

Unique Contributions of Dynamic Affect Indicators – Beyond Static Variability

Kenneth Koslowski and Jana Holtmann

Leipzig University

ABSTRACT

Indicators of affect dynamics (IADs) capture temporal dependencies and instability in affective trajectories over time. However, the relevance of IADs for the prediction of time-invariant outcomes (e.g., depressive symptoms) was recently challenged due to results suggesting low predictive utility beyond intraindividual means and variances. We argue that these results may in part be explained by mathematical redundancies between IADs and static variability as well as the chosen modeling strategy. In three extensive simulation studies we investigate the accuracy and power for detecting non-null relations between IADs and an outcome variable in different relevant settings, illustrating the effect of the length of a time series, the presence of missing values or measurement error, as well as of erroneously fixing innovation variances to be equal across persons. We show that, if uncertainty in individual IAD estimates is not accounted for, relations between IADs (i.e., autoregressive effects) and a time-invariant outcome are underestimated even in large samples and propose the use of a latent multilevel one-step approach. In an empirical application we illustrate that the different modeling approaches can lead to different substantive conclusions regarding the role of negative affect inertia in the prediction of depressive symptoms.

KEYWORDS

Affect dynamics; emotional inertia; affect variability; innovation variance; two-step approach; measurement-error vector-autoregressive models

Emotions are dynamic in nature (e.g., Kuppens, Oravecz, et al., 2010) and studying the temporal aspects of affect trajectories via means of ambulatory assessments (AA) is a promising way to broaden the understanding of inter-individual differences in emotion dynamics and their role in psychological health (Koval et al., 2015; Kuppens, Allen, & Sheeber, 2010; van de Leemput et al., 2014). That is, regularities in short-term fluctuations of affective experience, such as the tendency of emotional states to carry over from one moment to the next, referred to as *emotional inertia* (Suls et al., 1998), or inter-individual differences in emotional variability and instability potentially carry meaningful information on inter-individual differences associated with relevant psychological outcomes and psychopathological symptomatology (Hamaker et al., 2018; Wang et al., 2012). In this vein, high levels of emotional inertia were hypothesized to reflect a decreased ability to adapt to significant events and regulate emotions effectively and, consequently, to

indicate dysfunctional emotional responding (Kuppens, Allen, & Sheeber, 2010). Indeed, emotional inertia was observed to be linked with a variety of indicators of psychological health (Houben et al., 2015), such as concurrent levels of different clinical (e.g., depressive symptoms, Brose et al., 2015; Koval et al., 2012, 2013; Nelson et al., 2020; Wenzel & Brose, 2023) and non-clinical (e.g., self-esteem, Kuppens, Allen, & Sheeber, 2010) outcomes, and shown to function as a prospective predictor of major depressive disorder onset (Kuppens et al., 2012; van de Leemput et al., 2014; for an opposing view, see Houben & Kuppens, 2020). Similarly, van Roekel et al. (2018) found variations in a genetic risk factor for emotional dysregulation to be meaningfully linked with inertia but not mean levels of negative affect.

Emotional inertia is commonly measured as the autocorrelation of a time series or the autoregressive effect (AR) in an autoregressive model of order 1 (i.e., with time lag 1). Besides AR, a variety of indicators of

affect dynamics (IADs), aiming to capture unique aspects of regularities in affect trajectories, such as mean squared successive differences (MSSD), were introduced (for an overview, see Dejonckheere et al., 2019) and linked to clinical (e.g., Heller et al., 2019; Santangelo et al., 2017; Trull et al., 2008) and non-clinical outcomes (e.g., Houben et al., 2015).

Recent findings question the relevance of IADs as predictors of clinical (Bos et al., 2019; Dejonckheere et al., 2019; Houben & Kuppens, 2020) and non-clinical (Wendt et al., 2020) outcomes beyond *static* mean levels and intraindividual standard deviations (*ISD*). The authors concluded that after accounting for overlap between the more parsimonious static and the *dynamic* IADs, the latter are of little predictive value and thus, suggest to select *ISD* above dynamic IADs as a predictor of well-being. In this paper, we argue and show that results on the predictive power of IADs may be largely influenced by the chosen modeling strategy. In the following, we focus on IADs and *ISD* as measures of variability, instability, or (affective) dynamics, and do only shortly discuss questions regarding the predictive utility above mean levels (see, e.g., Dejonckheere et al., 2019).

The findings of little predictive value added by dynamic IADs impose strong implications for researchers investigating the relation between IADs and a (time-invariant) criterion outcome of interest. Specifically, the argument implies that randomly reordering the measurements in a subject's time-series is not concomitant with a loss of information (with respect to the predictive power for explaining a time-invariant outcome)—as random reordering only affects the dynamic but not the static IADs (Ebner-Priemer et al., 2009)—thereby assuming that instability (as opposed to variability) and inertia are of minor importance. Consequently, researcher and participant burden could be reduced to a substantial degree, as less measurement occasions are necessary to reliably estimate person-specific means and intraindividual variances as compared to dynamic indicators (Du & Wang, 2018). Given the effort invested in conducting AA studies, this conclusion calls for further exploration of factors influencing the predictive power of IADs beyond static indicators, to inform the planning and analysis stages of AA studies. As discussed by Dejonckheere et al. (2019), reasons for the reported lack of predictive power of affect dynamics for psychological well-being might be located in current research practices and the authors list several potentially relevant methodological and measurement-related aspects. In the following, we will focus on the

aspect of the chosen modeling strategy, that is, data-analytical considerations.

Factors limiting (unique) predictive power of dynamic parameters

In the present article, we investigate how common modeling strategies impact inferences made on the (unique) role of IADs. We argue that the test of IADs' unique contributions beyond their static counterparts (i.e., affect variability) in the studies questioning the relevance of IADs was potentially overly conservative. Based on the mathematical relations between the different IADs and the static variability of univariate time series, we will shed light on different factors that determine the reliability and the predictive power of these measures. There are several factors that are of primary interest in this context, which will be discussed in the following. We start with an illustration of the relations between IADs and static variability of univariate time series and the resulting implications for their combined use as predictors.

Statistical overlap between IADs and static variability

IADs aim at capturing distinct aspects of (affective) trajectories across time. Instability in affect is commonly measured by the MSSD (von Neumann et al., 1941). The $MSSD_i$ of a univariate time-series of length T of a person i is calculated by taking the mean of squared differences between all consecutive measurements x_{it} and $x_{i(t+1)}$,

$$MSSD_i = \frac{1}{T-1} \sum_{t=1}^{T-1} (x_{it} - x_{i(t+1)})^2. \quad (1)$$

Evidently, if the affective states of a person exhibit large moment-to-moment fluctuations (high instability), the process is also characterized by a large overall variability. However, instability and its formalization as $MSSD_i$ differ from mere (static) variability as they take the temporal ordering of measurements into account. Hence, given a constant affective variability, the (in)stability of the underlying process may vary (see Jahng et al., 2008). Consider a stationary time-series following a first-order autoregressive (AR(1)) process with

$$x_{it} = c_i + \varphi_i * x_{i(t-1)} + \zeta_{it} \quad (2)$$

where the index i denotes the person, t the time point, φ_i the AR effect, and the regression residuals ζ_{it} , referred to as innovations, capture fluctuation in x_{it} that is not predicted by $x_{i(t-1)}$. A process with high serial dependency (high φ_i) is characterized by a

prolonged refractory period before it returns to its equilibrium state. Hence, once a higher or lower than usual emotional state is entered, longer time is needed before returning to the habitual level. The second element in (2) that drives the process's dynamics is the innovations, ζ_{it} , also referred to as perturbations or system shocks (Jongerling et al., 2015). In practice, all internal and external/contextual influences the model does not account for end up as unexplained variation in the innovations (Jongerling et al., 2015). In this model, differences between ISD_i^2 and $MSSD_i$ emerge from the serial dependency inherent in the process (Jahng et al., 2008), with

$$MSSD_i = 2 * ISD_i^2 (1 - \varphi_i). \quad (3)$$

For $\varphi_i = 0$, the $MSSD_i$ is perfectly correlated with ISD_i^2 , i.e., equaling $2 * ISD_i^2$. Analogously, in the AR(1)-model, the intraindividual variance can be expressed as (Hamilton, 1994):

$$ISD_i^2 = \frac{\sigma_{\zeta_i}^2}{(1 - \varphi_i^2)} \quad (4)$$

with $\sigma_{\zeta_i}^2$ denoting the i -th person's variance of the innovations ζ_{it} . Equation (4) illustrates that differences between ISD_i^2 and $\sigma_{\zeta_i}^2$ emerge if observations are non-independent (i.e., $\varphi_i \neq 0$), with the static variability of a time-series being a function of two dynamic components, φ_i and $\sigma_{\zeta_i}^2$. Given an AR-process, individual differences in static variability can be driven by both aspects, with identical levels of ISD_i^2 being generated by qualitatively different underlying dynamic processes. Entering (4) into (3) illustrates that the $MSSD_i$ of a time-series following an AR(1) process can also be expressed as a function of φ_i and $\sigma_{\zeta_i}^2$

$$MSSD_i = 2 \left(\frac{\sigma_{\zeta_i}^2}{(1 - \varphi_i^2)} \right) (1 - \varphi_i) = \frac{2\sigma_{\zeta_i}^2}{(1 + \varphi_i)} \quad (5)$$

Consequently, both person-specific ISD_i^2 (and thereby ISD_i) and $MSSD_i$ are a function of the first-order AR effect and the $\sigma_{\zeta_i}^2$ in an AR(1) model, however, differ regarding the impact of increased levels of φ_i . Note that, Equations (3)–(5) apply to the population parameters in an AR(1) model but not necessarily to the sample statistics.

Implications for interindividual differences in IADs beyond ISD

As evident from Equations (3)–(5) and given an AR(1) model holds, inter-individual differences in $MSSD$ and ISD can be expressed as combinations of differences in AR and the innovation variance. In

empirical applications, φ_i and $\sigma_{\zeta_i}^2$ may show different associations with an outcome of interest which can be obscured if using $MSSD$ or ISD as predictors of the outcome. That is, decomposing combined IADs (i.e., $MSSD_i$) into their underlying components may reveal otherwise masked relations with third-variable outcomes (Wang et al., 2012).

Assumption of constant innovation variance

Given the AR(1) model, if φ_i is constant across subjects, person-specific ISD_i^2 and $MSSD_i$ are both linear transformations of $\sigma_{\zeta_i}^2$ and thereby perfectly correlated. Similarly, when $\sigma_{\zeta_i}^2$ is constant across subjects, the only source of inter-individual differences in ISD_i^2 (and ISD_i) lies in subjects' differences in φ_i (see Equation (4)). Note that this contradicts the practice to estimate individual φ_i parameters and test their incremental contribution beyond ISD for the prediction of a time-invariant outcome under the assumption of a constant $\sigma_{\zeta_i}^2$. In essence, whenever φ_i or $\sigma_{\zeta_i}^2$ are truly constant across individuals, high redundancies between IADs and ISD ought to be expected, limiting the possibility of incremental contributions made by IADs beyond ISD .

We argue that the assumptions of a constant $\sigma_{\zeta_i}^2$ is potentially too restrictive in applied settings and should (when indicated) be discarded (Jongerling et al., 2015). Besides capturing potentially relevant information regarding inter-individual differences in dynamics, falsely ignoring inter-individual differences in innovation variances negatively affects the recovery of individual φ_i estimates (Asparouhov et al., 2018; Jongerling et al., 2015), thereby limiting their predictive utility. Nevertheless, the possibility of inter-individual differences in $\sigma_{\zeta_i}^2$ was disregarded in studies questioning the predictive utility of IADs (Dejonckheere et al., 2019; Houben & Kuppens 2020; Wendt et al., 2020) as well as in simulation studies comparing the performance of person-specific and multilevel estimation approaches (Liu, 2017, 2018; Liu et al., 2021). Though closely linked to ISD in stationary time series, innovation variances can be regarded as conceptually different from ISD (Jongerling et al., 2015), as they capture the variation in emotional states that cannot be predicted by the previous states and thereby provide a measure of a time series' instability.

Two-step modeling strategies

A common modeling approach is to recover and save individual IAD estimates in a first step and subsequently use the saved values for the prediction of an

outcome in a second step, therefore referred to as *two-step approach*. This procedure comes with the shortcoming of treating estimates as observed variables and ignoring estimation uncertainty. The two-step approach was found to produce negatively biased regression parameter estimates and low coverage rates when using empirical bayes estimates of random effects from a multilevel model as predictor variables (Liu et al., 2021). The amount of bias was related to the reliability of individual parameter estimates and disappeared using a *one-step approach*. In a one-step approach, individual parameter estimation and outcome prediction are combined within the same model. Choosing a two-step or a one-step approach can lead to differing conclusion regarding the unique predictive utility of IADs, such as emotional inertia in the prediction of depressive symptoms (Wenzel & Brose, 2023). Nevertheless, the two-step approach was still widely adopted in recent studies of affect dynamics (e.g., Burchert et al., 2021; Nowak & Lincoln, 2023; Panaite et al., 2020; Wang et al., 2012). Given that differences between the two-step and one-step approaches arise from differences in handling uncertainty in IAD estimates, factors influencing recovery of individual IADs will be discussed in the following.

Individual parameter reliability

Several factors are of primary interest in this context. Du and Wang (2018) showed that the static components (i.e., intraindividual means and variances) of univariate time-series are generally better recovered than dynamic parameters (i.e., φ_i and $MSSD_i$). Regarding their predictive utility, this suggests that the latter are potentially more strongly affected by regression attenuation, namely the phenomenon that regression weights of low-reliability predictors are underestimated in simple regression and potentially distorted in multiple regression models.

Number of time points and persons

The reliability of individual φ_i estimates was found to decrease with decreasing numbers of time points (T) available per person (Du & Wang, 2018), with a substantially smaller impact of T on the estimates of individual means and ISD_i . In the $N=1$ setting, a general recommendation for the estimation of φ_i in univariate time-series is a minimum of $T=50$, however, in two-level AR(1) models less T may be sufficient as the person sample size (N) increases (Liu, 2017, 2018; Schultzberg & Muthén, 2018). Furthermore, prior results are based on the restrictive assumption of a

constant innovation variance across subjects (see Liu, 2017, 2018). Investigating multilevel AR models with random innovation variances, Schultzberg & Muthén (2018) find compensating effects of N and T as well as higher T requirements when person-specific φ_i and $\ln(\sigma_{\zeta_i}^2)$ serve as predictors as compared to serving as outcomes of external variables.

Missing values

Closely related to the number of available time points is a problem frequently observed in AA studies, namely missing data. When the number of T per individual is unbalanced, reliabilities of individual φ_i estimates will differ across individuals. In this case, multilevel models, which borrow information from the other cluster-level units (here, individuals), may be advantageous. Low compliance rates are especially relevant for the estimation of dynamic IADs (i.e., φ_i and $MSSD_i$), as a missing value at time t is also lacking as lagged predictor at $t+1$.

Measurement error

Lastly, the presence of measurement error in the observed time series may play an important role with respect to the IADs' reliability and the power to detect predictive effects of IADs. Measurement error is common in psychological data and the presence of measurement error in an observed times series may lead to the underestimation of individual φ_i (Schuurman et al., 2015). Similarly, reliabilities of individual IADs (especially φ_i) were found to be more sensitive to the presence of measurement error as compared to static intraindividual means and variances (Du & Wang, 2018). Consequently, accounting for measurement error in the observed variables when estimating IADs is tantamount. To this end, approaches combining latent variable modeling with multilevel time-series modeling can be used. Here, we propose to use one-step, multilevel models that specify φ_i effects on the level of within-person latent factors (see, e.g., Asparouhov et al., 2018; Schuurman & Hamaker, 2019). These models additionally provide the advantage that, to account for inter-individual differences in trait levels, variables are centered on the latent cluster-means (Asparouhov et al., 2018; Asparouhov & Muthén, 2019; Hamaker & Grasman, 2014) instead of relying on a lagged variable centered at the observed sample mean. The latter approach is known to produce a bias in the φ_i estimates that depends on the level of AR as well as T , termed Nickell's bias (Nickell, 1981). For "reasonably large values of T " (Nickell, 1981, p.1422), Nickells' bias is approximated

by $\frac{-(1+\varphi)}{T-1}$, which suggests a higher negative bias in positive autoregressive effects that decreases with higher T .

Current study

Taken together, studies questioning the predictive utility of measures of temporal dependency potentially underestimated the true relations with outcome variables (Dejonckheere et al., 2019; Houben & Kuppens, 2020; Wendt et al., 2020) due to low reliability in individual parameter estimates, the restrictive assumption of constant innovation variances, and ignoring measurement error in the time-series. Using data on negative affect and depressive symptoms, we illustrate the impact of the modeling choices discussed above on the substantive conclusions regarding the predictive power of IADs in an empirical example. In three simulation studies, we examine the factors outlined above for their (combined) impact on estimation accuracy and power to detect true, non-null relations between dynamic IADs and a time-invariant external criterion (EC) under different modeling strategies and AA study design factors (e.g., number of subjects and time points). Based on the presented relations between IADs and *ISD*, the interplay of individual φ_i and $\sigma_{\zeta_i}^2$ is expected to affect the predictive utility of IADs when testing for unique contributions above *ISD*.

Modeling strategies

We compared two commonly applied *two-step* approaches for regressing time-invariant outcomes on (multiple) IAD estimates with two alternative *simultaneous* or *one-step* approaches.

Two-step approaches

Person-specific two-step (PS_{2-step})

In this approach, individual parameters (such as autoregressive effects φ_i and innovation variances (transformed to $\ln[\sigma_{\zeta_i}^2]$) are estimated using separate linear regression models for each individual in a first step. Analogously, *ISD* and *MSSD* are also estimated from $N=1$ models. In a second step, the obtained individual point estimates are used as predictors in (multiple) regression models to predict a time-invariant outcome.

Multilevel two-step (ML_{2-step})

This approach is similar to PS_{2-step} in that it is a two-step approach in which individual estimates are

obtained in a first step, with the difference that individual φ_i effects are estimated as random effects in a two-level AR(1) model, e.g., using restricted maximum likelihood estimation as implemented in the *lme4*-package (Bates et al., 2015). This approach was shown to better compensate for low number of time points by leveraging on the information from other subjects in the multilevel model (e.g., for a comparison of PS_{2-step} and ML_{2-step}, see Liu, 2017, 2018). However, the model assumes a constant innovation variance across individuals, which may be overly restrictive in applied settings (Jongerling et al., 2015). Falsey disregarding interindividual differences in innovation variances was shown to negatively affect the recovery of φ_i estimates (Asparouhov et al., 2018; Jongerling et al., 2015). Individual φ_i estimates will further be affected by Nickels' bias caused by centering the lagged predictor on the observed mean. Note that for ML_{2-step}, outcome prediction models involving innovation variances as predictor are generally (and in the following) not considered, as they are assumed to be constant.

One-step approaches

Bayesian one-step (BAY)

To overcome limitations of PS_{2-step} and ML_{2-step}, we suggest using a one-step approach implemented using Bayesian Markov-Chain Monte-Carlo (MCMC) estimation. In this approach (1) all model parameters (including innovation variances) are estimated as person-specific in a multilevel AR(1) model, (2) variables are latent-mean centered, and (3) the respective random effects are included as predictors for the time-invariant outcome within the same model (see, e.g., Asparouhov et al., 2018). We estimated these models via MCMC sampling using the free software Stan (Carpenter et al., 2017) from R, via the interface provided in the *rstan*-package (Stan Development Team, 2023). An advantage of using Stan is that it allows researchers to flexibly set up user-defined models, for instance, two-level AR(1) models with random innovations which simultaneously estimate intraindividual variances.

Latent bayesian one-step (BAY_{Lat})

All of the above modeling strategies disregard potential measurement error in the time-series, with the dynamics being modeled for observed variables or using a composite score across multiple manifest items. As an extension of the BAY approach, we employed BAY_{Lat}, which is a two-level dynamic factor analysis (two-level DAFS) model (Asparouhov et al.,

2018; also see Molenaar, 1985; Zhang & Nesselroade, 2007). Assume that we have a k – dimensional vector Y_{it} containing k observed indicator variables for person i at time t . The observed variables are each decomposed into their latent person-specific mean (μ_i) across time and a latent-person-mean-centered, time-specific deviation, Y_{it}^W :

$$Y_{it} = \mu_i + Y_{it}^W \quad (6)$$

with μ_i and Y_{it}^W being the respective k – dimensional vectors. At both the time invariant between-person as well as the time-specific within-person level, a common latent factor is measured by the k indicators, with

$$Y_{it}^W = \Lambda_W \eta_{it}^W + \varepsilon_{it}^W \quad (7)$$

$$\mu_i = \alpha + \Lambda_B \eta_i^B + \varepsilon_i^B \quad (8)$$

with α being a k – dimensional vector of intercept parameters, Λ_W , and Λ_B being k – dimensional vectors of factor loading parameters on the within- and the between-level, respectively, and ε_{it}^W , and ε_i^B being k – dimensional vectors of (measurement) error variables, with $\varepsilon_{it}^W \sim N(0, \sigma_{\varepsilon W k}^2)$ being serially uncorrelated, and $\varepsilon_i^B \sim N(0, \sigma_{\varepsilon B k}^2)$. Note that for simplicity we here assume one common latent factor on the within-level (η_{it}^W) and the between-level (η_i^B), respectively. For more general formulations including several factors see, for instance, Asparouhov et al. (2018). For identification purposes, 1) the first loading parameter in Λ_W and Λ_B is set to one (or the variance of the respective factor is set to equal one) and 2) one intercept parameter α_k is set to zero (or the mean of the latent μ_i is set to zero). Note that the between-level factor structure in Equation (8) can be adapted to yield different factor structures. For instance, to model indicator-specific stable trait variables, we could refrain from specifying a between-level measurement equation and estimate a variance-covariance matrix for the separate μ_i . The autoregressive process is modeled for the within-level latent variables η_{it}^W , that is

$$\eta_{it}^W = \varphi_i \eta_{i(t-1)}^W + \zeta_{it} \quad (9)$$

with φ_i being the person-specific autoregressive effect and ζ_{it} denoting the dynamic residual or innovation, with person-specific innovation variances, $\zeta_{it} \sim N(0, \sigma_{\zeta i}^2)$.

At the between-person level, the latent variables and person-specific parameters η_i^B (or μ_i), φ_i , and the logarithm of $\sigma_{\zeta i}^2$ (i.e., $\ln[\sigma_{\zeta i}^2]$) are assumed to follow a multivariate normal distribution. Additionally, the estimation of first-order correlations between all person-specific parameters and the time-invariant

outcome is directly incorporated into the model, which can be used to obtain the model-implied (standardized) prediction parameter estimates. To support empirical researchers in applying a latent-variable one-step approach using Bayesian MCMC techniques in Stan, we provide a comprehensive set of ready-to-use materials. These include code templates for immediate implementation, user-friendly *R* functions tailored to handle common challenges such as missing data and overnight lags, and an explanation of the underlying Stan code. Our goal is to make it as easy as possible for researchers to adopt and apply these models in their own work. In addition, full analysis scripts for reproducing the simulation studies and the empirical example are available via the Open Science Framework at: <https://osf.io/bj7fq/>.

Empirical example

To illustrate that findings on the predictive utility of dynamic IADs can vary strongly with the choice of a data-analytic approach and the affect variability measure used as covariate, we reanalyzed an empirical dataset first published in Koval et al. (2013) and later made publicly available by Dejonckheere et al. (2019) as part of their meta-analysis. In this AA study, inter-individual differences in IADs of negative affect and their relationship with depressive symptoms measured using the CES-D (Radloff, 1977) were investigated for $N=94$ subjects (after data exclusion)¹ observed on 10 measurement occasions per day across seven days. For the reanalysis, we quantified negative affect either as the mean score or the latent factor measured by participants' responses to three negative affect items (*sad*, *anxious*, and *angry*)². The number of available observations varied across subjects from 43 to 73 ($Mdn=61$). In both previous studies, data were analyzed by adopting a two-step estimation procedure for regressing depressive symptoms on individual IADs (e.g., *ISD*, *AR*, *MSSD*) of negative affect. In Koval et al. (2013) individual *AR* estimates were derived from person-specific $N=1$ models (*PS_{2-step}*) and in Dejonckheere et al. (2019) as random effects from a

¹We excluded the data of one subject that showed no variation in at least one of the three included negative emotion items.

²An additional item (*depressed*) was excluded from the analyses as we encountered model convergence issues when using all four items as indicators of a common latent factor. Results using observed mean scores across all four indicators can be found in the *supplementary material* in Table S14. General trends reported for the three-indicator solution remain intact, with an increased amount of variance explained in the outcome when *depressed* was included, probably due to the substantial overlap with the criterion (e.g., depressive symptoms).

Table 1. Results of separate multiple regression models regressing depressive symptoms (CES-D) on indicators of affect dynamics (IADs) of negative affect as a function of estimation approach and affect variability measures ($N=94$).

IAD	Estimation	Controlling for ISD					Controlling for $\ln(\sigma_i^2)$				
		β_{IAD}	95% CI	β_{ISD}	95% CI	R^2	β_{IAD}	95% CI	$\beta_{\ln(\sigma_i^2)}$	95% CI	R^2
<i>MSSD</i>	PS _{2-step}	-0.06	[-0.42, 0.29]	0.59	[0.23, 0.94]	0.282	—				
<i>AR</i>	ML _{2-step}	0.08	[-0.12, 0.29]	0.49	[0.28, 0.69]	0.287	—				
	PS _{2-step}	0.07	[-0.12, 0.25]	0.51	[0.32, 0.70]	0.285	0.17	[-0.01, 0.34]	0.50	[0.32, 0.68]	0.305
	BAY	0.14	[-0.10, 0.38]	0.47	[0.27, 0.67]	0.318	0.21	[-0.01, 0.42]	0.47	[0.30, 0.63]	0.336
	BAY _{Lat}	—					0.31	[0.10, 0.51]	0.56	[0.38, 0.72]	0.401

Note. β : standardized regression weight; *MSSD*: mean squared successive differences; *AR*: first-order autoregressive effect; *ISD*: intra-individual variance; $\ln(\sigma_i^2)$: log-transformed innovation variance; PS_{2-step} = person-specific two-step approach; ML_{2-step}: multilevel two-step approach deriving individual *AR* estimates as random effects from a two-level AR(1) model; BAY: Bayesian one-step approach using mean of observed indicators; BAY_{Lat}: Bayesian one-step approach with inclusion of a measurement model to account for measurement error in indicators of negative affect. For BAY models, R^2 was calculated as the proportion of the variance in predicted values and the total variance in depressive symptoms. Bold cells highlight significant prediction parameter estimates.

two-level model (ML_{2-step}). Note that in the re-analysis, we did not account for unequal measurement intervals due to missed beeps or overnight-lags to mimic the analysis approach administered in Koval et al. (2013)³.

Results

Two-step individual ($N=1$) AR(1) modeling and multilevel modeling approaches with fixed innovation variance

First, we calculated individual IADs per person (PS_{2-step}), namely *ISD*, *MSSD*, and *AR*, the latter additionally using the multilevel ML_{2-step} approach. Subsequently, we examined the distribution of person-specific innovation variances (σ_i^2) derived from the $N=1$ AR(1) models, which ranged between 0.04 to 5.70 ($M=1.06$, $SD=0.89$). Note that these non-negligible interindividual differences suggest a misspecification of the two-level AR(1) with constant σ_i^2 . Furthermore, graphical inspection suggests that the innovation variances follow a log-normal distribution (see the extended report on the respective OSF-repository).

Using these individual IADs, bivariate associations between depressive symptoms (CES-D) and IADs were all positive and significant (*ISD*: $r=0.53$, 95% CI = [0.36, 0.66]; *MSSD*: $r=0.45$, 95% CI = [0.26, 0.59]; *AR* using PS_{2-step}: $r=0.25$, 95% CI = [0.04, 0.42], *AR* using ML_{2-step}: $r=0.34$, 95% CI = [0.14, 0.51]). Visual inspection of the bivariate associations between CES-D and IADs did not reveal any violations of the

linearity assumption (see Figure S10 in the online supplement). This was also supported by the results of additional regression models which we ran to examine a potential non-linear (i.e., quadratic or cubic) relationship between IADs and CES-D (see section E in the online supplement for the full model results). We further tested the robustness of the linearity assumption by running B-spline regression models with the breakpoints of the piecewise polynomials set to the second, third, and fourth quantile of the respective predictors' distribution, as well as to the first, third, fifth, seventh, and ninth decile (i.e., 5 breakpoints). In subsequent model comparisons, comparing the simple linear models with the results of the spline regressions also indicated that the linear term was sufficient ($ps \geq 0.357$) to describe the bivariate associations. Although we did not find any significant contributions of the higher-order polynomial terms, we recommend applied researchers to consider the possibility of non-linear relations with the outcome.

Furthermore, adhering to the notion that the *dynamic* IADs should not be used as predictors in isolation but tested for their unique contributions to the outcome prediction beyond static IADs (e.g., *ISD*), we ran multiple regression analyses controlling for linear overlap between IADs (i.e., *AR* and *MSSD*) and *ISD*. The full results including point estimates and CIs of standardized regression weights as well as the explained variance⁴ in CES-D are summarized in Table 1.

As expected, given the high correlation between *ISD* and *MSSD* ($r=0.87$), after controlling for *ISD*, the standardized regression weight of *MSSD* was lower and non-significant ($\beta = -0.06$). Similarly, when controlling for *ISD*, we observed no significant contributions of the *AR* predictor (PS_{2-step}, $\beta = 0.07$; ML_{2-step},

³For BAY, the single-indicator one-step approach, we checked that the main finding (i.e., significant unique contribution of negative affect inertia beyond the log innovation variance in the prediction of depressive symptoms) holds, when missing beeps are imputed, as well as when overnight-lags are accounted for by removing the last observation of a day as lagged predictor of the subsequent measurement. The results were found to be robust, regardless of the implemented corrections for unequal distances in time intervals between observations (see the extended report on the OSF-repository).

⁴For BAY, R^2 was calculated by standardizing the outcome and predictors and deriving R^2 as the posterior mean of the ratio of variance in predicted values to observed variance in depression scores.

$\beta = 0.08$). These results replicate the results which led authors to conclude that accounting for the temporal dimension in affect dynamics may provide little predictive utility over simple affect variability.

One-step modeling approaches

As a direct comparison, we examined unique contributions of *AR* over *ISD* when adopting a one-step approach (i.e., *BAY*). This resulted in a higher but non-significant standardized regression estimate of *AR* as predictor, $\beta = 0.14$. Based on the expected associations between *AR* and *ISD* on the within-level given in (4), combining them as joint predictors in regression models might mask potential relations that exist between an outcome and *AR*, as well the outcome and individual $\sigma_{\zeta_i}^2$. Indeed, when including the log innovation variance as a second predictor (instead of *ISD*) and using *BAY*, unique contributions of the *AR* predictor increased, $\beta = 0.21$, 95% CI $[-0.01, 0.42]$. A similar trend could be observed using *AR* and the log-transformed $\sigma_{\zeta_i}^2$ derived from *PS*_{2-step} ($\beta = 0.17$, 95% CI $[-0.01, 0.34]$). Thus far, all of the above analyses relied on a composite score of negative affect, thereby disregarding potential measurement error in the time-series. When *BAY* was extended to a multiple-indicator model (*BAY*_{Lat}), the *AR* prediction parameter further increased to $\beta = 0.31$, 95% CI $[0.10, 0.51]$.

To highlight the relevance of this finding we summarize that, given the same set of data, applying a common two-step approach (*PS*_{2-step} or *ML*_{2-step}) and controlling for *ISD* not only lead to the rejection of the predictive role of negative affect inertia but also to less explained variance in depressive symptoms ($R^2 = 0.29$). Using *BAY* while controlling for $\ln(\sigma_{\zeta_i}^2)$ explained an additional 5% of variance in depressive symptoms. When additionally accounting for measurement error (*BAY*_{Lat}) R^2 further increased to 0.40. These results illustrate that conclusions on the unique contribution of negative affect inertia as a predictor of depressive symptoms can vary strongly based on the applied analysis approach. In the three following simulation studies we aim to shed more light on the factors that play a role in the emergence of this result pattern.

Simulation studies

In three simulation studies, we aim to illustrate the effect of (1) the choice of an IAD as predictor, (2) the chosen modeling strategy, (c) different design factors in AA studies, and (d) the pattern of intercorrelations between the individual parameters and the outcome on

the estimation accuracy and power for predicting an external outcome. Each simulation study focuses on one or several of the different issues discussed above and illustrated in the real data application. The results of the simulations serve to illustrate that a widely adopted approach to test the relation between IADs and outcome variables is far from ideal as it (a) fails to account for interindividual differences in innovation variances when using multilevel models, potentially leading to biased *AR* estimates, (b) models the dynamics for manifest composite scores, disregarding measurement error, (c) treats estimates as observations and thereby ignores uncertainty in these estimates, although these are usually not estimated reliably, with (d) their reliability depending on different design factors. We systematically investigate the effect of these design factors, that is, the presence of measurement error and missing data, the length of the time series, modeling fixed vs. random innovation variances, and using a one-step vs. a two-step approach.

Note that the discussed shortcomings and design factors remain relevant even if the assumed IAD-outcome relationship was non-linear. Our simulation studies do not (and cannot) address the question of an adequate functional form of the relationship in the outcome model, which should always be thoroughly considered by the researcher for a given question and variables at hand.

In all three simulation studies, we generated data from a two-level AR(1) model with random innovation variances, with non-null relations between individual *AR* effects and/or innovation variances with an external criterion (EC). We considered this specific scenario, as it appears to be the most common parameterization of the multilevel AR(1) model with the benefit of φ_i and $\ln(\sigma_{\zeta_i}^2)$ being the primary model parameters (next to individual trait levels). For simplicity, we assume linear relations between the time-invariant criterion and the individual AR(1) parameters. In simulation study I, we test the data-analytic strategies commonly used in applied studies, that is, the two-step approaches as described above, under varying population parameter constellations. Here our motivation was two-fold. One, to examine the accuracy of regression estimates when using dynamic IADs as predictors, and further, to illustrate the effect that complex, non-linear relations between dynamic IADs and static variability in stationary time-series may have on their simultaneous use as predictors. Note that we purposely varied the relations between the outcome and the *dynamic* IADs (i.e., φ_i and $\ln[\sigma_{\zeta_i}^2]$) to inspect how controlling for the *static* *ISD*

Table 2. Simulation setup in simulation studies I, II, and III.

Study	Factors held constant	Varied factors			
		Data generation		Analytic strategy	
		Parameter	# Conditions	Prediction model	Estimation approach
I	$N = 100$ $T = 70$	$\rho_{EC, \varphi} = -0.3, 0.3$	2	Solo φ_i	PS _{2-step} , ML _{2-step}
		$\rho_{EC, \ln(\sigma_{\zeta_i}^2)} = -0.3, 0.3$	2	Solo $\ln(\sigma_{\zeta_i}^2)$	PS _{2-step}
		$\rho_{\varphi, \ln(\sigma_{\zeta_i}^2)} = -0.3, 0, 0.3$	3	Solo <i>ISD</i>	PS _{2-step}
		$\tau_{\ln(\sigma_{\zeta_i}^2)}^2 = 0.1, 0.3, 1$	3	Solo <i>MSSD</i>	PS _{2-step}
		$\gamma_{\ln(\sigma_{\zeta_i}^2)} = -0.5, 0, 1$	3	φ_i controlled for $\ln(\sigma_{\zeta_i}^2)$ φ_i controlled for <i>ISD</i>	PS _{2-step} , PS _{2-step} , ML _{2-step}
II	$\rho_{EC, \varphi} = 0.3$ $\rho_{EC, \ln(\sigma_{\zeta_i}^2)} = 0.3$ $\rho_{\varphi, \ln(\sigma_{\zeta_i}^2)} = 0$ $\tau_{\ln(\sigma_{\zeta_i}^2)}^2 = 0.3$ $\gamma_{\ln(\sigma_{\zeta_i}^2)} = 0$	$N = 70, 100, 200$	= 108	<i>MSSD</i> controlled for <i>ISD</i>	PS _{2-step}
		$T = 50, 70, 100, 200$		φ_i controlled for $\ln(\sigma_{\zeta_i}^2)$	BAY/BAY _{Lat} , PS _{2-step}
		Miss = 0%, 20%(l), 20%(r)		φ_i controlled for <i>ISD</i>	BAY/BAY _{Lat} , PS _{2-step}
		Rel = 1, 0.8			
III	$\rho_{EC, \varphi} = 0.3$ $\rho_{EC, \ln(\sigma_{\zeta_i}^2)} = 0.3$ $\rho_{\varphi, \ln(\sigma_{\zeta_i}^2)} = 0$ $\tau_{\ln(\sigma_{\zeta_i}^2)}^2 = 0.3$ $\gamma_{\ln(\sigma_{\zeta_i}^2)} = 0$	$N = 70, 100, 200$	3	Solo φ_i	BAY, PS _{2-step} , ML _{2-step}
		$T = 50, 70, 100, 200$	4	φ_i controlled for μ_i	BAY, PS _{2-step} , ML _{2-step}
		$\rho_{EC, \mu} = -0.3, 0, 0.3$	3		
		$\rho_{\mu, \varphi} = -0.3, 0, 0.3$	3		
			= 108		

N: person sample size; *T*: number of time points per subject; $\rho_{EC, \varphi}$: bivariate association between external criterion (EC) and autoregressive effects; $\rho_{EC, \ln(\sigma_{\zeta_i}^2)}$: bivariate association between EC and log innovation variances; $\rho_{\varphi, \ln(\sigma_{\zeta_i}^2)}$: random effect correlation between individual autoregressive effects and log innovation variances; $\tau_{\ln(\sigma_{\zeta_i}^2)}^2$: between-level variance of individual log innovation variances; $\gamma_{\ln(\sigma_{\zeta_i}^2)}$: fixed effect of log innovation variance; Miss: missing data conditions of 0% (no missing values) and 20% of N^*T set as missing by either cutting of the last values of a subjects time-series (20%[l]) or distributing missing values randomly within a subjects time-series (20%[r]), Rel = Within-level reliability of composite score across two indicator variables. Estimation approaches: PS_{2-step} = person-specific two-step approach, ML_{2-step} = individual φ_i estimates derived from two-level AR(1) model using restricted maximum likelihood estimation with observed mean centering, BAY = Bayesian latent variable one-step approach, BAY_{Lat} = Bayesian multiple-indicator model with correction for measurement error.

would affect inferences made on the unique predictive role of the dynamic IADs.

In simulation study II, we follow-up on the results of simulation study I (i.e., negative bias in prediction parameters of φ_i) and illustrate how limitations of the employed two-step approaches can be overcome by implementation of a one-step approach, considering several relevant design factors of AA studies.

In a third simulation study, we illustrate that bias in prediction parameters of φ_i can vary depending on the relations between mean levels (μ_i), φ_i , and a time-invariant criterion when using ML_{2-step}.

Details on the conditions investigated in each of the three simulation studies are described in the respective sections below and an overview on the investigated factors in the simulation studies is given in Table 2.

Data generation

For all simulation conditions in the simulation studies, 500 datasets were generated based on a two-level AR(1) model with random innovation variances (Jongerling et al., 2015), corresponding to the decomposition in Equation (6) with the autoregressive process according to

$$Y_{it}^w = \varphi_i Y_{i(t-1)}^w + \zeta_{it} \quad (11)$$

where Y_{it}^w , the time-specific deviation of the observed

value Y_{it} of a person i at time t from the individual trait level μ_i , is regressed on its' preceding value $Y_{i(t-1)}^w$. The residual term ζ_{it} is generated with a person-specific innovation variance as $\zeta_{it} \sim N(0, \sigma_{\zeta_i}^2)$. Person-specific parameters, that is, individual trait levels (μ_i), autoregressive effects (φ_i), log innovation variances ($\ln(\sigma_{\zeta_i}^2)$), and a time-invariant external criterion (EC_i), were drawn from a multivariate normal distribution:⁵

$$\begin{bmatrix} \mu_i \\ \varphi_i \\ \ln(\sigma_{\zeta_i}^2) \\ EC_i \end{bmatrix} \sim MVN \left(\begin{bmatrix} \gamma_{\mu} = 2 \\ \gamma_{\varphi} = 0.3 \\ \gamma_{\ln(\sigma_{\zeta_i}^2)} \\ \gamma_{EC} = 2 \end{bmatrix}, \begin{bmatrix} \tau_{\mu}^2 & & & \\ 0 & 0.02 & & \\ 0 & \tau_{\varphi, \ln(\sigma_{\zeta_i}^2)} & \tau_{\ln(\sigma_{\zeta_i}^2)}^2 & \\ 0 & \tau_{\varphi, EC}^2 & \tau_{\ln(\sigma_{\zeta_i}^2), EC}^2 & \tau_{EC}^2 \end{bmatrix} \right). \quad (12)$$

The average AR effect (γ_{φ}) was set to 0.3 with $\tau_{\varphi}^2 = 0.02$ to keep individual φ_i effects mainly positive. To ensure stationarity, sampling was repeated if any $|\varphi_i| > 1$ were generated. Consistent with Jongerling et al.

⁵Note that correlations between μ_i and the other random effects as well as the outcome variable were set to zero to reduce model complexity and minimize confounding effects in the present simulation study. In practical applications the association between trait levels and the external outcome variable is oftentimes of substantive interest. However, the aim of the present study is to investigate associations and overlap between IADs and the intraindividual variance, such that trait levels are not of primary interest. To nevertheless keep the estimated multilevel time series model as realistic as possible we did include inter-individual differences in trait levels in the data generating model and the estimated multilevel model.

(2015), the level of between-subject variance (τ_μ^2) was chosen to match the average within-subject variance, $\tau_\mu^2 = E(1SD^2)$, resulting in an expected intraclass correlation of $\rho_{ICC} = 0.5$. Additionally, τ_{EC}^2 was chosen to equal τ_μ^2 in each condition. The average within-subject variance, $E(1SD^2)$, was derived from simulated data sets of 10^6 subjects per condition. The expected standardized prediction parameter values for prediction models including *AR* and (log) innovation variances as predictor of the *EC* were calculated based on their bivariate associations. Expected prediction parameter values for *ISD* and *MSSD* were derived from simulated data sets of 10^6 subjects where we calculated individual *ISD* and *MSSD* according to (4) and (5), and then calculated the expected linear overlap based on the observed bivariate associations.

Investigating the effect of measurement error, the data generating model is a two-level dynamic factor analysis (two-level DAFS) model according to Equations (6)–(9). Specifically, we created two manifest indicator variables ($k = 1, 2$) by adding measurement error terms ε_{kit}^W to each observation. For simulation purposes, we assumed a common between-level (trait) factor μ_i across indicators and loading parameters on both levels were set to one (see Equations (7)–(8)). Based on the expected within-person variance of the latent process, $E(1SD^2)$, the error variance $\sigma_{\varepsilon Wk}^2$ was set to 0.63, yielding a within-level composite reliability (omega [ω], Geldhof et al., 2014) of 0.8.

Convergence diagnostics and performance evaluation

Estimation of one-step models in Stan was restricted to the first 200 generated data sets in each simulation condition to compromise between estimation time and reliability of performance indices (compare Li et al., 2022)⁶. MCMC estimation was run using two chains with the number of iterations (with 50% of iterations used as warm-up) set to 5.000 for *BAY*, and 7.500 for *BAY_{Lat}*.

Model convergence was monitored based on the cutoff criteria of $\hat{R} < 1.01$, and bulk-ESS and tail-ESS > 200 (Vehtari et al., 2021). Replications not meeting the criteria were rerun with twice the number of iterations and ultimately removed if criteria were still not

⁶In Li et al. (2022) for a more complex VAR(1) model, model performance was evaluated based on 100 replications per condition, and they reported that no substantial differences in model results could be observed when running additional 400 replications in one of their conditions. Therefore, we assume that a reasonable stability in model results was achieved when running 200 replications in each condition.

met. Conditions with convergence ratios below 0.50 were removed from further analyses. See Section A of the [supplementary material](#) for an overview on chosen parameter priors used in the MCMC estimation.

Model performance was assessed based on the following criteria. In each replication, the reliability of person-specific parameters was quantified as the squared correlation between true scores and their estimated values (r^2). For each condition, we then calculated the median reliability across replications. Empirical power to detect non-null relations was defined as the relative frequency of replications in which a parameter's 95% confidence or credibility interval (CI) did not cover zero. 95% coverage rates were calculated as the percentage of replications in which a parameter's 95%-CIs covered the true data-generating value. The average relative bias was calculated as the ratio of the difference between a parameter's average estimate across replications and its true value relative to the true value. For models estimated by MCMC, the posterior mean was used as point estimate and coverage as well as power calculations were based on the 2.5% and 97.5% quantiles of the posterior distribution.

Simulation study I – limitations of two-step strategies

As illustrated in the empirical example, controlling for linear overlap between dynamic IADs and static variability can affect conclusions regarding the role of the dynamic IADs in the prediction of time-invariant outcomes. We aim to shed light on this phenomenon by considering multiple population parameter constellations which, based on the presented mathematical relations, define the statistical overlap between IADs. In addition, we illustrate the (poor) performance of the common two-step modeling approaches under varying population parameter conditions. To this end, in simulation study I, we investigate the linear overlap between IADs and their relations with an *EC* which was generated to be linearly associated with φ_i and $\ln(\sigma_{\zeta i}^2)$. Based on (3) and (5) parts of the main effects of φ_i and $\ln(\sigma_{\zeta i}^2)$ in predicting the *EC* should be reflected in a linear overlap between *MSSD*, *ISD*, and the *EC*. Specifically, based on (3) we expect *ISD* to show large overlap with the *EC* in scenarios commonly observed in studies of affect dynamics, that is, bivariate associations between (dynamic) affect indicators and psychological health constructs of the same direction (Houben et al., 2015), such as depressive symptoms being positively correlated with more variable, instable, and inert negative affect (e.g., Bos et al.,

2019; Koval et al., 2013). When the associations between the EC and φ_i as well as the EC and $\ln(\sigma_{\zeta_i}^2)$ are *unidirectional* (both of the same sign), we assume that controlling for static variability will reduce incremental contributions of φ_i for predicting the EC, and lower power for detecting these effects. Nevertheless, to capture a range of different dynamic process and outcome variable constellations, we also consider the possibility of opposed main effects' directions. Critically, we intend to examine potential shortcomings of the two-step procedures, namely negative bias in and low coverage of regression weights of IADs under common AA conditions. To this end, in a fully crossed design, we generated data for $N=100$ and $T=70$ (median across 15 studies in Dejonckheere et al., 2019) and varied the following factors:

- a. Direction of associations between external criterion and inertia ($\rho_{EC,\varphi} = -0.30, 0.30$) and log innovation variance ($\rho_{EC,\ln[\sigma_{\zeta_i}^2]} = -0.30, 0.30$),
- b. Random effect correlation, $\rho_{\varphi_i,\ln[\sigma_{\zeta_i}^2]} = -0.30, 0, 0.30$,
- c. Between-person differences in $\sigma_{\zeta_i}^2$, reflected in the random effect variance of log innovation variances ($\tau_{\ln[\sigma_{\zeta_i}^2]}^2$) in margins of 0.1, 0.3, and 1. Higher levels of $\tau_{\ln[\sigma_{\zeta_i}^2]}^2$, were expected to result in lower reliability of φ_i estimates when using $ML_{2\text{-step}}$, due to misspecification. In contrast, increased $\tau_{\ln[\sigma_{\zeta_i}^2]}^2$ yields less linear overlap between φ_i and $ISD(\rho_{\varphi_i,ISD})$, consequently affecting power of β_{φ_i} when controlling for ISD .
- d. Average log innovation variance ($\gamma_{\ln[\sigma_{\zeta_i}^2]}$) in levels of $-0.5, 0$, and 1 .

Each generated data set was then analyzed using dynamic IADs and/or static variability as single and/or joint predictors of the external outcome using the person-specific two-step estimation approach ($PS_{2\text{-step}}$). In total, seven prediction models were employed which comprised four simple regressions (φ_i , $\ln(\sigma_{\zeta_i}^2)$, ISD , and $MSSD$ as single predictors) as well as three multiple regression models. The latter involved entering φ_i alongside $\ln(\sigma_{\zeta_i}^2)$ or ISD , and $MSSD$ next to ISD as joint predictors to examine unique contributions of the dynamic parameters. Prediction models that involved φ_i were additionally estimated using the multilevel two-step approach ($ML_{2\text{-step}}$).

Results

As we observed no effect of varying the average (log) innovation variance, $\gamma_{\ln(\sigma_{\zeta_i}^2)}$, on individual parameter

reliabilities or performance indices of prediction parameters, we only present results for $\gamma_{\ln(\sigma_{\zeta_i}^2)} = 0$. Further, regarding the prediction of the EC, the observed trends were mirrored for positive ($\rho_{EC,\varphi} = 0.30$) and negative ($\rho_{EC,\varphi} = -0.30$) relations between the EC and φ_i , and thus, only results of the former will be presented (for the results of $\rho_{EC,\varphi} = -0.30$, see Tables S1–S4 and Figures S1–S5). In the following, we refer to the standardized prediction parameters as β using subscripts for the respective IAD.

Reliability of individual IADs

Median reliabilities of ISD_i , $\ln(\sigma_i^2)$, and $MSSD_i$ were generally high and of similar magnitude across conditions and increased in conditions with higher inter-individual differences in the innovation variances $\tau_{\ln(\sigma_i^2)}^2$, (Figure 1A).

Regardless of the selected two-step approach, for φ_i , the highest median reliability was $r^2 = 0.59$, corroborating findings on comparatively low reliability of individual φ_i parameters (Du & Wang, 2018). With increasing $\tau_{\ln(\sigma_i^2)}^2$, the magnitude of misspecification for $ML_{2\text{-step}}$ increased, which was reflected in decreased recovery of φ_i estimates.

Accuracy and power of IADs as predictors

In Figures 1B and 2A, absolute values of average estimates are plotted against the true effects (grey points) between the EC and IADs (i.e., expected standardized regression weights in linear and multiple regression models) in each condition. Empirical power of the IADs in simple and multiple regressions is depicted in Figures 1C and 2C, respectively. The main findings are presented separately for each IAD predictor.

ISD Predictor (β_{ISD}). Recall that higher φ_i and higher $\ln(\sigma_{\zeta_i}^2)$ result in higher ISD_i (see Equation (4)) and thus, ISD was able to capitalize on the main effects of φ_i and $\ln(\sigma_{\zeta_i}^2)$ on the EC in the unidirectional case ($\rho_{EC,\varphi} = \rho_{EC,\ln[\sigma_{\zeta_i}^2]}$). This became apparent in the higher true effects of the ISD predictor in simple regression models (Figure 1B) which exceeded the chosen population values for the bivariate associations between the EC and φ_i and $\ln(\sigma_{\zeta_i}^2)$. Consequently, in the unidirectional case, empirical power of ISD was generally high and above .84. The observed pattern in terms of expected linear overlap between ISD and the EC as well as power is reversed when $\rho_{EC,\varphi}$ and $\rho_{EC,\ln[\sigma_{\zeta_i}^2]}$ have opposing signs. Note, that the combination of unidirectional effects is the one observed in the empirical example and commonly reported for relations of inertia and variability of affect with

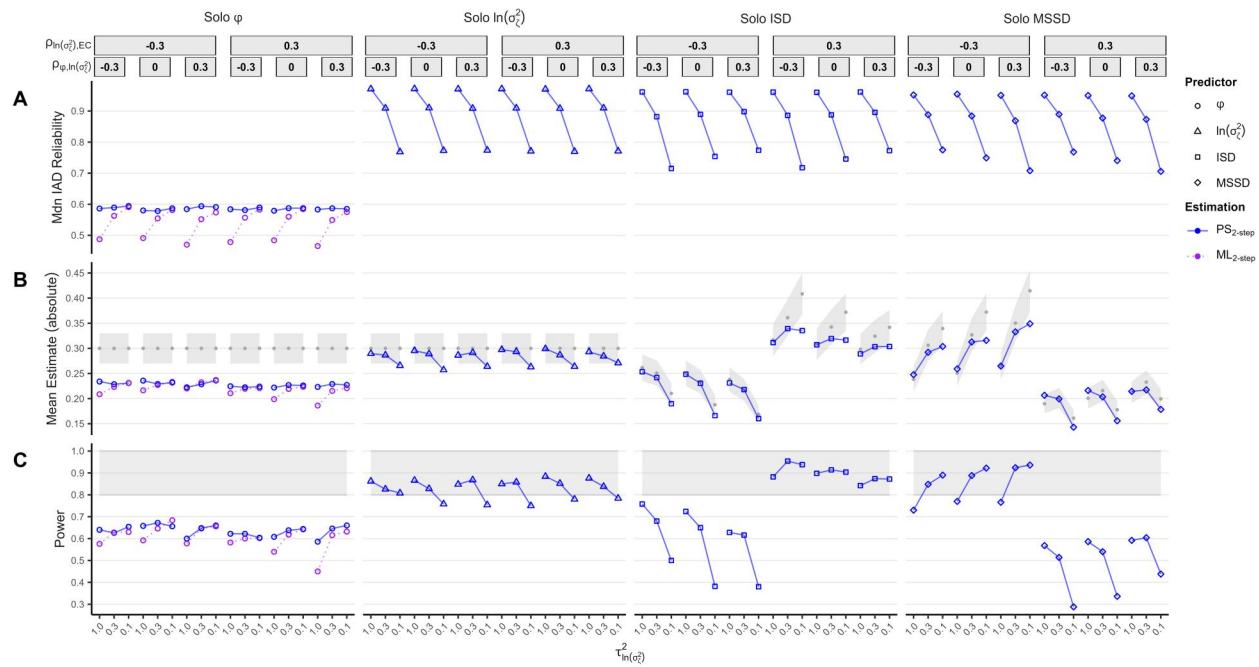


Figure 1. Simulation Study I: Individual Parameter Reliability, Mean Estimate, and Power of IAD Prediction Parameters in Simple Regression Models with $\rho_{EC, \varphi} = 0.30$ and $\gamma_{ln(\sigma_\zeta^2)} = 0$. φ : autoregressive effect; $ln(\sigma_\zeta^2)$: log innovation variance; ISD: intra-individual standard deviation; MSSD: mean squared successive difference. Median individual parameter reliability (Panel A), average absolute prediction parameter estimates (Panel B), and empirical power (Panel C) of IAD predictors in simple regression models (horizontal Panels). True values of absolute standardized regression weights are shown in Panel B as points with the area of 10% relative bias below and above highlighted in grey. Estimation approach ($PS_{2\text{-step}}$ = person-specific two-step [solid line], $ML_{2\text{-step}}$ = two-step with individual φ_i estimates derived from two-level AR[1] model [dotted line]) indicated by line type and color (see online version of the figure). Lines by estimation approach were added to the plots to highlight trends across simulated conditions.

depressive symptoms or psychological well-being (Houben et al., 2015).

AR Predictor (β_φ). Across conditions using $PS_{2\text{-step}}$, β_φ was underestimated with a substantial negative relative bias in simple (relative bias of -0.21 to -0.26) and multiple regression models (relative bias of -0.15 to -0.36), regardless of the selected second predictor. In line with the observed recovery of individual φ_i , for $ML_{2\text{-step}}$, increasing $\tau_{ln(\sigma_\zeta^2)}^2$ (i.e., the level of misspecification) resulted in higher relative bias of up to 38% in simple and up to 50% in multiple regression models (see Table S2). Given the expected linear overlap between *ISD* and the *EC* in conditions with $\rho_{EC, \varphi} = \rho_{EC, ln(\sigma_\zeta^2)}$, selecting *ISD* over $ln(\sigma_\zeta^2)$ as a second predictor resulted in generally lower power of β_φ . $\rho_{\varphi, ln(\sigma_\zeta^2)}$ had a moderating effect on β_φ regardless of the chosen covariate (i.e., *ISD* or $ln(\sigma_\zeta^2)$). The true scores (and therefore power) of β_φ over *ISD* also decreased with lower $\tau_{ln(\sigma_\zeta^2)}^2$ due to $\rho_{\varphi, ISD} > \rho_{\varphi, ln(\sigma_\zeta^2)}$ (see Figure 2). Note that the latter finding is to be expected given Equation (4), which suggests that lower $\tau_{ln(\sigma_\zeta^2)}^2$ results in higher overlap between φ_i and *ISD*. For the same reason, this pattern is reversed when the main effects of φ_i and $ln(\sigma_\zeta^2)$ on the *EC* are of opposing signs. In accordance with reliabilities of

individual parameter estimates, estimation accuracy and power were always higher for $\beta_{ln(\sigma_\zeta^2)}$ compared to β_φ although true absolute effects were of equal magnitude.

MSSD predictor (β_{MSSD}). Similar to *ISD*, expected associations between *MSSD* and the *EC* displayed a complex pattern as they were affected by the directionality of $\rho_{EC, \varphi}$ and $\rho_{EC, ln(\sigma_\zeta^2)}$, and further depended upon $\rho_{\varphi, ln(\sigma_\zeta^2)}$, as well as $\tau_{ln(\sigma_\zeta^2)}^2$. As expected, given the relation between φ_i and $MSSD_i$ (see Equation (5)), *MSSD* showed the highest overlap with the *EC* when the main effects of φ_i and $ln(\sigma_\zeta^2)$ on the *EC* were of opposing signs. In the respective conditions, the power of β_{MSSD} entered as single predictor was generally high (> 0.70) but tended to be underestimated as $\tau_{ln(\sigma_\zeta^2)}^2$ decreased (Figure 1B). As for *ISD*, these patterns were reversed when the sign of $\rho_{EC, \varphi}$ or $\rho_{EC, ln(\sigma_\zeta^2)}$ was changed. Furthermore, when controlling for *ISD*, true effects of β_{MSSD} displayed a complex, non-linear pattern. In case of opposing main effects, the expected linear overlap between *MSSD* and the *EC* and consequently the power of β_{MSSD} was strongly influenced by $\tau_{ln(\sigma_\zeta^2)}^2$ (see Figure 2C). Regardless of the main effects of φ_i and $ln(\sigma_\zeta^2)$ on the *EC*, β_{MSSD} was affected by increased negative bias when $\tau_{ln(\sigma_\zeta^2)}^2$ was

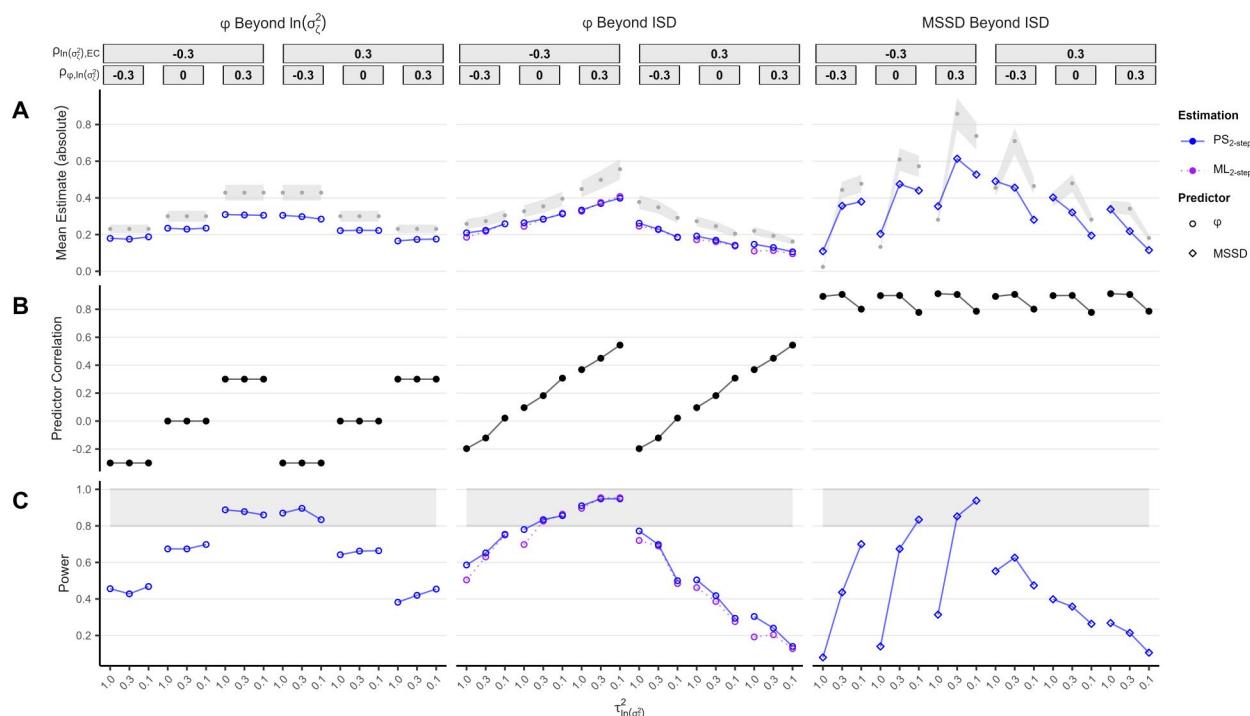


Figure 2. Simulation study I: Individual parameter reliability, mean estimate, and power of φ_i and MSSD as predictors in multiple regression models with $\rho_{EC, \varphi} = 0.30$ and $\gamma_{\ln(\sigma_i^2)} = 0$. φ : autoregressive effect; $\ln(\sigma_i^2)$: log innovation variance; ISD: intra-individual standard deviation; MSSD: mean squared successive difference. Absolute average prediction parameter estimates for φ_i and MSSD (Panel A), expected correlations between predictors (Panel B), and power (Panel C) across simulation conditions. True values of absolute standardized regression weights of β_φ and β_{MSSD} are shown in Panel A as points with the area of 10% relative bias below and above in grey. Estimation approach (PS_{2-step} = person-specific two-step [solid line], ML_{2-step} = two-step with individual φ_i estimates derived from two-level AR[1] model [dotted line]) indicated by line type and color (see online version of the figure). Lines were added to the plots to highlight trends across simulated conditions.

low. Further, in the unidirectional case, the power of β_{ISD} was always higher than that of β_{MSSD} when used as joint predictors (see Table S1). Nevertheless, substantial correlations between *MSSD* and *ISD* on the population level across conditions⁷, $\rho > 0.77$, indicated that results of multiple regression models with *MSSD* and *ISD* as predictors potentially suffered from multicollinearity (Figure 2B).

Conclusion

In view of the complexity and plethora of results, we briefly summarize the key findings of simulation study I that contribute to findings of low (added) predictive utility of dynamic IADs (i.e., *AR* and *MSSD*) after linearly accounting for *ISD*. Taken together, the estimation accuracy for the effects of dynamic IADs on a time-invariant outcome depends on how well the individual parameters are recovered. Critically, for inertia

as predictor, we found the common two-step approaches to produce negatively biased estimates (at $N=100$ and $T=70$). Furthermore, under a stationary AR(1) model, complex non-linear relations between dynamic IADs and static *ISD* emerge, which affect the unique predictive power of IADs when used as joint predictors with *ISD*. In practice, which IAD will be best suited as predictor of an EC may vary strongly with the studied context and a general recommendation is beyond what a simulation study can provide. However, if researchers encounter a typical scenario of unidirectional effects between IADs and the EC, it may be overly conservative to disregard the predictive value of emotional inertia (or the significance of accounting for the temporal dimension altogether) solely based on lower, non-significant regression estimates after controlling for the static variability.

Simulation study II – one-step to rule them all?

Using two-step approaches with $N=100$ and $T=70$, results of simulation study I revealed a considerable bias toward zero for β_φ across conditions, suggesting

⁷Note that high correlations between *ISD* and *MSSD* are to be expected when data are generated according to a stationary AR(1) model but can be lower in applied settings, e.g., in case of trends in person-specific mean levels over time (see, Jahng et al., 2008).

that regardless of the IAD-outcome-constellation, true relations between emotional inertia and an EC will be underestimated in simple and multiple regression models. In simulation study II, we did not further consider different outcome- φ_i - $\ln(\sigma_{\zeta_i}^2)$ associations and proceeded with $\rho_{EC, \varphi} = 0.3$, $\rho_{EC, \ln(\sigma_{\zeta_i}^2)} = 0.3$, and $\rho_{\varphi, \ln(\sigma_{\zeta_i}^2)} = 0$. In a fully crossed design, we examined factors related to AA research, such as *person sample size* ($N = 70, 100, 200$), *number of measurements* per person ($T = 50, 70, 100, 200$), and the choice of a data-analytic strategy. We consider two prediction models, (1) Model 1, the correct data generating model which involved regressing EC on φ_i and $\ln(\sigma_{\zeta_i}^2)$, and (2) Model 2, regressing EC on φ_i and *ISD*. Notably, the true expected contributions of φ_i beyond *ISD* (Model 2) were lower by design (i.e., the chosen population parameter values). Nevertheless, we were interested to assess how choosing *ISD* over $\ln(\sigma_{\zeta_i}^2)$ as a second predictor would affect the power to detect substantial relations between φ_i and an EC while controlling for *ISD*. The two-step approaches considered in simulation study I were compared with the performance of a one-step approach (BAY) using Bayesian MCMC estimation (Liu et al., 2021) and its extension to latent variable models that account for measurement error (BAY_{Lat}), as proposed above.

Further, to mimic typical scenarios encountered by applied researchers, we additionally considered different *missing data* and *measurement error* conditions. Next to a condition without missing values (0%), two conditions with a total of 20% missing data were chosen that resemble those typically observed in AA studies (e.g., Dejonckheere et al., 2019). The amount of missing values was generated to be distributed unequally across subjects.⁸ First, to mimic premature participant dropout, we introduced missing values by cutting off the last values of subjects' time-series (condition 20%[l]). In the second missing data condition, a more realistic scenario with missing values distributed randomly within the subjects' time-series (condition 20%[r]) was generated. Note that for 20%[r] using two-step approaches, no missing values were removed prior to data analyses, to keep equal distances between observations, which meant that the actual number of data points available for estimation of φ_i decreased compared to the condition 20%[l]. In contrast, for BAY and BAY_{Lat} at 20%[r], we implemented

an approach analogous to a Bayesian full-information likelihood approach, suitable for handling missing data (at random) as illustrated in Li et al. (2022). Aside from assuming perfect scale reliability, we considered a second condition where for two manifest indicators, measurement error was introduced to the within-part of the dynamic process (see Equations (6)–(9)) to achieve a within-level reliability of .80. Disregarding measurement error in the observed time series is expected to lower the accuracy of individual IAD estimates, thereby limiting their use as predictors of the EC. This negative effect should be resolved when measurement error is accounted for in latent variable models (BAY_{Lat}).

The accuracy of individual estimates is expected to be most sensitive to differing levels of T , with lower reliabilities of φ_i compared to those of variability estimates. Concerning the performance of the two-step approaches (PS_{2-step} vs. ML_{2-step}), expectations were two-fold. Enhanced individual parameter recovery was shown for the ML_{2-step} -approach when T is low (Liu, 2017)—as a compensating shrinkage effect of the multilevel approach. However, ignoring inter-individual differences in $\sigma_{\zeta_i}^2$ can lead to biased estimation of individual φ_i parameters in two-level AR(1) models (Asparouhov et al., 2018; Jongerling et al., 2015) which might result in better performance of PS_{2-step} compared to the mis-specified ML_{2-step} model. A multilevel approach (ML_{2-step}) might also be better suited in the presence of missing data as compared to PS_{2-step}, compensating for the differing numbers of observations across participants.

Results

For reasons of conciseness, we present a selection of the findings. The complete results including individual parameter reliabilities, power, coverage, relative bias, and MSE of $\beta_{\ln(\sigma_{\zeta_i}^2)}$ and β_{ISD} are summarized in Tables S9–S13.

Overall performance of the one-step approach

Model convergence ratios (CR) were generally acceptable but decreased with low N , lower T , and an increased number of missing values (see Table S5). For BAY, CR did not fall below 0.50 and exceeded 0.76 in all conditions with at least $N = 70$ and $T = 70$ or $N = 100$ and $T = 50$. A higher number of data points ($N = 70$ at $T = 100$, $N = 100$ at $T = 70$, or $N = 200$ at $T = 50$) was necessary to achieve similar CR above 0.70 for BAY_{Lat}. Additionally, five conditions ($N = 70$ at $T = 50$, and $N = 100$ at $T = 50$ with missing values) were removed due to CR < .50.

⁸In the 20%-missing-conditions, the percentages of missing values for $N = 200$ subjects (and varied proportionally for lower N) were 50% ($n = 16$), 40% ($n = 24$), 30% ($n = 36$), 20% ($n = 44$), 10% ($n = 28$), and 0% ($n = 52$), and in the 40%-missing-condition: 70% ($n = 16$), 60% ($n = 24$), 50% ($n = 44$), 40% ($n = 52$), 30% ($n = 32$), 20% ($n = 8$), 10% ($n = 4$), and 0% ($n = 20$).

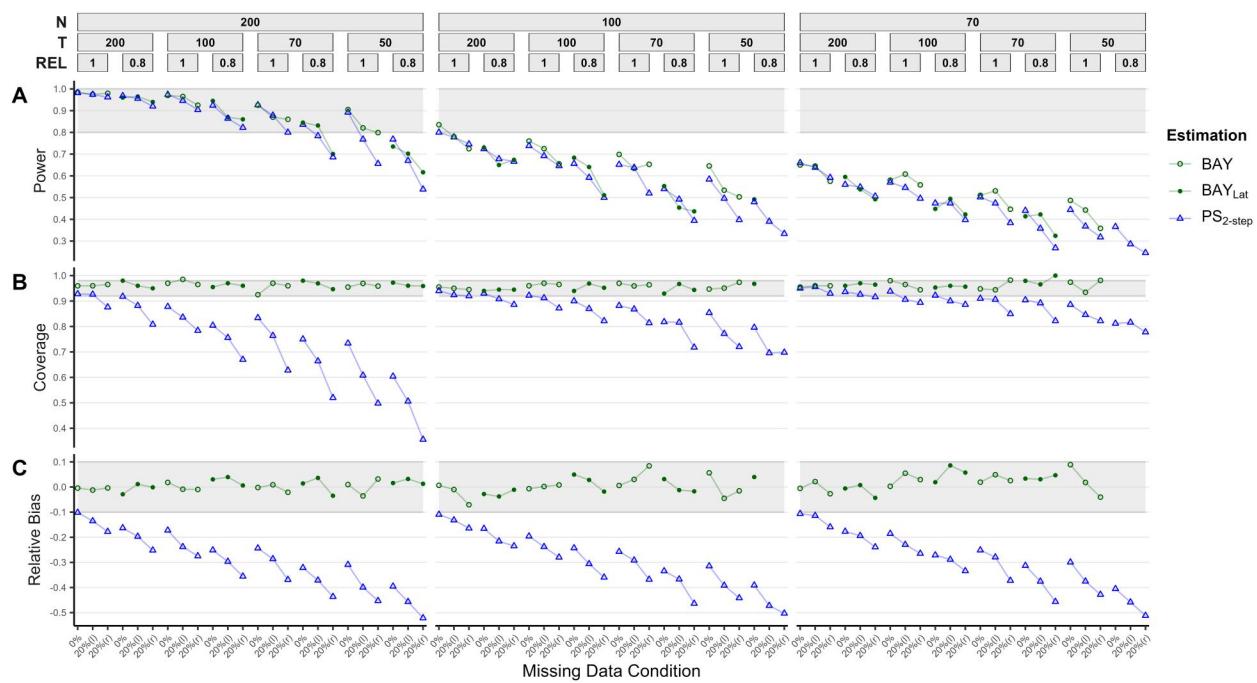


Figure 3. Simulation Study II. Power, Coverage, and Relative Bias of β_φ Controlling for $\ln(\sigma_{\zeta_i}^2)$. Power (Panel A), coverage (Panel B), and relative bias (Panel C) in prediction parameter of inertia (φ) controlled for log innovation variances across simulation conditions. Estimation approaches (BAY & BAY_{Lat} = one-step, PS_{2-step} = person-specific two-step) are indicated by point shapes (and color, see online version of the figure). Rel = within-level reliability of composite score. Missing data conditions: 0% = no missing values, 20%(l) = 20% of observations (N^*T) set as missing by cutting of the last observations of a subjects' time series, 20%(r) = 20% of observations (N^*T) set as missing with missing values distributed randomly within a subjects' time series. For BAY and BAY_{Lat}, results of conditions with a convergence rate below .5 were removed. Lines were added to the plots to highlight trends across simulated conditions. Target ranges of power $> .80$, coverage rates (of 95% CIs) between .925 and .975, and absolute relative bias < 0.10 are highlighted in grey.

Estimation accuracy of all population parameters (e.g., fixed effects and random effect variances) was generally high, except for random effect variances which tended to be overestimated in conditions with $N=70$ at $T \leq 100$ and $N=100$ at $T \leq 70$ (see Tables S6–S8). A drop in coverage rates of measurement model parameters (i.e., λ_W , σ_ε) in conditions with high N and low T was observed but deemed as unproblematic as estimates were unbiased with MSE below 0.004 throughout conditions.

Performance under correct prediction model specification (BAY vs. PS_{2-step})

Power of β_φ was affected by all varied factors but mostly depended on N . In practice unless $T > 200$, sample sizes larger than $N=100$ may be necessary for detection of medium sized effects between φ and the EC to achieve power $> .80$. BAY and PS_{2-step} performed similar but with minor gains in power for BAY/BAY_{Lat} in case of lower T as well as in the presence of missing values or measurement error, as illustrated in Figure 3.

In contrast, the power of $\beta_{\ln(\sigma_{\zeta_i}^2)}$ across all conditions (with CR > 0.50) and estimation approaches was higher (PS_{2-step} and BAY/BAY_{Lat}, average power = 0.79) than that of β_φ (PS_{2-step}, average power = 0.65; BAY/BAY_{Lat}, average power = 0.68). Examining individual parameter reliabilities, depicted in Figure 4A, revealed that reliability of φ_i tended to be low as T decreased, and in the presence of measurement error and missing values. The impact of the latter on $\ln(\sigma_{\zeta_i}^2)$ was lower, with reliability always higher than 0.64, explaining differences in power of the two predictors. BAY generally outperformed PS_{2-step} in terms of estimation accuracy for β_φ (Figure 3C) and in the presence of measurement error also for $\beta_{\ln(\sigma_{\zeta_i}^2)}$ (see Tables S11–S12). For PS_{2-step}, unique contributions of β_φ were again underestimated in all conditions with a minimum relative bias of -0.10 ($T=200$, $N=200$, with perfect reliability, and without missing data) ranging up to -0.51 ($T=50$, Miss = 20%(r), and within-level reliability of 0.80) and substantial drops in coverage rates (see Figure 3B). Bias for PS_{2-step} reduced solely by increasing the number of T (and reducing the number of missing values and

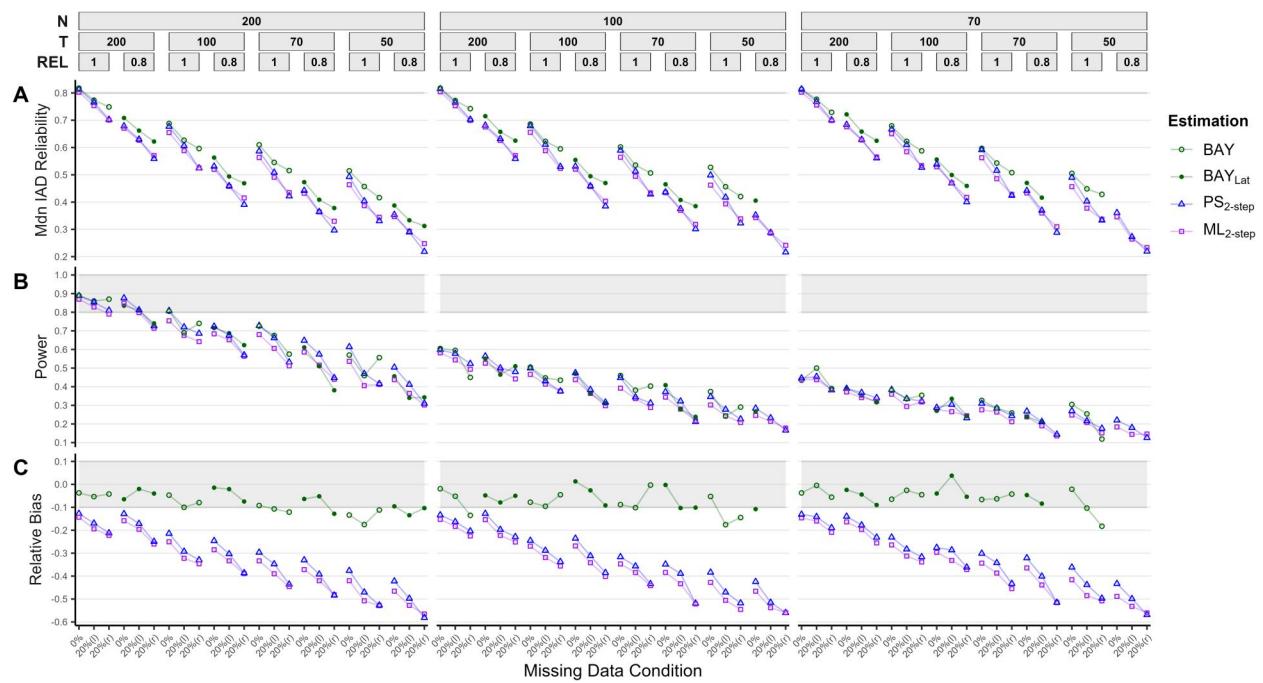


Figure 4. Simulation study II. Individual φ_i reliability, power, and relative bias of β_φ controlling for ISD. Median reliability of individual inertia (φ_i) estimates (Panel A), empirical power (Panel B), and relative bias (Panel C) of prediction parameter β_φ controlling for ISD across simulation conditions. Estimation approaches (BAY & BAY_{Lat} = one-step, PS_{2-step} = person-specific two-step, ML_{2-step} = multilevel two-step) are indicated by point shapes (and color, see online version of the figure). Rel = within-level reliability of composite score. Missing data conditions (Miss): 0% = no missing values, 20%(l) = 20% of observations (N*T) set as missing by cutting of the last observations of a subjects' time series, 20%(r) = 20% of observations (N*T) set as missing with missing values distributed randomly within a subjects' time series. For BAY and BAY_{Lat}, results of conditions with a convergence rate below 0.5 were removed. Lines were added to the plots to highlight trends across simulated conditions. Target ranges of power > 0.80 , coverage rates (of 95% CIs) between 0.925 and 0.975, and absolute relative bias < 0.10 are highlighted in grey.

measurement error). In contrast, BAY_{Lat} with latent variable modeling yielded unbiased parameter estimates and good coverage values in all conditions and tended to outperform PS_{2-step} in terms of MSE as N increases above 100 (see Table S13).

Controlling for ISD

Results for Model 2, controlling for ISD instead of $\ln(\sigma_{\varphi_i}^2)$, are summarized in Figure 4.

The general trends reported for β_φ in Model 1 remain intact, that is a considerable negative bias in β_φ across conditions when using a two-step strategy. Regardless of the estimation approach, the power of β_φ was lower (compared to Model 1) due to higher overlap between φ_i and ISD ($\rho_{\varphi, ISD} = 0.18$) compared to $\rho_{\varphi, \ln(\sigma_{\varphi_i}^2)} = 0$. Considering the results of simulation study I, these differences in power of the prediction parameters were largely driven by the choice of population parameters during data generation. Again, as for Model 1, the average power of the second predictor (here ISD) was substantially larger (0.78–0.81) compared to β_φ (0.43–0.47) regardless of the estimation approach. For Model 2 using BAY/BAY_{Lat},

relative bias in β_φ slightly increased and dropped below -0.10 in some conditions with $T \leq 70$ (Figure 4C) with a similar trend for PS_{2-step}. Coverage for β_φ associated with PS_{2-step} was below acceptable levels in most conditions, while BAY/BAY_{Lat} performed well (see Table S13). The results suggest that the two-step approach is generally not well suited to test relations between individual φ_i estimates and a time-invariant outcome even when $T = 200$ and $N = 200$, and perfect scale reliability, regardless of the chosen variability measure. In contrast, BAY/BAY_{Lat} generally showed superior performance, with non-negligible negative bias in β_φ only in some conditions with low T .

Ignoring random innovation variances (PS_{2-step} vs. ML_{2-step})

In a final step, we compared the estimation behavior of the two-step approaches (PS_{2-step} vs. ML_{2-step}) regarding β_φ when using ISD as a joint predictor. As shown in simulation study I, performance differences in terms of power and estimation accuracy are largely driven by the magnitude of misspecification, that is, the amount of inter-individual differences in

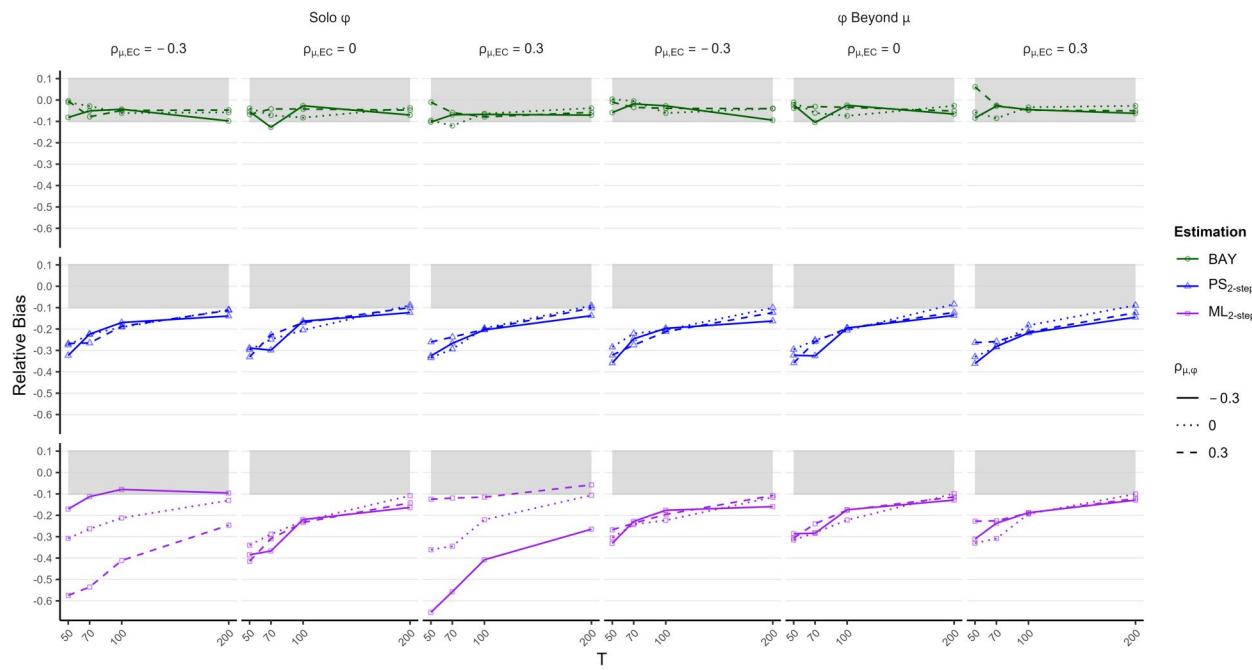


Figure 5. Simulation Study III. Relative Bias of β_φ in Simple and Multiple Regressions Controlling for μ_i . Estimation approaches (BAY = one-step, PS_{2-step} = person-specific two-step, ML_{2-step} = multilevel two-step) are indicated by point shapes (and color, see online version of the figure). $\rho_{\mu,EC}$ = correlation between individual means and the EC, $\rho_{\mu\varphi}$ random effects correlation indicated by line type. T = Number of time points. The area of absolute relative bias < 0.10 is highlighted in grey. Note that in all conditions the correlation between AR and the EC was kept constant at $\rho = .30$.

innovation variances, $\tau_{\ln(\sigma_i^2)}^2$. Nevertheless, the multi-level approach for estimation of individual φ_i estimates might be beneficial and potentially outperform PS_{2-step} in conditions with low T and increased number of missing values. Indeed, ML_{2-step} showed slightly better performance in conditions with $T = 50$ and the highest number of missing values (i.e., 20%[r]), such as increased recovery of φ_i , and similar power and bias compared to PS_{2-step} (see Figure 4). However, the gains were marginal, with an overall worse performance than PS_{2-step}.

Simulation study III – “regression to the mean”

In the above simulations, we assumed that individual means (μ_i) were uncorrelated with both AR and the EC. However, in applied settings, mean levels (e.g., of positive and negative affect) are often of interest and highly predictive of psychological well-being (e.g., Dejonckheere et al., 2019). In consideration of these findings, we conducted Simulation Study III to examine how the presence of non-zero relations between mean levels and AR ($\rho_{\mu,\varphi}$) as well as mean levels and the EC ($\rho_{\mu,EC}$) might impact the estimation accuracy of the AR prediction parameter.

As described by Asparouhov et al. (2018), the estimation accuracy of random effects in two-level models may benefit from higher random effect correlations. For a two-level AR(1) model with high correlations between individual φ_i and $\ln(\sigma_i^2)$, they observed an improved recovery of individual φ_i estimates. Based on this notion—multiple correlated random effects may inform one another during model estimation—we expected 1) an increased reliability of individual AR parameter estimates in the presence of correlated random effects (i.e., $\rho_{\mu,\varphi} \neq 0$), and 2) that this translates into more accurate AR predictor parameter estimates. However, we expected this “information flow” to be primarily unidirectional, with random intercepts (i.e., means) exerting a stronger influence on the accuracy of the random slopes (i.e., AR) than the other way around, due to the more reliable estimation of mean as compared to AR parameters. Consequently, the effect should be less prominent in longer time-series (i.e., with increasing φ_i reliability). This raises the question to what extent this flow of information may affect estimates of the relationship between AR and an EC.

Consider the (exaggerated) scenario of a perfect correlation between individual means and an EC ($\rho_{\mu,EC} = 1$) and a positive random effect correlation ($\rho_{\mu,\varphi} > 0$) and its consequences for estimates of the

AR-EC-relation. For $ML_{2\text{-step}}$, which underestimated the latter association in nearly all simulation conditions tested above, we may observe a compensating effect if the true relation between *AR* and the *EC* is positive. In contrast, the negative bias might further increase if we assume $\rho_{\mu, EC} = -1$.

In a fully crossed design, we varied $\rho_{\mu, \varphi}$ and $\rho_{\mu, EC}$ in margins of -0.3 , 0 , and 0.3 , and sample sizes (i.e., N and T) were selected similarly to simulation study II (see Table 2 for details). We focus on the accuracy of β_{φ} in simple linear and in multiple regression models controlling for the mean.

Results

Complete simulation results are provided in Tables S14–S17 and Figures S6–S9. Here, we focus on the relative bias in β_{φ} in conditions with $N = 100$. As observed in simulation study II, the relative bias of the prediction parameters remained consistent across different levels of N .

Figure 5 shows the relative bias in β_{φ} for varying levels of T , $\rho_{\mu, \varphi}$, $\rho_{\mu, EC}$ as a function of the chosen estimation approach. Notably, for $ML_{2\text{-step}}$, the bias in β_{φ} in simple regression models was moderated by the signs of $\rho_{\mu, \varphi}$ and $\rho_{\mu, EC}$ with the largest differences at low T . Recall that throughout, the *EC* was generated to be positively correlated with *AR* ($\rho = 0.3$). In conditions where the signs of both $\rho_{\mu, \varphi}$ and $\rho_{\mu, EC}$ were also positive, applying $ML_{2\text{-step}}$ yielded less biased estimates (less underestimation) even at low values of T , suggesting that the estimation of β_{φ} capitalizes on the unidirectional associations $\rho_{\mu, \varphi}$ and $\rho_{\mu, EC}$. However, if either $\rho_{\mu, \varphi}$ or $\rho_{\mu, EC}$ were of a negative sign, the pattern flipped, resulting in an even more pronounced negative bias in β_{φ} . Furthermore, if either $\rho_{\mu, \varphi} = 0$ or $\rho_{\mu, EC} = 0$, the bias in β_{φ} only depended on T . This interaction observed for $ML_{2\text{-step}}$ did not occur for *BAY* or $PS_{2\text{-step}}$ estimation, which both yielded result patterns consistent with simulation study II, with *BAY* showing the best performance.

Interestingly, when individual means were included as a second predictor alongside *AR*, the interaction pattern for $ML_{2\text{-step}}$ also vanishes. This conclusion was further supported by the generally consistent reliability of φ_i estimates across different levels of $\rho_{\mu, \varphi}$ and $\rho_{\mu, EC}$ (see Figure S6).

Simulation study III highlights a non-intuitive factor that can affect the accuracy of *AR* prediction parameters when using $ML_{2\text{-step}}$. This observation aligns with findings from the empirical example, where individual means of negative affect were

positively correlated with the outcome ($r = 0.56$) and *AR* ($r = 0.42$). Consequently, for $ML_{2\text{-step}}$ a higher first-order correlation between *AR* and depressive symptoms ($r = 0.34$) compared to $PS_{2\text{-step}}$ ($r = 0.25$) could be observed.

Overall, we conclude that the results further strengthen the recommendation to choose a one-step over two-step modeling approaches.

Discussion

The present study examines dynamic indicators for their use as single and joint predictors of a time-invariant outcome. We observed that associations between inertia and time-invariant outcomes are likely underestimated in common AA scenarios (i.e., $T < 200$) when applying a two-step estimation approach that fails to account for the low reliability of individual φ_i estimates. Based on the simulation results, we strongly advise against using two-step approaches, given the considerable negative bias observed in the respective regression weights for φ_i . This result is in line with previous studies promoting a one-step approach (Liu et al., 2021; Wenzel & Brose, 2023). Notably, the negative bias in β_{φ} when applying a two-step approach was reduced only by increasing the length of the time series (T), however, relative bias was still beyond acceptable levels even with as many as 200 time points. This result questions their use when considering participant burden. As an alternative to two-step approaches, we suggest the use of multilevel latent time-series models in a one-step approach. Note that recently, for similar reasons, Wenzel and Brose (2023) suggested a related one-step approach, which differs from the model proposed here with respect to the modeling strategy used to increase the reliability of individual φ_i estimates. While Wenzel and Brose (2023) estimate φ_i effects on the level of single items (e.g., “sad” and “angry”) and subsequently load them on a common factor, the model proposed here defines time-specific latent factors for the items at the within-person level and estimates the φ_i effects for these measurement-error free latent factors. Both approaches come with advantages and disadvantages. While the former approach allows items to have different levels of inertia, it still assumes that these show the same correlational pattern with the external criterion variable of interest. Additionally, φ_i effects are potentially underestimated due to specifying *AR* effects on the manifest items afflicted by measurement-error (and thereby potentially estimated less reliably in case of individually-varying item reliabilities; Schuurman & Hamaker, 2019). In contrast, the approach presented

here requires that the items measure a common underlying latent variable (e.g., different items measuring affective valence), with a single, common *AR* effect of this latent variable. This approach avoids underestimation of *AR* effects and additionally considers person-specific innovation variances defined on the level of latent variables, which might be used as additional predictor variables.

In contrast to prior studies comparing one-step and two-step approaches (Liu et al., 2021; Wenzel & Brose, 2023) as well as person-specific and multilevel approaches for estimation of individual dynamic parameters (Liu, 2017, 2018), we allowed innovation variances to differ between subjects. Differences in innovation variances are not only relevant from a statistical point of view but also provide an alternative indicator of affect variability that captures the effects of all internal and external influences that cannot be predicted by previous states alone (Jongerling et al., 2015). Interindividual differences in innovation variances were argued to reflect subjects' differences in responsiveness to or in the exposure to external factors (Hamaker et al., 2018), and in the perception of emotions as predictable or erratic (Simons et al., 2021). Given the potential bias created in individual φ_i estimates and between-level regression weights if falsely assuming constant innovation variances, testing for violations of this assumption seems advisable.

This study is not the first to illustrate statistical overlap between complex affect indicators, such as *MSSD* and inertia (*AR*) and static variability indices (e.g., Jahng et al., 2008; Wang et al., 2012). In contrast to previous studies, we provide extensive information on the conditions under which substantive associations of these IADs with an EC may be expected to be discovered in applications. In applied settings, for instance, examining individual differences in depressive symptoms (e.g., Bos et al., 2019; Koval et al., 2013) authors concluded that after accounting for interindividual differences in static variability of negative affect, unique contributions made by emotional inertia (*AR*) and affective instability (*MSSD*) were non-substantial. As previously argued, these findings might indicate that complex IADs are irrelevant as predictors of psychological health outcomes beyond individual means and variances (Dejonckheere et al., 2019). This question may not be resolved within the present paper. However, we believe that the underlying assumptions under which these results occur need to be made explicit to inform researchers on the selection of their measure of variability and analysis strategy. The present results suggest that multiple factors

afford consideration when testing complex IADs for their predictive utility (beyond measures of affect variability). First, the statistical overlap between *MSSD* and static variability measures (i.e., *ISD*, and $\ln(\sigma_{\zeta_i}^2)$), given an *AR(1)* model holds, might be of substantial magnitude, depending on the between-subject joint distributions of φ_i and $\ln(\sigma_{\zeta_i}^2)$. Note that this overlap might vanish in non-stationary time series (Jahng et al., 2008). Second, the choice of (static) affect variability indicator affects the unique contribution φ_i can make, as, due to mathematical dependencies, *ISD* carries parts of the dynamic information captured in φ_i . The respective statistical overlap depends on the amount of inter-individual differences in innovation variances, rendering the practice to estimate φ_i under the assumption of constant $\ln(\sigma_{\zeta_i}^2)$ questionable. Our findings suggest that, given the *AR* model holds, φ_i tends to show more positive relations with *ISD* compared to $\ln(\sigma_{\zeta_i}^2)$, reducing predictive power of φ_i in case of unidirectional relations between IADs and the outcome. Hence, true relations between an outcome with φ_i and $\ln(\sigma_{\zeta_i}^2)$ might be masked (Jongerling et al., 2015; Wang et al., 2012). This conclusion was corroborated by the reanalysis of an empirical data set, testing for the role of temporal dependency in negative affect for predicting depressive symptoms above and beyond static affect variability.

We would like to stress that, although not investigated in detail in the present study, assumption of a linear relationship between IADs and an external outcome might not be adequate in many applied scenarios. We therefore recommend researchers to always explore the form of relationship and rely, for instance, on spline-based regression approaches in the case that relationships are prone to be non-linear. Note, however, that the main findings from our simulation studies remain the same irrespective of the functional form used in the outcome model. That is, the effect of ignoring low reliability in the IAD estimates or measurement error in the observed time series will affect the predictive power of IAD estimates irrespective of the form of relationship that is assumed between IAD estimates and an outcome.

One shortcoming of the presented model and simulations is the assumption of constant item reliabilities across persons. That is, if measurement error variances vary across persons, disregarding these differences may result in less reliable φ_i estimates. However, models accounting for inter-individual differences in item reliabilities (see, e.g., Schuurmann & Hamaker, 2019) tend to be complex and may require larger *T* for adequate estimation accuracy. Second, we did not

consider the issue of potentially varying time intervals between observations. As these play an important role for the estimation and interpretation of *AR* effects (and innovation variances), considering varying time intervals across persons and time in the estimation of IADs is tantamount in empirical applications. Note, however, that the decision to not simulate differences in time intervals does not affect the conclusions of the simulation study.

In conclusion, our simulations show that person-specific dynamic indicators are more sensitive to the number of time points, common AA realities such as missing values and measurement error, and the choice of an estimation approach, as compared to their static counterparts, impacting statistical inferences made on their relevance as predictors of third-variable outcomes. These limitations can be circumvented by using latent variable multilevel time-series models and a one-step approach, which provide accurate regression estimates and are suitable to inform future meta-analyses on the role of measures of temporal dependency as predictors of time-invariant outcomes. Further, choosing individual innovation variances over indicators of static affect variability in the prediction of measures of psychological well-being may reveal otherwise masked relations between emotional inertia and the outcome of interest. Nevertheless, whether the decomposition of the overall variability of a time-series into the explained (inertia) and unexplained part (innovation variance) for their use as joint prediction of time-invariant outcomes is a fruitful endeavor remains a substantive question. We hope that the presented results can provide a promising way forward to resolving the question whether persons' differences in moment-to-moment fluctuations carry meaningful information that may be relevant for the prediction of outcomes beyond static variability.

CRediT author statement

Kenneth Koslowski: formal analysis, investigation, methodology, data curation, visualization, writing—original draft
 Jana Holtmann: conceptualization, methodology, supervision, writing—review & editing
 We have no known conflict of interest to disclose.

The analyses code for this study was made available at the Open Science Framework and can be accessed at <https://osf.io/bj7fq/>.

The data set chosen for the re-analysis can be accessed at <https://osf.io/zm6uw/#>.

Article information

Conflict of Interest Disclosures: Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical Principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: This work was not supported by a grant.

Role of the Funders/Sponsors: None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

References

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

Asparouhov, T., & Muthén, B. (2019). Latent variable centering of predictors and mediators in multilevel and time-series models. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(1), 119–142. <https://doi.org/10.1080/10705511.2018.1511375>

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G. (2015). lme4: Linear Mixed-Effects Models using 'Eigen' and S4 (1.1-8) [Computer software]. <https://cran.microsoft.com/snapshot/2015-08-15/web/packages/lme4/index.html>

Bos, E. H., Jonge, P., & Cox, R. F. A. (2019). Affective variability in depression: Revisiting the inertia-instability paradox. *British Journal of Psychology*, 110(4), 814–827. <https://doi.org/10.1111/bjop.12372>

Brose, A., Schmiedek, F., Koval, P., & Kuppens, P. (2015). Emotional inertia contributes to depressive symptoms beyond perseverative thinking. *Cognition & Emotion*, 29(3), 527–538. <https://doi.org/10.1080/02699931.2014.916252>

Burchert, S., Kerber, A., Zimmermann, J., & Knaevelsrud, C. (2021). Screening accuracy of a 14-day smartphone

ambulatory assessment of depression symptoms and mood dynamics in a general population sample: Comparison with the PHQ-9 depression screening. *PloS One*, 16(1), e0244955. <https://doi.org/10.1371/journal.pone.0244955>

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>

Dejonckheere, E., Mestdagh, M., Houben, M., Rutten, I., Sels, L., Kuppens, P., & Tuerlinckx, F. (2019). Complex affect dynamics add limited information to the prediction of psychological well-being. *Nature Human Behaviour*, 3(5), 478–491. <https://doi.org/10.1038/s41562-019-0555-0>

Du, H., & Wang, L. (2018). Reliabilities of Intraindividual Variability Indicators with Autocorrelated Longitudinal Data: Implications for Longitudinal Study Designs. *Multivariate Behavioral Research*, 53(4), 502–520. <https://doi.org/10.1080/00273171.2018.1457939>

Ebner-Priemer, U. W., Eid, M., Kleindienst, N., Stabenow, S., & Trull, T. J. (2009). Analytic strategies for understanding affective (in)stability and other dynamic processes in psychopathology. *Journal of Abnormal Psychology*, 118(1), 195–202. <https://doi.org/10.1037/a0014868>

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, 19(1), 72–91. <https://doi.org/10.1037/a0032138>

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

Hamaker, E. L., & Grasman, R. P. P. P. (2014). To center or not to center? Investigating inertia with a multilevel autoregressive model. *Frontiers in Psychology*, 5, 1492. <https://doi.org/10.3389/fpsyg.2014.01492>

Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press Princeton.

Heller, A. S., Fox, A. S., & Davidson, R. J. (2019). Parsing affective dynamics to identify risk for mood and anxiety disorders. *Emotion*, 19(2), 283–291. <https://doi.org/10.1037/emo0000440>

Houben, M., & Kuppens, P. (2020). Emotion dynamics and the association with depressive features and borderline personality disorder traits: Unique, specific, and prospective relationships. *Clinical Psychological Science*, 8(2), 226–239. <https://doi.org/10.1177/2167702619871962>

Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, 141(4), 901–930. <https://doi.org/10.1037/a0038822>

Jahng, S., Wood, P. K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods*, 13(4), 354–375. <https://doi.org/10.1037/a0014173>

Jongerling, J., Laurenceau, J.-P., & Hamaker, E. L. (2015). A multilevel AR(1) model: Allowing for inter-individual differences in trait-scores, inertia, and innovation variance. *Multivariate Behavioral Research*, 50(3), 334–349. <https://doi.org/10.1080/00273171.2014.1003772>

Koval, P., Kuppens, P., Allen, N. B., & Sheeber, L. (2012). Getting stuck in depression: The roles of rumination and emotional inertia. *Cognition & Emotion*, 26(8), 1412–1427. <https://doi.org/10.1080/02699931.2012.667392>

Koval, P., Pe, M. L., Meers, K., & Kuppens, P. (2013). Affect dynamics in relation to depressive symptoms: Variable, unstable or inert? *Emotion*, 13(6), 1132–1141. <https://doi.org/10.1037/a0033579>

Koval, P., Sütterlin, S., & Kuppens, P. (2015). Emotional inertia is associated with lower well-being when controlling for differences in emotional context. *Frontiers in Psychology*, 6, 1997. <https://doi.org/10.3389/fpsyg.2015.01997>

Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional Inertia and Psychological Maladjustment. *Psychological Science*, 21(7), 984–991. <https://doi.org/10.1177/0956797610372634>

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

Kuppens, P., Sheeber, L. B., Yap, M. B. H., Whittle, S., Simmons, J. G., & Allen, N. B. (2012). Emotional inertia prospectively predicts the onset of depressive disorder in adolescence. *Emotion*, 12(2), 283–289. <https://doi.org/10.1037/a0025046>

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(3), 452–475. <https://doi.org/10.1080/10705511.2021.1911657>

Liu, S. (2017). Person-specific versus multilevel autoregressive models: Accuracy in parameter estimates at the population and individual levels. *The British Journal of Mathematical and Statistical Psychology*, 70(3), 480–498. <https://doi.org/10.1111/bmsp.12096>

Liu, S. (2018). Accuracy and reliability of autoregressive parameter estimates: A comparison between person-specific and multilevel modeling approaches. In M. Wiberg, S. Culpepper, R. Janssen, J. González, & D. Molenaar (Eds.), *Quantitative psychology* (pp. 385–394) Springer International Publishing.

Liu, S., Kuppens, P., & Bringmann, L. (2021). On the use of empirical bayes estimates as measures of individual traits. *Assessment*, 28(3), 845–857. <https://doi.org/10.1177/1073191119885019>

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, 50(2), 181–202. <https://doi.org/10.1007/BF02294246>

Nelson, J., Klumparendt, A., Doebler, P., & Ehring, T. (2020). Everyday emotional dynamics in major depression. *Emotion (Washington, D.C.)*, 20(2), 179–191. <https://doi.org/10.1037/emo0000541>

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49(6), 1417–1426. <https://doi.org/10.2307/1911408>

Nowak, U., & Lincoln, T. M. (2023). An experience-sampling study on the relevance of affect dynamics to paranoid ideation. *Emotion (Washington, D.C.)*, 23(1), 111–123. <https://doi.org/10.1037/emo0000807>

Panaite, V., Rottenberg, J., & Bylsma, L. M. (2020). Daily affective dynamics predict depression symptom trajectories among adults with major and minor depression. *Affective Science*, 1(3), 186–198. <https://doi.org/10.1007/s42761-020-00014-w>

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1(3), 385–401. <https://doi.org/10.1177/014662167700100306>

Santangelo, P. S., Reinhard, I., Koudela-Hamila, S., Bohus, M., Holtmann, J., Eid, M., & Ebner-Priemer, U. W. (2017). The temporal interplay of self-esteem instability and affective instability in borderline personality disorder patients' everyday lives. *Journal of Abnormal Psychology*, 126(8), 1057–1065. <https://doi.org/10.1037/abn0000288>

Schultzberg, M., & Muthén, B. (2018). Number of subjects and time points needed for multilevel time-series analysis: A simulation study of dynamic structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(4), 495–515. <https://doi.org/10.1080/10705511.2017.1392862>

Schuurman, N. K., Houtveen, J. H., & Hamaker, E. L. (2015). Incorporating measurement error in $n = 1$ psychological autoregressive modeling. *Frontiers in Psychology*, 6, 1038. <https://doi.org/10.3389/fpsyg.2015.01038>

Schuurman, N. K., & Hamaker, E. L. (2019). Measurement error and person-specific reliability in multilevel autoregressive modeling. *Psychological Methods*, 24(1), 70–91. <https://doi.org/10.1037/met0000188>

Simons, J. S., Simons, R. M., Grimm, K. J., Keith, J. A., & Stoltenberg, S. F. (2021). Affective dynamics among veterans: Associations with distress tolerance and posttraumatic stress symptoms. *Emotion*, 21(4), 757–771. <https://doi.org/10.1037/emo0000745>

Stan Development Team (2023). *RStan: The R interface to Stan*. (2.33) [Computer software]. <http://mc-stan.org/>

Suls, J., Green, P., & Hillis, S. (1998). Emotional reactivity to everyday problems, affective inertia, and neuroticism. *Personality and Social Psychology Bulletin*, 24(2), 127–136. <https://doi.org/10.1177/0146167298242002>

Trull, T. J., Solhan, M. B., Tragesser, S. L., Jahng, S., Wood, P. K., Piasecki, T. M., & Watson, D. (2008). Affective instability: Measuring a core feature of borderline personality disorder with ecological momentary assessment. *Journal of Abnormal Psychology*, 117(3), 647–661. <https://doi.org/10.1037/a0012532>

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

van Roekel, E., Verhagen, M., Engels, R. C. M. E., & Kuppens, P. (2018). Variation in the serotonin transporter polymorphism (5-HTTLPR) and inertia of negative and positive emotions in daily life. *Emotion*, 18(2), 229–236. <https://doi.org/10.1037/emo0000336>

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

von Neumann, J., Kent, R. H., Bellinson, H. R., & Hart, B. I. (1941). The mean square successive difference. *The Annals of Mathematical Statistics*, 12(2), 153–162. <https://doi.org/10.1214/aoms/1177731746>

Wang, L., Hamaker, E., & Bergeman, C. S. (2012). Investigating inter-individual differences in short-term intra-individual variability. *Psychological Methods*, 17(4), 567–581. <https://doi.org/10.1037/a0029317>

Wendt, L. P., Wright, A. G. C., Pilkonis, P. A., Woods, W. C., Denissen, J. J. A., Kühnel, A., & Zimmermann, J. (2020). Indicators of affect dynamics: Structure, reliability, and personality correlates. *European Journal of Personality*, 34(6), 1060–1072. <https://doi.org/10.1002/per.2277>

Wenzel, M., & Brose, A. (2023). Addressing measurement issues in affect dynamic research: Modeling emotional inertia's reliability to improve its predictive validity of depressive symptoms. *Emotion*, 23(2), 412–424. <https://doi.org/10.1037/emo0001108>

Zhang, Z., & Nesselroade, J. R. (2007). Bayesian estimation of categorical dynamic factor models. *Multivariate Behavioral Research*, 42(4), 729–756. <https://doi.org/10.1080/00273170701715998>