STRUCTURAL EQUATION MODELING
2025, VOL. 32, NO. 2
ISSN: 1070-5511 | E-ISSN: 1532-8007

A Growth of Hierarchical Autoregression Model for Capturing Individual Differences in Changes of Dynamic Characteristics of Psychological Processes

Yanling Li[a], Lindy Williams[a], Chelsea Muth[b], Saeideh Heshmati[c], Sy-Miin Chow[a], and Zita Oravecz[a]

[a]The Pennsylvania State University; [b]Westhampton Family Psychologists; [c]Claremont Graduate University

ABSTRACT
Several methodological innovations have been advanced in the past decades that combine growth curve models (GCMs) with models of autoregressive (AR) processes. However, most of these approaches do not effectively capitalize on known (e.g., study design-related) information to structure the growth curves into meaningful between- and within-phase changes, while simultaneously accommodating interindividual differences in these intraindividual changes. We propose a Bayesian growth of hierarchical autoregression (GoHiAR) model, which combines AR and GCM to evaluate phase-to-phase changes in multifaceted dynamic characteristics (e.g., baseline, variability, and inertia) as well as individual differences in these changes. This approach allows for drawing conclusions in a way that the proposed data generating mechanisms are in line with the theoretical insights about psychological change and dynamics. Our Bayesian implementation of the GoHiAR model allows for all parameters to be estimated simultaneously. First, we evaluated GoHiAR's overall estimation accuracy and sampling efficiency, effects of model misspecifications, and sensitivity to effect sizes via a simulation study. Results showed reasonable performance. Then, we applied GoHiAR to an ecological momentary assessment (EMA) study that comprised data from pre-, during, and following an intervention, and investigated changes in the dynamic characteristics of individuals' psychological well-being (specifically in meaning of life) within and across phases.

KEYWORDS
Growth curve; hierarchical autoregression; mHealth intervention; psychological well-being; time-varying parameter

1. Introduction

In intensive longitudinal designs, individuals' characteristics are measured intensively across a time period and sometimes even several time periods (also called waves, phases or bursts). These designs allow researchers to study behavioral change at multiple timescales, including both fine-grained trajectory patterns (i.e., change on a fast timescale) and structural shifts in trajectory patterns (i.e., change on a slow timescale). Examples of these multi-timescale processes abound in the developmental, prevention/intervention, and health literature (see, e.g., Ebner-Priemer & Trull, 2009; Ram et al., 2013).

The current work contributes to these methods by combining autoregressive (AR) and growth curve models (GCMs) to capture three dynamic characteristics that have been shown in previous studies to link to meaningful interindividual differences, namely: (1) baseline, (2) intraindividual variability, and (3) inertia. The (1) baseline simply captures a person's average levels of experience. (2) Intraindividual variation (IIV) represents how dramatically or imperceptibly a person's level of experience fluctuates over time in terms of reactivity to events, with higher levels of IIV potentially indicating instability. (3) Inertia (or autoregression; AR) expresses the current measurement occasion as a function of one or more previous occasions, and quantifies the tendency for a process to linger in extreme values before returning to baseline. Higher inertia means the process will return to baseline level more slowly, therefore showing less effective homeostatic regulation. It is assumed that these dynamic features may fundamentally change when the entire system is altered, for example by an intervention. Together, these three dynamic characteristics provide a framework for modeling fundamental changes in the process system over time and individual differences therein.

Previous work in the GCM literature has considered multiple latent phases for modeling fundamental changes (Helm et al., 2016), and combined the GCM with AR processes. The latter includes, among other variations, the Autoregressive Latent Trajectory (ALT) model (Curran & Bollen, 2001; Bollen & Curran, 2004), which incorporates AR among the manifest observations directly, and other GCMs with AR in the residuals (Browne & Du Toit, 1991; Hamaker, 2005). Ou et al. (2017) showed that interpretations of the growth curve components in the ALT model can vary distinctly depending on the initial condition specification of the ALT process before the first observation. Other variations that include AR features in the residuals as opposed to the manifest observations may be characterized by greater interpretational ease, but these components may still suffer from empirical identifiability issues in the estimation process due to their overlapping roles in governing the change processes. Thus, appropriate inclusion of constraints through prior distributions or study design information to help isolate these dynamic characteristics from each other is critical to ensure their empirical identifiability and meaningful extraction of interindividual differences in these dynamic characteristics. Our proposed combined model is cast in a hierarchical Bayesian framework that allows for constraining the dynamic characteristics through prior specifications, this way facilitating the recovery of dynamical and growth cure components.

To capture the effects of interventions in terms of these three dynamic characteristics, we allow them to vary over time - on a slow timescale (e.g., week to week, year to year etc.). Extant time-varying parameter (TVP) models and applications have been restricted to model such slow timescale changes for only a subset of the three dynamic features described above (e.g., studies focusing on one feature: Chow et al., 2010; Del Negro & Otrok, 2008; Deschamps, 2003; studies focusing on two features: Bringmann et al., 2017). In fact, many intervention studies have focused almost exclusively on intervention-related changes in the baseline level of the outcome variable (e.g., Drozd et al., 2013; Roesch et al., 2010). Additionally, unlike commonly used models for economic or business time series data analysis (e.g., generalized autoregressive conditional heteroskedasticity (GARCH) models; Bollerslev, 1987) which assume stochastic volatility (i.e., time-varying process noise variances) over time, the majority of TVP models used in social and behavioral research tended to ignore such change in the IIV (or process noise variances). Failing to account for such stochastic volatility may lead to biased estimates of structural parameters such as intercepts and AR parameters in AR-type models. Our proposed model can allow for slow timescale changes in all three dynamical parameters simultaneously, this way providing a more comprehensive model that may be developed to enable an overarching modeling framework for capturing multi-timescale processes.

Another important consideration, especially when modeling changes due to interventions, is individual differences in the change process (i.e., clinical or treatment heterogeneity). Therefore, extending these models to be able to incorporate random effects is necessary. With some exceptions where between-person differences in the change of baseline and regulation levels were taken into consideration in the analysis (see, e.g., Bollen & Curran, 2004; Chen et al., 2021), few TVP models can simultaneously account for within-person differences between multiple bursts (i.e., intra-individual change on a slow timescale) and between-person differences in such slow timescale change.

We propose a growth of hierarchical autoregression (GoHiAR) model which combines AR and GCM to simultaneously evaluate intervention-related changes in three dynamic characteristics (i.e., baseline, IIV, and inertia) and individual differences therein. Specifically, the AR model is utilized to articulate dynamic characteristics on a *fast* timescale, where the intercepts, AR parameter, and process noise variances represent baselines, inertia, and IIV, respectively, and are estimated for each individual. Additionally, a growth curve structure is fitted to the person-specific parameters of the AR model (i.e., the intercept, AR parameter, and process noise variance) to capture changes in these parameters on a *slow* (long-term or phase-based) timescale. Given its multilevel formulation, the proposed approach can capture between-group (e.g., intervention vs. control) and/or between-individual differences in intraindividual changes in dynamic characteristics. Our approach adds to conventional dynamical systems models which were developed to capture either within-person variations in TVPs or between-person differences in static parameters, as the parameters in the GoHiAR model are made both person-specific and time-varying to capture both within- and between-person variations. In addition, compared with previous work that combined AR and GCM (e.g., ALT models and GCMs with AR in the residuals), the GoHiAR model allows for not only phase-varying intercept and AR parameters but also phase-varying process noise variances.

We fit the GoHIAR model in the Bayesian statistical framework. From a computational standpoint, this is a necessity, because of the hierarchical/multilevel nature of the model with a large number of person-specific model parameters (with some of them time-varying). Fitting such multilevel model leads to high-dimensional integration demands, which may exceed the capacity of classical estimation methods (such as maximum likelihood estimation). One possible solution could be a two-step approach, where we fit multiple single-person models individually and conduct follow-up regression analysis with the estimated parameters. However, this approach introduces problems such as accumulating estimation errors across multiple individual models (Pagan, 1984), sacrificing the accuracy of the mean estimate of the average dynamic, while a hierarchical estimate gains power from estimating multiple series at once (Song & Ferrer, 2012). The Markov chain Monte Carlo (MCMC) methods underlying the Bayesian estimation perform numerical integration to address the computational challenges and provide a one-step approach to estimate parameters in a highly complex hierarchical model, without accumulation of estimation errors.

The remainder of the article is organized as follows. We will start with an empirical data example in which intervention effects in terms of dynamic characteristics are in focus, followed by a detailed introduction about model and prior specifications, as well as the estimation procedures. Then we evaluate its performance via a simulation study and illustrate its application using our empirical data. Finally, we discuss the findings, limitations and future directions.

2. Motivating Example

The development of the GoHiAR model was motivated by an ecological momentary intervention (EMI; Heron & Smyth, 2010) study on the psychological well-being (PWB) of college-attending early adults. The study was designed as a 56-day long Randomized Control Trial (RCT) where 160 participants were randomly assigned to three study groups. All three groups were completing momentary PWB self-reports up to 6 times a day. The active control group ($N = 55$) had no intervention (but did some neutral activities timed similarly to the interventions), the first intervention groups had positive practice intervention ($N = 51$ after excluding four people who dropped out of the study) and the second intervention group ($N = 54$ after excluding one person who dropped out of the study) had positive practice and meditation intervention. To simplify this first introduction of the GoHiAR model, we focused only on comparing the active control group (labeled as *Control* from here on) and the group that received positive practice and meditation intervention (labeled as *Treatment*). All measures and details on the interventions can be found on the project's OSF page[1].

The study had four main phases: (P1) pre-intervention phase (14 days); (P2) intervention phase (15 days), during which the Treatment group received intervention, and the Control performed a working-memory task every day; (P3) post-intervention phase I (13 days); and (P4) post-intervention phase II (14 days). Phase membership was not estimated, but rather, was known a priori based on the study design. Theoretically, P3 was supposed to capture the effects right after the intervention, while P4 focused more on longer term benefits. The momentary assessments of PWB were based on Seligman's PERMA model (Seligman, 2012) which consists of five dimensions of PWB—positive emotions, engagement, relationship, meaning, and accomplishment.

The study aimed at investigating the PWB change mechanisms from a process-oriented perspective (Heshmati et al., 2024). It was hypothesized that effectively changing existing dynamic PWB characteristics was centrally important to the success of a PWB intervention. Previous research in adult populations indicated that improved PWB was related to directional changes in dynamic characteristics, such as increased baseline levels, decreased IIV, and increased regulation (lower inertia) of one's PWB (Kuppens et al., 2010; Röcke & Brose, 2013; Segrin & Taylor, 2007). For illustrating our proposed methods approach for testing these hypotheses, we focused on one dimension of PWB, Meaning of life, which is defined as having a sense of purpose in life, feeling that life is valuable, or connecting to something greater than ourselves (e.g., a religious faith, a spiritual calling, etc.) (Butler & Kern, 2016).

Figure 1 shows the dynamics of the meaning of life variable for four randomly selected participants, with the two on the left panel from the Treatment group and the two on the right panel from the Control group. For the Treatment group, the shaded area indicates the intervention phase. Both groups experienced four phases as separated by the three dash lines that represent cut-offs between phases. We can see substantial individual differences in terms of how meaning of life changed over the course of the study. For instance, participant 1 experienced an increase in baseline levels during the intervention and maintained this increased level following the intervention, suggesting a positive intervention effect. By contrast, participant 3 displayed a notable decrease in IIV during the intervention, indicating less fluctuations during the intervention. By looking at the example participants from the Control group, we can see that although participant 2 did not receive any intervention, they also displayed an increased baseline, indicating that the momentary monitoring itself might have had positive effects on the meaning of life for this particular person. Finally, participant 4 did not display substantial changes in baseline levels but a slight decrease in IIV following the intervention.

The overarching goal of the well-being study was to examine changes related to the EMA design itself and based on group memberships (i.e., control and treatment groups, for more information see Hypothesis on the OSF page). Our analysis here focuses on two specific questions: (1) how the dynamic characteristics of meaning of life changed over the course of the study, and (2) whether and how these dynamic characteristics changed differently between the Control and the selected Treatment groups. We hypothesized that the intervention effects could manifest via credibly increased baseline levels, decreased IIV, and increased regulation more so for the Treatment than the Control group.

3. Specifications of the GoHiAR Model

The GoHiAR model consists of an AR model with person- and phase-specific parameters on its first level (level-1) and an overarching GCM on its second level (level-2) to capture phase-to-phase (slow timescale) changes in the parameters of the AR model. Specifically, the AR component is specified as:

$$Y_{i,p,t} = \mu_{i,p} + \phi_{i,p}(Y_{i,p,t-1} - \mu_{i,p}) + \epsilon_{i,p,t}, \ \epsilon_{i,p,t} \sim N(0, IIV_{i,p}), \tag{1}$$

where $Y_{i,p,t}$ represents the value of the variable of interest (in our case, meaning of life) for person i ($i = 1, \ldots, N$), during phase p ($p = 1, \ldots, 4$), at time t ($t = 1, \ldots, T_i$). The person- and phase-specific intercept is denoted by $\mu_{i,p}$, which represents the baseline level around which the process of interest fluctuates. The person- and phase-specific AR parameter is denoted by $\phi_{i,p}$, which we limit to be into the range of (–1,1) to make sure the time series is stationary, namely, conforming to the same distributional properties, including mean, variance and covariance functions, over time. This is a conventional constraint that has been applied to AR processes to ensure that they do not exhibit unrealistic (e.g., exploding) dynamics (Harvey, 2001). Adding informative stationarity-related constraints such as this is instrumental to teasing apart the AR (stationary) from the growth curve (by definition non-stationary due to its time-varying means) components encapsulated in the over-phase changes in $\mu_{i,p}$. In affective literature, the AR parameter is also referred to as *inertia* given that a positive $\phi_{i,p}$ reflects a construct's relative resistance to change. In our case, a higher positive $\phi_{i,p}$ that is closer to 1 indicates that participants' meaning of life is less likely to change and it takes longer to recover after a change caused by external events. This in turn means that the process exhibits a lower level of self-regulation, that is more resistance to change. Given the nature of the PWB intervention, we would expect an increase in the regulation of PWB dynamics

[1] https://osf.io/7hvce/?view_only=3fd9b04b35244afdbbeab8567db3b1eb

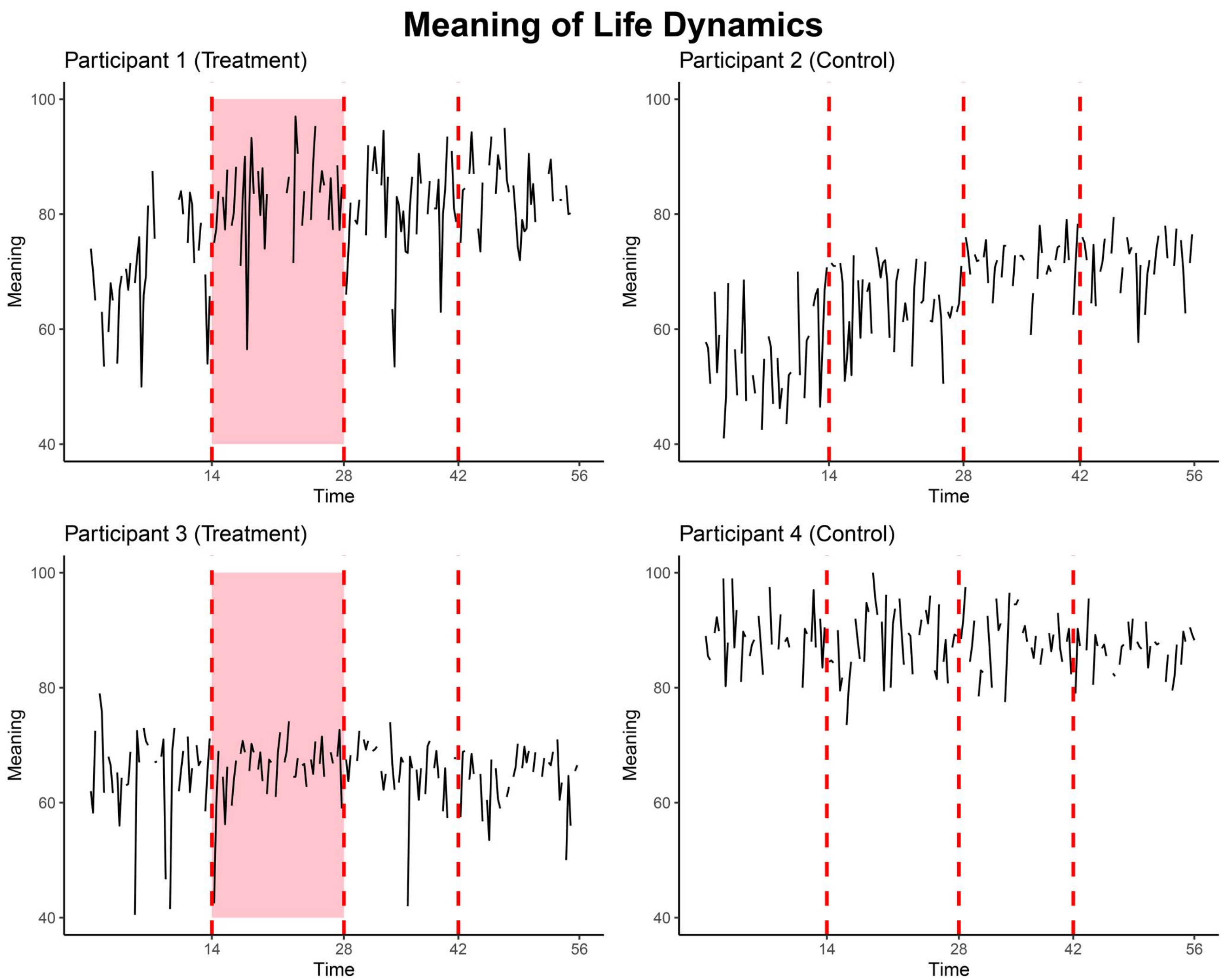

Figure 1. Meaning dynamics of four randomly selected participants. Participant 1&3 were from the treatment group; participant 2&4 were from the control group. The shaded area represents the intervention phase. Both groups experienced four phases as described in the main text.

and thus a decrease in inertia (lower values for the AR parameter). Finally, the process noise, $\epsilon_{i,p,t}$, follows a normal distribution with zero mean and a person- and phase-specific variance, $IIV_{i,p}$. This variance parameter captures the within-person fluctuations in meaning of life. In sum, all parameters quantifying dynamic characteristics (i.e., baseline, inertia/regulation, IIV) are both person-specific and time-varying (allowed to change over phases), allowing us to model both individual differences and changes of these characteristics over time.

For all three dynamic characteristics we impose an individual-specific growth curve model structure to quantify the nature of growth (slow timescale change). Based on the visual inspection of the time series plots of our empirical data, we assumed a linear trend in these TVPs, but the GoHiAR model could be easily extended for example to capture quadratic or other parametric shapes of change in future studies. The level-2 GCMs across the phases are specified as:

$$\mu_{i,p} = \beta_{\mu 0,i} + \beta_{\mu 1,i} Phase_{i,p} + e_{\mu i,p}, \quad e_{\mu i,p} \sim N(0, \sigma^2_\mu) \tag{2}$$

$$\phi_{i,p} = \beta_{\phi 0,i} + \beta_{\phi 1,i} Phase_{i,p} + e_{\phi i,p}, \quad e_{\phi i,p} \sim N(0, \sigma^2_\phi) \tag{3}$$

$$\log(IIV_{i,p}) = \beta_{IIV0,i} + \beta_{IIV1,i} Phase_{i,p} + e_{IIVi,p}, \quad e_{IIVi,p} \sim N(0, \sigma^2_{IIV}). \tag{4}$$

The "Phase" variable can take values of 0, 1, 2, or 3, representing the pre-, during, and two post-intervention follow-up phases described earlier. The person-specific intercepts (i.e., $\beta_{\mu 0,i}$, $\beta_{\phi 0,i}$, $\beta_{IIV0,i}$) represent person i's levels of the corresponding dynamic characteristics (i.e., $\mu_{i,p}$, $\phi_{i,p}$, $\log(IIV_{i,p})$) in the first phase (i.e., pre-intervention), and the person-specific slopes (i.e., $\beta_{\mu 1,i}$, $\beta_{\phi 1,i}$, $\beta_{IIV1,i}$) capture the corresponding rates of change across the phases of the study. The error term in each GCM model follows a normal distribution. Note that we log-transformed the IIV variable (which as a variance parameter was limited to take non-negative values) so that it takes values on the real line, which is more convenient for regression modeling.

Next, sources of individual differences in these person-specific intercepts and slopes are modeled through person-level predictors. These level-3 models of each level-2 GCM can be defined as:

$$\beta = z\Gamma + \zeta, \quad \zeta \sim N(0, \Sigma_\beta) \tag{5}$$

where β is an $N \times 2$ matrix containing the person-specific intercepts and slopes for each GCM (i. e., $[\beta_{\mu 0,i}, \beta_{\mu 1,i}]$ or $[\beta_{\phi 0,i}, \beta_{\phi 1,i}]$ or $[\beta_{IIV0,i}, \beta_{IIV1,i}]$); variable $\boldsymbol{z}$ is an $N \times k$ matrix of exogenous person-level predictors hypothesized to explain some of the individual differences in β, with the first column being 1s to define intercept terms, and Γ is the corresponding $k \times 2$ matrix of intercepts and regression coefficients (or fixed effects). Finally, ζ is an $N \times 2$ matrix of random effects representing deviations in the values of β not accounted for by the exogenous variables; these random effects are normally distributed with zero means and a random effect covariance matrix, $\Sigma_{\boldsymbol{\beta}}$.

Recall that one of the research questions presented in the motivating example section is to investigate whether and how dynamic characteristics changed differentially in the Treatment vs. Control group. To answer this question, we need to examine group-level differences in slope parameters in the above GCMs (i.e., $\beta_{\mu 1,i}$, $\beta_{\phi 1,i}$, $\beta_{IIV1,i}$). However, we might also want to test whether there were group-level differences in intercept parameters ($\beta_{\mu 0,i}$, $\beta_{\phi 0,i}$, $\beta_{IIV0,i}$) to check whether the two randomly assigned groups had similar pre-intervention baseline, IIV, and regulation levels. The predictor $\boldsymbol{z}$ in Equation (5) can be set as a person-specific group membership indicator, which can be constructed as a dummy variable with 1 representing the Treatment group. That said, using intercept and slope parameters related to baseline levels as an example, Equation (5) can be written out as:

$$\beta_{\mu 0,i} = \gamma_{\mu 0,Control} + \gamma_{\mu 0,Treatment} Group_i + \zeta_{\beta_{\mu 0,i}} \quad (6)$$

$$\beta_{\mu 1,i} = \gamma_{\mu 1,Control} + \gamma_{\mu 1,Treatment} Group_i + \zeta_{\beta_{\mu 1,i}} \quad (7)$$

With the above model specification, the group-level differences in intercept and slope parameters can be captured by the regression coefficients. For instance, the expectations of the growth rates of baseline levels over time for the Control and the Treatment group are $\gamma_{\mu 1,Control}$ and $\gamma_{\mu 1,Control} + \gamma_{\mu 1,Treatment}$, respectively, thus $\gamma_{\mu 1,Treatment}$ captures the group-level difference in terms of the growth rate (slow timescale change) of the baseline characteristic. A positive $\gamma_{\mu 1,Treatment}$ means that the Treatment group have a higher growth rate of baseline levels than the Control group. The unexplained errors in the variation of the intercepts and slopes are bivariate normally distributed per each of the three dynamic characteristics, specifically as:

$$\begin{bmatrix} \zeta_{\beta_{\mu 0,i}} \\ \zeta_{\beta_{\mu 1,i}} \end{bmatrix} \sim N(0, \Sigma_{\boldsymbol{\beta}_\mu} = \begin{bmatrix} \sigma^2_{\beta_{\mu 0}} & \sigma_{\beta_{\mu 0}\beta_{\mu 1}} \\ \sigma_{\beta_{\mu 0}\beta_{\mu 1}} & \sigma^2_{\beta_{\mu 1}} \end{bmatrix}). \quad (8)$$

The models for the other two dynamic features, IIV and inertia, can be specified in a similar way. A comprehensive summary of key model parameters and their definitions were summarized in Table 1.

3.1. Specifications of the Prior Distributions

Given the model specification, the unknown parameters that need to be estimated include: (1) residual standard deviations in GCM models (i.e., σ_μ, σ_ϕ, σ_{IIV} in Equations (2)–(4); (2) fixed effects parameters (e.g., regression coefficients in Equations (6) and (7)); and (3) random effect standard deviations and covariances (e.g., elements in $\Sigma_{\boldsymbol{\beta}_\mu}$ in Equations (8)).

Overall, we assigned diffuse or non-informative prior distributions for all parameters except for parameters related to the AR parameter, for which weakly informative priors were assigned to make sure the AR parameter did not exceed the stationary range. Specifically, we directly imposed a constraint to the latent AR parameters (i.e., $\phi_{i,p}$) to restrict their ranges to between -1 and 1. We then assigned standard normal distributions (i.e., $N(0,1)$) to the AR-parameter-related regression coefficients (see, e.g., $\gamma_{\phi 0,Control}$, $\gamma_{\phi 1,Control}$, $\gamma_{\phi 0,Treatment}$, $\gamma_{\phi 1,Treatment}$ in Table 1), and uniform distributions ($U(0,1)$) to their random effect standard deviations (see $\sigma_{\beta_{\phi 0}}$ and $\sigma_{\beta_{\phi 1}}$ in Table 1) to ensure that the time series would not exhibit unrealistic (e.g., exploding) dynamics. Aside from this specific group of priors, for all remaining parameters we used relatively non-informative priors that distributed over a wide range of possible values. For instance, for regression coefficients, we used $N(0,100)$; for standard deviation parameters, we used $U(0,100)$; for correlation parameters, we used $U(-1,1)$. All these prior distributions were relatively diffuse and thus would not introduce much information into the estimation process.

3.2. MCMC Estimation Procedures

The model was fitted using the default MCMC algorithms in the statistical software "Just Another Gibbs Sampler" (JAGS; Plummer, 2003). In general, JAGS adopts Gibbs sampling which iteratively draws samples from the conditional posterior distributions of model parameters. Some alternative sampling methods such as slice sampling (Neal, 2003) may be used to enhance sampling efficiency when dealing with more complex models. Based on the posterior samples, we can obtain point and standard error estimates, as well as credible intervals for a particular parameter by calculating the means or medians, standard deviations, and quantiles of the posterior samples for this parameter, respectively. Note

Table 1. A Summary of key model parameters.

Abbreviations	Symbols	Descriptions
IN/IIV/MU intercept (Control)	$\gamma_{\phi 0,Control}$ / $\gamma_{IIV0,Control}$ / $\gamma_{\mu 0,Control}$	Initial levels of Inertia/IIV/Baseline in the control group
IN/IIV/MU slope (Control)	$\gamma_{\phi 1,Control}$ / $\gamma_{IIV1,Control}$ / $\gamma_{\mu 1,Control}$	Growth rates of Inertia/IIV/Baseline in the control group
IN/IIV/MU intercept (Contrast)	$\gamma_{\phi 0,Treatment}$ / $\gamma_{IIV0,Treatment}$ / $\gamma_{\mu 0,Treatment}$	Between-group differences in the initial levels of Inertia/IIV/Baseline
IN/IIV/MU slope (Contrast)	$\gamma_{\phi 1,Treatment}$ / $\gamma_{IIV1,Treatment}$ / $\gamma_{\mu 1,Treatment}$	Between-group differences in the growth rates of Inertia/IIV/Baseline
Variance of IN/IIV/MU intercepts	$\sigma^2_{\beta_{\phi 0}}$ / $\sigma^2_{\beta_{IIV0}}$ / $\sigma^2_{\beta_{\mu 0}}$	Unexplained variance in the initial levels of Inertia/IIV/Baseline after accounting for group-level differences
Variance of IN/IIV/MU slopes	$\sigma^2_{\beta_{\phi 1}}$ / $\sigma^2_{\beta_{IIV1}}$ / $\sigma^2_{\beta_{\mu 1}}$	Unexplained variance in the growth rates of Inertia/IIV/Baseline after accounting for group-level differences

that the proposed model can also be implemented using alternative software such as Stan and Mplus. We opted for JAGS because its relative flexibility in writing out the likelihood functions and customizing priors, compared with Mplus, as well as less computational time, compared with Stan. Our previous study provided detailed comparisons about fitting complex multilevel VAR models using these software programs (Li et al., 2022).

The sampling quality and convergence problems were evaluated by two diagnostic statistics (Gelman et al., 2013): (1) the effective sample size (ESS), which describes how many posterior draws in the MCMC procedure can be regarded as independent, and (2) $\hat{R}$, which describes the ratio of the overall variance of posterior samples across chains to the within-chain variance, and is indicative of convergence issues.

4. Simulation Study

The goal of the simulation study was to answer the following questions.

1. *What is the estimation and sampling performance of the proposed model and estimation procedures?*
2. *How would the estimation results be affected if the model was misspecified?* One possible model misspecification may be that the phase-to-phase changes in TVPs were specified in the true data generation model but not captured by the fitted model. In the context of intervention, it means that, for example, there are actually group-level increases in baseline levels but the fitted model fails to capture such changes. Another scenario may be that a general model with TVPs is fitted when there are actually no time-varying changes in parameters in the true data generation model. In such scenario, it is important to evaluate the type I error rates of relevant parameters to see if there is any false detection of TVPs.
3. *To what extent are the proposed approaches sensitive to intervention effect sizes?* Recall that intervention effects are captured in the proposed model by parameters representing between-group differences in phase-to-phase changes in TVPs (e.g., $\gamma_{\mu 1, Treatment}$ which represents the between-group difference in growth rates of baseline levels). It can be hard for the model to detect small intervention effect sizes, so it would be of interest to investigate intervention effect sizes needed to successfully detect credible between-group differences in parameters. To evaluate this, we first define the effect size, manipulate the magnitude of the effect size in the data generation model (see below), and then evaluate how the proposed model would perform differently under different effect size conditions (e.g., the sensitivity (power) for detecting different intervention effect sizes).

Based on the simulation results, we will provide recommendations in practical settings where gradual changes in parameters are of key interest to applied researchers.

4.1. Simulation Designs

The sample size configurations were set to be realistic for EMI studies. Specifically, there were 200 subjects and for each subject, they were 240 time points, with 60 time points in each phase. The data were simulated based on the GoHiAR model specified in the previous section, where the true values of model parameters were mostly set according to the estimation results in our empirical study. An exhausted list of true values under each simulation condition can be found in Tables 2–6.

To evaluate research question 2, two scenarios were considered. First, we simulated data based on the full model and true values specified above, and then fitted a reduced model where the baseline growth parameter, $\beta_{\mu 1, i}$, was fixed to 0. Second, we simulated new data sets based on the reduced model so that there were no time-varying changes in the baseline level, and then fitted the full model. This allowed us to evaluate the type I error rates for detecting gradual changes in TVPs. For research question 3, we defined the effect size of between-group gradual change differences to be the standardized differences (in terms of standard deviations) between group means, following the conventional definition of the effect size measure (Cohen, 2013). For instance, for baseline levels, the corresponding effect size can be defined as $d = |\frac{\gamma_{\mu_1, Treatment}}{\sigma_{\beta_{\mu 1}}}|$. We then manipulated the values of $\gamma_{\mu_1, Treatment}$ to yield different effect size conditions. For instance, we used $d = 0.2$, 0.5, and 0.8 to reflect small, moderate, and large effect sizes based on benchmarks suggested by Cohen (2013).

4.1.1. Performance Measures

The following summary statistics were calculated to help evaluate the performance of the proposed approach and answer the research questions discussed above.

Suppose that H Monte Carlo replications are implemented in a simulation, θ is the true value of a particular parameter, and the point and standard error estimates of θ in the hth $(h = 1, \ldots, H)$ replication are $\hat{\theta}_h$ and $SE_{\hat{\theta}_h}$, respectively. Let the average of point estimates across H replications be $\bar{\theta}$, then the relative bias, SE, MCSE and RMSE are defined as follows:

$$\text{bias} = \frac{1}{H}\sum_{h=1}^{H}(\hat{\theta}_h - \theta), \tag{9}$$

$$\text{relative bias} = \frac{1}{H}\sum_{h=1}^{H}\frac{\hat{\theta}_h - \theta}{\theta}, \tag{10}$$

$$\text{SE} = \frac{1}{H}\sum_{h=1}^{H} SE_{\hat{\theta}_h}, \tag{11}$$

$$\text{MCSE} = \sqrt{\frac{1}{H-1}\sum_{h=1}^{H}(\hat{\theta}_h - \bar{\theta})^2}, \tag{12}$$

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

Table 2. Simulation results under the large effect size condition.

Par	True	Bias	RBias	SE	MCSE	RMSE	SENS	CR(%)	ESS
Group-level growth parameters									
Control Group									
IN intercept, $\gamma_{\phi0,Control}$	0.4	0	0	0.02	0.02	0.02	1	95	336
IIV intercept, $\gamma_{IIV0,Control}$	5	0	0	0.08	0.07	0.07	1	96	7054
MU intercept, $\gamma_{\mu0,Control}$	40	0.39	0.01	1.45	1.41	1.46	1	95	7469
IN slope, $\gamma_{\phi1,Control}$	−0.05	0	−0.01	0.01	0.01	0.01	1	96	193
IIV slope, $\gamma_{IIV1,Control}$	−0.3	0	0	0.02	0.02	0.02	1	94	2735
MU slope, $\gamma_{\mu1,Control}$	5	0.01	0	0.45	0.45	0.45	1	95	4539
Group-level contrast parameters									
IN intercept, $\gamma_{\phi0,Treatment}$	0.1	0	0.02	0.03	0.03	0.03	0.87	95	347
IIV intercept, $\gamma_{IIV0,Treatment}$	−0.1	−0.01	0.06	0.11	0.11	0.11	0.19	93	7045
MU intercept, $\gamma_{\mu0,Treatment}$	6	−0.58	−0.1	2.05	1.94	2.02	0.77	96	6361
IN slope, $\gamma_{\phi1,Treatment}$	0.03	0	−0.03	0.01	0.02	0.02	0.47	92	196
IIV slope, $\gamma_{IIV1,Treatment}$	0.05	0	0.02	0.03	0.03	0.03	0.42	95	2744
MU slope, $\gamma_{\mu1,Treatment}$	−2.4	0.01	0	0.65	0.63	0.63	0.97	96	3488
Random effect standard deviations									
SD of IN intercept, $\sigma_{\beta_{\phi0}}$	0.1	0	0.04	0.02	0.01	0.02	1	97	186
SD of IIV intercept, $\sigma_{\beta_{IIV0}}$	0.5	0.01	0.01	0.04	0.04	0.04	1	95	5730
SD of MU intercept, $\sigma_{\beta_{\mu0}}$	10	0.15	0.01	0.78	0.76	0.78	1	95	3974
SD of IN slope, $\sigma_{\beta_{\phi1}}$	0.02	0	0.14	0.01	0.01	0.01	1	98	71
SD of IIV slope, $\sigma_{\beta_{IIV1}}$	0.1	0	0.01	0.02	0.02	0.02	0.99	94	945
SD of MU slope, $\sigma_{\beta_{\mu1}}$	3	0.02	0.01	0.25	0.25	0.25	1	95	2027
Random effect correlations									
corr(IN intercept, IN slope)	0.2	−0.08	−0.41	0.41	0.25	0.26	0.03	98	95
corr(IIV intercept, IIV slope)	0.2	0.02	0.11	0.17	0.17	0.17	0.24	94	1139
corr(MU intercept, MU slope)	0.2	−0.01	−0.04	0.11	0.11	0.11	0.43	94	2118
Level-1 error standard deviations									
sdLevel1ErrorIN, σ_{ϕ}	0.1	0	−0.01	0.01	0.01	0.01	1	97	758
sdLevel1ErrorIIV, σ_{IIV}	0.1	−0.01	−0.07	0.03	0.03	0.03	1	88	268
sdLevel1ErrorMU, σ_{μ}	0.1	0.32	3.2	0.23	0.15	0.35	1	91	65

Note. MU = baseline; IIV = intraindividual variability; IN = inertia; SD = standard deviation; RBias = relative bias; SE = standard error; SENS = sensitivity; CR = coverage rate; ESS = effective sample size. Summary statistics were calculated based on 100 replications.

Table 3. Simulation results under model misspecifications where the slope parameters were incorrectly constrained to zero.

Par	True	Bias	RBias	SE	MCSE	RMSE	SENS	CR(%)	ESS
Group-level growth parameters									
Control Group									
IN intercept, $\gamma_{\phi0,Control}$	0.4	0	0.01	0.02	0.02	0.02	1	95	332
IIV intercept, $\gamma_{IIV0,Control}$	5	0	0	0.08	0.07	0.07	1	94	6973
MU intercept, $\gamma_{\mu0,Control}$	40	7.93	0.2	1.66	1.62	8.09	1	0	12648
IN slope, $\gamma_{\phi1,Control}$	−0.05	0	0.01	0.01	0.01	0.01	0.98	94	199
IIV slope, $\gamma_{IIV1,Control}$	−0.3	0	0	0.02	0.02	0.02	1	94	2688
MU slope, $\gamma_{\mu1,Control}$	5	–	–	–	–	–	–	–	–
Group-level contrast parameters									
IN intercept, $\gamma_{\phi0,Treatment}$	0.1	0	−0.05	0.03	0.04	0.04	0.79	93	346
IIV intercept, $\gamma_{IIV0,Treatment}$	−0.1	0	0.03	0.11	0.11	0.11	0.14	95	6948
MU intercept, $\gamma_{\mu0,Treatment}$	6	−4.09	−0.68	2.34	2.26	4.67	0.12	58	12607
IN slope, $\gamma_{\phi1,Treatment}$	0.03	0	0.01	0.02	0.02	0.02	0.51	94	203
IIV slope, $\gamma_{IIV1,Treatment}$	0.05	0	0.01	0.03	0.03	0.03	0.43	94	2704
MU slope, $\gamma_{\mu1,Treatment}$	−2.4	–	–	–	–	–	–	–	–
Random effect standard deviations									
SD of IN intercept, $\sigma_{\beta_{\phi0}}$	0.1	0	0.04	0.02	0.02	0.02	1	96	183
SD of IIV intercept, $\sigma_{\beta_{IIV0}}$	0.5	0.01	0.01	0.04	0.04	0.04	1	95	5515
SD of MU intercept, $\sigma_{\beta_{\mu0}}$	10	1.51	0.15	0.91	0.89	1.75	1	60	16833
SD of IN slope, $\sigma_{\beta_{\phi1}}$	0.02	0	0.18	0.01	0.01	0.01	1	97	73
SD of IIV slope, $\sigma_{\beta_{IIV1}}$	0.1	0	0	0.02	0.02	0.02	0.98	93	902
SD of MU slope, $\sigma_{\beta_{\mu1}}$	3	–	–	–	–	–	–	–	–
Random effect correlations									
corr(IN intercept, IN slope)	0.2	−0.1	−0.51	0.41	0.25	0.27	0.01	98	95
corr(IIV intercept, IIV slope)	0.2	0.02	0.08	0.18	0.16	0.16	0.23	94	1110
corr(MU intercept, MU slope)	0.2	–	–	–	–	–	–	–	–
Level-1 error standard deviations									
sdLevel1ErrorIN, σ_{ϕ}	0.1	0	−0.02	0.01	0.01	0.01	1	96	802
sdLevel1ErrorIIV, σ_{IIV}	0.1	−0.01	−0.06	0.03	0.03	0.03	1	89	255
sdLevel1ErrorMU, σ_{μ}	0.1	6.38	63.81	0.31	0.42	6.39	1	0	10765

Note. MU = baseline; IIV = intraindividual variability; IN = inertia; SD = standard deviation; RBias = relative bias; SE = standard error; SENS = sensitivity; CR = coverage rate; ESS = effective sample size. Summary statistics were calculated based on 100 replications.

Among these measures, a relative bias smaller than .1 indicates relatively good estimation performance, while other measures are subject to the scale of the true parameter value. In addition, we also considered (1) *sensitivity* (similar to power in the frequentist framework), defined as the proportion of replications whose credible intervals do not contain

Table 4. Simulation results under conditions where the slope parameters were freely estimated when their true values were zero.

Par	True	Bias	RBias	SE	MCSE	RMSE	SENS	CR(%)	ESS
Group-level growth parameters									
Control Group									
IN intercept, $\gamma_{\phi0,Control}$	0.4	0	0	0.02	0.02	0.02	1	96	342
IIV intercept, $\gamma_{IIV0,Control}$	5	0	0	0.08	0.08	0.08	1	93	7187
MU intercept, $\gamma_{\mu0,Control}$	40	0.28	0.01	1.44	1.35	1.37	1	95	3817
IN slope, $\gamma_{\phi1,Control}$	−0.05	0	0.01	0.01	0.01	0.01	1	96	197
IIV slope, $\gamma_{IIV1,Control}$	−0.3	0	−0.01	0.02	0.02	0.02	1	96	2778
MU slope, $\gamma_{\mu1,Control}$	0	0	–	0.12	0.11	0.11	0.05	95	134
Group-level contrast parameters									
IN intercept, $\gamma_{\phi0,Treatment}$	0.1	0	0.01	0.03	0.03	0.03	0.84	96	354
IIV intercept, $\gamma_{IIV0,Treatment}$	−0.1	0	0.01	0.11	0.11	0.11	0.15	96	7185
MU intercept, $\gamma_{\mu0,Treatment}$	6	−0.43	−0.07	2.03	1.89	1.94	0.8	96	3500
IN slope, $\gamma_{\phi1,Treatment}$	0.03	0	0.01	0.01	0.01	0.01	0.54	94	203
IIV slope, $\gamma_{IIV1,Treatment}$	0.05	0	−0.03	0.03	0.03	0.03	0.41	94	2793
MU slope, $\gamma_{\mu1,Treatment}$	0	0.01	–	0.19	0.18	0.18	0.04	96	115
Random effect standard deviations									
SD of IN intercept, $\sigma_{\beta_{\phi0}}$	0.1	0	0.03	0.02	0.02	0.02	1	97	184
SD of IIV intercept, $\sigma_{\beta_{IIV0}}$	0.5	0.01	0.01	0.04	0.04	0.04	1	96	5829
SD of MU intercept, $\sigma_{\beta_{\mu0}}$	10	0.1	0.01	0.76	0.74	0.75	1	94	4276
SD of IN slope, $\sigma_{\beta_{\phi1}}$	0.02	0	0.17	0.01	0.01	0.01	1	97	73
SD of IIV slope, $\sigma_{\beta_{IIV1}}$	0.1	0	−0.01	0.02	0.02	0.02	1	93	924
SD of MU slope, $\sigma_{\beta_{\mu1}}$	0	0.18	–	0.12	0.06	0.19	0.99	1	62
Random effect correlations									
corr(IN intercept, IN slope)	0.2	−0.07	−0.35	0.41	0.25	0.26	0.02	98	93
corr(IIV intercept, IIV slope)	0.2	0.03	0.13	0.18	0.16	0.17	0.28	95	1121
corr(MU intercept, MU slope)	0	−0.02	–	0.46	0.25	0.25	0.02	98	187
Level-1 error standard deviations									
sdLevel1ErrorIN, σ_{ϕ}	0.1	0	−0.02	0.01	0.01	0.01	1	96	776
sdLevel1ErrorIIV, σ_{IIV}	0.1	0	−0.05	0.03	0.02	0.03	1	90	263
sdLevel1ErrorMU, σ_{μ}	0.1	0.24	2.37	0.2	0.11	0.26	1	96	57

Note. MU = baseline; IIV = intraindividual variability; IN = inertia; SD = standard deviation; RBias = relative bias; SE = standard error; SENS = sensitivity (for parameter $\gamma_{\mu1,Control}$, $\gamma_{\mu1,Treatment}$, $\sigma_{\beta_{\mu1}}$ and corr(MU intercept, MU slope), sensitivity can also be regarded as type I error rates); CR = coverage rate; ESS = effective sample size. Summary statistics were calculated based on 100 replications.

Table 5. Simulation results under the medium effect size condition.

Par	True	Bias	RBias	SE	MCSE	RMSE	SENS	CR(%)	ESS
Group-level growth parameters									
Control Group									
IN intercept, $\gamma_{\phi0,Control}$	0.4	0	0	0.02	0.02	0.02	1	95	321
IIV intercept, $\gamma_{IIV0,Control}$	5	0	0	0.08	0.07	0.07	1	96	7081
MU intercept, $\gamma_{\mu0,Control}$	40	0.3	0.01	1.45	1.42	1.45	1	95	7504
IN slope, $\gamma_{\phi1,Control}$	−0.05	0	−0.01	0.01	0.01	0.01	1	93	186
IIV slope, $\gamma_{IIV1,Control}$	−0.3	0	0	0.02	0.02	0.02	1	96	2754
MU slope, $\gamma_{\mu1,Control}$	5	0.02	0	0.45	0.44	0.44	1	95	4395
Group-level contrast parameters									
IN intercept, $\gamma_{\phi0,Treatment}$	0.1	0	0	0.03	0.03	0.03	0.86	96	332
IIV intercept, $\gamma_{IIV0,Treatment}$	−0.1	0.01	−0.06	0.11	0.1	0.1	0.13	97	7091
MU intercept, $\gamma_{\mu0,Treatment}$	6	−0.47	−0.08	2.04	1.79	1.85	0.79	96	6468
IN slope, $\gamma_{\phi1,Treatment}$	0.03	0	0	0.01	0.02	0.02	0.52	94	190
IIV slope, $\gamma_{IIV1,Treatment}$	0.05	0	−0.02	0.03	0.03	0.03	0.45	97	2774
MU slope, $\gamma_{\mu1,Treatment}$	−1.5	−0.02	0.01	0.65	0.65	0.65	0.66	96	3439
Random effect standard deviations									
SD of IN intercept, $\sigma_{\beta_{\phi0}}$	0.1	0	0.04	0.02	0.02	0.02	1	95	173
SD of IIV intercept, $\sigma_{\beta_{IIV0}}$	0.5	0.01	0.02	0.04	0.04	0.04	1	96	5680
SD of MU intercept, $\sigma_{\beta_{\mu0}}$	10	0.14	0.01	0.78	0.78	0.79	1	95	3912
SD of IN slope, $\sigma_{\beta_{\phi1}}$	0.02	0	0.14	0.01	0.01	0.01	1	99	67
SD of IIV slope, $\sigma_{\beta_{IIV1}}$	0.1	0	0	0.02	0.02	0.02	0.99	95	922
SD of MU slope, $\sigma_{\beta_{\mu1}}$	3	0.02	0.01	0.25	0.25	0.25	1	96	1932
Random effect correlations									
corr(IN intercept, IN slope)	0.2	−0.09	−0.45	0.42	0.25	0.27	0.04	99	89
corr(IIV intercept, IIV slope)	0.2	0.01	0.07	0.18	0.17	0.17	0.23	94	1123
corr(MU intercept, MU slope)	0.2	−0.01	−0.03	0.11	0.11	0.11	0.45	94	2040
Level-1 error standard deviations									
sdLevel1ErrorIN, σ_{ϕ}	0.1	0	−0.02	0.01	0.01	0.01	1	96	720
sdLevel1ErrorIIV, σ_{IIV}	0.1	0	−0.04	0.03	0.03	0.03	1	90	269
sdLevel1ErrorMU, σ_{μ}	0.1	0.31	3.13	0.23	0.14	0.34	1	94	61

Note. MU = baseline; IIV = intraindividual variability; IN = inertia; SD = standard deviation; RBias = relative bias; SE = standard error; SENS = sensitivity; CR = coverage rate; ESS = effective sample size. Summary statistics were calculated based on 100 replications.

0. A sensitivity value close to 1 indicates a high level of sensitivity to the effect size (i.e., high power in detecting the effect size if conceptually understood in the frequentist framework); (2) *coverage rates*, defined as the percentages of replications whose credible intervals contain the true values. A coverage rate close to the nominal rate of 95% would be

Table 6. Simulation results under the small effect size condition.

Par	True	Bias	RBias	SE	MCSE	RMSE	SENS	CR(%)	ESS
Group-level growth parameters									
Control Group									
IN intercept, $\gamma_{\phi 0, Control}$	0.4	0	0	0.02	0.03	0.03	1	92	335
IIV intercept, $\gamma_{IIV0, Control}$	5	0	0	0.08	0.07	0.07	1	96	7243
MU intercept, $\gamma_{\mu 0, Control}$	40	0.34	0.01	1.44	1.48	1.52	1	94	7581
IN slope, $\gamma_{\phi 1, Control}$	−0.05	0	0	0.01	0.01	0.01	1	94	190
IIV slope, $\gamma_{IIV1, Control}$	−0.3	0	0	0.02	0.02	0.02	1	96	2831
MU slope, $\gamma_{\mu 1, Control}$	5	0.01	0	0.45	0.46	0.46	1	93	4532
Group-level contrast parameters									
IN intercept, $\gamma_{\phi 0, Treatment}$	0.1	0	0.01	0.03	0.03	0.03	0.85	95	350
IIV intercept, $\gamma_{IIV0, Treatment}$	−0.1	0	−0.01	0.11	0.11	0.11	0.15	94	7245
MU intercept, $\gamma_{\mu 0, Treatment}$	6	−0.52	−0.09	2.05	2.09	2.15	0.75	94	6387
IN slope, $\gamma_{\phi 1, Treatment}$	0.03	0	−0.01	0.02	0.01	0.01	0.49	95	195
IIV slope, $\gamma_{IIV1, Treatment}$	0.05	0	0.01	0.03	0.03	0.03	0.43	94	2833
MU slope, $\gamma_{\mu 1, Treatment}$	−0.6	0.01	−0.02	0.65	0.65	0.65	0.16	95	3467
Random effect standard deviations									
SD of IN intercept, $\sigma_{\beta_{\phi 0}}$	0.1	0	0	0.02	0.02	0.02	1	96	184
SD of IIV intercept, $\sigma_{\beta_{IIV0}}$	0.5	0.01	0.01	0.04	0.04	0.04	1	95	5912
SD of MU intercept, $\sigma_{\beta_{\mu 0}}$	10	0.16	0.02	0.78	0.81	0.82	1	95	3882
SD of IN slope, $\sigma_{\beta_{\phi 1}}$	0.02	0	0.13	0.01	0.01	0.01	1	98	70
SD of IIV slope, $\sigma_{\beta_{IIV1}}$	0.1	0	0.01	0.02	0.02	0.02	1	94	938
SD of MU slope, $\sigma_{\beta_{\mu 1}}$	3	0.02	0.01	0.25	0.25	0.25	1	95	1981
Random effect correlations									
corr(IN intercept, IN slope)	0.2	−0.09	−0.46	0.42	0.26	0.28	0.03	97	93
corr(IIV intercept, IIV slope)	0.2	0.02	0.11	0.17	0.16	0.17	0.23	94	1139
corr(MU intercept, MU slope)	0.2	0	−0.02	0.11	0.11	0.11	0.47	94	2100
Level-1 error standard deviations									
sdLevel1ErrorIN, σ_{ϕ}	0.1	0	−0.01	0.01	0.01	0.01	1	96	747
sdLevel1ErrorIIV, σ_{IIV}	0.1	0	−0.04	0.03	0.03	0.03	1	89	287
sdLevel1ErrorMU, σ_{μ}	0.1	0.32	3.16	0.23	0.15	0.35	1	90	65

Note. MU = baseline; IIV = intraindividual variability; IN = inertia; SD = standard deviation; RBias = relative bias; SE = standard error; SENS = sensitivity; CR = coverage rate; ESS = effective sample size. Summary statistics were calculated based on 100 replications.

considered ideal; and (3) *type I error rates* (only used for research question 4), defined as the proportion of replications whose credible intervals do not contain 0 when the true value is 0. A type I error rate higher than 5% indicates the false rejection of an actually true null effect.

4.2. Simulation Results

For each condition, we ran 500 Monte Carlo replications. For each replication, we ran two chains, each with 20000 iterations in total and a burn-in of 5000 (discarded) iterations. The computational time was about 1 hour for each replication under the present settings.

4.2.1. Overall Performance of the Proposed Model

In this subsection, we chose one simulation condition (the large effect size condition) to demonstrate the estimation accuracy and sampling efficiency of the proposed approach, for which the results were summarized in Table 2. Results under other conditions will be shown and discussed later.

Overall, most parameters were recovered well, with relative biases between −.1 and .1. Even for parameters with more biased estimates, the coverage rates were above 90%. In terms of sensitivity, the contrast parameters generally yielded lower sensitivity than other parameters, mainly because of the small true values (close to 0) and large SE estimates, such that the credible intervals were likely to cover zero (null effect). In fact, if calculated based on the effect size definition introduced before, group-level contrast parameters such as $\gamma_{\mu 0, Treatment}$, $\gamma_{IIV0, Treatment}$, and $\gamma_{IIV1, Treatment}$ actually reflected small to medium effect sizes and thus yielded lower sensitivity. However, since we specified a large effect size for the slope differences in baseline levels, the sensitivity for this particular contrast parameter (i.e., $\gamma_{\mu 1, Treatment}$) was high. Finally, in terms of the ESS, we found that for most key parameters of interests (e.g., group-level growth and contrast parameters), the ESS was acceptable. Compared with baseline and IIV-related parameters, AR-related parameters such as $\gamma_{\phi 1, Control}$ and $\gamma_{\phi 1, Treatment}$ seemed to consistently yield lower ESS, which can be increased by running more iterations in the MCMC process.

4.2.2. Effects of Model Misspecifications

First, we focused on the condition where the true model assumed time-varying changes in baseline levels, whereas the fitted model was misspecified such that the slope parameters were incorrectly constrained to 0. The simulation results in Table 3 showed that in this scenario, baseline-related parameters (see $\gamma_{\mu 0, Control}$, $\gamma_{\mu 0, Treatment}$, $\sigma_{\beta_{\mu 0}}$, and σ_{μ}) would be biased with low coverage rates, whereas parameters related to inertia and IIV were not largely affected. Specifically, the measurement error standard deviation was overestimated since time-varying changes were captured by the measurement error when the slope parameter was missing in the fitted model. In addition, the between-group difference in the initial baseline level was biased because without the slope parameter, the initial baseline level parameter actually captured the average baseline level across time.

In another scenario where the true model assumed no time-varying changes in baseline levels, whereas the fitted model contained slope parameters, Table 4 showed that our proposed approach yielded acceptable type I error rates (e.g., 0.05 for $\gamma_{\mu1,Control}$ and 0.04 for $\gamma_{\mu1,Treatment}$). That is, the probability of falsely detecting gradual changes was relatively low. However, the corresponding ESS was notably lower (e.g., around 100) compared to other conditions where the true values of slope parameters were non-zero, providing some indication that the corresponding fixed effects and in particular the variance of the slope parameter tended to show small variations near the value of 0.

4.2.3. Sensitivity to Intervention Effect Sizes

Results under the large effect size condition (see Table 2, discussed before) were compared to those under small (Table 6) and medium (Table 5) effect size conditions. As expected, reducing the effect size yielded lower sensitivity in detecting the effect (large: 0.97; medium: 0.66; small: 0.16). That is, a moderate to large effect size is needed to be able to reliably detect intervention effects with the chosen sample size configuration. Under small effect size conditions, a larger sample size is needed to detect credible intervention effects. Specifically, additional simulations (not shown in the tables) suggested that for small effect sizes we need 1000 subjects for acceptable sensitivity (which was 0.77 in that higher sample size setting).

Overall, the simulation results highlighted the importance of accounting for TVPs in the model when there were actually time-varying changes in dynamic features. When applied to examine intervention effects, our findings suggested satisfactory estimation accuracy across all three effect size conditions and acceptable sensitivity under moderate to large effect sizes. However, to achieve a high sensitivity level under the small effect size, more samples are required.

5 Empirical Illustration

5.1. Data Descriptions and Preprocessing

Part of the information about the data set has been provided in the motivating example section, so this subsection will focus on details not described before, along with some data pre-processing procedures.

5.1.1. Participants

Subjects who responded to the recruitment advertisement were provided a consent form during an introductory session. Having read the consent form, they were free to ask further information about the study. Once all of their questions were answered and if they were still interested in participating, they were asked to sign a consent form. The study visits were scheduled and procedures began after written consent was obtained. All participant interactions were overseen by the University's Institutional Review Board (see more details in protocol "8926"). Participants enrolled in the intervention study were undergraduate students aged 18–22 years, with middle to high socioeconomic status. Among 160 participants, 68% were females and 32% were males; 77% were white, 12% were Asian, 6% were black, 4% were Latino, and 1% did not disclose their race.

5.1.2. Measures

Meaning of life was measured via the "Meaning" subscale of the momentary PERMA (mPERMA) questionnaire, which was adapted from the PERMA-Profiler (Butler & Kern, 2016) to fit the repeated, momentary assessment nature of the study's ecological momentary assessments (EMAs) design (Heshmati et al., 2023). The study span over 56 days, during which time participants were prompted to complete self-reports on their Meaning of life 6 times per day, which were randomly distributed across the day (during their self-reported waking hours and at least 30 minutes apart). At each measurement point, participants responded to a set of questions on a continuous sliding scale from 0 to 100, with the labels "Not at all" and "Extremely" at the two endpoints. Example questions included the extent to which they felt that they led a purposeful and meaningful life, had a sense of direction in their life.

5.1.3. Data Preprocessing

After removing data from participants with limited within-person variability in data over time ($n = 1$), a final sample size of $n = 108$ was retained for analysis. In addition, the discrete-time nature of the GoHiAR model requires data to be equally spaced. Since the time intervals between assessments varied within individuals over time in our data set, following procedures adopted in previous studies (Chow & Zhang, 2013; Li et al., 2019), we aggregated data into four blocks (i.e., 12am–6am, 6am–12pm, 12 pm–6pm, 6 pm–12am) per day to represent the sleeping, morning, afternoon, and night periods. The final number of time points ranged from 212 to 224 (i.e., 56×4). The overall missing rate was 30% across individuals and time points. For the purpose of this first introduction of the GoHiAR model, missingness was assumed to be missing completely at random or missing at random, with no missing data model specified. The model was fitted in JAGS. When JAGS encounters missing data coded as NA, it imputes a value from the posterior predictive, which conforms to the missing completely at random or missing at random assumption we made. We ran two chains in JAGS, each with 200,000 iterations in total and a burn-in of 10,000 iterations.

5.2. Results

The diagnostic criteria for adequate sampling and convergence were set as ESS greater than 1000 and $\hat{R}$ below 1.1, respectively. Results showed that ESS was greater than 1000 for all parameters except for two random effect standard deviation and correlation parameters, for which the ESS was 307 and 559, respectively; the $\hat{R}$ was below 1.1 for all parameters. Scripts and data are provided on OSF[2].

[2]https://osf.io/vfps8/?view_only=fd1b3f23a7d64a37bf76fb6a8117be93

Table 7. Estimation results.

	Est	SE	95% CI (LL)	95% CI (UL)
Group-level growth parameters				
Control Group				
IN intercept, $\gamma_{\phi0,Control}$	0.14	0.02	0.10	0.18
IIV intercept, $\gamma_{IIV0,Control}$	3.70	0.14	3.42	3.99
MU intercept, $\gamma_{\mu0,Control}$	75.18	1.96	71.31	78.99
IN slope, $\gamma_{\phi1,Control}$	−0.03	0.01	−0.04	−0.01
IIV slope, $\gamma_{IIV1,Control}$	−0.23	0.06	−0.36	−0.11
MU slope, $\gamma_{\mu1,Control}$	0.19	0.53	−0.86	1.23
Treatment Group				
IN intercept, $\gamma_{\phi0,Control} + \gamma_{\phi0,Treatment}$	0.12	0.02	0.08	0.15
IIV intercept, $\gamma_{IIV0,Control} + \gamma_{IIV0,Treatment}$	3.64	0.14	3.36	3.93
MU intercept, $\gamma_{\mu0,Control} + \gamma_{\mu0,Treatment}$	76.07	1.99	72.16	79.99
IN slope, $\gamma_{\phi1,Control} + \gamma_{\phi1,Treatment}$	−0.02	0.01	−0.04	0.00
IIV slope, $\gamma_{IIV1,Control} + \gamma_{IIV1,Treatment}$	−0.22	0.06	−0.35	−0.09
MU slope, $\gamma_{\mu1,Control} + \gamma_{\mu1,Treatment}$	−0.36	0.53	−1.41	0.68
Group-level contrast parameters				
IN intercept, $\gamma_{\phi0,Treatment}$	−0.02	0.03	−0.08	0.03
IIV intercept, $\gamma_{IIV0,Treatment}$	−0.06	0.20	−0.46	0.34
MU intercept, $\gamma_{\mu0,Treatment}$	0.90	2.74	−4.45	6.31
IN slope, $\gamma_{\phi1,Treatment}$	0.01	0.01	−0.01	0.03
IIV slope, $\gamma_{IIV1,Treatment}$	0.01	0.09	−0.17	0.19
MU slope, $\gamma_{\mu1,Treatment}$	−0.55	0.75	−2.03	0.93
Random effect standard deviations				
SD of IN intercept, $\sigma_{\beta_{\phi0}}$	0.08	0.02	0.05	0.11
SD of IIV intercept, $\sigma_{\beta_{IIV0}}$	0.91	0.09	0.75	1.10
SD of MU intercept, $\sigma_{\beta_{\mu0}}$	14.61	1.07	12.69	16.88
SD of IN slope, $\sigma_{\beta_{\phi1}}$	0.02	0.01	0.00	0.04
SD of IIV slope, $\sigma_{\beta_{IIV1}}$	0.38	0.04	0.29	0.47
SD of MU slope, $\sigma_{\beta_{\mu1}}$	3.60	0.31	3.04	4.25
Random effect correlations				
corr(IN intercept, IN slope)	−0.18	0.49	−0.91	0.87
corr(IIV intercept, IIV slope)	0.29	0.16	−0.01	0.62
corr(MU intercept, MU slope)	−0.08	0.11	−0.28	0.14
Level-1 error standard deviations				
sdLevel1ErrorIN, σ_{μ}	0.09	0.01	0.08	0.11
sdLevel1ErrorIIV, σ_{ϕ}	0.59	0.03	0.53	0.66
sdLevel1ErrorMU, σ_{IIV}	3.06	0.19	2.71	3.45

Note. MU = baseline; IIV = intraindividual variability; IN = inertia; Est = Estimate; SE = standard error; SD = standard deviation; CI = credible interval; LL = lower limit; UL = upper limit. The point estimates, SEs, and 95% CIs were obtained by calculating the means, standard deviations, 2.5th and 97.5 percentiles of the posterior distributions for each parameter.

Parameter estimation results are shown in Table 7. The “Est” column shows the point parameter estimates (posterior means) and the “SE” (standard error) column shows their corresponding posterior standard deviations. The last two columns show the lower and upper limits of their 95% credible intervals, which were used to determine whether a parameter was credibly different from 0.

5.2.1. Between-Group Differences in Pre-Intervention Dynamic Characteristics

In the pre-intervention phase, participants in both Control and Treatment groups had relatively high baseline levels of meaning of life (e.g., around 76 out of 100, see “MU intercept” under both Control and Treatment Groups in Table 7) and displayed moderate inertia (see “IN intercept”) and relatively high variations in their meaning of life levels (IIV; see “IIV intercept”).

The contrast between the two groups’ intercept parameters for all dynamic characteristics indicated no credible differences in meaning of life dynamics between groups in the pre-intervention phase, as indicated by “IN/IIV/MU intercept” under “Group-level contrast parameters” in Table 7. This was expected because under the RCT design, participants were randomly assigned into different groups and would display similar Meaning dynamics before any digital monitoring or intervention content.

5.2.2. Within-Group Changes of Dynamic Characteristics

Based on the estimates in the “Group-level growth parameters” part of Table 7, we concluded that there were no credible changes in baseline levels of meaning of life in either groups (see “MU slope” under both Control and Treatment Groups). However, we found credible improvements in meaning of life in terms of decreased IIV (see “IIV slope”) and decreased inertia/increased regulation (see “IN slope”) for both groups. A graphical illustration of the phase-to-phase changes for Control and Treatment groups was shown in Figure 2.

5.2.3. Between-Group Differences in Changes of Dynamic Characteristics

The contrast between two groups’ slope parameters for all dynamic characteristics indicated no credible differences in changes of meaning of life dynamic characteristics between groups (see “IN/IIV/MU slope” under “group-level contrast parameters” in Table 7). That is, no credible intervention effects were found when comparing the Treatment group to the Control group. In fact, the estimated effect size was

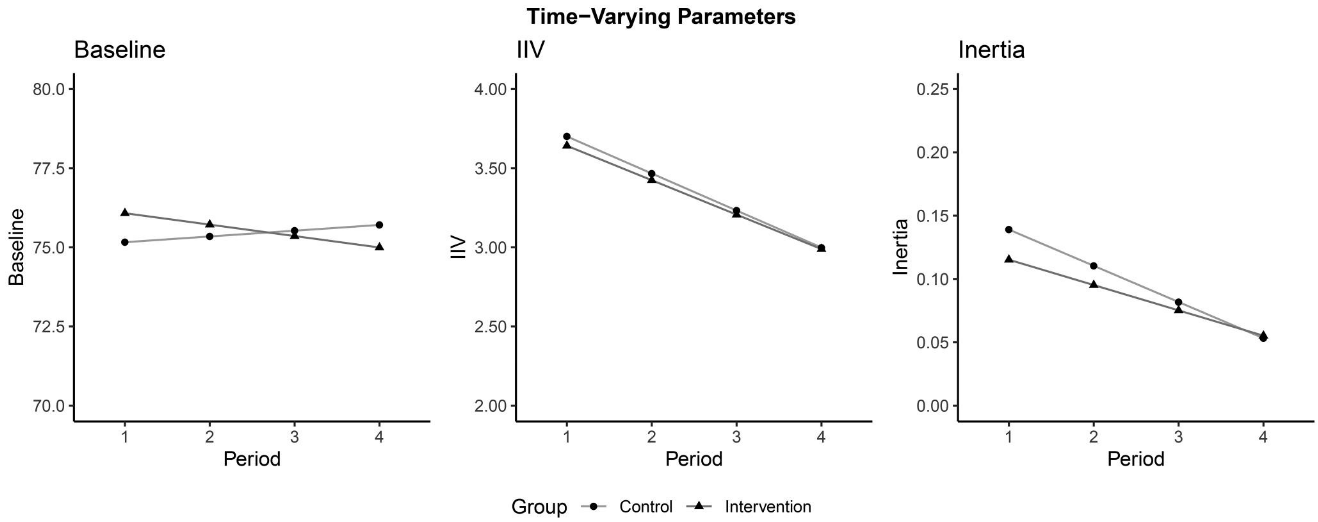

Figure 2. Between-group differences in changes of baseline (left column), IIV (Middle column), and inertia (right column) of meaning dynamics. The lines were plotted based on the estimated population mean of the corresponding intercept and slope parameters. For instance, the line indicating the change of baseline of meaning in the control group was plotted based on the estimated "MU intercept" and "MU slope" under control group in Table 7).

relatively small based on the definition in the simulation study. According to our simulation results under the small effect size condition, more samples are needed to detect smallish intervention effects, so the current study might have been underpowered for capturing small between-group differences in changes of dynamic characteristics related to meaning of life.

Finally, we also built an alternative TVP model, where we did not assume any trend in changes of baseline, IIV and inertia, but directly estimated their values in each of the four phases and then calculated contrast parameters. The results led to the same conclusion that there were no credible differences between groups in terms of changes of dynamic characteristics.

6. Discussion

In this study, we proposed the GoHiAR model, a comprehensive modeling framework for understanding individual differences in changes at multiple timescales. Our approach extracts dynamical features from intensive longitudinal data that map onto theoretical insights about psychological dynamics. At the same time, it also models slow timescale change in terms of these dynamics, allowing for testing of change mechanisms that are developmental or intervention triggered. While many studies would focus on change in mean levels when evaluating the effect of an intervention, our approach provides a more nuanced picture by being able to test changes in intra-individual variation and inertia as well at the same time.

The multilevel Bayesian implementation of the model offers additional advantages. Since all model parameters are estimated simultaneously, the uncertainty in the dynamical and the slow timescale parameters are propagated in a statistically sound manner. The Bayesian framework also allows for introducing prior information, that can help increase estimation accuracy, precision and efficiency. While introducing informative priors (i.e., information based on previous studies) is certainly possible in this framework, we did not focus on such priors given that the novelty of the approach makes it unlikely to have such information available. Instead, we tested prior specifications that can be considered weakly informative. These specifications leverage on user's understanding of the scales of the data and the characteristics of the parameters to impose some restrictions on the priors. The proposed weakly informative priors constrain the estimation in the sense that mean values that are outside of the range of the data are assigned minimal probability, variance parameters cannot be negative, and so on. Finally, the Bayesian framework provides for an intuitive interpretation of the results, in terms of quantifying the probability for ranges of parameter values. For example, we can easily calculate the probability that we have at least a medium-sized effect in terms of change in intra-individual variation based on the posterior distribution of the relevant parameter.

We conducted a simulation study to evaluate the performance of the proposed approach under different conditions. Our simulation results highlighted the importance of accounting for TVPs in the model when there were actually time-varying changes in dynamic features. When applied to examine intervention effects, our approach was shown to yield satisfactory estimation accuracy across all effect size conditions, as well as acceptable sensitivity under moderate to large effect sizes. However, to achieve a high sensitivity level under the small effect size, more samples are required.

We also illustrated its benefits of the GoHiAR model for studying multifaceted processes of PWB and PWB interventions using an empirical study. The empirical study suggested that participants generally experienced some improvement in their meaning of life (e.g., decreased IIV and inertia) over the course of the intervention. This might be due to the nature of the EMA study, where frequent assessments of experiences can yield changes in the participants' psychological states that are being measured (i.e., measurement reactivity (Eisele et al., 2023)). In our case, responding to repeated EMA prompts on well-being might impact someone's well-being. That is, the mere participation in such a study may have an effect on the participants' Meaning reports throughout the intervention – such changes in actions or feelings from repeated measurements throughout a study have been reported in previous studies (Clifford & Davis, 2012; Oravecz et al., 2020; Ottenstein et al., 2024; Runyan et al., 2013).

We were not able to detect any effect of the tested PPM intervention on meaning of life at the group level. These findings suggested that a larger sample size is needed if we are interested in capturing potentially small sized effects or more personalized digital interventions might be needed to improve the intervention efficacy (i.e., by increasing effect size). It is possible that the intervention itself might lack the intensity needed to lead to substantial changes, or the delivery of interventions might not be well suited for changing participants' Meaning. There might also be some measurement issues that biased these results, such as measurement burden due to frequent interruptions in the intervention group and/or unmeasured environmental factors. Also, the lack of credible group-level differences could be due to selection effects (Stone et al., 2023), as the participants in both the control and intervention groups at the pre-intervention phase were found to have high baseline levels of meaning of life. It is possible that the participants who enrolled into the study were already in a good place in terms of their psychological well-being, and there was not much room for improvement. Individuals with different dynamical characteristics could have responded differently to such an intervention.

Nevertheless, the GoHiAR model highlighted some alternative dynamic characteristics in the study of participants' intraindividual change processes beyond changes in baseline levels. Without the GoHiAR modeling, we would not have been able to see that PWB dynamics did change across the study, but not in terms of baseline but other dynamical characteristics of the individual.

Despite the promising use of the GoHiAR model to inform intervention studies, there were several limitations that need to be noted. First, our model was proposed based on strong assumptions about change patterns of dynamic characteristics, including known timing of change and number of change points. These assumptions were made mainly due to the nature of the study design, i.e., the pre-, during, and post-intervention design. Hence, we built highly theory-driven parametric models for TVPs. In more general cases where no contextual information can be used in model specifications, it is highly recommended that researchers fit a set of candidate models for TVPs (e.g., parametric, semi- and non-parametric models, etc.) and select appropriate models using model selection tools such as cross-validation, Bayes factor, and information criteria. Second, in the present study, we only considered linear models for phase-to-phase changes. One possible extension would be extending the linear GCM to nonlinear GCM to capture potential nonlinear changes. For instance, the gradual change may be approximated by the Gompertz growth curve where the growth rate decays exponentially over time, since participants may experience a higher rate of change during the intervention period, followed by a slower rate of change after the intervention. Third, our model can be regarded as a piecewise multilevel AR model, so the stationarity assumption needs to hold in each of the four phase. By plotting the raw time series data, we did not observe obvious trends within phases in our data set, so the current model specification would not violate the assumption. However, if the dynamic characteristics do change substantially within a certain phase, our model needs to be modified to capture such changes. Lastly, in this study the raw data were aggregated to be equally spaced to fit a discrete-time model. However, modeling continuous-time processes with discrete-time models can be problematic and continuous-time extensions to the current model may be needed.

Some future directions may be pursued. First, the proposed model can potentially be extended to a multivariate model (e.g., a growth of hierarchical vector autoregressive model (GoHiVAR)) to capture the relationships between variables as well as changes in these relationships over time. One challenge related to fitting a GoHiVAR model is the rapid increase in computational complexity caused by the increased number of model parameters. Second, instead of gradual changes in dynamic characteristics, some processes may be featured by frequent and sudden shifts in the magnitudes of these dynamic characteristics, in which case the regime-switching structure could be added to the current model to capture such shifts.

Overall, we proposed and demonstrated promising uses of GoHiAR models in analyzing multi-subject and multi-phase time series data. By investigating individual differences in intraindividual changes in multifaceted characteristics of psychological processes, the GoHiAR model helps address a number of substantive questions such as how individuals/groups benefit differentially from the intervention in terms of multiple characteristics as well as how interventions can be personalized to improve their efficacy.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Bollen, K. A., & Curran, P. J. (2004). Autoregressive latent trajectory (alt) models a synthesis of two traditions. *Sociological Methods & Research*, *32*, 336–383. https://doi.org/10.1177/0049124103260222

Segrin, C., & Taylor, M. (2007). Positive interpersonal relationships mediate the association between social skills and psychological well-being. *Personality and Individual Differences*, *43*, 637–646. https://doi.org/10.1016/j.paid.2007.01.017

Seligman, M. E. (2012). *Flourish: A visionary new understanding of happiness and well-being*. Simon and Schuster.

Song, H., & Ferrer, E. (2012). Bayesian estimation of random coefficient dynamic factor models. *Multivariate Behavioral Research*, *47*, 26–60. https://doi.org/10.1080/00273171.2012.640593

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

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, *69*, 542–547. https://doi.org/10.2307/1925546

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*, 409–425. https://doi.org/10.1037/met0000085

Browne, M. W., & Du Toit, H. C. (1991). Models for learning data. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (p. 47–68). American Psychological Association.

Butler, J., & Kern, M. L. (2016). The perma-profiler: A brief multidimensional measure of flourishing. *International Journal of Wellbeing*, *6*, 1–48. https://doi.org/10.5502/ijw.v6i3.526

Chen, M., Chow, S.-M., Hammal, Z., Messinger, D. S., & Cohn, J. F. (2021). A person-and time-varying vector autoregressive model to capture interactive infant-mother head movement dynamics. *Multivariate Behavioral Research*, *56*, 739–767. https://doi.org/10.1080/00273171.2020.1762065

Chow, S.-M., Haltigan, J. D., & Messinger, D. S. (2010). Dynamic infant–parent affect coupling during the face-to-face/still-face. *Emotion (Washington, D.C.)*, *10*, 101–114. https://doi.org/10.1037/a0017824

Chow, S.-M., & Zhang, G. (2013). Nonlinear regime-switching state-space (RSSS) models. *Psychometrika*, *78*, 740–768. https://doi.org/10.1007/s11336-013-9330-8

Clifford, P. R., & Davis, C. M. (2012). Alcohol treatment research assessment exposure: A critical review of the literature. *Psychology of Addictive Behaviors*, *26*, 773–781. https://doi.org/10.1037/a0029747

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. routledge.

Curran, P. J., & Bollen, K. A. (2001). The best of both worlds: Combining autoregressive and latent curve models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change.* (pp. 105–136). American Psychological Association.

Del Negro, M., & Otrok, C. (2008). Dynamic factor models with time-varying parameters: Measuring changes in international business cycles. FRB of New York Staff Report.

Deschamps, P. J. (2003). Time-varying intercepts and equilibrium analysis: An extension of the dynamic almost ideal demand model. *Journal of Applied Econometrics*, *18*, 209–236. https://doi.org/10.1002/jae.674

Drozd, F., Raeder, S., Kraft, P., & Bjørkli, C. A. (2013). Multilevel growth curve analyses of treatment effects of a web-based intervention for stress reduction: Randomized controlled trial. *Journal of Medical Internet Research*, *15*, e84. https://doi.org/10.2196/jmir.2570

Ebner-Priemer, U. W., & Trull, T. J. (2009). Ecological momentary assessment of mood disorders and mood dysregulation. *Psychological Assessment*, *21*, 463–475. https://doi.org/10.1037/a0017075

Eisele, G., Vachon, H., Lafit, G., Tuyaerts, D., Houben, M., Kuppens, P., Myin-Germeys, I., & Viechtbauer, W. (2023). A mixed-method investigation into measurement reactivity to the experience sampling method: The role of sampling protocol and individual characteristics. *Psychological Assessment*, *35*, 68–81. https://doi.org/10.1037/pas0001177

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

Hamaker, E. L. (2005). Conditions for the equivalence of the autoregressive latent trajectory model and a latent growth curve model with autoregressive disturbances. *Sociological Methods & Research*, *33*, 404–416. https://doi.org/10.1177/0049124104270220

Harvey, A. C. (2001). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press.

Helm, J. L., Ram, N., Cole, P. M., & Chow, S.-M. (2016). Modeling self-regulation as a process usinga multiple time-scale multiphase latent basis growth model. *Structural Equation Modeling*, *23*, 635–648. https://doi.org/10.1080/10705511.2016.1178580

Heron, K. E., & Smyth, J. M. (2010). Ecological momentary interventions: Incorporating mobile technology into psychosocial and health behaviour treatments. *British Journal of Health Psychology*, *15*, 1–39. https://doi.org/10.1348/135910709X466063

Heshmati, S., Kibrislioglu Uysal, N., Kim, S. H., Oravecz, Z., & Donaldson, S. I. (2023). Momentary PERMA: An adapted measurement tool for studying well-being in daily life. *Journal of Happiness Studies, 24*, 2441–2472.

Heshmati, S., Muth, C., Roeser, R., Smyth, J., Jamalabadi, H., & Oravecz, Z. (2024). Conceptualizing psychological well-being as a dynamic process: Implications for research on mobile health interventions. *Social and Personality Psychology Compass*, *18*, 933. https://doi.org/10.1111/spc3.12933

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

Li, Y., Ji, L., Oravecz, Z., Brick, T. R., Hunter, M. D., & Chow, S.-M. (2019). dynr.mi: An r program for multiple imputation in dynamic modeling. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, *13*, 302–311.

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*, *29*, 452–475. https://doi.org/10.1080/10705511.2021.1911657

Neal, R. M. (2003). Slice sampling. *Annals of Statistics*, *31*, 705–741.

Oravecz, Z., Dirsmith, J., Heshmati, S., Vandekerckhove, J., & Brick, T. R. (2020). Psychological well-being and personality traits are associated with experiencing love in everyday life. *Personality and Individual Differences*, *153*, 109620. https://doi.org/10.1016/j.paid.2019.109620

Ottenstein, C., Hasselhorn, K., & Lischetzke, T. (2024). Measurement reactivity in ambulatory assessment: Increase in emotional clarity over time independent of sampling frequency. *Behavior Research Methods*, *56*, 6150–6164. https://doi.org/10.3758/s13428-024-02346-y

Ou, L., Chow, S.-M., Ji, L., & Molenaar, P. C. (2017). (re) evaluating the implications of the autoregressive latent trajectory model through likelihood ratio tests of its initial conditions. *Multivariate Behavioral Research*, *52*, 178–199. https://doi.org/10.1080/00273171.2016.1259980

Pagan, A. (1984). Econometric issues in the analysis of regressions with generated regressors. *International Economic Review*, *25*, 221–247. https://doi.org/10.2307/2648877

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing* (Vol. 124, p. 1–10).

Ram, N., Brose, A., Molenaar, P. C. (2013). Dynamic factor analysis: Modeling person-specific process. *The Oxford Handbook of Quantitative Methods*, *2*, 441–457.

Röcke, C., & Brose, A. (2013). Intraindividual variability and stability of affect and well-being: Short-term and long-term change and stabilization processes. *GeroPsych*, *26*, 185–199. https://doi.org/10.1024/1662-9647/a000094

Roesch, S. C., Norman, G. J., Villodas, F., Sallis, J. F., & Patrick, K. (2010). Intervention-mediated effects for adult physical activity: A latent growth curve analysis. *Social Science & Medicine (1982)*, *71*, 494–501. https://doi.org/10.1016/j.socscimed.2010.04.032

Runyan, J. D., Steenbergh, T. A., Bainbridge, C., Daugherty, D. A., Oke, L., & Fry, B. N. (2013). A smartphone ecological momentary assessment/intervention "app" for collecting real-time data and promoting self-awareness. *PloS One*, *8*, e71325. https://doi.org/10.1371/journal.pone.0071325