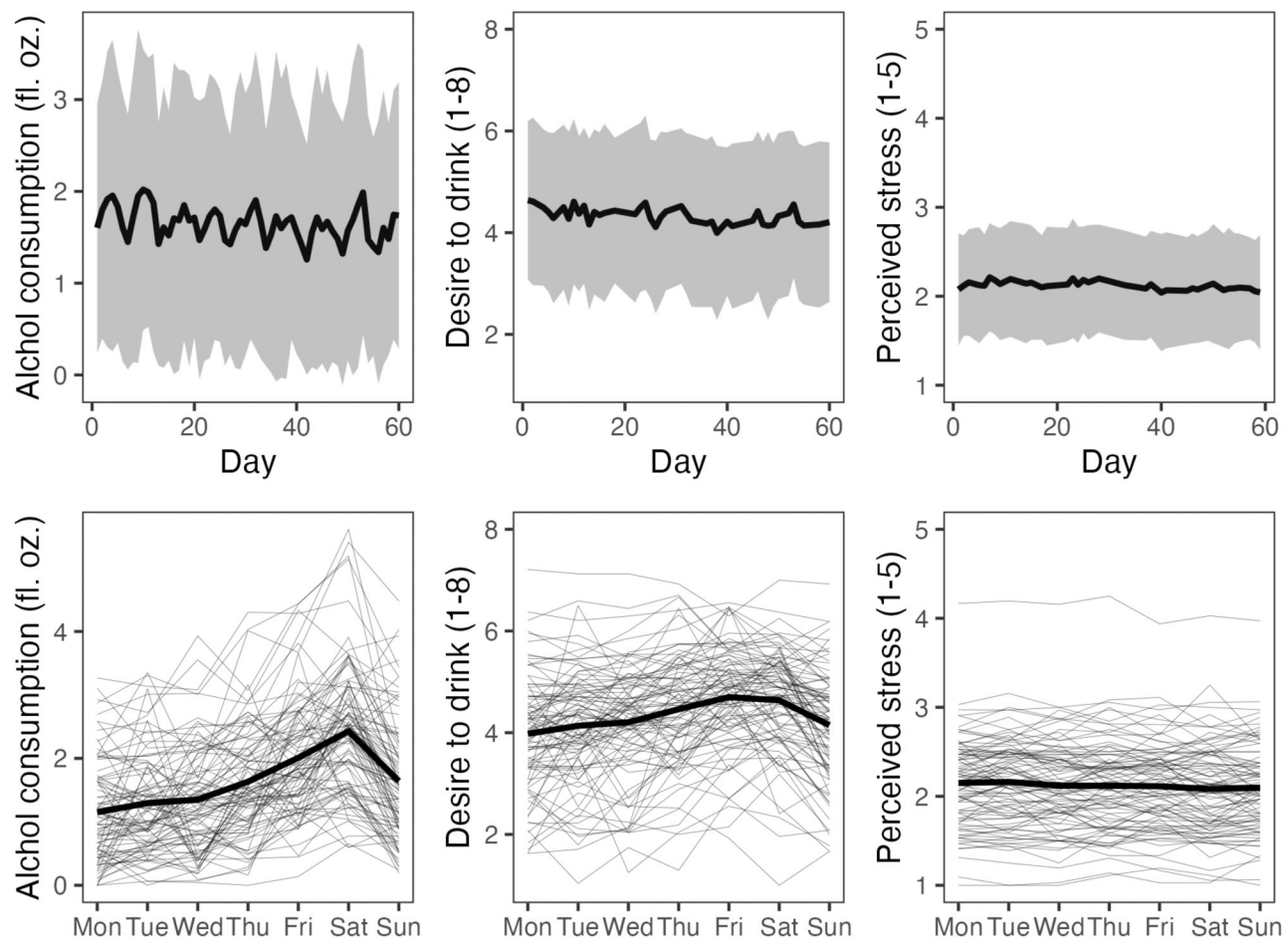
Figure 3. Descriptive plots of the data. The top row shows the average alcohol consumption, desire to drink, and perceived stress level across the 93 participants over the 60 days of the study. Shaded regions show ± one standard deviation. In the bottom row, the thin lines show the average across weekdays for each participant, and the thick black line shows the average over all participants.

variables of interest. Considering the top row, there seems to be a weak downward trend in alcohol consumption and desire to drink throughout the study period. The bottom row demonstrates cyclic patterns in alcohol consumption, with an increase throughout the week and a peak on Saturday. A weaker cycle is also apparent in desire to drink. Perceived stress, on the other hand, seems very stable both throughout the study period and during the week. Also note from Figure 3 that there are large individual variations in average level for all three variables. There might also be individual variations in the cyclic patterns, but this is harder to detect by the naked eye.

Carney et al. (2000) analyzed an extended version of the data using multilevel linear regression models, studying how alcohol consumption and desire to drink could be predicted by perceived stress and positive and negative events on the same day. Liu and West (2016) used the data to illustrate the

consequence of ignoring weekly cycles in daily report data, still in the context of multilevel regression models. Given the large number of timepoints per individual and the relatively dense sampling, we here instead focus on the temporal dynamics. We start by studying how alcohol consumption on a given day depends on alcohol consumption the day before, using an AR(1) model. We then consider the autoregressive and cross-lagged effects between alcohol consumption, desire to drink, and perceived stress, using a VAR(1) model.

AR(1) model with cyclic and long-term trends

Our level-1 model is given by (2) and the first three lines of the level-2 model are given by (3)–(5). However, to allow alcohol consumption to vary by weekday, we modify (6) so it contains both a long-term trend in alcohol consumption over the duration

of the study as well as a cyclic term varying by weekday. The smooth terms thus become

$$f_i(\boldsymbol{x}_{it}) = s_{\text{cycle}}(x_{1it}) + s_{\text{trend}}(t) + s_{\text{cycle},i}(x_{1it}) + s_{\text{trend},i}(t),$$

(23)

where $\boldsymbol{x}_{it} = (x_{1it}, t)'$ is a vector containing the weekday $x_{1it} \in \{1, ..., 7\}$. The day of the study $t \in \{1, ..., 60\}$ did not correspond to the same calendar day for each participant, so in general we have $x_{1it} \neq x_{1jt}$ for $i \neq j$.

In Equation (23), $s_{\text{cycle}}(x_{1it})$ is a cyclic cubic spline (Wood, 2017a, Ch. 5.3.2) constrained so that the endpoints match up to second derivatives, and $s_{\text{cycle},i}(x_{1it})$ are individual splines defined equivalently. That is, letting $f^{(d)}(x)$ denote the d'th derivative of some function $f(x)$, we require $s_{\text{cycle}}^{(d)}(1) = s_{\text{cycle}}^{(d)}(7)$ and $s_{\text{cycle},i}^{(d)}(1) = s_{\text{cycle},i}^{(d)}(7)$ for $d = 0, 1, 2$ and $i = 1, ..., N$. Considering the bottom left panel of Figure 3, Tuesdays and Wednesdays were chosen as endpoints, since the assumption of matching values up to second derivatives seemed more likely to be satisfied for these days rather than Mondays and Sundays, i.e., Wednesday corresponded to $x_{1it} = 1$ and Tuesdays to $x_{1it} = 7$.

Both the common and individual cyclic splines were set up using seven basis functions with knots placed on each weekday. In addition, $s_{\text{trend}}(t)$ is a smooth term representing any overall trend across the duration of the study, and $s_{\text{trend},i}(t)$ is an additional individual component for the overall trend. Both trend terms were set up using six cubic regression splines as basis functions. We do not have access to information about the date of data collection for each individual, and thus $s_{\text{trend}}(t)$ may represent a mix of seasonal effects (Cho et al., 2001) and potential reactivity effects, by which the measurement itself influences behavior (Shiffman et al., 2008).

Priors and the computational set-up were as described for the simulations in "Bayesian estimation", except that we now used the default value of 10 for the treedepth. Obtaining 12,000 post-burn-in draws from the posterior took about 15 min on a MacBook Pro. Supplementary Figure S1 shows diagnostics for the Hamiltonian Monte Carlo algorithm, which indicate that the chains mixed well. The effective sample size averaged 10,906 over all parameters and was always above the lower bound of 400 suggested by Vehtari et al. (2021), while the $\hat{R}$ statistic was very close to 1 and well below the upper limit of 1.01 which may indicate convergence problems (Vehtari et al., 2021).

Model selection

We first compared the full spline model (23) to six simpler models. First, models were fitted without either the individual cyclic term, the individual trend term, or without both individual terms. A model comparison using Bayes factors gave strong support for removing the individual trend term but keeping the individual cyclic term. Next, we performed the same analysis by removing the common cyclic term, the common trend term, or both. The Bayes factors still showed strong support for the model containing both common terms as well as individual cyclic terms, over all of these simpler models. Hence, in the chosen model the full smooth term (23) was reduced to

$$f_i(\boldsymbol{x}_{it}) = s_{\text{cycle}}(x_{1it}) + s_{\text{trend}}(t) + s_{\text{cycle},i}(x_{1it}).$$

The marginal likelihood of each model was computed using bridge sampling (Meng & Schilling, 2002), with the `bridgesampling` package (Gronau et al., 2020), from which the Bayes factors could be directly calculated. The marginal likelihoods of all models are shown in Supplementary Table S1.

Posterior estimates

We here present posterior estimates from the chosen model. Figure 4 (left) shows a cyclic pattern, whereby the consumption increases considerably on Fridays and Saturdays, and then drops steeply again on Sundays. Figure 4 (right) shows that alcohol consumption has a moderate downward trend throughout the study.

Figure 5 shows histograms of the individual intercepts, autoregressive coefficients, and noise levels. Note that there is large variability in intercepts, meaning that the average consumption varies considerably between people. Also the noise level has high variability, implying large variation in day-to-day fluctuation among individuals. The center plot in Figure 5 for the autoregressive coefficient shows that for most people, the deviation from the overall trend on a given day is weakly positively correlated with the deviation on the previous day.

Table 2 summarizes the population-level means and standard deviations. First, we see clear evidence that the mean intercept, autoregressive effect, and residual standard deviation are larger than zero. Also the standard deviations for the same quantities are nonzero with relatively narrow credible intervals, indicating the presence of individual variability. Considering the smooth terms, τ_{cycle} is the smoothing parameter for $s_{\text{cycle}}(x_{1it})$ and $\tau_{\text{cycle,ind}}$ the common smoothing parameter for all $s_{\text{cycle},i}(x_{1it})$. We see that
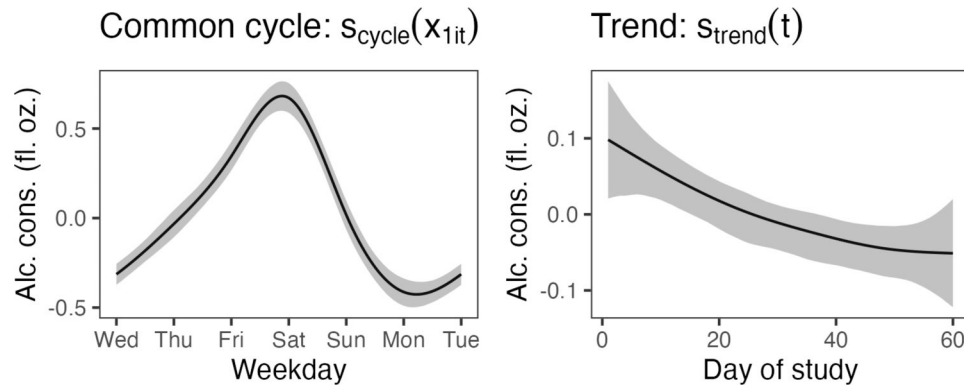
Figure 4. Smooth terms for the AR(1) model for alcohol consumption. The solid black lines are posterior means and the shaded regions are 95% credible intervals.
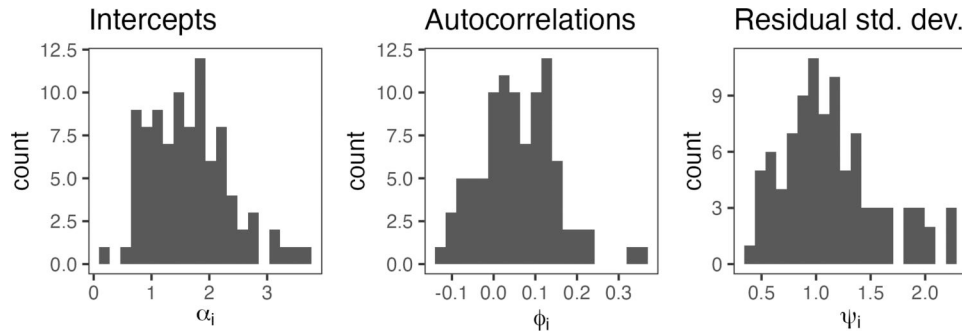

Figure 5. Histograms of the individual intercepts, autoregressive coefficients, and residual standard deviations.

the latter is larger than the former, showing that the common cyclic term's deviation from linearity is larger than the deviation from linearity of the individual terms around this common term. The smoothing parameter for the common trend term, τ_{trend} is small and contains zero in its credible interval, suggesting that a linear term would be sufficient here. This is also in accordance with what we see in the bottom row of Figure 4 and highlights a strength of the Bayesian approach with smoothing priors; when the data don't provide evidence of large nonlinearities, the posterior distribution will tend toward linearity.

VAR model for positive and negative daily events, perceived stress, and alcohol use

We next considered a VAR(1) model for alcohol consumption, y_{1it}, desire to drink, y_{2it}, and perceived stress, y_{3it}, with level one as defined in Equation (16). Based on the AR(1) analyses in the previous section and the plots in Figure 3, we modeled the cycles in alcohol consumption and desire to drink using hierarchical smooth terms and the long-term trends with linear effects. No cyclic term was included for perceived stress. The smooth function was thus

Table 2. Summary of the population parameters of the AR(1) model for alcohol consumption.

Parameter	Mean	95% credible interval
γ_α	1.65	(1.50, 1.80)
γ_ϕ	0.06	(0.02, 0.10)
γ_ψ	0.09	(−0.08, 0.26)
$E(\psi_i)$	1.24	(1.12, 1.38)
σ_α	0.73	(0.63, 0.86)
σ_ϕ	0.13	(0.09, 0.17)
σ_ψ	0.82	(0.70, 0.95)
τ_{cycle}	0.13	(0.06, 0.29)
$\tau_{cycle,ind}$	0.06	(0.05, 0.07)
τ_{trend}	0.02	(0.00, 0.06)

Note: $E(\psi_i) = \sqrt{\exp\{\gamma_\psi + \sigma_\psi^2/2\}}$ is the estimated population mean of the residual standard deviation.

defined by

$$f_i(x_{it}) = \begin{bmatrix} f_{1i}(x_{it}) \\ f_{2i}(x_{it}) \\ f_{3i}(x_{it}) \end{bmatrix} = \begin{bmatrix} s_{cycle,1}(x_{1it}) + s_{cycle,1i}(x_{1it}) + \beta_{1i}t \\ s_{cycle,2}(x_{1it}) + s_{cycle,2i}(x_{1it}) + \beta_{2i}t \\ \beta_{3i}t \end{bmatrix},$$

where $s_{cycle,1}(x_{1it})$ and $s_{cycle,2}(x_{1it})$ are cyclic terms for alcohol consumption and desire to drink. Since these are qualitatively different variables, measured in different units, we did not use a common smooth term as in (22) and used separate smoothing parameters, $\tau_{\beta,1}$ and $\tau_{\beta,2}$. Next, $s_{cycle,1i}(x_{1it})$ and $s_{cycle,2i}(x_{1it})$ are individual terms for each domain, with smoothing

parameters $\tau_{\beta,\mathrm{ind},1}$ and $\tau_{\beta,\mathrm{ind},2}$. The cyclic terms were set up as for the AR(1) model of the previous section.

The regression coefficients for the long-term trend, $\boldsymbol{\beta}_i = (\beta_{1i}, \beta_{2i}, \beta_{3i})'$, were assumed normally distributed with mean $\boldsymbol{\gamma}_\beta$ and allowed to be correlated with the intercepts, autoregressive and cross-lagged effects, and residual standard deviations. Hence we added an additional line $\boldsymbol{\beta}_i = \boldsymbol{\gamma}_\beta + \boldsymbol{u}_{\beta i}$ to the level-2 model (17)–(20), yielding the vector of individual deviations $\boldsymbol{u}_i = (\boldsymbol{u}'_{\alpha i}, \boldsymbol{u}'_{\beta i}, \boldsymbol{u}'_{\phi i}, \boldsymbol{u}'_{\psi i})'$ and the covariance matrix $\boldsymbol{T}$ expanded accordingly.

Both the computational set-up and the priors were identical to the previous section, and obtaining 12,000 post-burn-in samples took about an hour on a MacBook Pro. Model diagnostics indicated that the chains were mixing well, and is summarized in Supplementary Figure S2.

Imputation of missing values
Of the 5580 timepoint-individual combinations, there were 24 missing observations of desire to drink and 30 missing observations of perceived stress. We imputed these by imposing a weakly informative prior on the missing observations, a normal distribution with mean 4 and standard deviation 2 for desire to drink and mean and standard deviation 2 for perceived stress. We then sampled the missing values simultaneously with all other model parameters, conditional on the data. This ensures that uncertainty about the missing values is properly incorporated in the posterior distributions of all parameters and also that the posterior distribution of the missing values is marginalized with respect to all other parameters.

Posterior estimates
Figure 6 shows the common smooth terms and the individual smooth terms for alcohol consumption and desire to drink. In the left part, we see that alcohol consumption has a pattern similar to the one shown in Figure 4 (left), and that most individuals have relatively similar patterns. Figure 6 (right) indicates that desire to drink has a cyclic pattern similar to alcohol consumption, but in this case it is clear that a number of individuals deviate from the common trend. In particular, some curves are basically flat, indicating that these individuals have very small weekly variations in their desire to drink. Note however that with 60 timepoints we only have ≈ 8.5 weeks of observations per individual, and hence the individual curves are likely to be drawn strongly toward the common curve due to the pooling effect inherent in hierarchical models. With a higher number of timepoints, it might have been possible to detect larger individual deviations from the mean curve. Supplementary Table S4 shows the inverse smoothing parameters for the four cyclic terms, none of whose 95% credible intervals included zero, suggesting that both the common cyclic terms and the individual deviations from the cyclic terms are indeed nonlinear.

Supplementary Table S2 shows posterior means and credible intervals for the 18 population means $(\boldsymbol{\gamma}_\alpha, \boldsymbol{\gamma}_\beta, \boldsymbol{\gamma}_\phi, \boldsymbol{\gamma}_\psi)'$, and Supplementary Table S3 shows the same for the 18 standard deviations $(\boldsymbol{\sigma}_\alpha, \boldsymbol{\sigma}_\beta, \boldsymbol{\sigma}_\phi, \boldsymbol{\sigma}_\psi)'$ describing the individual variation around these means. From the tables, we note in particular that both desire to drink and perceived stress have relatively high autoregressive effects, with means $\gamma_{\phi,22} = 0.18\ (0.14, 0.22)$ and $\gamma_{\phi,33} = 0.20\ (0.15, 0.24)$, respectively, where the numbers in parentheses are 95% credible intervals. The autoregressive coefficient of alcohol consumption was down to $\gamma_{\phi,11} = 0.03\ (-0.01, 0.08)$, from $0.06\ (0.02, 0.10)$ in the AR(1) model. There is also evidence of individual variability in these parameters, with standard deviations $\sigma_{\phi,22} = 0.12\ (0.06, 0.17)$ and $\sigma_{\phi,33} = 0.17\ (0.14, 0.22)$, respectively. The only cross-lagged
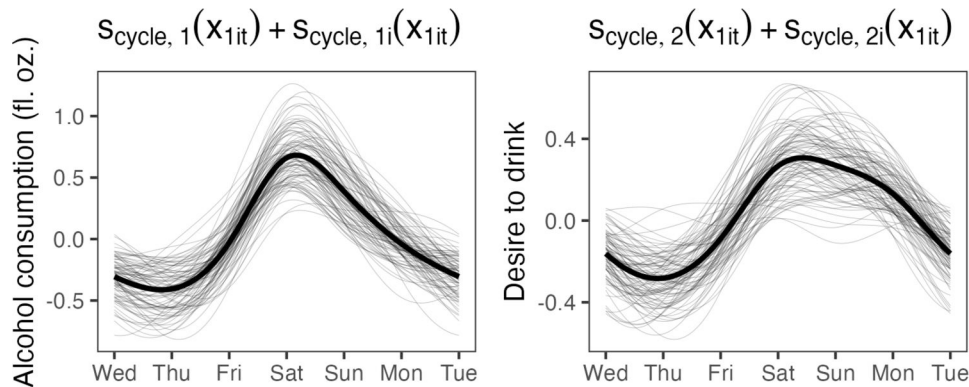

Figure 6. Posterior means of the common cyclic curves (solid black lines) and the individual cyclic curves (thin gray lines) for alcohol consumption (left) and desire to drink (right).

effect for which the 95% credible interval did not include zero was the effect of alcohol consumption on desire to drink the next day. This effect was negative, $\gamma_{\phi,12} = -0.20 \ (-0.25, -0.15)$, indicating that above average consumption on one day leads to a reduction in desire to drink the next day. This was also the only cross-lagged effect for which there was evidence for substantial individual variability, $\sigma_{\phi,12} = 0.12 \ (0.08, 0.17)$. The regression coefficients for long-term linear trends had small negative values for alcohol consumption and desire to drink, and a small positive value for perceived stress, with evidence for individual variability.

Discussion

We have proposed an extension of the DSEM framework using penalized regression splines which allows flexible modeling of nonlinear effects of time-dependent covariates both at the population level and the individual level. In contrast to alternative approaches using nonlinear models, regression splines require minimal assumptions about the functional forms, and are able to learn basically any smooth function from data. In multivariate models, splines allow borrowing strength across variables, as illustrated in "Multivariate hierarchical smooth terms". While we have considered AR(1) and VAR(1) models in particular, the extensibility of the Stan language combined with its highly efficient sampling algorithm allows our openly available code to be incorporated into the within-level model of basically any DSEM, as we outline in Appendix A.

Our simulation experiments in "Simulation experiment" demonstrated how ignoring nonlinear effects, or simply using too restrictive nonlinear models, may lead to biases in all parameters of interest, whereas allowing sufficient flexibility alleviated these issues. Conversely, when the truth is linear, the price to pay in terms of statistical power was very low in the settings considered here. This is likely due to the fact that the smoothing priors indeed put most of the prior probability mass on functions which are close to linear.

In "Analysis of daily diary data on alcohol consumption", we demonstrated how cyclic cubic splines can model weekly patterns in alcohol consumption and desire to drink, and be incorporated into a larger model also containing long-term trends. Compared to sine-cosine functions, which have perfectly symmetric shape within each period, cyclic splines only require that the endpoints match up to second derivatives. As can be seen in Figure 6 (left), this flexibility was particularly useful for modeling alcohol consumption, which was low during the weekdays and had a sharp peak on Fridays and Saturdays. Despite the larger number of parameters, obtaining samples of sufficient size from the posterior distributions required very moderate computational efforts.

One particularly interesting future development would be to let the level-2 parameters in Equations (17)–(19) depend smoothly on covariates of interest. For example, in cognitive testing or educational measurement the individual intercepts α_i and autoregressive effects Ψ_i might depend nonlinearly on age, or on exposures like participation in training programs. We will explore this in future work.

Article information

Data availability statement

Stan and R code for reproducing all simulation results are available in our OSF repository https://osf.io/qpkmg/. The alcohol consumption dataset considered in "Analysis of daily diary data on alcohol consumption" cannot be shared online; instead, we provide a simulated dataset which can be used to run the code and produce similar results.

References

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

Betancourt, M. (2018). *A conceptual introduction to Hamiltonian Monte Carlo.* arXiv:1701.02434 [stat]. https://doi.org/10.48550/arXiv.1701.02434

Bjørnarå, H. B., Berntsen, S., Te Velde, S. J., Fyhri, A., Isaksen, K., Deforche, B., Andersen, L. B., Stenling, A., & Bere, E. (2023). The impact of weather conditions on everyday cycling with different bike types in parents of young children participating in the CARTOBIKE randomized controlled trial. *International Journal of Sustainable Transportation*, 17(2), 128–135. https://doi.org/10.1080/15568318.2021.1999538

Brezger, A., & Lang, S. (2006). Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics & Data Analysis*, 50(4), 967–991. https://doi.org/10.1016/j.csda.2004.10.011

Brumback, B. A., & Rice, J. A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association*, 93(443), 961–976. https://doi.org/10.1080/01621459.1998.10473755

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software*, 80(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Carney, M. A., Armeli, S., Tennen, H., Affleck, G., & O'Neil, T. P. (2000). Positive and negative daily events, perceived stress, and alcohol use: A diary study. *Journal of Consulting and Clinical Psychology*, 68(5), 788–798. https://doi.org/10.1037/0022-006X.68.5.788

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Cho, Y. I., Johnson, T. P., & Fendrich, M. (2001). Monthly variations in self-reports of alcohol consumption. *Journal of Studies on Alcohol*, 62(2), 268–272. https://doi.org/10.15288/jsa.2001.62.268

Eilers, P. H. C., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2), 89–121. https://doi.org/10.1214/ss/1038425655

Feinn, R., Armeli, S., & Tennen, H. (2023). Individual differences in affect dynamics and alcohol-related outcomes. *Substance Use & Misuse*, 58(8), 967–974. https://doi.org/10.1080/10826084.2023.2201829

Gentle, J. E. (2024). *Matrix algebra: Theory, computations and applications in statistics.* Springer International.

Gronau, Q. F., Singmann, H., & Wagenmakers, E.-J. (2020). bridgesampling: An R package for estimating normalizing constants. *Journal of Statistical Software*, 92(10), 1–29. https://doi.org/10.18637/jss.v092.i10

Gu, C., & Wahba, G. (1991). Minimizing GCV/GML scores with multiple smoothing parameters via the Newton method. *SIAM Journal on Scientific and Statistical Computing*, 12(2), 383–398. https://doi.org/10.1137/0912021

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

Hippke, M., David, T. J., Mulders, G. D., & Heller, R. (2019). Wōtan: Comprehensive time-series detrending in python. *The Astronomical Journal*, 158(4):143.

Hoffman, M. D., & Gelman, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(47), 1593–1623.

Judd, N., Aristodemou, M., Klingberg, T., & Kievit, R. (2024). Interindividual differences in cognitive variability are ubiquitous and distinct from mean performance in a battery of eleven tasks. *Journal of Cognition*, 7(1), 45. https://doi.org/10.5334/joc.371

Kauermann, G., Krivobokova, T., & Semmler, W. (2011). Filtering time series with penalized splines. *Studies in Nonlinear Dynamics & Econometrics*, 15(2), 1–28. https://doi.org/10.2202/1558-3708.1789

Klesse, S. (2021). Critical note on the application of the "two-third" spline. *Dendrochronologia*, 65, 125786. https://doi.org/10.1016/j.dendro.2020.125786

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100(9), 1989–2001. https://doi.org/10.1016/j.jmva.2009.04.008

Li, Y., Wood, J., Ji, L., Chow, S.-M., & Oravecz, Z. (2022). Fitting multilevel vector autoregressive models in Stan, JAGS, and Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(3), 452–475. https://doi.org/10.1080/10705511.2021.1911657

Liu, Y., & West, S. G. (2016). Weekly cycles in daily report data: An overlooked issue. *Journal of Personality*, 84(5), 560–579. https://onlinelibrary.wiley.com/doi/pdf/10.1111/jopy.12182 https://doi.org/10.1111/jopy.12182

McNeish, D., & Hamaker, E. L. (2020). A primer on two-level dynamic structural equation models for intensive longitudinal data in Mplus. *Psychological Methods*, 25(5), 610–635. https://doi.org/10.1037/met0000250

McNeish, D., Mackinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2021). Measurement in intensive longitudinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(5), 807–822. https://doi.org/10.1080/10705511.2021.1915788

Meng, X.-L., & Schilling, S. (2002). Warp bridge sampling. *Journal of Computational and Graphical Statistics*, 11(3), 552–586. https://doi.org/10.1198/106186002457

Moore, E. H. (1920). On the reciprocal of the general algebraic matrix. *Bulletin of the American Mathematical Society*, 26(9), 394–395.

Muthén, B., Asparouhov, T., & Keijsers, L. (2024). Dynamic structural equation modeling with cycles. *Structural Equation Modeling: A Multidisciplinary Journal*, 0(0):1–23. https://doi.org/10.1080/10705511.2024.2406510

Oh, H., & Jahng, S. (2023). Incorporating measurement error in the dynamic structural equation modeling using a single indicator or multiple indicators. *Structural Equation Modeling: A Multidisciplinary Journal*, 30(3), 501–514. https://doi.org/10.1080/10705511.2022.2103703

Pedersen, E. J., Miller, D. L., Simpson, G. L., & Ross, N. (2019). Hierarchical generalized additive models in ecology: An introduction with mgcv. *PeerJ*, 7, E6876.

Penrose, R. (1955). A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51(3), 406–413. https://doi.org/10.1017/S0305004100030401

R Core Team (2024). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4, 1–32.

Tanabe, J., Miller, D., Tregellas, J., Freedman, R., & Meyer, F. G. (2002). Comparison of detrending methods for optimal fMRI preprocessing. *NeuroImage*, 15(4), 902–907. https://doi.org/10.1006/nimg.2002.1053

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved R ^ for assessing convergence of MCMC (with discussion). *Bayesian Analysis*, 16(2), 667–718. https://doi.org/10.1214/20-BA1221

Wang, L. P., & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods*, 20(1), 63–83. https://doi.org/10.1037/met0000030

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 65(1), 95–114. https://doi.org/10.1111/1467-9868.00374

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(1), 3–36. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2010.00749.x https://doi.org/10.1111/j.1467-9868.2010.00749.x

Wood, S. N. (2017a). *Generalized additive models: An introduction with R* (2nd ed.). Chapman and Hall/CRC.

Wood, S. N. (2017b). P-splines with derivative based penalties and tensor product smoothing of unevenly distributed data. *Statistics and Computing*, 27(4), 985–989. https://doi.org/10.1007/s11222-016-9666-x

Wu, Z., Huang, N. E., Long, S. R., & Peng, C.-K. (2007). On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proceedings of the National Academy of Sciences*, 104(38):14889–14894.

Appendix

Within-level smooths in DSEM

We here show how smooth terms can be incorporated into the within-level model in the general cross-classified DSEM framework of Asparouhov et al. (2018), and then derive the AR(1) model of "Two-level AR(1) models" section and the VAR(1) model of "Two-level VAR(1) models" section as special cases. The notation in this appendix deviates somewhat from that used in the rest of the paper.

A brief review of cross-classified DSEM

We start by decomposing the observed responses $y_{it} \in \mathbb{R}^P$ into an individual-specific part $y_{2,i}$, a timepoint-specific part $y_{3,t}$, and the deviation of individual i at timepoint t, $y_{1,it}$, yielding

$$y_{it} = y_{1,it} + y_{2,i} + y_{3,t}. \quad (A1)$$

We next specify a measurement model for each term on the right-hand side of (A1) and a corresponding structural model for latent variables $\eta_{3,t}$, $\eta_{2,i}$, and $\eta_{1,it}$ varying between timepoints, between individuals, and between individuals and timepoints, respectively.

The between-timepoints model is

$$y_{3,t} = v_3 + \Lambda_3 \eta_{3,t} + K_3 X_{3,t} + \epsilon_{3,t} \quad (A2)$$

$$\eta_{3,t} = \alpha_3 + B_3 \eta_{3,t} + \Gamma_3 X_{3,t} + \xi_{3,t}, \quad (A3)$$

where v_3 and α_3 are intercepts, Λ_3 is a loading matrix, K_3 and Γ_3 are matrices of regression coefficients for covariates $X_{3,t}$ which vary between timepoints but not between individuals, B_3 are regression coefficients between latent variables, and $\epsilon_{3,t}$ and $\xi_{3,t}$ are noise terms. The between-individuals model is similarly

$$y_{2,i} = v_2 + \Lambda_2 \eta_{2,i} + K_2 X_{2,i} + \epsilon_{2,i} \quad (A4)$$

$$\eta_{2,i} = \alpha_2 + B_2 \eta_{2,i} + \Gamma_2 X_{2,i} + \xi_{2,i}, \quad (A5)$$

where $X_{2,i}$ is a matrix of covariates varying between individuals but not between timepoints. All other parameters in (A4)–(A5) have similar interpretations as those in (A2)–(A3).

In the residual DSEM formulation, the within-level model has a structural part

$$y_{1,it} = v_1 + \Lambda_{1,0it} \eta_{1,i,t} + R_{0it} y_{1,i,t} + K_{1it} X_{1,it} + \hat{y}_{1,it} \quad (A6)$$

$$\eta_{1,it} = \alpha_1 + B_{1,0it} \eta_{1,i,t} + Q_{0it} y_{1,i,t} + \Gamma_{1it} X_{1,it} + \hat{\eta}_{1,it}, \quad (A7)$$

where R_{0it} and Q_{0it} are regression coefficients between the components of $y_{1,it}$ and $\eta_{1,it}$, respectively, and $X_{1,it}$ is a matrix of regression coefficients allowed to vary both between timepoints and between individuals. The other parameters have the same interpretation as in the between-level model, except that the residuals are now written $\hat{y}_{1,it}$ and $\hat{\eta}_{1,it}$. The it subscripts indicate that the within-level parameters can vary both between individuals and between timepoints, cf. Equation (8) in Asparouhov et al. (2018). The autoregressive parts relate the residuals through the equations

$$\hat{y}_{1,it} = \sum_{l=1}^{L} \Lambda_{1,lit} \hat{\eta}_{1,i,t-l} + \sum_{l=1}^{L} R_{lit} \hat{y}_{1,i,t-l} + \epsilon_{1,it} \quad (A8)$$

$$\hat{\eta}_{1,it} = \sum_{l=1}^{L} B_{1,lit} \hat{\eta}_{1,i,t-l} + \sum_{l=1}^{L} Q_{lit} \hat{y}_{1,i,t-l} + \xi_{1,it}, \quad (A9)$$

where the integer $L \geq 1$ represents the extent of the lag and $\Lambda_{1,lit}$, R_{lit}, $B_{1,lit}$, and Q_{lit} are matrices of autoregressive effects and cross-correlations relating the residuals at time $t - l$ to the residuals at time t.

Within-level smooth terms in DSEM

Within-level smooth terms occur most naturally in the structural part (A6)–(A7) of the residual DSEM formulation. We can simply replace the linear regression terms $K_{1it}X_{1,it}$ and $\Gamma_{1it}X_{1,it}$ with vectors $f_{yi}(x_{it})$ and $f_{\eta i}(x_{it})$ of smooth terms predicting the measurements or the latent variables, respectively. This gives the modified structural part

$$y_{1,it} = \nu_1 + \Lambda_{1,0it}\eta_{1,i,t} + R_{0it}y_{1,i,t} + f_{yi}(x_{it}) + \hat{y}_{1,it} \quad (A10)$$

$$\eta_{1,it} = \alpha_1 + B_{1,0it}\eta_{1,i,t} + Q_{0it}y_{1,i,t} + f_{\eta i}(x_{it}) + \hat{\eta}_{1,it}. \quad (A11)$$

Estimating the resulting DSEM would require smoothing priors as described in "Two-level AR(1) models" and "Two-level VAR(1) models".

Recovering the AR(1) and VAR(1) models

The models considered in "Two-level AR(1) models" and "Two-level VAR(1) models" are two-level DSEMs with no between-timepoint model (A2)–(A3), and hence $y_{3,t} = 0$ and $\eta_{3,t} = 0$. We also assume the latent variables are directly observed, i.e., $\eta_{2,i} = y_{2,i}$, and can hence ignore the measurement model (A4) and plug $y_{2,i}$ directly into (A5) to get $y_{2,i} = \alpha_2 + B_{2,i}y_{2,i} + \Gamma_2 X_{2,i} + \xi_{2,i}$. We further assume no regressions between variables at a given timepoint, so we set $B_{2,i} = 0$ and we have no level-2 covariates so also $\Gamma_2 X_{2,i}$ can be dropped, yielding the between-individual model $y_{2,i} =$

$\alpha_2 + \xi_{2,i}$, which in the notation of "Two-level VAR(1) models" section becomes $y_{2,i} = \gamma_\alpha + u_{\alpha i}$.

Since we also assume $\eta_{1,it} = y_{1,it}$ at the within-level, we can ignore the measurement models (A10) and (A8), set $\alpha_1 = 0$ for identifiability, and plug $y_{1,it}$ directly into (A11) and (A9) to get the structural and autoregressive parts as

$$y_{1,it} = B_{1,0it}y_{1,i,t} + f_i(x_{it}) + \hat{y}_{1,it} \quad (A12)$$

$$\hat{y}_{1,it} = \sum_{l=1}^{L} B_{1,lit}\hat{y}_{1,i,t-l} + \xi_{1,it}, \quad (A13)$$

where the matrix Q_{lit} has been dropped since it in this case becomes identical to $B_{1,lit}$ and we have omitted the η subscript on the smooth term. We have no regressions between variables at a given timepoint, and can hence set $B_{1,0it} = 0$ and our lag is $L = 1$, so (A12)–(A13) simplify further to

$$y_{1,it} = f_i(x_{it}) + \hat{y}_{1,it} \quad (A14)$$

$$\hat{y}_{1,it} = B_{1,1it}\hat{y}_{1,i,t-1} + \xi_{1,it}. \quad (A15)$$

Plugging (A15) into (A14) and using $y_{1,it} = y_{it} - \alpha_i$ and $\hat{y}_{1,i,t-1} = y_{1,i,t-1} - \Gamma_{1,0it}X_{1,i,t-1}$ we get

$$y_{it} - \alpha_i = f_i(x_{it}) + B_{1,1it}\{y_{it} - \alpha_i - f_i(x_{i,t-1})\} + \xi_{1,it} \quad (A16)$$

which can be easily reorganized to yield (16), recognizing $B_{1,1,it}$ as Φ_i and $\xi_{1,it}$ as δ_i. The AR(1) model (2) is recovered as the special case where y_{it} is univariate.