To center or not to center? Investigating inertia with a multilevel autoregressive model

Ellen L. Hamaker[1]* and Raoul P. P. P. Grasman[2]

[1] Methodology and Statistics, Faculty of Social and Behavioural Sciences, Utrecht University, Utrecht, Netherlands
[2] Psychological Methods, University of Amsterdam, Amsterdam, Netherlands

Whether level 1 predictors should be centered per cluster has received considerable attention in the multilevel literature. While most agree that there is no one preferred approach, it has also been argued that cluster mean centering is desirable when the within-cluster slope and the between-cluster slope are expected to deviate, and the main interest is in the within-cluster slope. However, we show in a series of simulations that if one has a multilevel autoregressive model in which the level 1 predictor is the lagged outcome variable (i.e., the outcome variable at the previous occasion), cluster mean centering will in general lead to a downward bias in the parameter estimate of the within-cluster slope (i.e., the autoregressive relationship). This is particularly relevant if the main question is whether there is on average an autoregressive effect. Nonetheless, we show that if the main interest is in estimating the effect of a level 2 predictor on the autoregressive parameter (i.e., a cross-level interaction), cluster mean centering should be preferred over other forms of centering. Hence, researchers should be clear on what is considered the main goal of their study, and base their choice of centering method on this when using a multilevel autoregressive model.

Keywords: centering, autoregressive models, multilevel models, dynamics, inertia

Longitudinal data are characterized by a nested structure, in which occasions are clustered within individuals. While such data are traditionally analyzed using repeated measures ANOVA, this approach is restrictive in that it requires an equal number of observations for each participant. A further limitation associated with repeated measures ANOVA is that the results pertain to the aggregate and may not be meaningful for any particular individual. A more sophisticated approach—which overcomes these limitations—is *multilevel modeling* (Singer and Willett, 2003; Hox, 2010; Snijders and Bosker, 2012; also known as mixed modeling, see Verbeke and Molenberghs, 2000; hierarchical modeling, see Raudenbush and Bryk, 2002; or random-effects modeling, see Laird and Ware, 1982): This approach can be used for (highly) unbalanced longitudinal data, and it allows for individual trajectories over time. The latter implies we can study between-person (or interindividual) differences in within-person (or intraindividual) patterns of change.

It is not uncommon for the residuals of a longitudinal multilevel model to be autocorrelated, meaning that residuals are related to each other over time. Failing to account for this may bias the estimates of the standard errors, and as a result affect the inferences based on them. Therefore, multilevel software packages include the option to control for autocorrelation through specifying diverse structures for the errors, such as a Toeplitz matrix, or a first order autoregressive process. Alternatively, autocorrelation can be modeled explicitly through the inclusion of the lagged outcome variable (that is, the outcome variable at the previous occasion) as a covariate. Such models have been referred to as (prospective) change models (e.g., Larson and Almeida, 1999),

and are used to investigate the—potentially causal—effect of a (lagged) predictor on the outcome variable, while "controlling" or "adjusting" for the previous level of the outcome variable (e.g., Bolger and Zuckerman, 1995; Gunthert et al., 2007; Moberly and Watkins, 2008; Henquet et al., 2010).

While autocorrelation is typically considered a nuisance in longitudinal multilevel modeling, there are a few multilevel studies that focus specifically on the autoregressive relationship between consecutive observations, and on individual differences therein (cf., Suls et al., 1998; Rovine and Walls, 2006; Kuppens et al., 2010; Koval and Kuppens, 2012; Wang et al., 2012; Brose et al., 2014). The interest in an individual's autoregressive parameter comes from the fact that this parameter is related to the time it takes the individual to recover from a perturbation and restore equilibrium: While an autoregressive parameter close to zero implies that there is little carryover from one measurement occasion to the next and recovery is thus instant, an autoregressive parameter close to one implies that there is considerable carryover between consecutive measurement occasions, such that perturbations continue to have an effect on subsequent occasions. For this reason, the autoregressive parameter can also be considered as a measure of *inertia* or *regulatory weakness*.

Empirical studies have shown that individual differences in inertia in emotions and affect are positively related to neuroticism and depression, in that people higher on neuroticism or depression take longer to restore equilibrium than others (Suls et al., 1998; Kuppens et al., 2010; Wang et al., 2012). In addition, women tend to have higher inertia than men in both their daily affect (Wang et al., 2012), and their daily drinking behavior (Rovine and

Walls, 2006). In a prospective study, Kuppens et al. (2012) showed that affective inertia at age 9–12 was predictive of the onset of depression two and a half years latter, corresponding to the idea that high inertia is reflective of a maladaptive regulation mechanism. Similarly, Wang et al. (2012) showed that inertia is positively related to later detrimental health outcomes. Furthermore, inertia has been shown to be related—but not identical—to rumination (Koval et al., 2012) and perseverative thoughts (Brose et al., 2014), and is positively related to depression even after these related characteristics are taken into account. Taken together, these studies show that inertia is a meaningful individual characteristic that is reflective of a maladaptive regulatory mechanism that is associated with both current *and* future well-being.

To model individual differences in inertia, the above studies all relied on multilevel modeling based on a first-order autoregressive process: In this model, the level 1 predictor is formed by the lagged outcome variable, and its random slope thus represents individual differences in inertia. A pressing question in this context is whether the autoregressive predictor should be centered per person or not. This is a rather fundamental issue, as it is well-known from the multilevel literature that the centering method used for a level 1 predictor (i.e., no centering, centering with the grand mean, or centering per cluster), affects the results (cf. Kreft et al., 1995; Raudenbush and Bryk, 2002; Hox, 2010; Snijders and Bosker, 2012). The consensus seems to be that there is no one preferred method and that the choice should depend on the specific situation and the research question (cf. Kreft et al., 1995; Nezlek, 2001; Snijders and Bosker, 2012). One such specific situation is described by Raudenbush and Bryk (2002), who indicate that if the within-cluster and between-cluster slopes differ, centering per cluster should be preferred, because failing to do so will lead to results that are "uninterpretable" (p. 135). Furthermore, Enders and Tofighi (2007) argue that if there is a clear interest in the within-cluster slope, centering per cluster is recommendable.

With this latter advice in mind, centering the lagged autoregressive predictor per person seems the right approach, because: (a) we are interested in the within-person slope; and (b) we expect the within-person slope to differ from the between-person slope.[1] The aim of the current paper is therefore to investigate whether the advice formulated by Raudenbush and Bryk (2002) and Enders and Tofighi (2007) also applies to the multilevel autoregressive model with a random slope that represents individual differences in inertia. To this end, we begin by presenting the multilevel autoregressive model and discuss its interpretation. To make the model compatible with standard multilevel software, we discuss two parameterizations—based on different centering methods—and we show through an empirical application that these lead to different results for the inertia parameter. In the second section we draw from several key publications in the multilevel literature and discuss the effects of centering a level 1 predictor. The third section contains simulations based on the standard multilevel model to verify some of the claims made in the literature. Additionally, we simulate the multilevel

autoregressive model to investigate how centering affects the estimation of inertia. In the fourth section we apply the insights obtained from the simulation study to the empirical data set. We end by presenting recommendations to the researcher interested in studying inertia using the multilevel autoregressive model, either with or without level 2 predictors.

1. MULTILEVEL AUTOREGRESSIVE MODEL

Many applications of longitudinal multilevel modeling consist of modeling deterministic trajectories over time, for instance a linear or quadratic trend. While such models are extremely useful for studying developmental processes (cf., Curran and Bauer, 2011), they may be less useful when the longitudinal data comprise daily affective or symptom measurements, or affective ratings in an observation study: Then the interest may be not so much in overall trends (as they are likely to be absent from the data), but rather in the dynamics of a stationary process, that is, a process that is characterized by changes *over time*, while these changes are not directly a *function of time*. A promising model for this purpose is the multilevel autoregressive model, which has been successfully applied in an increasing number of studies (e.g., Suls et al., 1998; Rovine and Walls, 2006; Kuppens et al., 2010; Koval and Kuppens, 2012; Wang et al., 2012; Brose et al., 2014).

We begin this section by presenting the multilevel autoregressive model using a parametrization that we consider to be most useful from a substantive viewpoint. However, since this parametrization is not compatible with standard multilevel software, we also present two alternative parametrization of this model, and discuss their advantages and disadvantages. We apply both parameterizations to an empirical data set consisting of daily measurement of positive and negative affect.

1.1. A MODEL TO STUDY INDIVIDUAL DIFFERENCES IN MEAN AND INERTIA

Let y_{ti} be the observation for individual i at occasion t, for instance the person's negative affect or self-esteem measured at a daily basis, with $i = 1, \ldots, N$ and $t = 1, \ldots, T_i$. The most basic model for such nested data would be a model which allows for individual differences in means. At level 1 the observations are then modeled as

$$y_{ti} = \mu_i + a_{ti} \qquad (1)$$

where μ_i represents the individual's mean score, which can be interpreted as his/her trait score or equilibrium, while a_{ti} is the individual's temporal deviation from this equilibrium; and at level 2 the individual means are then modeled as

$$\mu_i = \mu + u_{0i} \qquad (2)$$

where μ is the grand mean, and u_{0i} is the individual's deviation from the grand mean. These deviations are assumed to be normally distributed, that is, $u_{0i} \sim N(0, \sigma_{u0}^2)$.[2]

---

[1] As we will show later on, the between-person slope will always be (essentially) 1 in this model, while the within-person slope is expected to lie between -1 and 1.

[2] Note that the model presented in Equations 1 and 2 corresponds to what is known as a random intercept model or empty model, and is typically considered as one of the options in longitudinal multilevel modeling.

If repeated measures are taken (relatively) close in time, the current measurement is likely to be predictable from the preceding measurement. That is, the individual's deviation from his/her equilibrium at a particular occasion is likely to affect the deviation at the next occasion, which can be expressed as a first order autoregressive model, that is,

$$a_{ti} = \phi_i a_{t-1,i} + e_{ti} \tag{3}$$

where the residuals e_{ti} are independently and identically distributed, with $e_{ti} \sim N(0, \sigma_e^2)$. This residual e_{ti} can be thought of as representing everything that influences the process under investigation: For instance, if we are measuring negative affect, factors that are likely to influence this process include the occurrence of negative or stressful events, the appraisal of these events and the associations and memories that they trigger, but also psychophysiological factors like caffeine or alcohol consumption, et cetera.

The autoregressive parameter ϕ_i relates the outcome variable to itself at the preceding occasion, and thus represents the *inertia* of the person. For an autoregressive process to be stationary, the autoregressive parameter has to lie between -1 and 1 (e.g., Hamilton, 1994). Note however that this does not imply that the autoregressive parameter is truly restricted to this range: Values larger than 1 (or smaller than -1) are possible, but the resulting process would no longer be a stationary process. In psychological research, this parameter typically lies somewhere between 0 and 0.6 (e.g., Rovine and Walls, 2006; Wang et al., 2012), and we are therefore not concerned about boundary constraints when estimating this model.

The individual differences in the autoregressive parameter can be modeled at level 2 as

$$\phi_i = \phi + u_{1i} \tag{4}$$

where ϕ denotes the average autoregressive parameter across people, and u_{1i} denotes the individual's deviation from this average, with $u_{1i} \sim N(0, \sigma_{u1}^2)$. Furthermore, the individuals' means and their autoregressive parameters may be correlated, as represented by the covariance between u_{0i} and u_{1i}, which is denoted as $\sigma_{u0,u1}$. Wang et al. (2012) for instance found a significant positive correlation of 0.40 between the individuals' means μ_i and their autoregressive parameters ϕ_i based on daily measurements of negative affect.

1.2. MAKING THE MODEL COMPATIBLE WITH STANDARD MULTILEVEL SOFTWARE

The model in Equations 1–4 represents the multilevel autoregressive model, where Equations 1 and 3 form level 1, while Equations 2 and 4 form level 2. However, most multilevel software packages do not allow for formulating a level 1 model using more than one equation. We consider two solutions for this.

The first solution consists of specifying the model at level 1 as

$$y_{ti} = c_i + \phi_i^n y_{t-1,i} + e_{ti}, \tag{5}$$

and at level 2 as

$$c_i = \gamma_{00}^n + u_{0i}^n \tag{6}$$

$$\phi_i^n = \gamma_{10}^n + u_{1i}^n, \tag{7}$$

where the superscript n indicates that in this approach *no centering* (NC) was used (i.e., the raw data were used). The relationship between the model specified in Equations 1–4 and the model specified in Equations 5–7 is shown in Appendix 1; however, while the multilevel models presented here are *structurally* the same, the current formulation is based on the assumption that c_i is normally distributed, which necessarily implies that μ_i will not have a normal distribution (as it is a function of c_i and ϕ_i, see Appendix 1). This is detrimental, as we are typically interested in μ_i as representing an individual's average or trait score, and assume these trait scores to be normally distributed in the population. In contrast, c_i is a rather arbitrary score (i.e., the expected score when the individual scored zero on the preceding occasion), that is of limited (or no) substantive interest, and for which we do not have a particular distributional expectation. Also, if we are interested in including predictors at level 2, we would prefer to use these as predictors of μ_i, rather than of c_i.

Therefore, we consider a second solution, which is based on using the individually centered lagged autoregressive predictor $(y_{t-1,i} - \mu_i)$, such that the model at level 1 is

$$y_{ti} = \mu_i + \phi_i^c (y_{t-1,i} - \mu_i) + e_{ti} \tag{8}$$

and at level 2 it is

$$\mu_i = \gamma_{00}^c + u_{0i}^c \tag{9}$$

$$\phi_i^c = \gamma_{10}^c + u_{1i}^c \tag{10}$$

where the superscript c implies that the level 1 predictor was subjected to *cluster mean centering* (CMC; also referred to as within-group or within-person centering). The advantage of the current approach over the previous one is that it results in μ_i and ϕ_i being the random coefficients that are subsequently modeled at level 2. However, it also presents us with a catch-22: To center the lagged predictor, we need an estimate of μ_i, which we actually need to estimate using this model. We will consider several solutions to this problem in our simulations, including the use of the sample mean per person.

1.3. APPLICATION: PART 1

To investigate whether the two approaches proposed above lead to the same or different results for the inertia parameter, we apply the two parameterizations of the multilevel autoregressive model to an empirical data set that was obtained as part of the Dynamics of Dyadic Interactions Project at the University of California, Davis (Ferrer and Widaman, 2008; Ferrer et al., 2012). The data used here consist of daily measurements of relationship specific positive and negative affect. We analyzed these data for men and women separately (sample sizes 193 and 192, respectively), using multilevel autoregressive models with random intercepts (i.e., c_i, based on NC) or means (i.e., μ_i, based on CMC), and random autoregressive parameters (ϕ_i). The estimates for the fixed effects parameters γ_{00} and γ_{10} are presented in **Table 1**.

Table 1 | Results for multilevel autoregressive model (with random effects).

			NC		CMC	
Males	PA	γ_{00}	2.167	[2.044, 2.290]	3.518	[3.425, 3.611]
		γ_{10}	0.387	[0.357, 0.417]	0.353	[0.322, 0.384]
	NA	γ_{00}	0.971	[0.923, 1.020]	1.344	[1.300, 1.389]
		γ_{10}	0.268	[0.235, 0.301]	0.242	[0.208, 0.275]
Females	PA	γ_{00}	2.220	[2.095, 2.346]	3.491	[3.392, 3.590]
		γ_{10}	0.370	[0.340, 0.399]	0.341	[0.311, 0.370]
	NA	γ_{00}	0.978	[0.935, 1.021]	1.348	[1.304, 1.392]
		γ_{10}	0.255	[0.222, 0.288]	0.225	[0.192, 0.258]

Estimates for the fixed effects parameters in a multilevel autoregressive model (with random intercept and slope). The 95% confidence intervals are given between brackets. Estimation was based on using NC or CMC (with the sample means) for the lagged autoregressive predictor. Fixed effects are: (a) γ_{00}, which represents the averaged intercept when using NC, or the grand mean when using CMC; and (b) γ_{10}, which represents the averaged (i.e., fixed effects) autoregressive parameter.

It shows that the parameter estimates obtained with the two models are not identical. This is not surprising as we are already aware that the two parameterizations differ with respect to the meaning of γ_{00}. However, it also shows that the parameter estimates for γ_{10}—which represents the average inertia in both parameterizations—differ from each other. Especially when considering relationship specific PA in males, it can be seen that CMC and NC lead to estimates of the inertia that are not covered by the 95% confidence interval of the alternative parametrization (implying these estimates are relatively different).

The question thus arises, which approach should be preferred—NC or CMC—when the interest is in obtaining an appropriate estimate of the average autoregressive parameter. As this touches upon the more general topic of whether level 1 predictors should be centered or not in multilevel models, we first consult the multilevel literature with respect to centering level 1 predictors.

2. TO CENTER OR NOT TO CENTER: A PERSISTING QUESTION IN MULTILEVEL MODELING

Centering a level 1 predictor in multilevel modeling is a complicated affaire. While there are several sources that provide excellent coverage of this topic (e.g., Kreft et al., 1995; Snijders and Bosker, 2012), it still seems to create much confusion, especially amongst the more novice users. A fundamental issue when dealing with a level 1 predictor is the fact that the relationship between a predictor and an outcome variable may differ across levels. For instance, consider the hypothetical example in the left panel of **Figure 1**, representing the relationship between typing speed and number of typos. This relationship is likely to be positive within individuals (i.e., at level 1), in that a person tends to make more mistakes if he/she types faster. However, the relationship across individuals (i.e., at level 2) is likely to be negative, because individuals who tend to type fast *on average*, also tend to be more experienced and therefore make fewer mistakes *on average* (cf. Hamaker, 2012; see also Nezlek, 2001; Enders and Tofighi, 2007; Kievit et al., 2013).

In this section we discuss the effects of different centering methods, when there are different slopes at the two levels. Our main interest is in obtaining an appropriate estimate for the within-cluster slope, as this is most informative with respect to the within-person process. To facilitate the transition to the multilevel autoregressive model, we will present the issue based on repeated measures within individuals (rather than individuals organized in groups). In following Raudenbush and Bryk (2002) and Enders and Tofighi (2007), we begin by considering models with a fixed slope only. Subsequently, we discuss contextual models, in which the cluster means are included as a predictor at level 2. Then we discuss extensions that allow for random slopes. We end this section by speculating on the effects of centering in the context of the multilevel autoregressive model.

2.1. THE WITHIN-CLUSTER AND BETWEEN-CLUSTER SLOPES IN MULTILEVEL DATA WITH A FIXED SLOPE

Suppose that x_{ti} is the predictor, such as typing speed or the occurrence of a negative event, and that y_{ti} is the outcome variable, such as number of typos or negative affect. Let $i = 1, \ldots, N$ denote the individual, and $t = 1, \ldots, T_i$ denote the measurement occasion within individual i. Raudenbush and Bryk (2002) discuss how to obtain estimates of the between-person slope, relating the trait scores on the outcome variable to the trait scores on the predictor, and of the averaged or pooled within-person slope, describing the process that operates within individuals, using ordinary least squares (OLS). To this end, we first need the individual means on the predictor and the outcome variable, that is,

$$\bar{x}_{\cdot i} = \frac{1}{T_i} \sum_{i=1}^{T_i} x_{ti} \quad \text{and} \quad \bar{y}_{\cdot i} = \frac{1}{T_i} \sum_{i=1}^{T_i} y_{ti}. \quad (11)$$

Then the between-person or *between-cluster* slope β_B can be obtained by analyzing these individual means using the regression equation

$$\bar{y}_{\cdot i} = \beta_0 + \beta_B \bar{x}_{\cdot i} + e_i. \quad (12)$$

Additionally, the averaged within-person or *within-cluster* slope β_W can be obtained through applying CMC to both the predictor and the outcome variable, and analyze these data for individuals simultaneously, that is

$$y_{ti} - \bar{y}_{\cdot i} = \beta_W \left(x_{ti} - \bar{x}_{\cdot i} \right) + e_{ti}. \quad (13)$$

Clearly, β_W and β_B need not be the same.

Raudenbush and Bryk (2002) discuss how the slopes from diverse multilevel approaches are related to these two basic slopes. There are three approaches that can be used, that is: no centering (NC), grand-mean centering (GMC), and cluster-mean centering. GMC is simply a linear transformation of the data, and leads to a model that is statistically equivalent to NC (cf. Kreft et al., 1995; Raudenbush and Bryk, 2002; Snijders and Bosker, 2012). Therefore, we do not discuss this approach separately, and only focus on the comparison between NC and CMC below.

FIGURE 1 | Illustration of within-person and between-person relationships between two variables. Each ellipse represents the data from a single person. Dashed lines represent the between-person slope (i.e., β_B), which may have a different sign as the within-person slope (**Left panel**), may be similar to the (average or fixed) within-person slope (**Middle panel**), or may be larger than the (average or fixed) within-person slope (**Right panel**).

The model based on NC—with a random intercept and a fixed slope—can be expressed as

$$
\begin{aligned}
y_{ti} &= \alpha_i^n + \beta_i^n x_{ti} + e_{ti} \\
\alpha_i^n &= \gamma_{00}^n + u_{0i}^n \\
\beta_i^n &= \gamma_{10}^n.
\end{aligned}
\tag{14}
$$

whereas the corresponding model based on CMC of the predictor results in

$$
\begin{aligned}
y_{ti} &= \alpha_i^c + \beta_i^c (x_{ti} - \bar{x}_{\cdot i}) + e_{ti} \\
\alpha_i^c &= \gamma_{00}^c + u_{0i}^c \\
\beta_i^c &= \gamma_{10}^c.
\end{aligned}
\tag{15}
$$

Note that the fixed slope γ_{10}^c from the CMC model is analogous to β_W in Equation 13, with $y_{ti} - \bar{y}_{\cdot i}$ being replaced by $y_{ti} - \alpha_i^c$. Hence, CMC leads to an estimate of the within-cluster slope. The question is whether the within-cluster slope can also be obtained from the model in Equation 14.

To this end, we enter the level 2 expressions into the level 1 expressions, such that the model based on NC can be expressed as

$$
y_{ti} = \gamma_{00}^n + \gamma_{10}^n x_{ti} + u_{0i}^n + e_{ti},
\tag{16}
$$

and the model based on CMC can be expressed as

$$
y_{ti} = \gamma_{00}^c + \gamma_{10}^c x_{ti} - \gamma_{10}^c \bar{x}_{\cdot i} + u_{0i}^c + e_{ti}.
\tag{17}
$$

From these expressions it becomes clear that these models are not equivalent, as one cannot be considered an alternative parametrization of the other (cf. Kreft et al., 1995). This also implies that the within-cluster slope cannot be derived based on the results obtained from NC. Raudenbush and Bryk (2002) indicate that the slope of the level 1 predictor obtained with NC (γ_{10}^n) is "an uninterpretable blend" (p. 139) of the averaged within-cluster slope β_W and the between-cluster slope β_B. This has led them to formulate the advice to use CMC whenever the interest is in obtaining an unbiased estimate of the within-cluster relationship, and the within-cluster and between-cluster relationships are

expected to differ from each other (see also Enders and Tofighi, 2007).[3]

One could argue that the models above are not correct, because the between-cluster relationship is not explicitly modeled. Raudenbush and Bryk (2002) discuss the option of obtaining estimates of both β_W and β_B in a single multilevel model, through including the cluster means on the predictor as a level 2 predictor for the intercept. In case of NC, this results in

$$
\begin{aligned}
y_{ti} &= \alpha_i^n + \beta_i^n x_{ti} + e_{ti} \\
\alpha_i^n &= \gamma_{00}^n + \gamma_{01}^n \bar{x}_{\cdot i} + u_{0i}^n, \\
\beta_i^n &= \gamma_{10}^n.
\end{aligned}
\tag{18}
$$

while in case of CMC this gives

$$
\begin{aligned}
y_{ti} &= \alpha_i^c + \beta_i^c (x_{ti} - \bar{x}_{\cdot i}) + e_{ti} \\
\alpha_i^c &= \gamma_{00}^c + \gamma_{01}^c \bar{x}_{\cdot i} + u_{0i}^c, \\
\beta_i^c &= \gamma_{10}^c.
\end{aligned}
\tag{19}
$$

In the latter approach, the within-cluster slope is again represented by γ_{10}^c, and now the between-cluster slope is represented by γ_{01}^c.

To see whether NC and CMC lead to equivalent models in this case, we substitute the level 2 expressions in the level 1 expression. For NC this results in

$$
y_{ti} = \gamma_{00}^n + \gamma_{01}^n \bar{x}_{\cdot i} + \gamma_{10}^n x_{ti} + u_{0i}^n + e_{ti},
\tag{20}
$$

---

[3]Note that there may be situations in which the within-person slope and the between-person slope do not differ that much, such that failing to separate them does not affect the results very much: For instance, in the middle panel of **Figure 1** the relationship between negative events and negative affect is represented, which shows that within individuals, there is (on average) a positive relationship (i.e., people tend to experience more negative affect on days that more negative events occur), and the between-person relationship is very similar (i.e., people who tend to experience more negative events on average, also tend to have higher levels of negative affect on average). In this case, using CMC or NC/GMC will not change the estimate of the fixed slope very much. However, the point remains that to obtain an adequate estimate of the averaged within-cluster slope, CMC should be preferred.

and for CMC it results in

$$y_{ti} = \gamma_{00}^c + (\gamma_{01}^c - \gamma_{10}^c)\bar{x}_{.i} + \gamma_{10}^c x_{ti} + u_{0i}^c + e_{ti}, \qquad (21)$$

showing the models are equivalent. Furthermore, it becomes clear that actually both models provide an estimate of the within-cluster slope, that is, $\gamma_{10}^n = \gamma_{10}^c = \beta_W$. Additionally, we have $\gamma_{10}^n = \gamma_{01}^c - \gamma_{10}^c = \beta_B - \beta_W$, that is, γ_{01}^n represents the *difference* in the between-cluster and the within-cluster slopes. This is also referred to as the *contextual* or *compositional effect* (cf. Raudenbush and Bryk, 2002 p. 141).[4] , and these models are referred to as contextual models.

In sum, when there is a fixed slope, NC, and CMC lead to equivalent models if one includes the cluster means for the level 1 predictor as a level 2 predictor for the intercept (Kreft et al., 1995). However, if the cluster means are not included, these two models are not equivalent.

2.2. EFFECTS OF NC AND CMC WHEN THERE IS A RANDOM SLOPE

While the model equivalence above is interesting, it is of limited value in practice, as we are often interested in models with random slopes. For instance, consider the middle panel of **Figure 1**, representing the hypothetical relationship between the number of negative events and negative affect in daily measurements: It shows that the strength of the within-person relationship differs across individuals.

Snijders and Bosker (2012) show that if there is a random slope, the model equivalence presented above no longer holds. Allowing for a random slope in the NC model in Equation 18, implies we have $\beta_i^n = \gamma_{10}^n + u_{1i}^n$, and we can thus write

$$y_{ti} = \gamma_{00}^n + \gamma_{01}^n \bar{x}_{.i} + \gamma_{10}^n x_{ti} + u_{1i}^n x_{ti} + u_{0i}^n + e_{ti}. \qquad (22)$$

For the CMC model in Equation 19, a random slope implies we have $\beta_i^c = \gamma_{10}^c + u_{1i}^c$, such that the model can be expressed as

$$\begin{aligned} y_{ti} = {}& \gamma_{00}^c + (\gamma_{01}^c - \gamma_{10}^c)\bar{x}_{.i} + \gamma_{10}^c x_{ti} + u_{1i}^c x_{ti} \\ & - u_{1i}^c \bar{x}_{.i} + u_{0i}^c + e_{ti}. \end{aligned} \qquad (23)$$

This shows that—once there is a random slope—these models are no longer statistically equivalent, as they differ with respect to the term $(-u_{1i}^c \bar{x}_{.i})$. However, Kreft et al. (1995) pointed out that the fixed effect within-cluster slope is still the same across these two models: That is, $\gamma_{10}^n = \gamma_{10}^c = \beta_W$ (see Kreft et al., 1995, p. 13). Hence, when the goal is to obtain an estimate of the within-cluster slope, and the within-cluster and between-cluster slope are expected to differ, it seems that one can chose either use CMC, or the contextual versions of CMC or NC/GMC: Although the contextual models are not equivalent when a random slope is included, they will result in the same within-cluster slope estimate.

For the sake of completeness, we also provide the expression for the models that include a random slope but without the cluster means as a level 2 predictor—as these are more common than the contextual models and the fixed slope models discussed above. In that case, NC leads to

$$y_{ti} = \gamma_{00}^n + \gamma_{10}^n x_{ti} + u_{1i}^n x_{ti} + u_{0i}^n + e_{ti}. \qquad (24)$$

and CMC leads to

$$y_{ti} = \gamma_{00}^c + \gamma_{10}^c x_{ti} - \gamma_{10}^c \bar{x}_{.i} + u_{1i}^c x_{ti} - u_{1i}^c \bar{x}_{.i} + u_{0i}^c + e_{ti}. \qquad (25)$$

As expected based on what was discussed above, both the fixed and the random parts of these models differ, and only CMC leads to an estimate of the average within-cluster slope, while NC leads to a slope that represents some mix of the within-cluster and between-cluster slopes.

2.3. PRELIMINARY THOUGHTS ON CENTERING IN THE MULTILEVEL AUTOREGRESSIVE MODEL

Before turning to our simulation study, we speculate briefly on the effects of NC and CMC in case of the multilevel autoregressive model. The contextual model would imply that we include the within-person means as a predictor for the intercept at level 2, that is

$$\begin{aligned} y_{ti} &= \mu_i + \phi_i(y_{t-1,i} - \mu_i) + e_{ti} \\ \mu_i &= \gamma_{00} + \gamma_{01}\mu_i + u_{0i} \\ \phi_i &= \phi + u_{1i}, \end{aligned} \qquad (26)$$

where $\phi = \beta_W$ is the average within-cluster relationship, and $\gamma_{01} = \beta_B$ is the between-cluster relationship. Note however that now μ_i appears on both sides of the equality sign, and it follows that $\gamma_{01} = 1$, $\gamma_{00} = 0$ and $u_{0i} = 0$ (and subsequently $\sigma_{u0}^2 = 0$)[5].

We can draw two conclusions from this. First, including the within-person means as a level 2 predictor in a multilevel autoregressive model is not logical, and therefore the results for contextual models presented above are less relevant in the current context. Second, the within-cluster slope will—without exception—differ from the between-cluster slope in multilevel autoregressive models: That is, while the between-cluster slope is (essentially) 1, the within-cluster slope is identical to the autocorrelation and will thus have to lie between -1 and 1 for a stationary process (Hamilton, 1994). This is illustrated in the right panel of **Figure 1**, which shows that the between-cluster slope is equal to 1, while the within-cluster slope (averaged across individuals) is smaller (in this case between 0 and 1).

Applying the reasoning offered by Raudenbush and Bryk (2002) and Enders and Tofighi (2007) about the effects of NC/GMC vs. CMC in case of standard multilevel models to

---

[4]Here it represents the expected difference in number of typos when comparing two participants who type the same number of words, while they differ one unit on the number of words they type per minute on average (meaning they have different levels of experience).

[5]In practice, the means on the outcome variable y_{ti} are virtually identical to the cluster means on the predictor $y_{t-1,i}$, as it is the same variable; slight difference may arise however, because the outcome runs over $t = 2, \ldots, T_i$ while the predictor runs over $t = 1, \ldots, T_i - 1$. As T becomes larger, these differences will become smaller.

the multilevel autoregressive model, one may thus be inclined to think that: (a) NC/GMC will result in an *overestimation* of the fixed effect (i.e., average) autoregressive parameter, since $\beta_W < \beta_B = 1$; and (b) CMC will remove the "contamination" of β_B, such that the fixed effect autoregressive parameter adequately represents β_W, that is, the (averaged or pooled) within-person autoregression ϕ. This would imply that CMC should be the preferred form of centering in a multilevel autoregressive model.

Note that we already discussed two other reasons for preferring CMC in case of the multilevel autoregressive model, that is, it allows us to model μ_i as a random effect, rather than the less meaningful $c_i = \mu_i(1 - \phi_i)$, and it allows us to include predictors for μ_i (rather than for c_i). Taken together, these seem very convincing reasons for preferring CMC over GMC/NC in a multilevel autoregressive model.

3. SIMULATIONS
We performed a series of simulations to investigate the effect of NC vs. CMC on the estimation of the within-cluster slope. We begin with the standard multilevel model to verify the claims made by Raudenbush and Bryk (2002) and Enders and Tofighi (2007), and to determine whether these also generalize to models with a random slope (as presented in Equations 24 and 25). Following this, we consider the effects of NC and CMC in the multilevel autoregressive model, both with a fixed and a random autoregressive parameter. In addition, we consider the effects of diverse factors, that is: sample sizes, the sign and strength of the autoregressive parameter, and a level 2 predictor for the autoregressive parameter. All our simulations were performed in R (R Development Core Team, 2009). To estimate the multilevel models, we used the function `lmer()` from the R-package lme4 (Bates and Sarkar, 2007).

3.1. SIMULATIONS FOR THE STANDARD MULTILEVEL MODEL
We begin with simulating data from the standard multilevel model with different within-cluster and between-cluster slopes, using Equation 19 for a model with a fixed within-cluster slope, and Equation 23 for a model with a random within-cluster slope. Our specific interest is in obtaining an appropriate estimate of the within-cluster slope, when this differs from the between-cluster slope. Hence, we want to verify that when the cluster means are not included as a level 2 predictor, the slope estimate obtained with NC is indeed a blend of the within-cluster and between-cluster slopes, while CMC (based on Equation 15) leads to a pure within-cluster slope estimate.

We used the following model parameter values: (a) the variance of the predictor x_{ti} within each cluster is 1, and the variance of the cluster means between the clusters is also 1; (b) the fixed effect within-cluster slope γ_{10} is 0.3; (c) the standard deviation of the within-cluster slope β_i is either 0 (i.e., fixed slope only model), or 0.1 (i.e., random slope model); (d) the between-cluster slope γ_{01} is 1; (e) the grand mean γ_{00} is zero; (f) the level 1 residual variance σ_e^2 was either 1 or 3; and (g) the level 2 residual variance for the intercept σ_{u0}^2 was either 1 or 0. The reason we considered 0 as well here, is because this would make the model more comparable to the multilevel autoregressive model we consider later on

(see Equation 26). We set the number of clusters to 100, and the number of observations per cluster to 20.

The results are presented in **Table 2**: It includes the OLS estimate of the between-cluster slope (based on Equation 12), the OLS estimate of the within-cluster slope (based on Equation 13), and the fixed effects slope obtained with CMC and with NC. These confirm the point made by Raudenbush and Bryk (2002) and Enders and Tofighi (2007): While CMC leads to a slope estimate that is almost identical to the OLS within-cluster estimate and which adequately represents the actual within-cluster slope, the estimate obtained with NC is a blend of the within-cluster and the between-cluster slopes. Specifically, if the level 2 residual variance (i.e., σ_{u0}^2) becomes smaller relative to the level 1 residual variance (i.e., σ_e^2), the slope estimate is more strongly affected by the between-cluster slope. Furthermore, the results are very similar for models and data without a random slope (left part of **Table 2**), and with a random slope (right part of **Table 2**).

3.2. SIMULATIONS FOR THE MULTILEVEL AUTOREGRESSIVE MODEL
To determine whether the results reported above generalize to the multilevel autoregressive model, we considered the following scenarios. We simulated data using the model defined in Equations 1–4, with: (a) a fixed effects within-cluster slope of $\phi = 0.3$; (b) a standard deviation of the individual within-cluster slope ϕ_i of either 0 (for a model with a fixed autoregressive parameter only) or 0.1 (for a model with a random autoregressive parameter); (c) a level 2 variance of the intercept μ_i of 1, 3 or 9; and (d) a grand mean of 0. We used the same number of observations as in the previous simulations, that is, 100 clusters (i.e., persons here) and 20 observations per cluster (i.e., repeated measurements here). The results based on 1000 replications are presented in **Table 3**.

Table 2 | Estimates for fixed effect slope γ_{10}.

Estimation method	σ_{u0}^2	σ_e^2	Fixed slope only	Random slope
OLS between	1	1	0.963	0.965
	0	1	0.966	0.967
	0	3	0.965	0.966
OLS within	1	1	0.300	0.299
	0	1	0.298	0.300
	0	3	0.299	0.302
CMC (sample)	1	1	0.300	0.299
	0	1	0.298	0.300
	0	3	0.299	0.303
NC	1	1	0.323	0.323
	0	1	0.372	0.373
	0	3	0.492	0.489

Mean point estimates for fixed effects slope γ_{10} in a standard multilevel model with either a fixed slope only (left; $\beta_i = \gamma_{10}$), or with a random slope (right; $\beta_i = \gamma_{10} + u_{1i}$). True fixed effect within-cluster slope is $\gamma_{10} = 0.3$, and true between-cluster slope is $\gamma_{01} = 1$. Number of observations per cluster is 20; number of clusters is 100; number of replications is 1000. Estimation methods are: OLS between and within (Equations 12 and 13); centering per cluster (CMC) using the sample mean; and no centering (NC).

Table 3 | Estimates for fixed effect autoregressive parameter ϕ.

Estimation method	σ_e^2	Fixed slope only	Random slope
OLS within	1	0.230	0.233
	3	0.229	0.233
	9	0.228	0.233
CMC (sample)	1	0.231	0.229
	3	0.230	0.229
	9	0.229	0.229
NC	1	0.304	0.307
	3	0.304	0.306
	9	0.303	0.304

Mean point estimates for the fixed effects autoregressive parameter ϕ in a multilevel autoregressive model, with a fixed slope only (left; $\phi_i = \phi$), and with a random slope (right; $\phi_i = \phi + u_{1i}$). True fixed effect autoregressive parameter (i.e., the true within-cluster slope) is $\phi = 0.3$. Number of observations per person is 20; number of persons is 100; number of replications is 1000. Estimation methods are: OLS within (Equation 13); centering per cluster (CMC) using the sample mean; and no centering (NC).

As before, CMC leads to estimates that are very close to the OLS within-cluster estimates. However, for the multilevel autoregressive model, these are biased: That is, they underestimate the actual fixed effect autoregressive parameter (i.e., estimated bias between 0.069 and 0.071 for CMC). Surprisingly, NC leads to an estimate that is less biased (i.e., estimated bias between 0.003 and 0.007). In Appendix 2, this downward bias for the OLS within-cluster estimate in multilevel autoregressive model is confirmed analytically. Note further that whether or not inertia was random, did not affect the results substantially.

3.3. INVESTIGATING THE INFLUENCE OF OTHER FACTORS

To gain more insight in this matter, we considered three additional factors that may affect the estimation of the within-cluster slope in a multilevel autoregressive model. First, in addition to using the individual sample means in CMC (i.e., $\bar{y}_{.i}$), we also considered the empirical Bayes estimator (also referred to as shrinkage estimator) of the individuals' means (i.e., $\hat{\mu}_i$, obtained with estimating the empty model first), and the true person means that were used to generate the data (i.e., μ_i; we considered this option here to see to what extent the results for CMC can be attributed to having to use an estimate of the individual's mean). Second, we considered different samples sizes, both with respect to number of persons N, and the number of repeated measures T. Third, we considered different strengths and signs of the fixed effects autoregressive parameter. Throughout we used the level 1 residual variance $\sigma_e^2 = 3$, the level 2 intercept variance $\sigma_{u0}^2 = 3$, and the level 2 slope variance $\sigma_{u1}^2 = 0.01$.

Based on the results presented in **Table 4**, we can conclude the following. First, CMC of the autoregressive predictor leads to bias, regardless of the kind of mean that is used (i.e., the sample estimate $\bar{y}_{.i}$, the empirical Bayes estimate $\hat{\mu}_i$, or the true value μ_i). It is noteworthy that even using the true mean results in bias that is about the same as the bias obtained with the empirical Bayes

estimate of the mean, while using the sample mean leads to only slightly more bias. In contrast, NC does not lead to (considerable) bias. Second, when using CMC, increasing the number of observations per person (i.e., T) leads to a decrease in bias, whereas the number of individuals N does not affect the bias. Third, the bias for CMC reported in **Table 4** is *always negative*, regardless of the actual value of ϕ, although the bias is largest when $\phi = 0.3$, and smallest when $\phi = -0.3$. This implies that in general, ϕ will be underestimated when CMC is used, and the bias is larger when ϕ is positive (which will often be the case in practice). This is also confirmed by the analytical results in Appendix 2.

With respect to the coverage rates of the 95% confidence intervals, we make the following two observations. First, while in general they are too low, for NC most coverage rates are above 0.900, while for all three forms of CMC they are much lower (which is not surprising, given the bias of CMC estimates). Second, while increasing T leads to higher coverage rates for the CMC approaches, increasing N actually leads to lower coverage rates. This result is explained by the fact that the standard errors decrease when N increases, while the bias remains unaffected by changes in N. Note that the pattern for the coverage rates obtained with NC is less clear.

3.4. INCLUDING A LEVEL 2 PREDICTOR OF THE AUTOREGRESSIVE PARAMETER

An important question when applying the multilevel autoregressive model is whether other variables predict individual differences in the autoregression (cf. Suls et al., 1998; Kuppens et al., 2010). Therefore, we performed an additional simulation study to determine the effect of CMC and NC on the estimation of the effect of a level 2 predictor on the autoregressive parameter.

Let z_i be a level 2 predictor that may have an effect on the individuals' average score μ_i, but more importantly, may have an effect on the individuals' autoregressive parameter ϕ_i. We assume this level 2 predictor is centered across people. When using NC, the model can be expressed as

$$y_{ti} = c_i + \phi_i^n y_{t-1,i} + e_{ti}$$
$$c_i = \gamma_{00}^n + \gamma_{01}^n z_i + u_{0i} \qquad (27)$$
$$\phi_i^n = \gamma_{10}^n + \gamma_{11}^n z_i + u_{1i}$$

where γ_{00}^n is the overall intercept, and γ_{10}^n is the average autoregressive parameter (assuming the level 2 predictor z_i is centered). The regression coefficients γ_{01}^n and γ_{11}^n represent the effects of the level 2 predictor on the individuals' intercept c_i and their autoregressive parameter ϕ_i^n, respectively.

In contrast, when using CMC for the autoregressive predictor, the model can be defined as

$$y_{ti} = \mu_i + \phi_i^c (y_{t-1,i} - \mu_i) + e_{ti}$$
$$\mu_i = \gamma_{00}^c + \gamma_{01}^c z_i + u_{0i} \qquad (28)$$
$$\phi_i^c = \gamma_{10}^c + \gamma_{11}^c z_i + u_{1i}$$

where γ_{00}^c now represents the grand mean, and γ_{10}^c is again the average autoregressive parameter (assuming the level 2 predictor

Table 4 | Bias and coverage rates for fixed autoregressive parameter ϕ in multilevel autoregressive model under diverse scenarios.

AR parameter	Sample size		Bias				CR$_{0.95}$			
	N	T	NC	C($\bar{y}_{.i}$)	C($\hat{\mu}_i$)	C(μ_i)	NC	C($\bar{y}_{.i}$)	C($\hat{\mu}_i$)	C(μ_i)
$\phi_i \sim N(0.3, 0.1)$	20	20	0.002	−0.072	−0.069	−0.068	0.928	0.762	0.785	0.787
		50	0.000	−0.027	−0.027	−0.026	0.940	0.900	0.901	0.898
		100	0.000	−0.013	−0.013	−0.013	0.932	0.932	0.932	0.932
	50	20	0.005	−0.071	−0.069	−0.067	0.893	0.480	0.512	0.518
		50	0.001	−0.027	−0.026	−0.026	0.936	0.800	0.804	0.805
		100	0.000	−0.013	−0.013	−0.013	0.946	0.902	0.902	0.903
	100	20	0.006	−0.070	−0.068	−0.066	0.892	0.196	0.227	0.242
		50	0.001	−0.027	−0.027	−0.027	0.930	0.623	0.630	0.637
		100	0.000	−0.013	−0.013	−0.013	0.930	0.851	0.854	0.851
$\phi_i \sim N(0, 0.1)$	20	20	0.001	−0.053	−0.050	−0.050	0.923	0.844	0.858	0.851
		50	−0.000	−0.020	−0.020	−0.020	0.944	0.912	0.915	0.911
		100	0.000	−0.010	−0.009	−0.009	0.929	0.926	0.926	0.925
	50	20	0.003	−0.052	−0.049	−0.049	0.922	0.700	0.727	0.725
		50	−0.001	−0.021	−0.021	−0.021	0.942	0.860	0.862	0.861
		100	0.000	−0.010	−0.010	−0.010	0.939	0.910	0.910	0.909
	100	20	0.003	−0.053	−0.051	−0.050	0.929	0.431	0.479	0.477
		50	0.000	−0.021	−0.021	−0.021	0.931	0.775	0.785	0.785
		100	0.000	−0.010	−0.010	−0.010	0.942	0.892	0.896	0.896
$\phi_i \sim N(-0.3, 0.1)$	20	20	0.003	−0.034	−0.031	−0.032	0.943	0.907	0.916	0.913
		50	0.000	−0.014	−0.013	−0.014	0.940	0.932	0.934	0.928
		100	0.001	−0.005	−0.005	−0.005	0.928	0.928	0.929	0.929
	50	20	0.000	−0.038	−0.035	−0.036	0.940	0.783	0.802	0.795
		50	0.000	−0.014	−0.014	−0.014	0.932	0.894	0.896	0.896
		100	0.000	−0.007	−0.006	−0.006	0.927	0.914	0.914	0.914
	100	20	0.000	−0.039	−0.036	−0.037	0.932	0.597	0.639	0.624
		50	0.000	−0.015	−0.015	−0.015	0.928	0.848	0.851	0.851
		100	0.000	−0.007	−0.007	−0.007	0.942	0.908	0.911	0.911

Bias and coverage rates of 95% confidence intervals (CR$_{0.95}$) based on 1000 replications. N refers to number of persons, T refers to number of observations per person. The random coefficient ϕ_i comes from a normal distribution, with mean ϕ (either 0.3, 0, or −0.3), and standard deviation 0.1. Results are obtained for: NC of the autoregressive predictor; C($\bar{y}_{.i}$) is CMC using the sample mean; C($\hat{\mu}_i$) is CMC using the empirical Bayes estimate; and C(μ_i) is CMC using the true mean per person (for comparison).

z_i is centered). The regression coefficients γ_{01}^c and γ_{11}^c represent the effects of the level 2 predictor on the individuals' means μ_i and their autoregressive parameters ϕ_i^c, respectively.

Based on the results from the previous simulations, we expect that CMC (as in Equation 28) will lead to a downward bias in the estimation of the average autoregressive parameter ϕ (i.e., γ_{10}^c will be an underestimate), while NC (as in Equation 27) is not associated with such bias (i.e., γ_{10}^n is an unbiased estimate of ϕ). However, the question here is how CMC and NC affect the estimation of the level 2 predictor on ϕ_i, that is, γ_{11}^c and γ_{11}^n.

We created a level 2 predictor with a mean of zero and a variance of 0.01. We chose this rather small variance for numerical reasons: Because the variance of ϕ_i is necessarily small (say about 0.01), having a level 2 predictor with a large variance may lead to numerical problems in estimating the regression coefficient γ_{11}. The mean autoregressive parameter ϕ was set to 0.3. The effect of the level 2 predictor z_i on the individual inertia parameters ϕ_i was set to 0.4. The other parameters were chosen such that the

correlations between μ_i, ϕ_i and z_i were not unrealistically high (μ_i and ϕ_i were both correlated 0.37 with z_i, and 0.14 with each other). After generating μ_i and ϕ_i from z_i, we used Equations 1 and 3 to generate the data. The results for this simulation study are presented in **Table 5**.

The left part of the **Table 5** contains the results that reflect the bias. In line with our previous results, the average autoregressive parameter $\gamma_{10} = \phi$ is characterized by a downward bias when CMC is used for the autoregressive predictor, while NC leads to unbiased estimates. However, when considering the effect of CMC vs. NC on the estimation of γ_{11}, we see that CMC actually leads to *less bias* than NC. Note also that while increasing T reduces the bias obtained with NC, the effect of increasing N is not that clear (i.e., when $T = 20$, increasing N actually increases the bias).

The right part of **Table 5** contains the coverage rates of the 95% confidence intervals. As before, the coverage rates for the average autoregressive parameter obtained with CMC are lower than

Table 5 | Results for average autoregressive parameter ϕ and the effect of a level 2 predictor z_i on the autoregressive parameter ϕ_i.

N	T	Bias				$CR_{0.95}$			
		γ_{10}		γ_{11}		γ_{10}		γ_{11}	
		NC	CMC	NC	CMC	NC	CMC	NC	CMC
20	20	−0.007	−0.076	−0.050	−0.007	0.897	0.720	0.944	0.958
	50	−0.005	−0.030	−0.045	−0.026	0.922	0.856	0.939	0.944
	100	0.000	−0.013	−0.019	−0.008	0.923	0.909	0.942	0.944
50	20	0.004	−0.071	−0.068	−0.022	0.885	0.476	0.950	0.959
	50	0.001	−0.026	−0.032	−0.014	0.904	0.781	0.945	0.948
	100	0.000	−0.013	−0.018	−0.006	0.924	0.890	0.953	0.949
100	20	0.004	−0.071	−0.084	−0.036	0.918	0.170	0.921	0.940
	50	0.000	−0.027	−0.024	−0.003	0.907	0.628	0.944	0.955
	100	0.001	−0.012	−0.019	−0.008	0.928	0.832	0.939	0.942

Bias and coverage rate of 95% confidence intervals ($CR_{0.95}$) based on 1000 replications. Results for $\gamma_{10} = 0.3$, that is, the average autoregressive parameter ϕ, and for $\gamma_{11} = 0.4$, that is, the effect of a level 2 predictor on the autoregressive parameter, using NC and CMC (with sample mean) for the autoregressive predictor.

those obtained with NC. For γ_{11} the coverage rates obtained with NC are in general lower than those obtained with CMC (which was to be expected given the results for the bias).

3.5. CONCLUSION

The first set of simulations presented in this section clearly illustrated the point made by Raudenbush and Bryk (2002) and Enders and Tofighi (2007) in case of a standard multilevel model. In addition it was shown that the claims regarding the within-cluster slope generalize to the model with a random slope, in that CMC leads to an estimate of the within-cluster slope, whereas NC results in a blend of the within-cluster and the between-cluster slope. The second set of simulations was based on the multilevel autoregressive model and showed that while CMC still leads to results that are almost identical to the OLS-within estimate, both of these are biased with respect to the actual within-cluster slope (i.e., the autoregressive relationship).

Additional simulations showed that there is a downward bias regardless of the sign of ϕ, and that this bias is most severe when T is small, while N has little (if any) influence. This bias could not be attributed to the quality of the estimate of the individuals' means (as very similar results are obtained when using the true means μ_i for centering). Furthermore, these results were supported by the derived relationship between the OLS within-cluster slope estimate and the value of ϕ in Appendix 2. In contrast, NC does not lead to bias in the estimation of the autoregressive parameter, which implies that the obtained result is actually *not contaminated* by the between-cluster relationship, as is the case in regular multilevel analysis. Finally, when adding a level 2 predictor to the model, the results described above for the average autoregressive parameter remain intact, but for the effect of the level 2 predictor on the autoregressive parameter, NC actually results in bias, whereas CMC does not.

4. APPLICATION: PART 2

Returning to the empirical data that we introduced in the beginning of this paper, we are now able to study inertia in daily relationship specific PA and NA, and include a level 2 predictor for the individual differences in the means and the inertia. We used Relationship Satisfaction, which was obtained prior to the diary study, and standardized this level 2 predictor to facilitate interpretation (i.e., we subtracted the grand mean, and divided it by the grand standard deviation). We used the model based on CMC (see Equation 28), and summarized the results for all the fixed effects in **Table 6**. We also included the estimate of the fixed effects inertia obtained with NC in this table, as the simulations reported in this paper showed that this is an unbiased estimate of the average inertia, whereas the corresponding estimate obtained with CMC is negatively biased.

It shows that on average there is significant inertia in relationship specific PA and NA for both males and females (see γ_{10}^n). In addition, Relationship Satisfaction proved a significant positive predictor of mean levels of relationship specific PA in both males and females, and a significant negative predictor of mean levels of relationship specific NA in both males and females (see γ_{01}^c). Furthermore, Relationship Satisfaction is a significant negative predictor of inertia in relationship specific PA in males (but not in females), and in relationship specific NA in males and females (see γ_{11}^c). This implies that individuals who are less satisfied with their relationship, are characterized by more carryover of relationship specific NA, than individuals who are more satisfied with their relationship. In addition, males who are less satisfied with their relationship, are also characterized by more carryover in their relationship specific PA. While the latter may seem surprising at first—as it implies that elevated relationship specific PA tends to persist over time for males who are less satisfied with their relationship—it also implies that attenuated relationship specific PA tends to prevail, which could be considered undesirable. These results are in agreement with other findings regarding inertia reported by Koval et al. (2013), who found that the inertias of PA and NA are positively correlated, and Kuppens et al. (2012), who found that inertia of angry and dysphoric behavior, but also of happy behavior all predicted the onset of depression. Taken together, these results seem to confirm the

Table 6 | Results for multilevel autoregressive model with a level 2 predictor (with random effects).

		Males			Females		
		est.	SD	t-value	est.	SD	t-value
PA	γ_{00}^c	3.514	0.043	81.98	3.498	0.043	80.93
	γ_{01}^c	0.303	0.046	6.58	0.383	0.045	8.43
	γ_{10}^c	0.354	0.016	22.46	0.340	0.015	22.50
	γ_{10}^n	0.385	0.015	25.41	0.369	0.015	24.69
	γ_{11}^c	−0.045	0.017	−2.67	−0.015	0.016	−0.98
NA	γ_{00}^c	1.346	0.022	59.96	1.346	0.020	5.75
	γ_{01}^c	−0.072	0.024	−2.99	−0.132	0.022	−6.15
	γ_{10}^c	0.242	0.017	14.31	0.224	0.016	13.64
	γ_{10}^n	0.267	0.017	16.14	0.254	0.017	15.39
	γ_{11}^c	−0.046	0.017	−2.63	−0.060	0.016	−3.71

Parameter estimates, standard errors and t-values for the fixed effects parameters in a multilevel autoregressive model (with random intercept and slope), for males and females. Parameters include: (a) the grand mean (i.e., γ_{00}^c); (b) the effect of Relationship Satisfaction on the individuals' means (i.e., γ_{01}^c); (c) the average inertia obtained with CMC (i.e., γ_{10}^c), and with NC (i.e., γ_{10}^n); and (d) the effect of Relationship Satisfaction on the individuals' inertias (i.e., γ_{11}^c).

idea that inertia—whether in pleasant or unpleasant emotions—is a detrimental property of affect regulation, reflective of some maladaptive process.

5. DISCUSSION

Over the past two decades we have witnessed an exponential increase in the number of studies based on intensive longitudinal data in the social sciences. This development is triggered by the rapid development of electronic data collection methods based on hand-held computers, the internet, and—more recently—smart phones (Trull and Ebner-Priemer, 2013): As a result it has become relatively easy to gather large numbers of repeated measurements from a large sample of individuals. Such data differ from more traditional longitudinal data in two important ways: (1) intensive longitudinal data contain many more measurements per individual (i.e., often $T > 20$) than traditional longitudinal data (i.e., often $T < 10$); and (2) the measurements in intensive longitudinal data are typically spaced relatively close to each other in time (e.g., measurements are taken at a daily basis using a daily diary method, or even multiple times a day using experience method sampling), whereas traditional longitudinal data are characterized by much larger intervals between measurements (e.g., annual measurements are not uncommon). These differences reflect a different focus on part of the researchers: Whereas the purpose of many studies based on traditional longitudinal data is to discover broad underlying increasing or decreasing trends, the purpose of studies based on intensive longitudinal data is to gain more insight into the patterns of fluctuations in affect, behavior, and cognition in daily life (Bolger et al., 2003; Mehl and Conner, 2012).

One particular aspect of such patterns is referred to as inertia or autoregression, and represents the amount of carryover from one measurement occasion to the next (Suls et al., 1998). Diverse empirical studies have now shown that individual differences in inertia are meaningful with respect to the way people differ in their regulation of emotions and behavior. As the popular method for studying inertia is through a multilevel autoregressive model, an important research question in this area is whether the autoregressive predictor included at level 1 should be centered per person or not.

The current study shows through a series of simulations that CMC should be preferred if: (a) one wishes to obtain a meaningful intercept (i.e., an intercept that represents the individual's mean score over time, which can be interpreted as his/her trait score); and (b) the interest is in how the autoregressive parameter depends on a level 2 predictor. However, CMC should *not* be used when the interest is in whether or not there is an autoregressive relationship *on average* (i.e., across individuals).

In practice, researchers using a multilevel autoregressive model to study inertia are likely to be interested in various aspects of the model, including the individuals' means, the average autoregressive parameter, and the effect of a level 2 predictor on the individuals' means and autoregressive parameters. In that case, it may be wise to use both estimation procedures, as we did in the empirical application, and to use CMC for the estimation of the grand mean and the effect of the level 2 predictor on the individual means and autoregressive parameters, while NC results are used for determining whether there is an autoregressive effect on average. While this may be unconventional advice, it is based on the rather clear simulation results presented in this paper.

Given the recent interest in inertia, and its emerging recognition as a separate and valuable property of regulation that is related to but does not coincide with more traditionally studied process features such as the tendency to ruminate or the persistence of negative thoughts, we expect to see more work in this area. Hence, it is important to improve our ways to estimate average inertia, and individual differences therein. Specifically, the current study has shown that many of the inertia estimates reported in the literature may actually be underestimates of the true inertias, simply because the lagged autoregressive predictor was centered per person (e.g., Koval et al., 2012; Brose et al., 2014). Although this may not come as a surprise to those familiar with time series literature, as it has been known for a long time that estimates of autoregressive parameters are biased (cf., Orcutt, 1948; Marriott and Pope, 1954), it is an unexpected result from a multilevel perspective. Furthermore, it is of interest that the bias disappears when the lagged autoregressive predictor is not centered; in fact, this may be considered an important advantage of the multilevel approach over a two-step procedure in which during the first step individual time series models are estimated, while in the second step the individual parameters are combined into a population model.

Additional improvements in the study of inertia may come from taking measurement error into account—which is also likely to obscure the actual inertia of a process—and developing appropriate techniques for handling unequal intervals between the observations—which are a feature of certain intensive longitudinal data, and which may lead to less precise estimates when not taken into account, and therefor to more difficulty in detecting relationships between inertia and other person characteristics. When these issues are handled in an appropriate way, inertia may prove to be an even more important feature of regulatory processes in psychology than the existing studies already suggest.

Finally, note that the advice given here regarding CMC vs. NC or GMC *exclusively* applies to an autoregressive level 1 predictor: That is, if one includes other level 1 predictors, the common results based on Raudenbush and Bryk (2002) apply to them, meaning that CMC of these predictors should be preferred over NC or GMC if the within-cluster and between-cluster slopes are expected to differ, and one wants to obtain an estimate of the within-cluster slope.

ACKNOWLEDGMENT

REFERENCES

Bates, D. and Sarkar, D. (2007). *lme4: Linear mixed-effects models using S4 classes. R package*. Available online at: http://CRAN.R-project.org/

Bolger, N., Davis, A., and Rafaeli, E. (2003). Diary methods: capturing life as it is lived. *Annu. Rev. Psychol.* 54, 579–616. doi: 10.1146/annurev.psych.54.101601.145030

Bolger, N. and Zuckerman, A. (1995). A framework for studying personality in the stress process. *J. Pers. Soc. Psychol.* 69, 890–902. doi: 10.1037/0022-3514.69.5.890

Brose, A., Schmiedek, F., and Kuppens, P. (2014). Emotional inertia contributes to depressive symptoms beyond perseverative thinking. *Cogn. Emot.* 13:1–12. doi: 10.1080/02699931.2014.916252. [Epub ahead of print].

Curran, P. J. and Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annu. Rev. Psychol.* 62, 583–619. doi: 10.1146/annurev.psych.093008.100356

Enders, C. K. and Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychol. Methods* 12, 121–138. doi: 10.1037/1082-989X.12.2.121

Ferrer, E., Steele, J. S., and Hsieh, F. (2012). Analyzing the dynamics of affective dyadic interactions using patterns of intra- and interindividual variability. *Multivar. Behav. Res.* 47, 136–171. doi: 10.1080/00273171.2012.640605

Ferrer, E., and Widaman, K. (2008). "Dynamic factor analysis of dyadic affective processes with inter-group differences," in *Modeling Dyadic and Interdependent Data in the Developmental and Behavioral Sciences*, eds N. A. Card, J. P. Selig, and T. D. Little (New York, NY: Taylor & Francis Group), 107–137.

Gunthert, K. C., Cohen, L. H., Butler, A. C., and Beck, J. S. (2007). Depression and next-day spillover of negative mood and depressive cognitions following interpersonal stress. *Cogn. Ther. Res.* 31, 521–532. doi: 10.1007/s10608-006-9074-1

Hamaker, E. L. (2012). "Why researchers should think 'with in-person' a paradigmatic rationale," in *Handbook of Research Methods for Studying Daily Life*, eds M. R. Mehl and T. S. Conner (New York, NY: Guilford Publications), 43–61.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Henquet, C., Van Os, J., Kuepper, R., Delespaul, P., Smits, M., à Campo, J., et al. (2010). Psychosis reactivity to cannabis use in daily life: an experience sampling study. *Br. J. Psychiatry* 196, 447–453. doi: 10.1192/bjp.bp.109.072249

Hox, J. J. (2010). *Multilevel Analysis: Techniques and Application, 2nd Edn.* New York, NJ: Taylor & Francis Group.

Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J., and Borsboom, D. (2013). Simpsonŝs paradox in psychological science: a practical guide. *Front. Psychol.* 4:513. doi: 10.3389/fpsyg.2013.00513

Koval, P. and Kuppens, P. (2012). Changing emotion dynamics: individual differences in the effect of anticipatory social stress on emotional inertia. *Emotion* 12, 256–267. doi: 10.1037/a0024756

Koval, P., Kuppens, P., Allen, N. B., and Sheeber, L. (2012). Getting stuck in depression: the roles of rumination and emotional inertia. *Cogn. Emot.* 26, 1412–1427. doi: 10.1080/02699931.2012.667392

Koval, P., Pe, M., Meers, K., and Kuppens, P. (2013). Affective dynamics in relation to depressive symptoms: variable, unstable or inert? *Emotion* 13, 1132–1141. doi: 10.1037/a0033579

Kreft, I. G. G., de Leeuw, J., and Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivar. Behav. Res.* 30, 1–21. doi: 10.1207/s15327906mbr3001_1

Kuppens, P., Allen, N. B., and Sheeber, L. B. (2010). Emotional inertia and psychological maladjustment. *Psychol. Sci.* 21, 984–991. doi: 10.1177/0956797610372634

Kuppens, P., Sheeber, L. B., Yap, M. B. H., Whittle, S., Simmons, J., and Allen, N. B. (2012). Emotional inertia prospectively predicts the onset of depression in adolescence. *Emotion* 12, 283–289. doi: 10.1037/a0025046

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* 38, 963–974. doi: 10.2307/2529876

Larson, R. W. and Almeida, D. M. (1999). Emotional transmission in the daily lives of families: a new paradigm for studying family process. *J. Marriage Fam.* 61, 5–20. doi: 10.2307/353879

Marriott, F. H. C. and Pope, J. A. (1954). Bias in the estimation of autocorrelations. *Biometrika* 41, 390–402. doi: 10.1093/biomet/41.3-4.390

Mehl, M. R. and Conner, T. S., editors (2012). *Handbook of Research Methods for Studying Daily Life*. New York, NY: The Guilford Press.

Moberly, N. J. and Watkins, E. R. (2008). Ruminative self-focus and negative affect: An experience sampling study. *J. Abnorm. Psychol.* 117, 314–323. doi: 10.1037/0021-843X.117.2.314

Nezlek, J. B. (2001). Multilevel random coefficient analyses of event- and interval-contingent data in social and personality psychology research. *Pers. Soc. Psychol. Bull.* 27, 771–785. doi: 10.1177/0146167201277001

Orcutt, G. H. (1948). A study of the autoregressive nature of the time series used for Tinbergen's model of the economic system of the United States, 1919-1932. *J. R. Stat. Soc. Series B* 10, 1–53.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods, 2nd Edn.* Thousand Oaks, CA: Sage Publications.

Rovine, M. J., and Walls, T. A. (2006). "Multilevel autoregressive modeling of interindividual differences in the stability of a process," in *Models for Intensive Longitudinal Data*, eds T. A. Walls and J. L. Schafer (New York, NY: Oxford University Press), 124–147.

Singer, J. D. and Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195152968.001.0001

Snijders, T. A. B. and Bosker, R. J. (2012). *Multilevel Analysis: An introduction to basic and advanced multilevel modeling, 2 Edn.* London: Sage Publishers.

Suls, J., Green, P., and Hillis, S. (1998). Emotional reactivity to everyday problems, affective inertia, and neuroticism. *Pers. Soc. Psychol. Bull.* 24, 127–136. doi: 10.1177/0146167298242002

Trull, T. J. and Ebner-Priemer, U. (2013). Ambulatory assessment. *Annu. Rev. Clin. Psychol.* 9, 151–176. doi: 10.1146/annurev-clinpsy-050212-185510

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York, NY: Springer-Verlag.

Wang, L., Hamaker, E. L., and Bergman, C. S. (2012). Investigating inter-individual difference in short-term intra-individual variability. *Psychol. Methods* 17, 567–581. doi: 10.1037/a0029317

APPENDIX 1

To show that the model expressed in Equations 5, 6 and 7, is structurally equivalent to the model in Equations 1–4, we first make use of the fact that $y_{t-1,i} - \mu_i = a_{t-1,i}$, such that we can rewrite Equation 3 as

$$
\begin{aligned}
a_{ti} &= \phi_i a_{t-1,i} + e_{ti} \\
&= \phi_i(y_{t-1,i} - \mu_i) + e_{ti} \\
&= \phi_i y_{t-1,i} - \phi_i \mu_i + e_{ti}.
\end{aligned} \tag{A1}
$$

Entering this in Equation 1, we can write

$$
\begin{aligned}
y_{ti} &= \mu_i - \phi_i \mu_i + \phi_i y_{t-1,i} + e_{ti} \\
&= (1 - \phi_i)\mu_i + \phi_i y_{t-1,i} + e_{ti},
\end{aligned} \tag{A2}
$$

which shows that

$$
c_i = (1 - \phi_i)\mu_i. \tag{A3}
$$

This is a standard result from the time series literature on the first-order autoregressive model (cf. Hamilton, 1994).

APPENDIX 2

The rather unexpected results regarding CMC (i.e., centering the autoregressive predictor per person actually leads to bias in estimating the autoregressive parameter), was confirmed by the near identical results when using OLS. Below we show that OLS indeed leads to bias in the estimation of the autoregressive parameter ϕ. For simplicity of the presentation we will not make notational distinction between a random variable and it's (observed) value here.

Note that the OLS estimate of the regression model $y_i = b_0 + b_1 x_i + e_i$ can be expressed as $\hat{b}_1 = \text{cov}(x_i, y_i)/\text{var}(x_i)$. In a similar fashion, the OLS estimate of the within-person relationship ϕ in the autoregressive multilevel model can be expressed as the covariance between the person-centered predictor variable y_{ti} and the person-centered outcome variable $y_{i,t+1}$, divided by the variance of the person-centered predictor variable. To this end, let $T_i^* = T_i - 1$, and let

$$
\bar{y}_{i\cdot(1)} = \frac{1}{T_i^*} \sum_{t=1}^{T_i^*} y_{ti} \quad \text{and} \quad \bar{y}_{i\cdot(2)} = \frac{1}{T_i^*} \sum_{t=1}^{T_i^*} y_{i,t+1} \tag{A4}
$$

represent the estimated person means of the predictor variable y_{ti} and the outcome variable $y_{i,t+1}$, respectively. Then the OLS estimator of ϕ can be expressed as

$$
\begin{aligned}
\hat{\phi} &= \frac{\sum_{i=1}^{n}\left\{\frac{1}{T_i^*}\sum_{t=1}^{T_i^*}\left(y_{ti} - \bar{y}_{i\cdot(1)}\right)\left(y_{i,t+1} - \bar{y}_{i\cdot(2)}\right)\right\}}{\sum_{i=1}^{n}\left\{\frac{1}{T_i^*}\sum_{t=1}^{T_i}\left(y_{ti} - \bar{y}_{i\cdot(1)}\right)^2\right\}} \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{T_i^*}\sum_{t=1}^{T_i-1}\left(y_{ti}y_{it+1} - T_i^*\bar{y}_{i\cdot(1)}\bar{y}_{i\cdot(2)}\right)\right\}}{\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{T_i^*}\sum_{t=1}^{T_i}\left(y_{ti}^2 - T_i^*\bar{y}_{i\cdot(1)}^2\right)\right\}},
\end{aligned} \tag{A5}
$$

To derive the asymptotic bias of this estimator, we begin by deriving the numerator of Equation A5. To this end, we first consider the conventional estimate of the covariance between y_{ti} and $y_{i,t+1}$ per person, that is,

$$
s_i(t+1, t) = \frac{1}{T_i^*} \sum_{t=1}^{T_i^*}\left\{y_{ti}y_{it+1} - T_i^*\bar{y}_{i\cdot(1)}\bar{y}_{i\cdot(2)}\right\}.
$$

Taking the expectation of this covariance, conditional on i, gives

$$
\begin{aligned}
E[s_i(t+1, t)\,|\,i] = \frac{1}{T_i^*} \sum_{t=1}^{T_i^*} \Big\{ & E[y_{ti}y_{i,t+1}\,|\,i] \\
& - T_i^* E[\bar{y}_{i\cdot(1)}\bar{y}_{i\cdot(2)}\,|\,i]\Big\}.
\end{aligned} \tag{A6}
$$

Focussing on the first expectation on the right-hand side of Equation A6, we make use of the fact that $y_{i,t+1} = c_i + \phi_i y_{ti} + e_{i,t+1}$, such that we can write

$$
\begin{aligned}
E[y_{ti}y_{i,t+1}|i] &= E\Big[y_{ti}\{c_i + \phi_i y_{ti} + e_{i,t+1}\}|i\Big] \\
&= E[y_{ti}c_i|i] + E[\phi_i y_{ti}^2|i] + E[y_{ti}e_{i,t+1}|i] \\
&= (1 - \phi_i)\mu_i E[y_{ti}|i] + \phi_i E[y_{ti}^2|i] \\
&= (1 - \phi_i)\mu_i^2 + \phi_i(\sigma_i^2 + \mu_i^2) \\
&= \mu_i^2 + \phi_i \sigma_i^2,
\end{aligned} \tag{A7}
$$

where $c_i = \mu_i(1 - \phi_i)$, $\mu_i = E(y_{ti}|i)$ and $\sigma_i^2 = \text{Var}(y_{ti}|i) \propto \frac{1}{1-\phi_i^2}$.

The second expectation on the right-hand side of Equation A6 can be rewritten (using the geometric series), to obtain

$$
\begin{aligned}
E[\bar{y}_{i\cdot(1)}\bar{y}_{i\cdot(2)}\,|\,i] &= \frac{1}{T_i^* T_i^*} \sum_{t=1}^{T_i^*}\sum_{\tau=1}^{T_i^*} E[y_{ti}y_{i,\tau+1}\,|\,i] \\
&= \frac{1}{T_i^* T_i^*} \sum_{t=1}^{T_i^*}\sum_{u=1}^{T_i^*}\left\{\phi_i^{|t-u-1|}\sigma_i^2 + \mu_i^2\right\} \\
&= \sigma_i^2 \frac{T_i^*(1-\phi_i^2) - (1+\phi_i^2)(1-\phi_i^{T_i^*})}{T_i^{*2}(1-\phi_i)^2} \\
&\quad + \mu_i^2.
\end{aligned} \tag{A8}
$$

Inserting the expression in Equations A7 and A8 in Equation A6, the expected value for the covariance conditional on person i can be expressed as

$$
E[s_i(t+1, t)\,|\,i] = \sigma_i^2 \left\{\phi_i - \frac{T_i^*(1-\phi_i^2) - (1+\phi_i^2)(1-\phi_i^{T_i^*})}{T_i^{*2}(1-\phi_i)^2}\right\} \tag{A9}
$$

In a similar way, the expected value of the variance conditional on person i can be obtained, resulting in

$$E\big[s_i(t,t)\,|\,i\big] = \sigma_i^2 \left\{ 1 - \frac{T_i(1-\phi_i^2) - 2\phi_i(1-\phi_i^{T_i})}{T_i^2(1-\phi_i)^2} \right\}. \text{ (A10)}$$

By the law of large numbers, as the number of participants $n \to \infty$, the numerator on the right-hand side of Equation A5 converges in probability to $E[s_i(t+1,t)] = E\{E[s_i(t+1,t)\,|\,i]\}$, while the denominator converges in probability to $E[s_i(t,t)] = E\{E[s_i(t,t)\,|\,i]\}$. Therefore,

$$\hat\phi \xrightarrow{p} \frac{E\big[s_i(t+1,t)\big]}{E\big[s_i(t,t)\big]} = \frac{E\left[\sigma_i^2\left\{\phi_i - \frac{T_i^*(1-\phi_i^2)-(1+\phi_i^2)(1-\phi_i^{T_i^*})}{T_i^{*2}(1-\phi_i)^2}\right\}\right]}{E\left[\sigma_i^2\left\{1 - \frac{T_i(1-\phi_i^2)-2\phi_i(1-\phi_i^{T_i})}{T_i^2(1-\phi_i)^2}\right\}\right]}.$$
$$\text{(A11)}$$

To show that in general the asymptotic bias will be negative, for simplicity, we assume T_i is large enough to treat $T_i^* \approx T_i$. Then we have to show that

$$\frac{E\left[\sigma_i^2\left\{\phi_i - \frac{T_i^*(1-\phi_i^2)-(1+\phi_i^2)(1-\phi_i^{T_i^*})}{T_i^{*2}(1-\phi_i)^2}\right\}\right]}{E\left[\sigma_i^2\left\{1 - \frac{T_i(1-\phi_i^2)-2\phi_i(1-\phi_i^{T_i})}{T_i^2(1-\phi_i)^2}\right\}\right]} \le \phi.$$

As the denominator is always positive[6] when $T_i \ge 2$, this is equivalent to showing that

$$E\left[\sigma_i^2(\phi_i - \phi)\right]$$
$$+ E\left[\sigma_i^2 \frac{T_i(1-\phi)(1-\phi_i^2) - (1-2\phi\phi_i+\phi_i^2)(1-\phi_i^{T_i})}{T_i^2(1-\phi_i)^2}\right] \ge 0.$$

The first term on the left-hand side is the sum of the covariance between autoregressive parameter ϕ_i and the variance of the series σ_i^2. Note that $\sigma_i^2 = \sigma_e^2/(1-\phi_i^2)$, which implies that σ_i^2 and ϕ_i are correlated. For symmetric distributions of ϕ_i around ϕ, using a Taylor expansion of $\frac{\phi_i - \phi}{1-\phi_i^2}$, we can show[7] that the correlation

---

[6]This follows if $[T(1-\phi^2)-2(1-\phi^T)]/[T^2(1-\phi)^2] < 1 \iff T(1-\phi^2)-2(1-\phi^T) < T^2(1-\phi)^2$, or $0 < T(T-1)-2(T^2-1)\phi+T(T+1)\phi^2-2\phi^{T+1}$ for all ϕ. The latter may be seen to be true by graphing the function, or by differentiation. To this end, let's denote the right-hand side by Q. The second derivative is $Q'' = 2T(T+1)(1-\phi^{T-1})$ which is clearly always positive when $\phi \in (-1,1)$. Hence, the first order derivative $Q' = -2(T^2-1)+2T(T+1)\phi-2(T+1)\phi^T$ is increasing on $(-1,1)$ and reaches a maximum of 0 at $\phi = 1$. At any ϕ lower than $\phi = 1$ therefore, $y' < 0$ (for instance at $\phi = 0$, $Q' = -2(T^2-1) < 0$ assuming $T > 1$) which implies that Q must be strictly decreasing on $(-1,1)$. In fact, Q reaches a minimum at $Q' = 0 \iff \phi = 1$, at which point $Q = 0$. Therefore, the minimum of the right-hand side of the inequality is always positive, and so the inequality holds for all $\phi \in (-1,1)$, as required.

[7]The Maclaurin series of $\frac{\phi_i - \phi}{1-\phi_i^2} = \frac{u_{1i}}{1-(\phi+u_{1i})^2} = \sum_{k=0}[(1-\phi)^{-k}-(-1-\phi)^{-k}]u_{1i}^k$, hence $E\{\frac{u_{1i}}{1-(\phi+u_{1i})^2}\} = \sum_{k=0}[(1-\phi)^{-k}-(-1-\phi)^{-k}]E\{u_{1i}^k\}$.
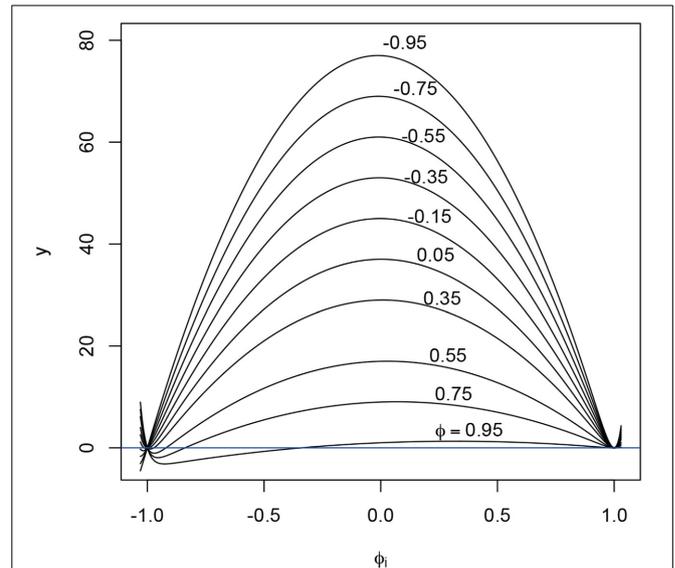
FIGURE A1 | Numerator of expectation, [that is, $y = T(1-\phi)(1-\phi_i^2) - (1-2\phi\phi_i+\phi_i^2)(1-\phi_i^T)$] plotted against ϕ_i, for $T = 40$ and different values of ϕ (i.e., average ϕ_i). Note that only for $\phi = 0.95$ the numerator becomes negative on a substantial portion of the interval $(-1,1)$. See text for implications.

is positive if $\phi > 0$ and negative if $\phi < 0$. Therefore, assuming a density $f_\Phi(\phi_i)$ that is (approximately) symmetric about ϕ, the first term should be deemed positive if $\phi > 0$, and negative, if $\phi < 0$.

Regarding the second term, since the denominator in this term is always positive, the only way this expected value can be negative is if on a substantial portion of the support of the density $f_\Phi(\phi)$ of ϕ_i the numerator is negative. Since no closed form expression for this region can be found in terms of T and ϕ, below we plot the numerator for $T = 40$ and different values of ϕ in **Figure A1**. The interval for ϕ is limited to $(-1,1)$ by the requirement of stationarity. The picture is slightly different for uneven T, but since the term $1-\phi^T$ only really matters near the edges of the interval, it has little effect on the global shape. It is clear that only for extreme values of ϕ ($\phi > 0.9$) a substantial portion of the numerator is negative, but then also only at the lower end of the interval. This means that for any reasonable $f_\Phi(\phi)$ (which incidentally must have $\int_{-1}^{1} \phi f_\Phi(\phi)d\phi = E(\phi_i) > 0.9$ and therefore cannot have a large probability mass in the area where the numerator is negative), the numerator will be positive and hence the second term will be positive. As T_i grows larger, the always positive term $T_i(1-\phi)(1-\phi_i^2)$ in the numerator becomes much larger than the term $(1-2\phi\phi + \phi^2)(1-\phi^T)$. Hence for values of T_i larger than 40, the negative region of the numerator in the expectation vanishes. As a result, the estimator $\hat\phi$ will in most cases underestimate the real value of ϕ.

For u_{1i} symmetrically distributed about 0, the odd moments are zero and the even moments are positive, and so the sign of $E\{\frac{u_{1i}}{1-(\phi+u_{1i})^2}\}$ only depends on the sign of $(1-\phi)^{-k}-(-1-\phi)^{-k}$ where $k = 2m$ even. It is readily verified that this is negative if and only if $\phi < 0$.

From Equation A11 it can be seen that as the lengths T_i of the observed series increase without bound, $\hat{\phi}$ converges in probability to $E\{(\sigma_i^2 + \mu_i^2)\phi_i\}/E\{(\sigma_i^2 + \mu_i^2)\}$. The disconcerting consequence is that the OLS estimator may be biased, even if an infinity number of samples is obtained!

THE CASE OF $\sigma_{u1}^2 = 0$

In the first set of simulations for the multilevel autoregressive model, $\sigma_{u1}^2 = 0$—that is, all individuals were characterized by the same autoregressive parameter (i.e., $\phi_i = \phi$ with probability 1). In this case, setting $T_i = T$ without loss of generality, the above inequality simplifies to

$$E\left[\sigma_i^2 \frac{T(1-\phi)(1-\phi^2)-(1-\phi^2)(1-\phi^T)}{T^2(1-\phi)^2}\right]$$

$$= \left(1 - \phi^2\right)\frac{T(1-\phi)-(1-\phi^T)}{T^2(1-\phi)^2} E[\mu_i^2] \geq 0,$$

where the first expectation dropped out because $E[\sigma_i^2(\phi_i - \phi)] = 0$ (since $\phi_i - \phi = 0$ when $\phi_i = \phi$).

The above inequality is satisfied if $g = T(1 - \phi) - (1 - \phi^T) \geq 0$. This is true for all $-1 \leq \phi \leq 1$, as then $g' = -T + T\phi^{T-1} \leq 0$ for all $T = 1, 2, ...,$ and g achieves a minimum of 0 at $\phi = 1$. Hence, in this case, the estimator of ϕ is always biased downwards.