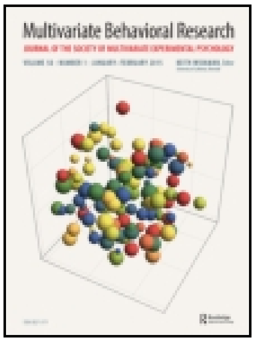

Multivariate Behavioral Research

ISSN: 0027-3171 (Print) 1532-7906 (Online) Journal homepage: http://www.tandfonline.com/loi/hmbr20

A Multilevel AR(1) Model: Allowing for Inter-Individual Differences in Trait-Scores, Inertia, and Innovation Variance

Joran Jongerling, Jean-Philippe Laurenceau & Ellen L. Hamaker

To cite this article: Joran Jongerling, Jean-Philippe Laurenceau & Ellen L. Hamaker (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, DOI: 10.1080/00273171.2014.1003772

To link to this article: http://dx.doi.org/10.1080/00273171.2014.1003772

Published online: 19 Jun 2015.

Multivariate Behavioral Research, 50:334–349, 2015

ISSN: 0027-3171 print / 1532-7906 online
DOI: 10.1080/00273171.2014.1003772

A Multilevel AR(1) Model: Allowing for Inter-Individual Differences in Trait-Scores, Inertia, and Innovation Variance

Joran Jongerling
Department of Methodology and Statistics, Utrecht University

Jean-Philippe Laurenceau
Department of Psychological and Brain Sciences, University of Delaware

Ellen L. Hamaker
Department of Methodology and Statistics, Utrecht University

In this article we consider a multilevel first-order autoregressive [AR(1)] model with random intercepts, random autoregression, and random innovation variance (i.e., the level 1 residual variance). Including random innovation variance is an important extension of the multilevel AR(1) model for two reasons. First, between-person differences in innovation variance are important from a substantive point of view, in that they capture differences in sensitivity and/or exposure to unmeasured internal and external factors that influence the process. Second, using simulation methods we show that modeling the innovation variance as fixed across individuals, when it should be modeled as a random effect, leads to biased parameter estimates. Additionally, we use simulation methods to compare maximum likelihood estimation to Bayesian estimation of the multilevel AR(1) model and investigate the trade-off between the number of individuals and the number of time points. We provide an empirical illustration by applying the extended multilevel AR(1) model to daily positive affect ratings from 89 married women over the course of 42 consecutive days.

Over the past few decades, there has been a growing interest in the study of processes as they unfold over time. This is accompanied by an increased need for longitudinal models that both capture the essence of these intra-individual processes, as well as allow for investigating any individual differences therein. While the study of *developmental processes* has blossomed with the introduction of techniques like latent growth curve modeling (Bollen & Curran, 2004, 2006; Meredith & Tisak, 1990) and latent transition models (Schmittmann, Dolan, Maas, & Neale, 2005), the statistical techniques for studying *stable processes* have only recently started to gain the attention of a wider audience of psychological researchers.

Stable processes can be roughly defined as processes that are characterized by within-person reversible variability over time in the absence of a gross underlying trend (Nesselroade, 1991). Examples include individuals' daily fluctuations in affect or the interaction between dyadic partners during a conversation. An innovative and promising modeling approach to the study of stable processes is *dynamic multilevel modeling*, which is based on modeling the repeated measures of an individual at level 1 using a time series model, while allowing for individual differences in the model parameters at level 2.

Suls, Green, and Hillis (1998) were the first to use this approach. They used a first-order autoregressive [AR(1)] model at level 1, in which each observation is regressed upon the preceding observation using an autoregressive parameter. The part that cannot be predicted from the previous observation is referred to as the *innovation* (also known as perturbation, random shock, or residual). Suls et al. (1998)

Correspondence concerning this article should be addressed to Joran Jongerling, Department of Methodology and Statistics, Faculty of Social and Behavioral Sciences, Utrecht University, P.O. Box 80140, 3508 TC, Utrecht, The Netherlands. E-mail: j.jongerling@uu.nl.

conceptualized the autoregressive parameter as a measure of spillover or carryover, as it indicates the degree to which prior states affect current states. They also proposed an interpretation of the autoregressive parameter as a measure of *inertia*, because the further it is away from zero, the longer it takes the individual to restore equilibrium after being perturbed by an innovation. Hence, the autoregressive parameter can be thought of as indicating a person's regulatory weakness, being inversely related to attractor strength (Hamaker, 2012). At the second level of their dynamic multilevel model, Suls et al. (1998) established a positive relationship between inertia and neuroticism and a negative relationship between inertia and agreeableness. Recently, this innovative work by Suls et al. (1998) has received attention from Kuppens and his colleagues, who have performed a series of studies focused on emotional inertia, showing that there is a positive relationship between inertia and depression (Kuppens, Allen, & Sheeber, 2010), that emotional inertia prospectively predicts the onset of depression (Kuppens et al., 2012), and that emotional inertia is related to rumination, although both factors separately contribute to depression (Koval, Kuppens, Allen, & Sheeber, 2012).

Across all these studies, the models were characterized by a random intercept and a random autoregressive (i.e., inertia) parameter, while the residual variance at level 1 (i.e., the innovation variance) was restricted to be the same across individuals. In contrast, Wang, Hamaker, and Bergeman (2012) considered a multilevel AR(1) model that also included a random innovation variance to allow for individual differences in this aspect of the process. However, they did not consider this issue in depth, neither from a substantive nor from a statistical point of view. Therefore, the current article is focused on the need for including a random innovation variance in the multilevel AR(1) model. We will argue that individual differences in residual variances are meaningful from a substantive point of view and may contain important information about regulatory processes. We then investigate what the effect is of ignoring this potential source of individual differences in a simulation study. In addition, we consider the trade-off between the number of observations within each person, and the number of people in the sample, as this is clearly of interest to applied researchers.

The remainder of this article is organized as follows. First, we introduce a multilevel AR(1) model that allows for individual differences in means, inertias, and innovation variances. Second, we discuss five different estimation methods for this model—three Maximum Likelihood (ML)-based methods that can be run using standard multilevel software and two Bayesian methods that can be run using WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000). Third, we present a simulation study in which the five estimation methods are compared. Fourth, we provide an illustration in which we apply the multilevel AR(1) model to an empirical data set, consisting of 42 daily emotion ratings obtained from a larger questionnaire administered to 96 married couples in a study on intimacy in marriage (for details, see Laurenceau, Feldman Barrett, & Rovine, 2005). We end with a discussion of the findings and recommendations for applied researchers who wish to make use of this model in their own work.

A MULTILEVEL AR(1) MODEL

In this section, we begin with presenting the level 1 or *within-person part* of the multilevel AR(1) model. This is comprised of an AR(1) model (cf. Hamilton, 1994; Chatfield, 2003) that can be expressed in two ways. We discuss the roles of the various model components in substantive terms. This is followed by the presentation of the level 2 or *between-persons part* of the multilevel AR(1) model, which allows us to model individual differences in the level 1 parameters. We explain what such differences may reflect, and why they could be of interest to psychological researchers.

Level 1: Within-Person

An AR(1) process can be expressed by using either one or two equations. Below, we present both, and discuss how the expressions are related. Let y_{it} be the observed score of individual i at time point t. If we express the AR(1) process with a single equation, we regress the observed score directly on the preceding score, that is,

$$y_{it} = c_i + \phi_i y_{i,t-1} + \epsilon_{i,t}, \tag{1}$$

where c_i is the individual's intercept (i.e., the expected score, when $y_{i,t-1} = 0$), ϕ_i is the AR-parameter, and ϵ_{it} is the unpredictable part, also referred to as the innovation, residual, or random shock. It is assumed that ϕ_i lies between –1 and 1 to ensure stationarity (that is, a situation in which the mean and variance of the process do not change over time; see Hamilton, 1994; Chatfield, 2003). Furthermore, it is assumed that the innovations are independent and normally distributed with mean 0 and variance σ_i^2.

Alternatively, when using the two-equation specification, we can think of the individual's score as consisting of two parts: a mean score μ_i, which represents an individual's trait score (i.e., his/her long-run tendency, equilibrium, or long-term preferred state) and a temporal deviation from this mean, which we denote as ζ_{it}, that is,

$$y_{it} = \mu_i + \zeta_{it}. \tag{2}$$

The temporal deviations (or states) themselves also may be characterized by autocorrelation and can be modeled with the AR(1) model

$$\zeta_{it} = \phi_i \zeta_{i,t-1} + \epsilon_{it}. \tag{3}$$

The two models expressed above are simply reparametrizations of each other, meaning that the actual process they describe is exactly the same. The equivalence between these two expressions can be seen by relating the mean in Equation (2) to the intercept in Equation (1) through

$$\mu_i = \frac{c_i}{1 - \phi_i}, \tag{4}$$

which is a standard result in time series literature (cf. Hamilton, 1994; Chatfield, 2003). Despite the equivalence of the two expressions, we feel that the latter two-equation specification is the model that researchers would typically want to estimate, because it provides estimates of the AR parameter, innovation variance, and the mean (instead of a less meaningful intercept). With standard maximum likelihood software, however, this model cannot be estimated, because both equations need to be combined into one. As a result, maximum likelihood estimation leads to the single equation specification and an inability to model individual means directly.

The inertia parameter ϕ_i in Equations (1) and (3) reflects the degree to which previous scores or states carry over into current scores or states. Suppose we have a number of daily measurements of negative affect for an individual. If the inertia parameter is close to zero, this implies that there is little or no carryover from the level of negative affect yesterday on the level of negative affect today. In contrast, when the inertia parameter is close to 1, this implies that an increased level of negative affect yesterday is likely to persist into today (and subsequent days), while decreased levels also tend to persist for several days. This is where the interpretation of inertia comes from.

The innovation ϵ_{it} represents the part of the process that cannot be predicted based on previous scores or states. Thus, it can be thought of as the collection of all unobserved (or omitted) factors that influence the process under investigation. For instance, today's negative affect not only depends on yesterday's negative affect, but also on sleep quality, recent stress experiences, caffeine and alcohol intake, hormonal levels, social obligations and interactions, et cetera. Furthermore, individuals may be more or less sensitive to these factors. While it is possible to include measurements of such factors in our level 1 model (e.g., Suls et al., 1998), and to model an individual's sensitivity to such a factor (e.g., Wichers et al., 2009; Wichers, Lothmann, Simons, Nicolson, & Peeters, 2012; Wichers et al., 2010; Wichers, Peeters, et al., 2012), there will always be additional factors that influence the process but that were not observed and therefore cannot be modeled explicitly. These effects are absorbed into the innovation term, and thus influence the innovation variance parameter σ_i^2.

Level 2: Between-Person

The fixed-effect parameters of the within-person part of the model, that is, the mean μ_i, the inertia ϕ_i, and the innovation variance σ_i^2, may be characterized by individual differences, which we can model at level 2 by including random effects. Before presenting the level 2 model, we take a more detailed look at the effect of individual differences in the level 1 parameters. To this end, we make use of four simulated AR(1) processes that are presented in Figure 1.

To draw links between the parameters of the model and the behavior of the outcome, there are several aspects of Figure 1 that are worth noting. First, the two upper panels of this figure show that differences in means are just indicative of differences in the the vertical position of the series (i.e., the equilibria or preferred states) of different individuals, but that they do not give any information about the individual dynamics. Second, comparing the two left panels, it can be seen that differences in the autoregressive parameter result in differences in the dynamics, that is, the pattern of fluctuations over time, as well as in the amount of total variance on the outcome variable. An AR-parameter closer to 1 (e.g., .9 for lower-left panel) leads to more carryover and therefore less random fluctuations over time and a wider range of fluctuations than an AR-parameter closer to 0 (e.g., .2 for upper-left panel). Third, comparing the upper-left panel with the lower-right panel, we can see that differences in innovation variances also result in differences in observed variances: when the innovation variance is larger (e.g., 3 for the upper-left panel), the total variance for the observations is also larger compared to the case where the innovation variance is smaller (e.g., 1 for the lower-right panel), despite the AR-parameter being the same.

The fact that both the innovation variance σ^2 and the AR-parameter ϕ affect the total variance also becomes apparent from the relationship between these three characteristics (e.g., Hamilton, 1994; Chatfield, 2003), that is,

$$\psi_i^2 = \frac{\sigma_i^2}{1 - \phi_i^2}, \tag{5}$$

where ψ_i^2 is the variance of the observed variable for individual i.

The need to allow for individual differences in means is obvious: Different individuals have different trait levels or preferred states, and this can be captured by individual differences in μ_i. In addition, the importance of allowing for individual differences in the AR-parameter has been the focus of a small number of studies, which have shown that this measure of inertia can be meaningfully related to other person characteristics such as gender (Rovine & Walls, 2006), neuroticism (Suls et al., 1998; Wang et al., 2012), depression (Kuppens et al., 2010; Koval et al., 2012), and rumination (Koval et al., 2012). Furthermore, Koval and Kuppens (2012) have shown that inertia can be state-dependent, and that it decreases more under stress in persons vulnerable to stress than in others. In addition, it has been shown that inertia can prospectively predict the onset of depression in adolescence (Kuppens et al., 2012) and health outcomes (Wang et al., 2012). Note that for most psychological processes individuals will be characterized by a positive AR-parameter, but this is not necessarily the case. For example, Rovine and Walls

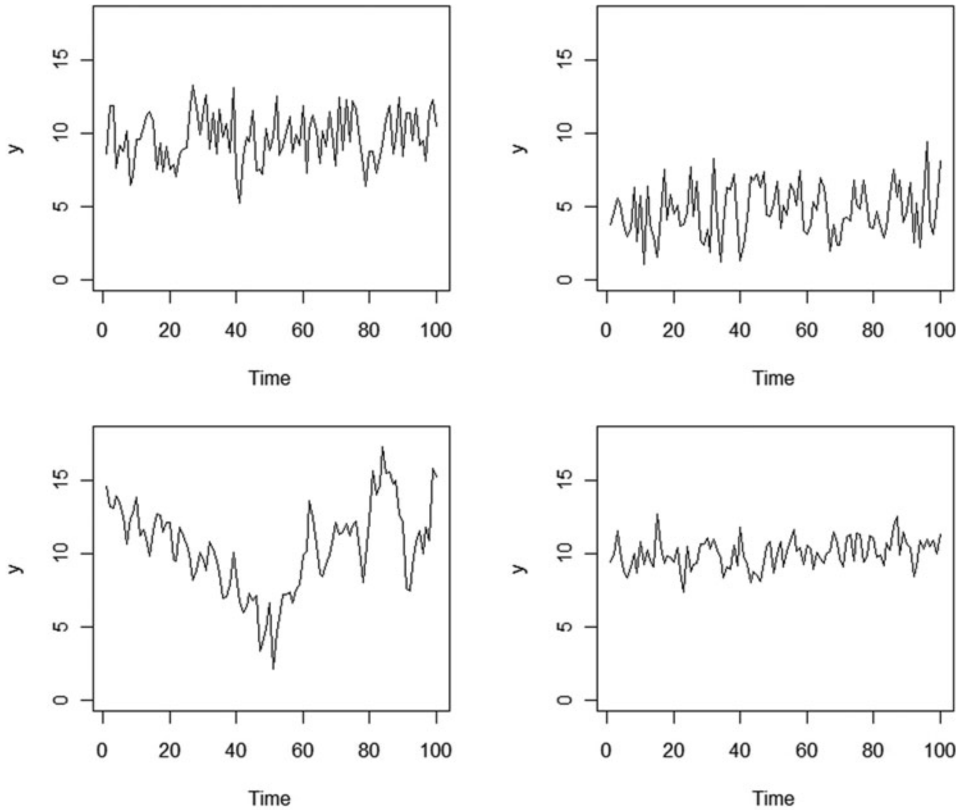

FIGURE 1 Timeseries with different means, inertias, and innovation variances, where y denotes the outcome variable. The series in the upper-left panel is characterized by a mean of 10, an AR-parameter of .2, and an innovation variance of 3; the series in the upper-right panel is characterized by a mean of 5, an AR-parameter of .2, and an innovation variance of 3; the series in the lower-left panel is characterized by a mean of 10, an AR-parameter of .9, and an innovation variance of 3; and the series in the lower-right panel is characterized by a mean of 10, an AR-parameter of .2, and an innovation variance of 1.

(2006) found that daily drinking behavior of some individuals was actually characterized by a negative AR-parameter. This implies a regulatory process that follows a sawtooth pattern: days on which these persons drank more than their average are typically followed by days on which they drink less than their average, and vice versa.

The possibility of individual differences in the innovation variance has been largely ignored in the literature thus far. While Wang et al. (2012) included a person-specific innovation variance, they have not considered the need for this potentially important feature of the model in depth. From a substantive point of view, we would like to argue that the existence of individual differences in innovation variances are to be expected for two reasons. First, there are probably individual differences in the range of fluctuation of the unobserved or omitted factors that influence the process under investigation, and this can be reflected by individual differences in the innovation variance. Second, individuals are likely to differ from each other with respect to their responsiveness to such factors. For instance, Rottenberg (2005) has indicated that depressed individuals are characterized by *emotion context insensitivity*, that is, a reduced emotional responsiveness to the environment, while Wichers et al. (2010) have shown that depressed individuals respond less strongly to positive events and more strongly to negative events. These types of individual differences in responsiveness or sensitivity to unobserved factors can also be reflected by individual differences in innovation variance.

Therefore, from a substantive point of view, we believe individual differences in innovation variance are likely the norm, rather than the exception. Relating individual differences in innovation variance to other individual differences is likely to help us to obtain more insight in the process under investigation, and in the possible sources of individual differences. Furthermore, there is also a statistical motivation for including the innovation variance as a random effect. Although ignoring random effects typically does not bias the estimates of the fixed effects in multilevel models (cf. Hox, 2002), we believe that in the current context the situation may be different because the observed variance is a function of both the innovation variance and the AR-parameter [as shown in Equation (5)]. If the innovation variance is not allowed to vary across individuals (i.e., $\sigma_i^2 = \sigma^2$), individual differences in the variance of the observed process (ψ_i) can only be accounted for by individual differences in the AR-parameter (ϕ_i). Thus, ignoring the possibility of individual differences in the innovation variance may lead to bias in the estimate of the AR-parameter.

With this line of reasoning in mind, we decided to specify the individual means, AR-parameters, and innovation variances as random effects, which may also be related to each other using a multivariate normal distribution, that is,

$$\begin{bmatrix} \mu_i \\ \phi_i \\ \sigma_i^2 \end{bmatrix} \sim MVN\left(\begin{bmatrix} \mu \\ \phi \\ \sigma^2 \end{bmatrix}, \begin{bmatrix} \tau_\mu^2 & & \\ \tau_{\mu\phi} & \tau_\phi^2 & \\ \tau_{\mu\sigma^2} & \tau_{\phi\sigma^2} & \tau_{\sigma^2}^2 \end{bmatrix} \right) \quad (6)$$

where MVN stands for multivariate normal.[1] The correlated random effects imply that the parameters may be influenced by the same unobserved person characteristics. For instance, Wichers et al. (2009) showed that the affective experience of individuals suffering from depression was more sensitive to the occurrence of negative events, whereas Kuppens et al. (2010) illustrated that depressed individuals are characterized by stronger emotional inertia. If negative events are not explicitly measured and included as a predictor of affect, the effect of negative events will be absorbed into the innovation term, leading to a larger innovation variance for more sensitive (i.e., depressed) individuals. As a result, the effect of depression on both the AR-parameter and the innovation variance will lead to a positive correlation between these two random effects at the second level.

It should be noted that the level 2 model presented here is very basic, but could be readily extended to include level 2 predictors such as neuroticism or gender (see Wang et al., 2012).

ESTIMATION

In this section, two estimation approaches are considered that can be used to estimate a multilevel AR(1) model: ML estimation using standard multilevel software and Bayesian estimation using WinBUGS. Thus far, most studies employing a multilevel AR(1) model used ML estimation in standard multilevel software. We also consider Bayesian estimation methods as a way to overcome some of the limitations associated with the use of standard multilevel software.

ML Estimation with Standard Multilevel Software

There are two problems associated with estimating the multilevel AR(1) model defined in Equations (2), (3), and (6) with standard multilevel software. First, most multilevel software packages only allow for the single equation formulation at level 1, such that defining the model as in Equations (2) and (3) is not possible. Second, by default the level 1 residual variance (i.e., the innovation variance) is identical across individuals in standard multilevel software. We elaborate on both limitations below.

Focusing on the first limitation, in standard multilevel software, the researcher cannot use our preferred specification of the AR-process at level 1 based on the two equations [Equations (2) and (3)]. Instead, the equivalent specification in Equation (1) can be used, but this has the disadvantage that it includes estimation of the intercept c_i, not the mean μ_i: As explained above, the intercept c_i is generally less interesting and intuitive from a substantive point of view than μ_i. Based on the relationship expressed in Equation (4), we can derive an estimate of μ_i, if we first obtain individuals' shrinkage estimates for c_i and ϕ_i (Bryk & Raudenbush, 1992; Hox, 2002). However, these shrinkage estimates are only available after model estimation, and as a result, one could only model individual differences in means by conducting a second modeling step.

As a solution to the inability to model individuals' mean scores, we could center the predictor $y_{i,t-1}$ per person: this implies we are centering level 1 predictor variables within the higher level units, and as a result the level 1 intercepts are equivalent to the level 1 cluster (i.e., person) means on the outcome variable (cf. Enders & Tofighi, 2007; Kreft, de Leeuw, & Aiken, 1995). The reason for this now follows. Since an individual's true mean on $y_{i,t-1}$ is identical to his/her mean on y_{it} (i.e., it is the individual's mean over time μ_i), the person-mean centered lagged predictor can be written as $y_{i,t-1} - \mu_i = \zeta_{i,t-1}$. Using this centered predictor, and making use of the fact that $c_i = \mu_i(1 - \phi_i)$, we can write

$$\begin{aligned} y_{it} &= c_i + \phi_i\mu_i + \phi_i(y_{i,t-1} - \mu_i) + \epsilon_{it} \\ &= \mu_i(1 - \phi_i) + \phi_i\mu_i + \phi_i\zeta_{i,t-1} + \epsilon_{it} \\ &= \mu_i + \phi_i\zeta_{i,t-1} + \epsilon_{it}, \end{aligned} \quad (7)$$

which shows that person-mean centering does indeed allow for the direct modeling of μ_i with a single equation specification of an AR(1) process at level 1. However, there is a problem with this approach. To center $y_{i,t-1}$, we need μ_i, but we don't know the value of this parameter. In fact, the aim is to estimate μ_i using Equation (7) [i.e., the whole purpose of expressing the model as in Equation (7) is to obtain an estimate of μ_i]. In short, person-mean centering requires simultaneously estimating both an individual's mean and already knowing it so it can be used for the actual centering of the lagged predictor, which is obviously impossible.

We will consider two solutions to the catch-22 we are in. First, we simply compute an individual's sample mean (i.e., the ordinary least squares estimate), and use this as an estimate for μ_i. This conforms to the usual approach when using cluster-mean centering in multilevel modeling. Second, we

[1]Instead of using the innovation variance in the multivariate normal distribution, we could have decided to use the logarithm of this variance to ensure that no negative variances can occur. However, we believe this to be less intuitive than considering the variance itself, and moreover we do not expect computational problems because innovation variances are expected to be clearly larger than zero in the data.

will consider a two-step procedure, where we begin with an empty model, with the level 1 model being $y_{it} = \mu_i + \zeta_{it}$. From this model, shrinkage estimates of the individuals' μ_i are obtained, which are then used to center the predictor variable at level 1, such that in the second step the model in Equation (7) can be estimated. These two procedures are referred to as *MLpc1* and *MLpc2* respectively (where pc stands for person-centered). In addition, we will also consider estimation based on Equation (1), that is, without centering the predictor, which we refer to as *MLuc*. In this case, estimates of μ and μ_i will be obtained using estimates of c and ϕ and the shrinkage estimates of c_i and ϕ_i, respectively [see Equation (4)].

The second limitation of standard ML software is that most multilevel software packages do not allow for individual differences in the residual variance at level 1, and if they do allow for individual differences, these differences would need to be fully accounted for by a level 2 predictor. That is, while it may be possible to model some individual differences, randomness of the innovation variance is not included as an option.[2] As argued in the previous section, individual differences in innovation variance are expected to be the norm, rather than the exception, so the assumption that this variance is the same for everyone (i.e., a fixed effect), an assumption that is implicitly made in standard software by only allowing a single error variance term, is unrealistic and undesirable: not only does this assumption prevent us from studying individual differences in this part of the process, but since the variance of an AR(1) process is a function of both the AR-parameter and the innovation variance as shown in Equation (5), it may also lead to bias in the estimation of the AR-parameter. This possible source of bias is what we will investigate in the simulation study below.

Bayesian Estimation with WinBUGS

WinBUGS is a free software package that can be used for Bayesian estimation (Lunn et al., 2000). In contrast to standard multilevel software, the WinBUGS program allows for a lot of freedom in specifying a model. As a result, we can define the multilevel AR(1) model using the two-equation structure at level 1 [Equations (2) and (3)], and relate individual differences in individual means to other individual differences at level 2 [Equation (6)]. Furthermore, it allows us to include the innovation variance as a random effect that may be related to other random effects.

We will consider two models when using WinBUGS: in the first Bayesian estimation method (*B1*), all the level 1 parameters of the multilevel AR(1) model (μ_i, ϕ_i, and σ_i^2) will be included as random effects, while in the second Bayesian estimation method (*B2*), only μ_i and ϕ_i will be random, thus implying that all individuals have the same innovation variance (i.e., $\sigma_i^2 = \sigma^2$). This latter model is included in the simulation study because it is the Bayesian equivalent of the ML estimation methods, which also contain a single residual variance term. Therefore, by not only comparing the Bayesian methods to the ML methods, but also comparing the two Bayesian methods to each other, we can differentiate between performance differences resulting from (erroneously) modeling the innovation variance as a fixed effect, and differences resulting from the use of Bayesian versus ML analyses.

Since WinBUGS is based on Bayesian estimation of the model, several steps are required before the model can be estimated by the program. While a thorough discussion of Bayesian statistics is beyond the scope of this article [interested readers are referred to Gelman, Carlin, Stern, and Rubin (2004), Hamaker and Klugkist (2011), and Hoijtink (2009)], there is one feature of Bayesian analysis that needs to be discussed here briefly—the prior distribution. In Bayesian statistics, researchers need to specify prior distributions for all model parameters, where these prior distributions represent a researcher's prior beliefs or knowledge about these parameters by assigning probabilities to their different possible values. These prior distributions are then combined with the distribution of the data using Bayes theorem in the following way:

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{f(y)}, \tag{8}$$

where $f(\theta|y)$ is the posterior distribution of a parameter that represents the combined information from both the prior and the data about this parameter, $f(y|\theta)$ is the distribution of the data (y) conditional on parameter θ, $f(\theta)$ is the prior distribution for parameter θ, and $f(y)$ is the distribution of the data. Posterior distributions of parameters of interest are subsequently used for model estimation. That is, the mean, median, or mode of a posterior distribution can be used as the point estimate of a parameter, while the standard deviation of the posterior distribution can be seen as a measure of the sample variability of this estimate (analogues to the standard error in standard maximum likelihood estimation). If one has little or no prior knowledge, *uninformative* priors can be used, which are characterized by assigning low and (approximately) equal probabilities to a very large range of possible values of a parameter. The results obtained with such priors depend almost exclusively on the data, and are therefore often close to ML estimates.

Specifically, for estimation method *B1*, we need to specify priors for all nine parameters that are defined in Equation (6), that is, for the three fixed effect (i.e., μ, ϕ, and σ^2), for the three random effects (i.e., τ_μ^2, τ_ϕ^2, and $\tau_{\sigma^2}^2$), and the three covariances between these random effects (i.e., $\tau_{\mu\phi}$, $\tau_{\mu\sigma^2}$, and $\tau_{\phi\sigma^2}$). For the fixed effects, normal distributions with 0 means and variances equal to 10,000 were chosen as priors.

[2]In MLWin (Rasbash, Charlton, Browne, Healy, & Cameron, 2009), this second limitation can be circumvented by using syntax, but this implies one can no longer make use of the user friendly interface of the program.

That is,

$$\mu \sim \mathcal{N}(0, 10, 000) \tag{9}$$

$$\phi \sim \mathcal{N}(0, 10, 000) \tag{10}$$

$$\sigma^2 \sim \mathcal{N}(0, 10, 000). \tag{11}$$

These large variances spread the normal distribution out over a very large range of values, making the distribution uninformative. In addition, we use an Inverse Wishart (IW) distribution as the prior for the covariance matrix in Equation (6) containing the random effects and their covariances, so that,

$$\begin{bmatrix} \tau_\mu^2 & & \\ \tau_{\mu\phi} & \tau_\phi^2 & \\ \tau_{\mu\sigma^2} & \tau_{\phi\sigma^2} & \tau_{\sigma^2}^2 \end{bmatrix} \sim IW(R, df), \tag{12}$$

where R is a scale matrix that positions the distribution in multivariate space, and df are the degrees of freedom of the distribution that determine how informative it is. Usually, an identity matrix is used for scale matrix R, but given the fact that the variance of the AR-parameter is expected to be much smaller than 1 (since most individuals will have a AR-parameter between, say, 0 and .5), this may not be an appropriate prior in this case. In fact, preliminary analyses showed that the identity matrix is not appropriate here. Therefore, we will use ML estimates of the random effects τ_μ^2 and τ_ϕ^2 in the scale matrix of the IW prior, setting the off-diagonal elements to 0.[3] Since there is no ML estimate of the random effect $\tau_{\sigma^2}^2$, we set the value for this parameter to 1, as would have been the case had an identity matrix been used. Finally, to make the IW an uninformative prior, the degrees of freedom need to be set to the number of random effects in the model, which in this case is 3.

For estimation method *B2*, the same priors are used, with the exception that the IW distribution is now only for the random effects τ_μ^2 and τ_ϕ^2, and their covariance $\tau_{\mu\phi}$, thus the degrees of freedom are set to 2.

SIMULATION STUDY

The purpose of the simulation study is twofold. On the one hand, we wanted to compare the performance of the different estimation methods in order to determine the effect of modeling a random innovation variance as a fixed effect. On the other hand, we were also interested in identifying the minimum number of individuals and/or time points required by the different methods for accurate parameter estimation. To this end, we varied: (a) the number of individuals in the sample N, (b) the number of time points T, and (c) the correlation between the innovation variance and the AR-parameter.

Data Generation and Model Estimation

Specifically, we generated data with 20, 50, and 100 individuals, and with 10, 20, and 50 time points. These samples sizes were considered realistic for these kinds of models based on the literature.

For the fixed and random effects parameters we used

$$\begin{bmatrix} \mu_i \\ \phi_i \\ \sigma_i^2 \end{bmatrix} \sim MVN\left(\begin{bmatrix} 10 \\ .2 \\ 3 \end{bmatrix}, \begin{bmatrix} \tau_\mu^2 & & \\ 0 & .01 & \\ 0 & \tau_{\phi\sigma^2} & 1 \end{bmatrix} \right). \tag{13}$$

Note that a key focus in this simulation is on whether ignoring randomness in the innovation variance results in bias in the estimation of ϕ_i. Because the relationship between ϕ_i and σ_i^2 may have an effect on this bias, the correlation between these random effects $\rho_{\phi\sigma^2}$ was varied from –.6, 0, to .6. The mean and variance for the autoregressive parameter were chosen based on Wang et al. (2012). The mean of the innovation variance is arbitrary in the sense that rescaling the observed variable y_{it} will also result in rescaling of the innovations, and thus in its variance. However, this is not independent of the variance of the innovation variance: we chose the variance of the innovation variance such that individual negative variances are unlikely to occur.[4] Finally, the fixed effect for the mean is also an arbitrary value (here set at 10), and the variance of the mean proved arbitrary in preliminary simulations. In the current simulations, the variance of μ_i was chosen such that the total between-person variance was the same as the average within-person variance.[5]

Data sets were created by randomly drawing N cases from the distribution defined in Equation (13), and subsequently generating a time series of T observations for each of the N cases using the AR(1) model defined in Equations (2) and (3).[6] In each scenario, 1,000 replications were created. The three ML methods were run in R, using restricted maximum likelihood (REML) estimation in the `lme4` package (Bates, Maechler, & Bolker, 2011). The two Bayesian estimation

[3]Note that this approach of using the ML estimates in the IW prior is equivalent to the training sample approach suggested by O'Hagan (1995), where part of the data is used to obtain a prior for the analysis of the rest of the data.

[4]This is relevant in the creation of the data sets as negative variances will make it impossible to generate data. However, it has no further practical consequences.

[5]We do not have an analytical expression of the within-person variance. Instead, we simulated a data set consisting of 10,000,000 persons, given a particular set of parameter values, and determined the average within-person variance. This way we determined that when $\phi = .20$, $\tau_\phi^2 = .01$, $\sigma^2 = 3.00$, and $\tau_{\sigma^2} = 1.00$, the total within-person variation is equal to 3.198, 3.170, and 3.141 for $\rho_{\phi\sigma^2}$ values of .6, 0, and -.6 respectively.

[6]During data generation it was evaluated whether all parameter values fell into a permissible range, that is, no individual innovation variances smaller than 0 and no values of ϕ_i greater than $|\ 1\ |$ to ensure stationarity. If parameter values fell outside these ranges, those data were discarded and a new data set was generated; however this was rarely needed (i.e., for less than 4–5% of generated data sets).

methods were run by calling WinBUGS from R using the `R2WinBUGS` package (Sturtz, Ligges, & Gelman, 2005). Based on preliminary convergence checks, the number of iterations for the Bayesian estimation procedures was set to 10,000 with a burn-in of 5,000.[7]

Evaluating Performance

We evaluate performance with respect to the fixed effects (i.e., μ, ϕ, and σ^2), the random effects (i.e., τ_μ^2, τ_ϕ^2, and $\tau_{\sigma^2}^2$), and their correlations [i.e., $\rho_{\mu\phi} = \tau_{\mu\phi}/(\tau_\mu \tau_\phi)$, $\rho_{\mu\sigma^2} = \tau_{\mu\sigma^2}/(\tau_\mu \tau_{\sigma^2})$, and $\rho_{\phi\sigma^2} = \tau_{\phi\sigma^2}/(\tau_\mu \tau_{\sigma^2})$], and the individual parameter estimates (i.e., μ_i's, ϕ_i's and σ_i^2's). Performance with respect to the fixed effects is based on evaluating: (a) the bias, which is determined by taking the difference between the true parameter value and the average parameter estimate; and (b) the coverage rate of the 95% confidence or credibility interval (CI), which is determined by computing the proportion of replications for which the true parameter value lies inside the associated interval. These coverage rates should be about .95; coverage rates lower than .90 will be considered to be too low, while coverage rates over .99 will be considered to be too high. In addition, performance with respect to the random effects and their correlations is assessed on the basis of bias, while performance with respect to the individual parameter estimates is based on: (a) the coverage rates of individual 95% CIs; and (b) the average correlation between the true individual parameters and the estimated individual parameters.

Results

Fixed Effects

The results reported on the left side of Table 1 show that while the five methods perform similarly with respect to bias in μ, the two ML methods based on person-centering the lagged predictor (i.e., methods *MLpc1* and *MLpc2*) lead to considerable bias in the estimation of ϕ and σ^2. This is especially true when T becomes smaller, while the effect of sample size at the between-person level (i.e., N) has a negligible effect. This issue has been studied in more detail by Hamaker and Grasman (2013), and we will therefore refrain from pursuing this issue here. Instead, we conclude that the bias for the ϕ-parameter obtained with methods *MLpc1* and *MLpc2* disqualifies the two centering procedures as a proper approach to multilevel AR(1) modeling.

When comparing the remaining three methods, it can be seen that in general, the bias reported for the uncentered ML method (i.e., *MLuc*) and its Bayesian equivalent in which the innovation variance is also modeled as a fixed effect (i.e., *B2*), is of a similar size, unless $T = 10$, when the Bayesian estimation procedure outperforms the ML method with respect to ϕ. The bias obtained with the true model (i.e., method *B1*) is much smaller in comparison to the other two methods when ϕ is considered, but larger when σ^2 is considered. Overall, the three methods seem to overestimate σ^2 regardless of the correlation between ϕ_i and σ_i^2. For ϕ, a negative correlation tends to result in a negative bias, while a positive correlation is associated with a positive bias.

On the right side of Table 1, the coverage rates for the 95% CIs of the fixed effects are reported. The two centered ML options resulted in extremely low coverage rates for the CI of ϕ (e.g., even .014 and .040 when $N = 100$ and $T = 10$), which could be expected given the bias in this estimate. In contrast, the coverage rates for μ obtained with *MLuc* are too high (i.e., it is equal to 1), which is the result of extremely large standard errors for this estimate.[8] The ML approach does not provide a standard error for the level 1 residual variance (here, the innovation variance σ^2), which is the reason we could not obtain coverage rates for the innovation variance obtained with the ML methods. Finally, the coverage rates of the Bayesian methods clearly outperformed the ML results, with the *B1* method (which was based on the true model) resulting in coverage rates close to the target value of .95.

Random Effects

The bias in the estimation of the random effects and their correlations are summarized in Tables 2 and 3, respectively. Table 2 contains the parameters estimated by all five procedures, that is, τ_μ^2, τ_ϕ^2, and the correlation $\rho_{\mu\phi}$. Table 3 contains the additional parameters which are only estimated with B1, that is, $\tau_{\sigma^2}^2$, $\rho_{\mu\sigma^2}$, and $\rho_{\phi\sigma^2}$. Table 2 shows that the ML methods performed quite similarly, and that they are quite comparable to the Bayes methods when the bias for τ_μ^2 is considered. With respect to the bias for τ_ϕ^2, the Bayesian methods performed less well than the ML methods, especially when both N and T are small. The more complex (and correct) model estimated with method *B1* led to slightly more bias in the estimation of τ_ϕ^2 than the incorrect model estimated with *B2*. The bias associated with the Bayesian methods is always positive.

[7]A problem we encountered with the use of the ML estimates in the IW prior is that when the ML estimates are very inaccurate, the WinBUGS analysis crashes. This occurred in one out of every 300 to 500 data sets. In practice, if this problem occurs, the user should change the scale values; in the current simulation study, we solved this by preventing the τ_ϕ^2 estimate in the scale matrix (R) of the IW prior to become too small (by substituting the value .005 for the ML estimate of τ_ϕ^2 if this estimate is smaller than this boundary value) and by producing a new data set in case WinBUGS crashed.

[8]The reason for this large standard error is that it had to be computed from the standard errors of c and ϕ, since μ is not directly estimated in this approach: To this end, we used the following equation from Mood, Graybill, and Boes (1985) for the variance of a quotient $\text{var}[\frac{X}{Y}] = (\frac{\mu_X}{\mu_Y})^2(\frac{\text{var}[X]}{\mu_X^2} + \frac{\text{var}[Y]}{\mu_Y^2} - \frac{2\text{cov}[X,Y]}{\mu_X \mu_Y})$. This extra estimation step forms an additional source of uncertainty, leading to large standard errors and thus coverage rates that are always 1.

TABLE 1
Bias and Coverage Rates of Fixed Effect Estimates

		Bias									Coverage Rates								
		μ			ϕ			σ^2			μ			ϕ			σ^2		
	$\rho_{\phi\sigma^2}$	.6	0	−.6	.6	0	−.6	.6	0	−.6	.6	0	−.6	.6	0	−.6	.6	0	−.6
$N = 100$	MLpc1	.010	.003	.012	−.011	−.025	−.039	.001	−.002	.000	.961	.954	.946	.904	**.690**	**.397**			
$T = 50$	MLpc2	.010	.003	.012	−.010	−.025	−.038	.001	−.002	.000	.959	.951	.946	.910	**.702**	**.407**			
	MLuc	.011	.003	.012	.016	.001	−.013	.007	.003	.003	**1.00**	**1.00**	**1.00**	**.827**	.932	**.888**			
	B1	.013	.005	.014	.001	.000	−.001	.008	.006	.006	.966	.961	.948	.955	.950	.955	.943	.945	.945
	B2	.011	.003	.013	.016	.001	−.013	.007	.003	.003	.967	.963	.952	**.864**	.944	**.898**	**.704**	**.680**	**.723**
$T = 20$	MLpc1	.004	−.006	.006	−.052	−.063	−.081	−.018	−.017	−.007	.950	.932	.946	**.477**	**.295**	**.117**			
	MLpc2	.004	−.006	.006	−.048	−.059	−.078	−.017	−.020	−.007	.949	.930	.943	**.534**	**.341**	**.149**			
	MLuc	.004	−.008	.006	.020	.007	−.012	.003	.000	.005	**1.00**	**1.00**	**1.00**	**.829**	.905	**.877**			
	B1	.008	−.003	.010	.002	.002	−.004	.009	.006	.016	.954	.942	.950	.957	.945	.932	.955	.950	.941
	B2	.005	−.005	.008	.017	.005	−.014	.004	.000	.008	.954	.940	.949	.903	.941	.900	**.821**	**.824**	**.812**
$T = 10$	MLpc1	.005	−.003	.005	−.126	−.139	−.154	−.074	−.065	−.055	.939	.950	.953	**.070**	**.039**	**.014**			
	MLpc2	.005	−.004	.005	−.112	−.124	−.139	−.072	−.064	−.055	.937	.949	.954	**.139**	**.076**	**.040**			
	MLuc	.004	−.003	.007	.041	.026	.009	.009	.010	.010	**1.00**	**1.00**	**1.00**	**.696**	**.810**	**.860**			
	B1	.010	.002	.009	.007	.002	−.005	.012	.021	.021	.952	.957	.960	.936	.950	.940	.945	.947	.954
	B2	.008	.000	.007	.020	.006	−.012	−.007	.000	.000	.952	.955	.959	**.889**	.929	.923	**.870**	**.887**	**.886**
$N = 50$	MLpc1	.007	.004	.007	−.011	−.025	−.038	−.002	.008	.003	.951	.957	.941	.923	**.847**	**.662**			
$T = 50$	MLpc2	.007	.004	.007	−.011	−.024	−.038	−.002	.008	.006	.950	.956	.937	.926	**.853**	**.674**			
	MLuc	.007	.005	.006	.016	.001	−.013	.003	.012	.006	**1.00**	**1.00**	**1.00**	**.875**	.937	.900			
	B1	.008	.005	.008	.001	−.001	−.002	.011	.022	.016	.965	.963	.959	.957	.954	.954	.933	.935	.931
	B2	.008	.005	.008	.015	.001	−.013	.005	.014	.008	.959	.962	.954	.912	.955	.923	**.691**	**.657**	**.660**
$T = 20$	MLpc1	.000	.010	.007	−.052	−.066	−.083	−.013	−.009	−.010	.954	.953	.946	**.692**	**.558**	**.387**			
	MLpc2	.000	.010	.007	−.048	−.062	−.079	−.013	−.008	−.010	.952	.949	.944	**.723**	**.599**	**.421**			
	MLuc	.001	.011	.004	.019	.004	−.014	.006	.008	.003	**1.00**	**1.00**	**1.00**	**.859**	.908	**.898**			
	B1	.002	.010	.008	.000	−.003	−.010	.028	.030	.027	.961	.961	.955	.949	.944	.958	.936	.945	.948
	B2	.001	.010	.007	.017	.002	−.016	.011	.010	.007	.957	.955	.950	.910	.936	.934	**.816**	**.822**	**.814**
$T = 10$	MLpc1	.002	.011	.016	−.130	−.143	−.156	−.070	−.056	−.060	.942	.952	.936	**.315**	**.214**	**.155**			
	MLpc2	.003	.011	.016	−.115	−.128	−.141	−.068	−.055	−.059	.940	.947	.930	**.415**	**.300**	**.240**			
	MLuc	.005	.011	.018	.036	.020	.006	−.068	.020	.008	**1.00**	**1.00**	**1.00**	**.796**	**.848**	**.863**			
	B1	.006	.011	.017	−.001	−.009	−.015	.045	.056	.048	.957	.959	.952	.954	.946	.938	.952	.943	.936
	B2	.004	.010	.015	.018	.003	−.011	.003	.013	.005	.956	.956	.948	.913	.937	.932	**.893**	**.889**	**.862**
$N = 20$	MLpc1	.015	.015	−.015	−.012	−.025	−.040	.004	.002	.005	.942	.919	.935	.928	.902	**.804**			
$T = 50$	MLpc2	.015	.015	−.015	−.011	−.024	−.040	.004	.002	.005	.942	.919	.935	.928	.902	**.806**			
	MLuc	.015	.017	−.014	.015	.001	−.016	.009	.006	.008	**1.00**	**1.00**	**1.00**	**.893**	.935	.913			
	B1	.013	.014	−.016	.005	.003	−.002	.036	.035	.038	.965	.942	.951	.965	.965	.972	.947	.951	.956
	B2	.017	.017	−.013	.012	−.001	−.016	.014	.010	.013	.963	.938	.950	.947	.958	.938	**.669**	**.698**	**.700**
$T = 20$	MLpc1	.007	−.001	.012	−.053	−.067	−.083	−.029	−.039	−.023	.938	.921	.942	**.856**	**.795**	**.718**			
	MLpc2	.007	−.001	.012	−.050	−.063	−.079	−.029	−.039	−.023	.937	.919	.938	**.869**	**.816**	**.735**			
	MLuc	.008	.000	.010	.016	.001	−.016	−.010	−.024	−.009	**1.00**	**1.00**	**1.00**	.906	.918	.924			
	B1	.006	−.001	.012	.011	.006	−.002	.062	.047	.064	.957	.949	.962	.976	.955	.965	.963	.953	.962
	B2	.010	.002	.015	.009	−.004	−.020	.002	−.012	.004	.953	.944	.960	.965	.946	.947	**.832**	**.832**	**.816**
$T = 10$	MLpc1	−.005	.017	.014	−.130	−.148	−.157	−.068	−.075	−.054	.943	.935	.924	**.677**	**.585**	**.557**			
	MLpc2	−.005	.017	.013	−.114	−.132	−.141	−.067	−.074	−.053	.938	.934	.925	**.744**	**.652**	**.622**			
	MLuc	.000	.020	.004	.035	.012	.002	.017	−.001	.016	**.994**	**.994**	**.995**	**.837**	**.865**	**.886**			
	B1	−.004	.015	.016	.022	.009	.005	.145	.128	.151	.965	.965	.958	.963	.955	.970	.955	.961	.961
	B2	.000	.019	.020	.008	−.013	−.022	.022	.006	.026	.955	.955	.954	.952	.943	.962	**.894**	**.879**	**.881**

Note. Bias and coverage rates of the fixed effects estimates of the five estimation methods *MLpc1* (maximum likelihood estimation method in which the lagged predictor was person-centered using sample means), *MLpc2* (maximum likelihood estimation method in which the lagged predictor was person-centered using shrinkage estimates from an empty model), *Mluc* (maximum likelihood estimation method in which the lagged predictor was not centered), *B1* (Bayesian estimation method in which the mean, AR-parameter, and innovation variance are modeled as random effects), and *B2* (Bayesian estimation method in which the mean and the AR-parameter are modeled as random effects, while the innovation variance is modeled as a fixced effect). N and T are the number of individuals and the number of timepoints in the generated data respectively, μ is the mean parameter, ϕ is the AR-parameter, σ^2 is the innovation variance, and $\rho_{\phi\sigma}{}^2$ is the correlation between the AR-parameter and the innovation variance. The real values for μ, ϕ, and σ^2 were 10, .2 and 3 respectively. The coverage rates are for the 95% confidence (ML methods) and credibility intervals (Bayesian methods). Values lower than .90 and values of .99 or higher are printed in bold.

TABLE 2
Bias in Variance and Correlation Estimates for Mean and AR-Parameter

		τ^2_μ			τ^2_ϕ			$\rho_{\mu\phi}$		
	$\rho_{\phi\sigma^2_\epsilon}$	.6	0	−.6	.6	0	−.6	.6	0	−.6
$N = 100$	MLpc1	.044	.045	.029	−.001	.000	.000	−.008	−.003	−.004
$T = 50$	MLpc2	.020	.023	.009	−.001	.000	.000	−.008	−.003	−.000
	MLuc	−.048	−.043	−.054	−.002	−.001	.000	−.008	−.004	−.000
	B1	.096	.100	.086	.001	.000	.002	−.006	.001	.003
	B2	.073	.075	.061	−.001	.001	.001	−.003	.002	.001
$T = 20$	MLpc1	.098	.084	.074	.000	.000	.000	.005	.007	.010
	MLpc2	.050	.040	.037	.000	.000	.000	.004	.006	.010
	MLuc	−.079	−.100	−.093	−.001	.000	.002	.004	.006	.011
	B1	.087	.080	.079	.004	.003	.005	.004	.009	.002
	B2	.063	.054	.055	.001	.003	.003	.005	.008	.001
$T = 10$	MLpc1	.211	.191	.175	.000	.000	−.001	.014	.005	−.010
	MLpc2	.156	.144	.138	.000	.000	−.001	.016	−.004	−.003
	MLuc	−.154	−.170	−.174	.002	.005	.007	.012	.006	−.012
	B1	.084	.080	.087	.008	.009	.009	.001	.007	−.001
	B2	.058	.055	.066	.005	.007	.007	.003	.009	−.003
$N = 50$	MLpc1	.024	.039	.052	.000	.000	.000	−.002	−.010	.016
$T = 50$	MLpc2	.000	.017	.032	.000	.000	.000	−.002	−.012	.016
	MLuc	−.070	−.047	−.025	−.001	−.001	.000	−.001	−.010	.016
	B1	.181	.200	.215	.003	.001	.003	.004	−.004	.007
	B2	.116	.134	.149	.000	.001	.001	.001	−.007	.007
$T = 20$	MLpc1	.124	.111	.101	.000	.000	.000	.001	−.003	.012
	MLpc2	.076	.068	.064	.000	.000	.000	.004	.001	.012
	MLuc	−.060	−.059	−.060	.000	.000	.001	.001	−.003	.012
	B1	.226	.217	.216	.006	.006	.003	.006	.007	.017
	B2	.155	.148	.150	.003	.004	.005	−.002	−.002	.013
$T = 10$	MLpc1	.220	.188	.174	.002	.002	.002	−.029	.026	.027
	MLpc2	.166	.143	.137	.001	.002	.002	−.028	.020	.026
	MLuc	.166	−.183	−.183	.001	.006	.008	−.028	.023	.028
	B1	.228	.201	.206	.011	.012	.012	.003	.006	.006
	B2	.143	.120	.124	.009	.010	.011	−.008	−.005	−.006
$N = 20$	MLpc1	−.016	.025	.067	.000	.000	.000	−.007	−.017	−.010
$T = 50$	MLpc2	−.040	.003	.047	.000	.000	.000	−.006	−.018	−.010
	MLuc	−.102	−.059	−.017	−.001	.000	.000	−.007	−.017	−.010
	B1	.423	.469	.515	.007	.006	.006	−.011	−.016	−.015
	B2	.308	.355	.402	.004	.004	.004	.007	−.004	−.001
$T = 20$	MLpc1	.110	.123	.085	.002	.002	.003	.016	−.000	.001
	MLpc2	.063	.081	.048	.002	.002	.003	.013	−.002	−.000
	MLuc	−.079	−.051	−.086	.001	.002	.004	.015	−.000	−.000
	B1	.515	.536	.489	.013	.012	.015	−.018	−.017	−.011
	B2	.393	.414	.373	.010	.011	.012	.005	.003	.007
$T = 10$	MLpc1	.204	.190	.165	.007	.006	.006	.005	−.002	.056
	MLpc2	.149	.146	.128	.006	.006	.006	−.014	.013	.059
	MLuc	−.248	−.227	−.222	.008	.010	.011	.006	−.003	.056
	B1	.534	.519	.489	.023	.022	.022	−.022	−.017	−.015
	B2	.402	.373	.367	.019	.020	.020	.003	.003	.008

Note. Bias in the variance and correlation estimates for μ and ϕ of the five estimation methods *MLpc1* (maximum likelihood estimation method in which the lagged predictor was person-centered using sample means), *MLpc2* (maximum likelihood estimation method in which the lagged predictor was person-centered using shrinkage estimates from an empty model), *Mluc* (maximum likelihood estimation method in which the lagged predictor was not centered), *B1* (Bayesian estimation method in which the mean, AR-parameter, and innovation variance are modeled as random effects), and *B2* (Bayesian estimation method in which the mean and the AR-parameter are modeled as random effects, while the innovation variance is modeled as a fixced effect). Where N and T are the number of individuals and the number of timepoints in the generated data respectively, ${\tau^2}_\mu$ is the variance of the mean parameter, ${\tau^2}_\phi$ is the variance of the AR-parameter, $\rho_{\mu\phi}$ is the correlation between the mean parameter and the AR-parameter, and ${\rho_{\phi\sigma}}^2$ is the correlation between the AR-parameter and the innovation variance. The real values for ${\tau^2}_\mu$ are 2.188, 2.160, and 2.131 for ${\rho_{\phi\sigma}}^2$ values of .6, 0, and -.6 respectively. The real values for ${\tau^2}_\phi$ and $\rho_{\mu\phi}$ were always equal to .01 and 0 respectively.

TABLE 3
Bias in Variance and Correlation Estimates for the Innovation Variance

		$\tau^2_{\sigma^2}$			$\rho_{\mu\sigma^2}$			$\rho_{\phi\sigma^2}$		
	$\rho_{\phi\sigma^2}$	.6	0	−.6	.6	0	−.6	.6	0	−.6
$N = 100$	T = 50	.037	.026	.038	.004	.005	.004	−.124	.004	.127
	T = 20	.021	.025	.039	.007	.002	.003	−.286	.018	.300
	T = 10	.005	.003	.015	.004	.001	−.002	−.427	.030	.456
$N = 50$	T = 50	.078	.061	.088	.001	−.011	.004	−.210	.008	.214
	T = 20	.076	.072	.093	.001	.007	.002	−.377	.004	.392
	T = 10	.074	.088	.108	−.003	.005	.000	−.492	.028	.530
$N = 20$	T = 50	.286	.270	.289	.005	−.000	−.007	−.335	−.007	.355
	T = 20	.313	.313	.313	−.002	.002	−.006	−.466	.021	.502
	T = 10	.146	.477	.495	−.004	−.006	.002	−.514	.029	.581

Note. Bias in the variance and correlation estimates of estimation methods B1 (Bayesian estimation method in which the mean, AR-parameter, and innovation variance are modeled as random effects) for σ^2. Where *N* and *T* are the number of individuals and the number of timepoints in the generated data respectively, ${\tau^2}_\sigma{}^2$ is the variance of the innovation variance, $\rho_{\mu\sigma}{}^2$ is the correlation between the mean parameter and the innovation variance, and $\rho_{\phi\sigma}{}^2$ is the correlation between the AR-parameter and the innovation variance. The real value of $\rho_{\phi\sigma}{}^2$ differs between scenarios as indicated in the Table. The real values for $\tau_\sigma{}^2$ and $\rho_{\mu\sigma}{}^2$ were always equal to 1 and 0 respectively.

With respect to the bias in $\rho_{\mu\phi}$, the five methods performed quite similarly when $N = 100$ and $T = 50$. When T decreases, the bias obtained with the ML methods increases. Of the Bayesian methods, method *B2* seemed least affected by changes in *N* and/or *T*, while method *B1* performed quite similarly when *N* is 100 or 50, but clearly performs less well when $N = 20$, especially when this is also combined with a smaller T.

Table 3 shows that method *B1* resulted in little bias in the estimation of $\rho_{\mu\sigma^2}$. In addition, it can be seen that the estimates of $\tau^2_{\sigma^2}$ generally show positive bias, with the amount of bias being mostly affected by *N*, while *T* has little effect. For $\rho_{\phi\sigma^2}$ there generally is a bias towards zero (i.e., a positive bias when the correlation is negative and a negative bias when the correlation is positive). For this random effect, the amount of bias is especially affected by *T*, while decreasing *N* also has a detrimental, although less stark, effect.

Individual Parameters

In addition to the model parameters discussed above, we also consider the individual parameter estimates. That is, for each individual, an estimate of μ_i and ϕ_i can be obtained, and in case of method *B1*, also an estimate of σ_i^2. In the Bayesian analyses, we obtained posterior distributions of both the model parameters (as discussed above) and the individual parameters. From these posterior distributions, we can obtain point estimates as well as CIs. For the ML analyses, individual parameter estimates can be obtained after the analysis is run, using the function `ranef()` from the R-package `lme4` to obtain individual point estimates, and the function `se.ranef()` from the R-package `arm` (Gelman et al., 2011), to obtain the individual standard errors with which to construct the individual CIs.

First, we computed the correlation between the estimated and the true individual parameter values in order to see to what extent the rank order of individuals would be correct if these estimates were used. In order to save space, these results are only briefly summarized here.[9] The correlations obtained for the ML estimation methods were always higher than the ones obtained with the Bayesian methods. For μ_i, the minimum correlation obtained with the centered ML methods was .89 for both method MLpc1 and MLpc2, while the minimum correlation obtained with the MLuc method was equal to .87. The lowest correlations with the Bayesian methods for this estimate were .78 for method *B1* and .79 for method *B2*. In addition, when $T \geq 20$ all three the ML estimation methods show correlations between the real and estimated values of μ_i that are higher than .90, while the Bayesian methods need 50 time points for this correlations to exceed .90.

For ϕ_i, the correlations between the estimated and the true individual values were much lower than for μ_i. The maximum correlation obtained with the three ML methods was .53 (when $N = 100$ and $T = 50$), while the minimum correlation was .06 (at $N = 20$ and $T = 10$). The maximum correlation for the Bayesian methods was .43 (for method *B1* at $N = 100$ and $T = 50$), while the minimum value was .05 (for method *B2* when $N = 100$ and $T = 10$). Estimation method *B1* also provided correlations between the true and estimated values of σ_i^2: The maximum correlation was .74 (when $N = 100$ and $T = 50$), and the minimum correlation was .32 (when $N = 20$ and $T = 10$).

Taken together, the results show that for the ML estimation methods, the correlations between the true and estimated values of μ_i and ϕ_i increase as *N* and (particularly) *T* increases. For the Bayesian estimation methods, these correlations also increase when *T* increases, however the effect of an increase

[9]Tables containing these correlations and the individual coverage rates can be obtained from the first author.

in N is less consistent. The correlations obtained with method *B1* tend to decrease with N, except for the estimates obtained when $\rho_{\phi\sigma^2} = 0$. If the AR-parameter and innovation variance are not correlated, the correlation between the true and estimated values of the parameters tend to increase when N goes from 50 to 20. For method *B2* this trend of increased performance at lower sample sizes is even stronger, and the correlations tend to increase as N decreases regardless of the value of $\rho_{\phi\sigma^2}$.

Second, we considered the coverage rates of the individual 95% CIs. With respect to these, the ML methods performed less well than the Bayesian methods. The two centered ML methods always led to coverage rates below .90 for both ϕ_i and μ_i. Notably, the coverage rates obtained with *MLuc* for ϕ_i were also rather low (i.e., always below .90), and quite similar to the the ones obtained with the other two ML methods. The coverage rates for μ_i obtained with *MLuc* were acceptable however (i.e., always above .90). In contrast, the Bayesian methods resulted in coverage rates of the individual CIs that were always above .90.

Taken together, these results indicate that the ML estimation methods were a little better at retaining the rank order of the individual estimates with respect to ϕ_i, while the Bayesian methods were better for making individual inferences for all individual parameters.

Conclusion

The first aim of the simulation study was to determine whether ignoring the randomness in innovation variance leads to bias in the estimation of the other parameters, particularly the AR-parameter. Because the observed variance is a function of both the innovation variance and the AR-parameter, it could be expected that ignoring randomness in the former leads to problems concerning the latter. The second aim was to determine the tradeoff between T and N.

We compared five estimation methods, of which only one included the innovation variance as a random effect. When comparing the results from these five methods, we can conclude the following. First, person-mean centering the pre-

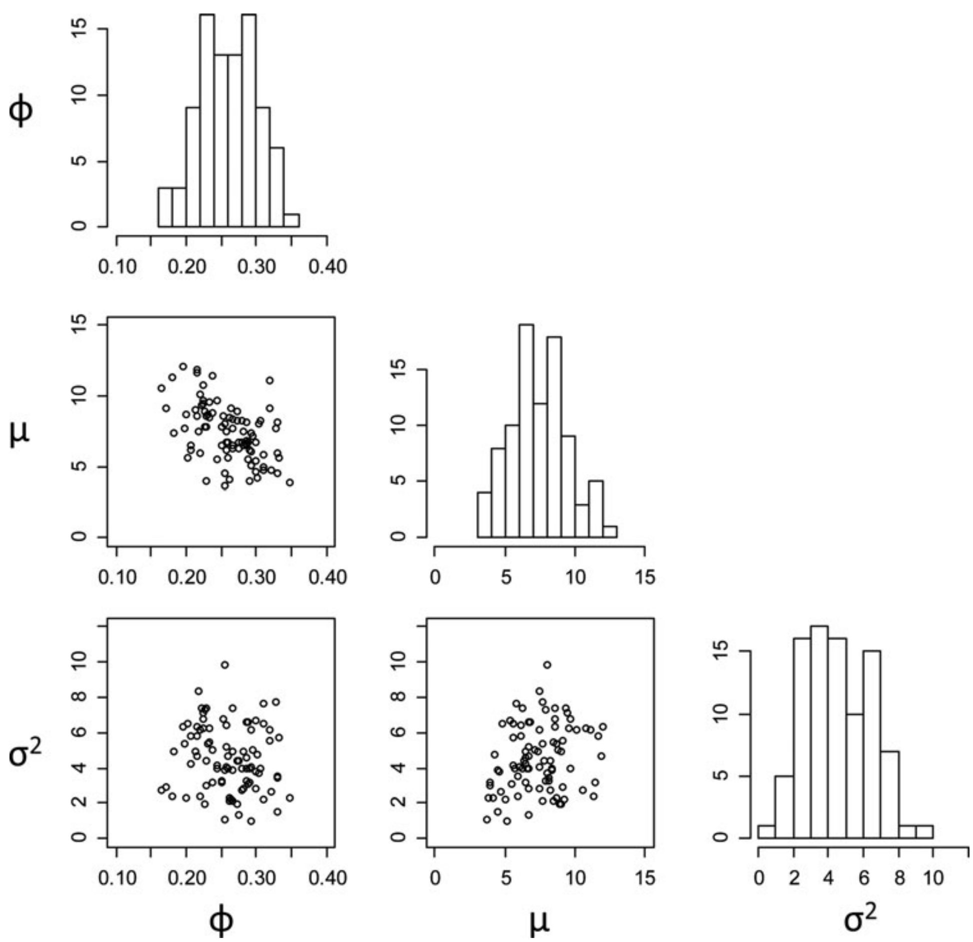

FIGURE 2 The histograms show the estimated posterior distributions of the three random parameters ϕ (the AR-parameter), μ (the mean parameter), and σ^2 (the innovation variance). The scatterplots show the bivariate relation between the random variables of the corresponding row and column (e.g, the scatterplot on the second row of the first column shows the relation between μ and ϕ).

TABLE 4
Results from the Empirical Application of the Multilevel AR(1) Model to Data from Laurenceau et al. (2005)

	Parameter Estimate	95% Credibility Interval
Fixed Effects		
μ	7.406 (.230)	6.955–7.857
ϕ	.260 (.020)	.220–.300
σ^2	4.533 (.249)	4.052–5.034
Random Effects		
τ^2_μ	4.442 (.711)	3.260–6.026
τ^2_ϕ	.008 (.004)	.002–.018
$\tau^2_{\sigma^2}$	4.389 (.880)	2.948–6.391
$\tau_{\mu\phi}$	−.234 (.216)	−.625–.216
$\tau_{\mu\sigma^2}$	.199 (.115)	−.036–.417
$\tau_{\phi\sigma^2}$	−.057 (.246)	−.532–.419

Note. The table shows the parameter estimates and 95% Credibility Intervals of the Bayesian analysis of the daily positive affect data from Laurenceau et al. (2005). The standard deviations of the posterior distributions of the parameters are given between brackets. μ is the mean parameter, ϕ is the AR-parameter, σ^2 is the innovation variance, ${\tau^2}_\mu$ is the variance of the mean parameter, ${\tau^2}_\phi$ is the variance of the AR-parameter, ${\tau^2}_{\sigma}{}^2$ is the variance of the innovation variance, $\tau_{\mu\phi}$ is the correlation between the mean parameter and the AR-parameter, and $\tau_{\mu\sigma}{}^2$ is the correlation between the AR-parameter and the innovation variance, and $\tau_{\phi\sigma}{}^2$ is the correlation between the AR-parameter and the innovation variance.

dictor leads to considerable bias in the estimation of ϕ, and should therefore not be used (see, for a more elaborate discussion, Hamaker & Grasman, 2013). Second, when comparing the results obtained from methods *MLuc* and *B2* to the results obtained with estimating the true model (method *B1*), we can conclude that ignoring the randomness in innovation variance leads to bias in the estimation of ϕ. The direction of this bias depends on the actual correlation between ϕ_i and σ_i^2: When there is a positive correlation, ϕ tends to be overestimated, and when there is a negative correlation, ϕ tends to be underestimated. These results are also reflected by coverage rates that regularly drop below .90. Third, including the innovation variance as a random effect (i.e., method *B1*) is not associated with a specific pattern of bias, and in comparison to the other methods, the estimated bias for ϕ is very small, although the random effects are generally overestimated. Furthermore, the coverage rates obtained with method *B1* are always above .90, and are often close to the target value .95. This is also true for the coverage rates of the individual CIs. The only criterion on which the ML methods outperformed method *B1* was with respect to the correlation between the true and estimated individual ϕ_is.

Focusing on the effect of sample size on the results obtained with method *B1*, we can conclude the following. For the fixed effects, the bias increases when either N or T decreases, while the bias for the random effects, which is always positive, seems more strongly affected by N than by T. Still, the coverage rates were always above .90, indicating that this approach can be effectively used for making inferences even with small sample sizes such as $N = 20$ and $T = 10$.

EMPIRICAL APPLICATION

To further illustrate the Bayesian estimation of a multilevel AR(1) model, we apply estimation method B1 to data collected in a study by Laurenceau et al. (2005). In this study, spouses from 96 married couples independently completed a structured diary each evening over a period of 42 consecutive days. Based on the partial overlap in the affective items in this data set and the items of the PANAS-X (Watson & Clark, 1999), we selected four items (i.e., excited, enthusiastic, energetic, and happy rated on 5-point Likert scales), to comprise a single positive affect (PA) score. Focusing on the women only, there were 127 out of the total of 4,032 ($= 96 \times 42$) PA scores missing. Based on individual sequence plots (i.e., plots of the repeated measurements of each woman), we removed seven women who had none or very little variability over time, such that the final data set contained 89 female participants.

To analyze the data using method B1, we began by analyzing the data using standard ML analysis with person centering based on observed mean scores and listwise deletion to get estimates for τ_μ^2 and τ_ϕ^2, which are needed for the scale matrix of the IW prior. Next, the scores of the 89 females were analyzed using estimation method B1. To evaluate whether the analysis converged, we ran three separate MCMC chains with different starting values and considered the mixing of the trace plots and the values of the Gelman-Rubin statistic for each parameter. Starting values for the fixed effects and the covariance matrix were based on random draws from a standard normal distribution (chain 1 and 2), or based on the ML analysis (chain 3). We used a burn-in of 5,000 iterations and total number of 10,000 iterations. Following initial convergence checks, we decided to use a thinning rate of 10. As a result, we ran the analysis for a total of 100,000 iterations ($10{,}000 \times 10$). With these settings, the analysis of the trace plots and the Gelman-Rubin statistic indicated convergence.

The results obtained with method B1 are summarized in Table 4 and Figure 2. The first column contains the point estimates (i.e., the means of the posterior distribution) and the standard deviation between parentheses (i.e., the standard deviation of the posterior distribution), while the second column contains the lower and upper bounds of the 95% CIs. Since the 95% CI for ϕ lies above zero, we can conclude that–on average–the women are characterized by a carryover of yesterday's PA on today's PA.

The point estimate of the variance of the mean (i.e., τ_μ^2) is equal to 4.442 (95% CI ranges from 3.260 to 6.026) indicating there is considerable variation in the average PA of individuals over time. The point estimate of the variance of the inertia (i.e., τ_ϕ^2) is .008 (95% CI ranges from .002 to .018). While this may seem like a small variance, it should be noted that the ϕ parameter itself is likely to be small as for stationary processes it must lie in the range of –1 to 1; in practice it will be much more often between 0 and .5 or so (cf. Wang et al., 2012). The point estimate of the variance of the innovation variance (i.e., $\tau^2_{\sigma^2}$) is 4.389 (95% CI ranging from 2.948 to 6.391), suggesting there is considerable between-person variation in this source of variance. This corresponds to the idea that individuals differ in their sensitivity, reactivity, and exposure to external events that influence the process under investigation. Here, it seems to imply that individuals differ in the amount and/or severity of positive and negative events that they encounter in daily life, as well as their sensitivity and reactivity to such events. Note that the 95% CIs of the variances cannot include zero because we are using an IW prior, such that we cannot use the CIs as an informal test of whether the parameter should be considered to differ from zero. However, since the lower bounds are (relatively) far away from zero, we believe it is safe to conclude that all three parameters are characterized by a meaningful level of individual differences.

Finally, when considering the covariances between the random effects, each of the CIs includes zero, such that we cannot conclude that these parameters are truly different from zero. This would imply that the unobserved factors that influence the individuals' means, their inertias, and their innovation variances do not overlap. For example, if the trait extraversion were to have a positive effect on the average PA level of individuals (and thus be predictive of μ_i), it is unlikely to affect the individuals' inertia or their exposure and/or reactivity to time-varying factors that influence PA (i.e., ϕ_i and σ_i^2). Although we have not considered the CIs for correlations or covariances in the current study in detail, preliminary results suggested to us that these tend to be too wide, such that they may not be that appropriate for the current purpose. Thus, at this stage it is too early to conclude that the individual means, inertias, and innovation variances are affected by the same factors, even though we have found no evidence for relatedness between these random effects.

DISCUSSION

In this article, we presented a multilevel extension of the AR(1) model and compared several ways to estimate it. The model we considered here is more extensive than typically considered in the literature as it includes a random (rather than fixed) innovation variance as well as a random autoregressive parameter. We argued that there are both substantive and statistical reasons for preferring this extended multilevel AR(1) model. First, between-person differences in innovation variances may form an important source of information. The innovation can be conceptualized as a collection of all unobserved temporal factors that influence the process under investigation, both internally (e.g., hormonal levels, alcohol intake, cognitions, associations, appraisal of events) and externally (e.g., social obligations, personal interactions, weather, political developments). Allowing for individual differences in the innovation variance implies we allow not only for individual differences in sensitivity and/or responsiveness to these factors, but also for individual differences in exposure to these factors (or, more specifically, individual differences in the variability of these factors).

Second, using a simulation study, we showed that when the innovation variance is in fact random, ignoring this source of individual differences leads to bias in the estimation of the AR-parameter (where the direction of the bias depends on the correlation between the innovation variance and the AR-parameter). This can be explained by the fact that the variance of an AR(1) process is a function of both the innovation variance and the AR-parameter, and when one of these is fixed across individuals, the other is the only random source that can account for individual differences in observed variance. The impact, or cost, of this bias in the AR-parameter depends on the amount of autocorrelation. Our simulation study showed that the maximum bias is likely around –.12, so if the true value of the AR parameter is far away from 0, the consequences of the bias are probably not that severe. If the true value is close to 0 however, the bias could change the estimate from positive to negative. The latter possibility is a more severe problem, as a negative AR parameter describes a qualitatively different process than a positive one.

Based on these arguments, we advise researchers interested in applying a multilevel AR(1) model to use an approach that allows for the inclusion of the innovation variance as a random effect. This can be done in WinBUGS, which has the additional advantage that it allows for defining the AR(1) process in two equations, such that we can estimate the individual mean rather than the intercept. The mean has a more meaningful interpretation in terms of an individual's long-term tendency (i.e., it is the score a person would turn to if there would be no more random input to the process), whereas the intercept is generally less meaningful (i.e., the expected score when the score at the preceding occasion is equal to zero). Furthermore, while person-mean centering the AR predictor implies that the intercept at level 1 becomes the individual's mean on the dependent variable, it should be discouraged as it leads to a negative bias in the estimation of the AR-parameter (as has been shown in the simulation study).

The results from our simulation study also indicated that it was difficult to retain the rank order in the individual innovation variances and the individual AR-parameters (especially if fewer than 50 time points are available). Future research should focus on how well between-person differences in innovation variances and AR-parameters can be predicted at the second level using person characteristics. This is particularly important because regressing the innovation variance at the second level on measurements of, for instance, sensation seeking behavior, neuroticism, and/or sensitivity (e.g., sensory-processing sensitivity, Aron & Aron, 1997) should help determine what factors play a role in the (individual differences in) variability of a particular process. Such an approach could also be used as a first step in determining which factors should be considered as candidates to be included as level 1 predictors in subsequent studies in order to model their effects on the process more explicitly.

Also, it should be noted that we focused on one particular form of heterogeneity in this article: inter-individual variability. However, researchers might also be interested in other forms of variability, like variability in parameters across time, or (qualitative) differences in the kind of process that best describes the repeated measurements of individuals. As an example of the first of these other types of variability, one could think of a situation in which the process under investigation depends on the current state of an individual, with different states leading to different parameter values (e.g., different amounts of inertia). Data from this type of scenario can be analyzed with Treshold Autoregressive (TAR) models (De Haan-Rietdijk, Gottman, Bergeman, & Hamaker, 2014), in which additional variability in model parameters is possible through regime switching. An example of the second alternative type of variability would be a sample in which the repeated measures of some individuals can be characterized by an AR(1) process, while others may be better described by an AR(2). In this case, researchers might choose to run separate analyses for each individual in the sample, or use a mixed model approach in which different (level 1) processes are allowed for different individuals within the larger multilevel model.

Note that these alternative types of variability can be combined with the form of inter-individual variability examined in this article. The TAR model could be extended by allowing different amounts of innovation variance for different individuals for example, with the amount of innovation variance of each individual also varying across time. This could be a important extension since the reason that erroneously modeling the innovation variance as fixed leads to bias in the AR-paramater likely applies to this type of model as well. Similarly, random model parameters (i.e., the innovation variance) can also be incorporated into mixed models that allow different types of processes for different individuals. This can be done by specifying random innovation variances for every individual in the sample to prevent bias in the parameter estimates of these mixed models, or by specifying random innovation variances for a subset of the sample to distinguish between individuals for who all factors of interest are explicitly modeled, and those for who some factors are still unknown or unmeasured.

ARTICLE INFORMATION

Conflict of Interest Disclosures: Each author signed a form for disclosure of potential conflicts of interest. Dr. Jongerling and Dr. Hamaker report a grant from the Netherlands Organization for Scientific Research (NWO; VIDI Grant 452-10-007), during the conduct of the study.

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 supported by Grant 452-10-007 from the Netherlands Organization for Scientific Research (NWO).

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.

Acknowledgments: The authors would like to thank Herbert Hoijtink for his advice during the preparation of this manuscript. The ideas and opinions expressed herein are those of the authors alone, and endorsement by the authors' institutions or the NWO is not intended and should not be inferred.

REFERENCES

Aron, E., & Aron, A. (1997). Sensory-processing sensitivity and its relation to introversion and emotionality. *Journal of Personality and Social Psychology*, *73*, 345–368.

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes [Computer software manual]. Retrieved from http://CRAN.R-project.org/package=lme4 (R package version 0.999375-39).

Bollen, K., & Curran, P. (2004). Autoregressive latent trajectories (ALT) model: A synthesis of two traditions. *Sociological Methods and Research*, *32*, 336–383.

Bollen, K., & Curran, P. (2006). *Latent curve models: A structural equation perspective*. New York, NY: Wiley.

Bryk, A., & Raudenbush, S. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.

Chatfield, C. (2003). *The analysis of time series: An introduction* (6th ed.). Washington, DC: Chapman & Hall/CRC.

De Haan-Rietdijk, S., Gottman, J., Bergeman, C., & Hamaker, E. L. (2014). Get over it! A multilevel threshold autoregressive model for state-dependent affect regulation. *Psychometrika*. doi:10.1007/s11336-014-9417-x

Enders, C., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*, 121–138.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2004). *Bayesian data analysis*. New York, NY: Chapman & Hall.

Gelman, A., Su, Y.-S., Yajima, M., Hill, J., Pittau, M. G., Kerman, J., & Zheng, T. (2011). arm: Data analysis using regression and multilevel/hierarchical models [Computer software manual]. Retrieved from http://CRAN.R-project.org/package=arm (R package version 1.4-13).

Hamaker, E. L. (2012). Why researchers should think within-person: A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life*. New York, NY: Guilford.

Hamaker, E. L., & Grasman, R. P. P. P. (2013). Studying inertia in the multilevel autoregressive model: Should the lagged predictor be cluster mean centered? [Submitted for publication].

Hamaker, E. L., & Klugkist, I. (2011). Bayesian estimation of multilevel models. In K. Roberts & J. Hox (Eds.), *Handbook of advanced multilevel analysis* (pp. 137–161). New York: Taylor and Francis.

Hamilton, J. (1994). *Time series analysis*. Princeton, NJ: Princeton University.

Hoijtink, H. (2009). Bayesian data analysis. In A. Millsap & R. E. Maydeu-Olivares (Eds.), *The Sage handbook of quantitative methods in psychology* (pp. 423–443). London, UK: Sage.

Hox, J. (2002). *Multilevel analysis: Techniques and applications* (2nd ed.). Mahwah, NJ: Erlbaum.

Koval, P., & Kuppens, P. (2012). Changing emotion dynamics: Individual differences in the effect of anticipatory social stress on emotional inertia. *Emotion*, *12*, 256–267.

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

Kreft, I., de Leeuw, J., & Aiken, L. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research*, *30*, 1–21.

Kuppens, P., Allen, N., & Sheeber, L. (2010). Emotional inertia and psychological maladjustment. *Psychological Science*, *21*, 984–991.

Kuppens, P., Sheeber, L. B., Yap, M. B., Whittle, S., Simmons, J. G., & Allen, N. B. (2012). Emotional inertia prospectively predicts the onset of depressive disorder in adolescence. *Emotion*, *12*, 283–289.

Laurenceau, J. P., Feldman Barrett, L., & Rovine, M. J. (2005). The interpersonal process model of intimacy in marriage: A daily-diary and multilevel modeling approach. *Journal of Family Psychology*, *19*, 314–323.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). Winbugs, a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, *10*, 325–337.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107–122.

Mood, A. M., Graybill, F. A., & Boes, D. C. (1985). *Introduction to the theory of statistics*. London, UK: McGraw-Hill.

Nesselroade, J. R. (1991). Interindividual differences in intraindividual change. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions*. Washington, DC: American Psychological Association.

O'Hagan, A. (1995). Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society Series B (Methodological)*, *57*, 99–138.

Rasbash, J., Charlton, C., Browne, W., Healy, M., & Cameron, B. (2009). *MLwiN Version 2.1* [Computer software manual]. Centre for Multilevel Modelling, University of Bristol.

Rottenberg, J. (2005). Mood and emotion in major depression. *Current Directions in Psychological Science*, *14*, 167–170.

Rovine, M. J., & Walls, T. A. (2006). A multilevel autoregressive model to describe interindividual differences in the stability of a process. In J. L. Schafer & T. A. Walls (Eds.), *Models for intensive longitudinal data*. New York, NY: Oxford University Press.

Schmittmann, V. D., Dolan, C. V., Van der Maas, H. L. J., & Neale, M. C. (2005). Discrete latent Markov models for normally distributed response data. *Multivariate Behavioral Research*, *40*, 461–488.

Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software*, *12*, 1–16. Retrieved from http://www.jstatsoft.org

Suls, J., Green, P., & Hillis, S. (1998). Emotional reactivity to everyday problems, affective inertia, and neuroticism. *Personality and Social Psychology Bulletin*, *24*, 127–136.

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

Watson, D., & Clark, L. A. (1999). The PANAS-X: Manual for the Positive and Negative Affect Schedule - Expanded Form [Computer software manual].

Wichers, M., Barge-Schaapman, D. Q. C. M., Nicolson, N. A., Peeters, F., de Vries, M., Mengelers, R., & van Os, J. (2009). Reduced stress-sensitivity or increased reward experience: The psychological mechanism of response to antidepressant medication. *Neuropsychopharmacology*, *34*, 923–931.

Wichers, M., Lothmann, C., Simons, C. J. P., Nicolson, N. A., & Peeters, F. (2012). The dynamic interplay between negative and positive emotions in daily life predicts response to treatment in depression: A momentary assessment study. *British Journal of Clinical Psychology*, *51*, 206–222.

Wichers, M., Peeters, F., Geschwind, N., Jacobs, N., Simons, C. J. P., Derom, C., . . . van Os, J. (2010). Unveiling patterns of affective responses in daily life may improve outcome prediction in depression: A momentary assessment study. *Journal of Affective Disorders*, *124*, 191–195.

Wichers, M., Peeters, F., Rutten, B. P. F., Jacobs, N., Derom, C., Thiery, E., . . . van Os, J. (2012). A time-lagged momentary assessment study on daily life physical activity and affect. *Health Psychology*, *31*, 135–144.