A Mixed Model to Disentangle Variance and Serial Autocorrelation in Affective Instability Using Ecological Momentary Assessment Data

DOI: 10.1080/00273171.2016.1159177
Handle: http://hdl.handle.net/1942/24009

A Mixed Model to Disentangle Variance and Serial Autocorrelation in Affective Instability using Ecological Momentary Assessment Data

blind

blind

Author Note

Abstract

Affective instability, the tendency to experience emotions that fluctuate frequently and intensively over time, is a core feature of several mental disorders including borderline personality disorder. Currently, affect is often measured with Ecological Momentary Assessment protocols, which yield the possibility to quantify the instability of affect over time. A number of Linear Mixed Models are proposed to examine (diagnostic) group differences in affective instability. The models contribute to the existing literature by estimating simultaneously both the variance and serial dependency component of affective instability when observations are unequally spaced in time with the serial autocorrelation (or emotional inertia) declining as a function of the time interval between observations. In addition, the models can eliminate systematic trends, take between subject differences into account and test for (diagnostic) group differences in serial autocorrelation, short-term as well as long-term affective variability. The usefulness of the models is illustrated in a study on diagnostic group differences in affective instability in the domain of eating disorders. Limitations of the model are that they pertain to group (and not individual) differences and do not focus explicitly on circadian rhythms or cycles in affect.

Keywords: affective instability; serial dependency; autocorrelation; variance; Ecological Momentary Assessment (EMA);Linear Mixed Model (LMM)

A Mixed Model to Disentangle Variance and Serial Autocorrelation in Affective Instability using Ecological Momentary Assessment Data

Introduction

Affective instability in humans is the tendency to experience emotions that fluctuate frequently and intensively over time. Measurement of the temporal instability of affect is important in psychology as it is a defining characteristic of several mental disorders including, amongst others, mood cycling disorders and borderline personality disorder (American Psychiatric Association, 2013; Linehan, 1993). For example, in comparison with healthy controls, subjects with a unipolar mood disorder are supposed to show less affective instability whereas subjects with a borderline personality disorder are supposed to show more affective instability. In psychology, affective instability is often quantified using data obtained by Ecological Momentary Assessment (EMA) protocols. EMA has many names (e.g., diary studies, experience sampling, ambulant monitoring) and exists in many shapes (e.g., time contingent, event contingent, and so forth) but essentially consists of the intensive repeated measurements of individuals in their natural circumstances at specific moments in time. For an overview of diary methods, designs, and intensive longitudinal methods, see Bolger and Laurenceau (2013), and Mehl and Conner (2012). An example of a typical data set in EMA research consists of a number of subjects who reported on their momentary affect (e.g., I feel sad) at random occasions during the day for several days contingent to a signal generated by a smartphone. Originally, the probably by far most used measure to quantify affective instability in psychological research is the within person variance (WPV) (Eid & Diener, 1999; Farmer, Nash, & Dance, 2004; Hoffman, 2007; Zeigler-Hill & Abraham, 2006). This measure simply is the variance of an affective variable of a subject i $(1 \dots i \dots I)$ over Oi $(1 \dots o \dots Oi)$ repeated occasions in time. However, there are a number of problems associated with the WPV as a quantification of affective instability as argued by several authors (Ebner-Priemer et al., 2007; Jahng, Wood, & Trull, 2008; Wang, Hamaker, & Bergeman, 2012).

In the next sections, we first discuss these problems and other model requirements

that are important in the study of affective instability together with the existing statistical models in literature. In addition, we will also highlight the novel contributions and limitations of the Linear Mixed Models (LMM) we will propose. Next, we describe a data set on affective instability in eating disorders. Finally, we will propose a LMM that deals with the problems and issues discussed in the first section and illustrate it in an application in the domain of eating disorders. In the discussion, the merits and limitations of our proposed model in relation to existing models will be discussed.

Model Requirements in the Study of Affective Instability

Problems associated with the within person variance

A first problem with the WPV as an index of affective instability (Ebner-Priemer et al., 2007; Jahng et al., 2008; Larsen, 1987; Wang et al., 2012) relates to the fact that affective instability actually comprises two components: variability on the one hand and serial dependency on the other hand. Failing to distinguish between both components of affective instability may lead to confusion of two characteristics of the time series process (the repeated measurements of affect over time). To demonstrate both components of affective instability –variability and serial autocorrelation– Jahng et al. (2008) generated a time series of 100 values from an autoregressive process of order 1 with a WPV of 1 and an autocorrelation of .50. Next, these authors generated a second time series simply by rearranging randomly the 100 values of the first series resulting in the absence of any serial autocorrelation. It is obvious that both series have the same WPV but a different serial autocorrelation. When we assume that both series are affect scores from two different subjects, the WPV fails to distinguish both series and fails to detect the differences in affective instability. Next, Jahng et al. (2008) generated a third time series by multiplying each value of the first time series by 2 resulting in two time series with the same serial autocorrelation but a different WPV. In this case, the serial autocorrelation fails to distinguish both series but the WPV can detect the differences in affective instability between subjects. These three time series clearly illustrate that both variability and serial dependency (or autocorrelation) are distinctive components

of affective instability and we aim for a model that allows one to examine group differences in affective instability by the estimation of both components. Moreover, research has proved that it is useful to consider both components separately as they have different predictability for health outcomes (Wang et al., 2012). In this context, Suls, Green, and Hillis (1998) translated serial autocorrelation into the concept of affective inertia –the extent to which affect at one particular moment is carried over to subsequent moments– and several studies have shown that higher levels of emotional inertia are associated with more neuroticism, more depression and lower psychological adjustment (Brose, Schmiedek, Kova, & Kuppens, 2015; Koval & Kuppens, 2012; Koval, Kuppens, Allen, & Sheeber, 2012; Kuppens, Allen, & Sheeber, 2010; Suls et al., 1998). Note, however, that the model we will propose is developed to examine (diagnostic) group differences in affective instability with the latter being the dependent variable. As a result, the model is not developed to predict outcome variables (e.g. health outcomes) as have been done in these studies.

Models for affective instability, variability and serial correlation

The last decade, availability of EMA data has stimulated the development of new modeling approaches for EMA data and affect in general (see Bolger & Laurenceau, 2013; Mehl & Conner, 2012) and affective instability in particular. For example, Jahng et al. (2008) have proposed to calculate Mean Squared Successive Difference (MSSD) of successive affective states as an index to quantify affective instability. Using this index, diagnostic group differences in MSSDs are modeled using a non-linear mixed model with gamma error distribution and log link (For more details, see Jahng et al., 2008). These authors have argued that this index is sensitive for both the variability and serial autocorrelation of affective instability. However, the MSSD measure is only an index that is sensitive for variability and serial autocorrelation but does not allow for the estimation of both components separately. Moreover, Wang et al. (2012) have pointed to serious limitations of the MSSD showing that subjects with exactly the same MSSD

can be characterized by obviously different first-order autoregressive processes. These authors themselves have proposed an alternative method using a Bayesian estimation approach that models interindividual differences in intra individual variation of affect by separately considering the variability and temporal dependency component (Wang et al., 2012). In this model, the stationary time series for different subjects, after detrending, is modeled using an autoregressive model of lag k (AR(K)), in which the AR parameters as well as the error/innovation variances may be subject-specific. These subject-specific parameters can be predicted on the basis of covariates using a model with log link function and gamma distribution. Moreover, this model makes it possible to predict outcome variables using both the variability and serial dependency component of affective instability.

In addition, other models have been proposed that focus on one of the two components of affective instability. Heterogeneous mixed models, for example, focus on the variability component to examine the effect of covariates on between and within subject variance (Hedeker, Berbaum, & Mermelstein, 2006; Hedeker, Demirtas, & Mermelstein, 2009; Hedeker, Mermelstein, & Demirtas, 2008, 2012; Hoffman, 2007). In these models, the between subject and within subject variance on the natural logarithm scale can be predicted on the basis of a linear combination of covariates. Moreover, Hedeker and colleagues recently extended their approach by allowing the inclusion of random subject effects permitting the within subject variance to vary at the subject level, above and beyond the influence of covariates on this variance (Hedeker et al., 2008, 2012). Next, research on emotional inertia has focused on the serial correlation component of affective instability (Brose et al., 2015; Koval & Kuppens, 2012; Koval et al., 2012; Kuppens et al., 2010; Suls et al., 1998) and related emotional inertia to outcome variables. In this research, the serial autocorrelation is examined using autoregressive multilevel models by regressing affect at one moment in time by affect at the previous moment in time and relating the autocorrelation parameters to covariates.

A somewhat different approach, with focus on within person affective dynamics rather than instability, are the differential equation models (Deboeck, 2012; Oravecz &

Tuerlinckx, 2011; Oravecz, Tuerlinckx, & Vandekerckhove, 2011; Oud & Jansen, 2000; Voelkle & Oud, 2013), which model variability in affect around a static or dynamic (changing) equilibrium state. In this context, Oravecz et al. (2011) have proposed a continuous-time state-space stochastic model with following key parameters: (a) a home base, an ideal latent position in the two dimensional (valence versus arousal) core affect space subjects are assumed to be drawn to, (b) variances and covariances representing fluctuations of subjects around the home base, and (c) a regulatory process that governs the strength and direction with which subjects return to their home base.

When the research interest is specifically on cycles like circadian rhythms in affect (e.g., affect associated with menstrual cycles), seasonal changes in mood or more complex time series (than basic AR(1) processes) specific models (rather than the LMM with random effects we will propose) can be considered including frequency-domain time series methods, spectral densitry, Fourier analysis, parametric sinusoidal curve fitting, or ARMA time series models (Baehr, Revelle, & Eastman, 2000; Browne & Nesselroade, 2005; Ram et al., 2005). Finally, in mixture latent Markov models (Crayen, Eid, Lischetzke, Courvoisier, & Vermunt, 2012; Rijmen, Vansteelandt, & De Boeck, 2008), affective variability measured with categorical variables is modeled within and between days with stability and change being represented by transition probabilities between latent affective states that are measured with multiple observed indicators.

Contribution of the proposed model

In this paper, we want to further contribute to these existing models by proposing a LMM to examine diagnostic group (not individual) differences in the variability and serial autocorrelation of affective instability when the data consist of observations that are not equally spaced in time. Indeed, in EMA data, signals are often generated at random occasions during the day to avoid anticipation or actor-observer phenomena. The latter means that subjects' reports on their affect may change because they know in advance (or can predict) the moment that they will have to report on their affect. Consequently, in EMA data, time intervals between reports are often unequally spaced.

As a result, to estimate the serial autocorrelation for such data, it is sensible to consider some kind of time-series covariance structure, where the correlation of the repeated measurements are assumed to be smaller for observations that are further apart in time (Littell, Miliken, Stroup, Wolfinger, & Schabenberger, 2006). Moreover, it has been shown that the autoregressive parameter (and error term) in autoregressive models of order one depend on the length of the time interval (Voelkle & Oud, 2013; Voelkle, Oud, Davidov, & Schmidt, 2012) and simulation studies (Oravecz & Tuerlinckx, 2011) show that the true serial correlation is inaccurately estimated when the length of the time interval is ignored. As a result, many of the time-series covariance structures available are inappropriate because they assume equal spacing (Littell et al., 2006). In this context, it may be noted that most of the models discussed above are only applicable for equally spaced data. Note that Jahng et al. (2008) were aware that equal MSSD values based on a different time interval do not have the same meaning and these authors have proposed a heuristic adaptation procedure to deal with the unequally spaced observations over time. However, as mentioned above, the MSSD is an index that is sensitive for both components of affective instability but does not allow for the estimation of both components. Further, also the continuous-time state-space stochastic model of Oravecz et al. (2011) can deal adequately with unequally spaced measures in time but this model is developed to model intraperson affective dynamics and does not focus explicitly on the concept of affective instability as, for example, used in the Diagnostic and Statistical Manual of Mental Disorders-V (DSM-V) (American Psychiatric Association, 2013).

In addition, the LMM we propose takes a number of other issues into account that are important in the study of affective instability. First, it is well-known that the variance of a time series that systematically decreases or increases over time will overestimate the actual dispersion of scores around the general trend line (Shumway & Stoffer, 2006). For example, when a linear trend is added to a stationary time series, the variance of the obtained times series is substantially larger than the original stationary time series. As a result, the estimation of the serial autocorrelation requires

a stationary time series without trends. Failing to remove trends may result in the estimation of spurious high autocorrelations (Shumway & Stoffer, 2006). As a result, systematic trends should be eliminated before estimating the serial autcorrelation as recognized by many authors (Jahng et al., 2008; Tennen, Affleck, & Armeli, 2005; Wang et al., 2012; West & Hepworth, 1991).

Second, another consideration in the quantification of affective instability has to do with the general time frame (Jahng et al., 2008). Some individuals may be characterized by very short-term instability reflected by hourly fluctuations within days whereas other subjects may show a fairly stable affect for several days. For example, a diagnostic criterion for borderline personality disorder is affective instability that is due to a marked reactivity of mood like intense episodic dysphoria, irritability, or anxiety usually lasting a few hours and only rarely more than a few days (American Psychiatric Association, 2013). On the other hand, in patients with mood disorders, a major depressive episode is described as a depressed mood that persists for most of the day, nearly every day, for at least two consecutive weeks (American Psychiatric Association, 2013). In general, we want to point to the fact that it is useful to make a distinction between short-term within day and long-term between day variability and that both sources of variability may occur in all kinds of combinations.

For example, in Figures 1a-b-c-d, we see the affective instability of several hypothetical subjects with measurements on a pleasure-displeasure valence dimension on 10 occasions during 7 days (each vertical dashed line indicates a new day). In Figure 1a, we generated two time series of two hypothetical subjects with the same variability and the same serial autocorrelation within days but added a day effect for the subjects with full line. In Figure 1a, it can be seen that both subjects show a different between day variability. In Figure 1b, however, after removal of the between day effect, it is clear that both subjects show the same within day variability. In Figures 1c-d, we did something similar but with two hypothetical subjects who differ from one another in both between day and within day variability. After removal of the between day effect, it can be seen that the subject with dashed line shows more within

day variability. The LMM we propose wants to further contribute to the existing literature by taking short-term within day and long-term between day variability into account as this in not the case in most of the models discussed above (for an exception, see Jahng et al., 2008; Rijmen et al., 2008).

Finally, beside within day and between day variability, it is also important to take the between subject variance into account. When comparing different diagnostic groups, variability in subjects' mean affect may also differ between diagnostic groups. For example, between subject variability may be larger in samples of subjects with borderline personality disorders than in healthy controls. Such between subject differences may be related to all kinds of known and unknown subject characteristics (e.g., gender, age, biological factors) that are related to affect scores and such variability should be taken into account when quantifying affective instability.

Summarizing, we want to propose a LMM to examine diagnostic group differences in affective instability (a) that allows for the explicit estimation of both the variability and serial autocorrelation component of affective instability, (b) that explicitly models the serial autocorrelation as a function of the length of the time interval between successive observations, and (c) that deals with all remaining issues mentioned above.

Affective Instability in Eating Disorders: Design and Data

To illustrate our models, we will use a dataset in the domain of eating disorders. For reasons that go beyond the scope of this article (For literature on this topic, see Corstorphine, Waller, Ohanian, & Baker, 2006; Mayer, Waller, & Walters, 1998; Vansteelandt et al., 2012; Waller, Kennerley, & Ohanian, 2007), we will examine whether three groups of eating disorders differ from one another in terms of affective instability. The dataset consists of 21 patients with Anorexia Nervosa-Restrictive Type (AN-RT), 17 patients with Anorexia Nervosa-Binge Purging Type (AN-BPT), and 20 patients with Bulimia Nervosa-Binge Purging Type (BN-BPT). The study design is an Ecological Momentary Assessment (EMA) design with patients receiving a hand held computer that generated signals at 9 random times a day (one random signal in nine blocks of 90 minutes with each minute in the block having the same probability of being

selected) during a one week period (9 signals/day x 7 days= 63 repeated measurements).

Affect is measured in terms of core affective states (Russell, 2003), which are simple affective states like 'I feel angry', or 'I feel excited', and so forth that are consciously accessible for a subject at any time of the day (without further reflections on the reasons why you are in that state). There is a vast amount of empirical (factor analytical) research on core affect (e.g., Feldman Barrett & Russell, 1998; Kring, Feldman Barrett, & Gard, 2003) that revealed that a large part of the information conveyed in people's daily affective experiences can be captured by two dimensions: First, the valence or pleasure-displeasure dimension which varies from one extreme positive pole (e.g., ecstasy) over a neutral midpoint to the opposite negative extreme pole (e.g., agony). Second, the activation-deactivation dimension that varies from the extreme deactive pole (e.g., sleep) over a neutral midpoint to the opposite extreme active pole (e.g., excitement). At each quadrant of the core affect space, an emotion word ('stressed','sad','elated', and 'relaxed) was selected inspired by the factor analytic research on core affect (Diener & Emmons, 1984; Russell, 2003; Russell & Barrett, 1999). These emotion words were translated into Dutch and synonyms with a very similar meaning (e.g., nervous and stressed) were sought for. At each signal, participants had to report on their momentary affect using these emotion words (using a 0–4 scale). For reasons of parsimony, we will restrict the analysis to the valence dimension. At each signal, two valence scores were calculated by adding and subtracting affect words with respectively positive and negative valence. As a result, all scores have a natural neutral midpoint of 0 and vary from extreme negative valence (-8) to extreme positive valence ($+8$). In this dataset, patients had on average 112 (out of 2 x 9 x 7 = 126) valid reports (sd= 18.13, min= 54, max= 126) on valence. As a result, as is typical for EMA studies, many repeated measurements per subject are obtained yielding the possibility to model variances (see Hedeker et al., 2008, 2012).

Model

In the next section, we will first discuss a LMM to estimate affective instability in one group. Subsequently, we will discuss how the model can be adapted for different types of datasets (e.g., data with only one measurement per signal nested within days, data with end-of -the-day measurements, and so forth) and what the implications are for the estimation of affective instability. Finally, in a third section, we explain how diagnostic group differences in affective instability can be examined. First and foremost, however, we want to mention that we have used the word variability rather loosely until now referring to its general meaning as well as to its specific meaning of variance as mathematically defined. When presenting the models below, it may be clear that variability refers to variance as mathematically defined.

Model to Estimate Affective Instability in One Group

We propose a LMM (Verbeke & Molenberghs, 2000) to model affective instability using EMA data, which allows for the estimation of both the variance and serial autocorrelation component of affective instability and which deals with the other issues mentioned above. In this model, Y_{msdi} is Measurement m ($m = 1 \dots m \dots M_{sdi}$) of affect variable Y, for example valence, at Signal s ($s = 1 \dots s \dots S_{di}$) nested in Day d ($d = 1 \dots d \dots D_i$) of Subject i ($i = 1 \dots i \dots I$). The measurements m are considered to be exchangeable items measuring the same construct at a particular signal. For example, in the dataset described above, 58 subjects ($I = 58 = (21 + 17 + 20)$) have two valence measures ($M_{sdi} = 2$) at 9 random signals ($S_{di} = 9$) during 7 days ($D_i = 7$). Note that, due to missing data, the number of measurements may vary over signals, days, and subjects; the number of signals may vary over days and subjects, and the number of days may vary over subjects. The model is a four-level LMM with measurements nested within signals, which are nested within days, which in turn are nested within subjects. The model can be written as follows:

$$Y_{msdi} = \beta_{0000} + r_{000i} + r_{00di} + r_{0sdi} + \epsilon_{(1)msdi} + \epsilon_{(2)msdi} \tag{1}$$

with β_{0000} an overall intercept. A representation of Model (1) with separate equations for each level is available in Appendix *A*. At this point, we want to emphasize that Diggle, Heagerty, Liang, and Zeger (2002) have clarified that, theoretically, the total covariance structure of a LMM like Model (1) can be decomposed in at least three sources of random variation:(a) random effects, (b) serial correlation, and (c) measurement error. In the next sections, we will explain the model by discussing these three sources of random variation.

Random effects: Between subject, between day, and within day variance. First, when units –subjects, days, and signals– are randomly sampled from a population, various aspects of their behavior may show stochastic variation between these units (e.g., there may be high and low responders and/or high and low response days). In Model (1) these random effects are denoted by r_{000i}, r_{00di}, and r_{0sdi} being respectively subject specific, day by subject specific, and signal by day by subject specific random effects. These random effect are all assumed to be normally distributed with mean zero and variances $\sigma^2_{r_{000i}}$, $\sigma^2_{r_{00di}}$, and $\sigma^2_{r_{0sdi}}$ respectively. Note that these variances are the parameters of interest in the study of affective instability: $\sigma^2_{r_{000i}}$ reflects between-subject differences in mean affect, $\sigma^2_{r_{00di}}$ reflects the long-term between day variance, and $\sigma^2_{r_{0sdi}}$ reflects the short-term within-day variance. To make this model more comprehensible, a graphical representation of Model (1) is shown in Figure 2. In this figure, the small black points represent the data of a hypothetical subject with 2 measurements of valence at 10 signals/day for one week. The vertical dashed lines indicate the start of a new day. The thickest line represents the overall intercept β_{0000} of the model; this is the overall population average valence. The two long horizontal lines indicate subject-specific mean valence scores (averaged over measurements, signals and days) for two subjects, which are obtained by adding/subtracting random subject effects r_{0001} and r_{0002} to the overall intercept (indicated by arrows starting from the overall intercept). The variance of all these random subject effects r_{000i} is the between subject variance ($\sigma^2_{r_{000i}}$). The short straight horizontal lines indicate the day-specific, subject-specific effects (r_{00di}) which vary over days. They are obtained by adding

random day by subject effects (r_{00di}) to the subject-specific long horizontal lines (as indicated by arrows). The variance of all these random day by subject effects is the long-term between day variance ($\sigma^2_{r_{00di}}$). Further, a signal by day by subject random effect r_{0sdi} is added to predict a subject's valence score at a particular signal on a particular day. These random signal by day by subject specific effects (r_{0sdi}) are the distances from the small fluctuating line to the (short) horizontal day-specific lines (at each separate signal); note that these distances may be larger or smaller at different signals (indicated by the arrows for Subject 1 on Day 6). Or stated in words, these signal by day by subject random effects are the distances from affect at a particular signal (averaged over measurements) within a day to the mean affect (averaged over all measurements and signals) of that day. The variance of all these random signal by day by subject effects is the short-term within-day variance ($\sigma^2_{r_{0sdi}}$). This component models the variance of affect (averaged over measurements) within a day.

Note at this point that, conditionally on all these random effects, there are three theoretical possibilities to model the remaining residual (or error) variation (Diggle et al., 2002): First, measurements at the same signal are perfectly correlated and measurements at different signals are less correlated with the correlation declining as a function of the time interval. This is a model with serial correlation but no measurement error ($\epsilon_{(1)msdi}$ but not $\epsilon_{(2)msdi}$ in Model (1)). Second, all measurements within a day of a subject (at the same signal or not) are uncorrelated. This is a model with only measurement error but no serial autocorrelation ($\epsilon_{(2)msdi}$ but not $\epsilon_{(1)msdi}$ in Model (1)). Third, measurements at the same signal are not perfectly correlated and measurements at different signals are correlated with the correlation declining as a function of the time lag. This is a model with both serial correlation and measurement error ($\epsilon_{(1)msdi}$ and $\epsilon_{(2)msdi}$ in Model (1)). Note that it is common in statistical literature on LMMs (Diggle et al., 2002; Verbeke & Molenberghs, 2000) to represent the decomposition of the residual variance –conditionally on the random effects– in residuals that are correlated on the one hand and residuals that are independent on the other hand by respectively $\epsilon_{(1)msdi}$ and $\epsilon_{(2)msdi}$ as is done in Model (1). In the next sections,

we will discuss the last model with both the serial correlation and measurement error.

Serial Autocorrelation. We start with the serial correlation, which is a second source of random variation that is related to time-varying stochastic processes that are operating within subjects. This type of stochastic variation results in a correlation between pairs of affect measured at different signals within the same subject. Typically, the correlation becomes weaker as the time interval increases and this serial correlation is represented in Model (1) by $\epsilon_{(1)msdi}$. The serial correlation involves two parameters, θ, and τ^2 that will be explained later on.

In Figure 2, the serial autocorrelation between valence scores within days is visualized by the fact that the small black lines do not fluctuate completely randomly around the short, straight horizontal day-specific lines but remain some time below or above these lines (indicating serial correlation between successive signals). For this serial autocorrelation, all elements $\epsilon_{(1)msdi}$ are normally distributed with mean zero and variance τ^2, and elements corresponding to measurements taken on the same Day d in Subject i, e.g., at signals with time points T_{sdi} and $T_{s'di}$, are allowed to be correlated where the correlation is modeled as a function $g(|T_{sdi} - T_{s'di}|)$ of the time-lag between the measurements. The variance parameter τ^2 is the variance of the serial autocorrelation (or time series) and can be interpreted in Figure 2 as the distance that the small black lines go away from the straight horizontal day-specific lines with larger (smaller) values of τ^2 implying larger (smaller) deviations from the horizontal day-specific lines. Further, T_{sdi} and $T_{s'di}$ denote the time passed since the start of the day (e.g., hours passed since 7 : 30 a.m.) at Signal s and s' nested in Day d in Subject i. Meaningful options for $g(.)$ are an exponential or Gaussian serial autocorrelation (Littell et al., 2006; Verbeke & Molenberghs, 2000), respectively:

$$h_{ss'di} = \exp\left(\frac{-|T_{sdi} - T_{s'di}|}{\theta}\right) \tag{2}$$

$$h_{ss'di} = \exp\left(\frac{-\left(T_{sdi} - T_{s'di}\right)^2}{\theta}\right) \tag{3}$$

with $h_{ss'di}$ being the (s, s')-element of a (S_{di} x S_{di})-correlation matrix.

These expressions clearly show that the within day serial autocorrelation is a function of the length of the time interval between successive signals. Consequently, the unequally spaced observations in time are explicitly taken into account. In Figure 3, both types of serial autocorrelations are depicted as a function of the time lag between signals for different values of θ. In this figure, it can be seen that θ models the strength of the decrease in autocorrelation as a function of the time lag between signals. This parameter may be related to the concept of emotional inertia (Brose et al., 2015; Koval & Kuppens, 2012; Koval et al., 2012; Kuppens et al., 2010; Suls et al., 1998), which indicates the extent to which affect at one particular moment is carried over to subsequent moments. As a result, higher values of θ imply higher emotional inertia. Note that the values of θ corresponding to (2) and (3) may be quite different when estimated using the same dataset; as a result, it is only meaningful to compare both fitted functions using the same data-set.

Further, note that for measurements taken at the same signal, the time interval is zero, and expressions (2) and (3) are all one (simply indicating that the serial autocorrelation of lag zero is one). This implies that in a model without measurement error (Model (1) without $\epsilon_{(2)msdi}$), measurements at the same signal are assumed to be perfectly correlated. This also shows that the serial correlation operates between successive signals within days. The serial correlation does not operate between measurements as measurements occur at the same signal (and one cannot have a serial correlation as there is no variation in time). Further, the serial *covariance* –the covariance between observations– equals τ^2 multiplied by expression (2) and (3). Consequently, the serial covariance for measurements at the same signal (with time interval 0) reduces to τ^2, which represents the variance of the within-day serial autocorrelation. As a result, the within day instability can be decomposed into the short-term within day variance, $\sigma^2_{r_{0sdi}}$ on the one hand, and the within day serial

autocorrelation τ^2 on the other hand. By comparing both, it can be concluded whether the within day instability is dominated by within day variance or within day serial autocorrelation. In other words, the within day instability is the amount by which valence at a particular signal deviates from the average valence of that specific day. It can be instantaneous or serially correlated. Instantaneous within day variance (also called short-term within day variance) occurs when the deviation at a particular signal is not related to the deviation at some other signal on the same day. Serially correlated within day variance is present when the deviation at two signals on the same day are correlated with a correlation that decreases with the time lag between both signals. Our model incorporates both components. As such, the short-term within day variance ($\sigma^2_{r_{0sdi}}$) –the variance of the signal by day by subject random effects (r_{0sdi}'s)– is the within day variance that does not involve any serial correlation. Conditionally on this signal by day by subject random effect (and the other random effects in the model) –when the short-term within day variance is removed–, part of the residuals may still show a serial correlation between successive signals. This is within day variance that shows a serial correlation and this serial correlation is based on a time series of successive signals that also has a variance, which is denoted by τ^2. Consequently, both sources of variation within days –short-term within day variance and serial autocorrelation –may coexist with their own particular variance, respectively $\sigma^2_{r_{0sdi}}$ and τ^2. For more information on this decomposition, we refer the interested reader to Diggle et al. (2002).

Further, the exponential serial autocorrelation (2) provides a direct generalization of the first-order Autoregressive (AR(1)) structure for unequally spaced data (Littell et al., 2006) and is equivalent to the power spatial structure for unequally spaced observations (Bolger & Laurenceau, 2013; Littell et al., 2006) written below:

$$h_{ss'di} = \rho^{|T_{sdi} - T_{s'di}|} \tag{4}$$

where ρ is a scale dependent parameter with $|\rho| < 1$. In terms of (2), $\rho = \exp\left(\frac{-1}{\theta}\right)$ (τ^2 remains the same) and corresponds to the serial autocorrelation between two signals

with one hour time interval. Finally, it is worth mentioning that alternative forms for the serial correlation are possible. For an overview and extended discussion on possible forms, see Littell et al. (2006) and Verbeke and Molenberghs (2000).

Measurement Error. Until now, Model (1) with only random effects and the serial autocorrelation ($\epsilon_{(1)msdi}$) assumes that measurements taken at the same signal are perfectly correlated as indicated by Equations (2) and(3). This is not realistic for real data and therefore, a last source of variation, the so-called measurement error, is included as the measurement process itself may add variation to the data. For example, when the same construct is measured twice at the same signal, one would expect identical values but some variability may occur by the measurement process itself, leading to different measured values. This error $\epsilon_{(2)msdi}$ of Model (1) models the variation between measurements at the same signal and is represented In Figure 2, by the distances from the data points at the same signal to the small fluctuating line. All these error terms $\epsilon_{(2)msdi}$ are independent and identically normally distributed with mean zero and variance $\sigma^2_{\epsilon_{(2)msdi}}$. This error is assumed independent of all other sources of variance (between subjects, between days, short-term within day) and the serial autocorrelation. Further, note that the total variation (in the broad sense of the word) in affect within days consists of the short-term within day variance, the serial correlation between successive signals, and the measurement error variance.

Finally, it is worth mentioning that in LMMs, the emphasis is on modeling random effects and their variances first, with the remaining residual variability in the data being modeled in terms of serial correlation and measurement error. Moreover, in the statistical literature on LMMs, is has been demonstrated that there is often strong competition between different stochastic sources like random effects, serial correlation and measurement error. Consequently, variability in one part of the model may disappear in another part of the model and vice versa. In addition, research on LMMs (For an overview, see Chapter 10 in Verbeke & Molenberghs, 2000) has shown that data are often not capable of making a distinction between different serial autocorrelation functions except when many measurements are available in a very short time interval,

which is rarely the case. As a result, conditionally on the random effects, the remaining serial correlation can often be modeled by a rather simple serial autocorrelation (see also Discussion).

As a final note, it is important to realize that when Model (1) is estimated, not all stochastic individual elements ($r_{000i}, r_{00di}, r_{0sdi}$, and $\epsilon_{(2)msdi}$) but only the fixed regression coefficients and the total variance-covariance matrix with all variances parameters (including τ^2 and θ) mentioned above are estimated. Individual predictions for single subjects (conditionally or not on their history at previous signals) can only be performed posthoc making use of the total variance-covariance matrix. For example, individual estimates of the random effects $r_{000i}, r_{00di}, r_{0sdi}$ are obtained posthoc using Empirical Bayes estimates (Verbeke & Molenberghs, 2000).

To facilitate the presentation of the model, we did not include systematic trends in the model. However, when such trends are present, the model can be extended in a straightforward way. For example, when systematic trends over time within days are present, linear (or higher order) effects for time (e.g., a variable indicating the time passed since the start of the day) can be added. These effects can be fixed if the time trends within days are uniformly present in all subjects or can even be made day-specific and subject-specific by including day by subject and subject specific random effects. In a similar way, linear (or higher order terms) for days (e.g., a variable indicating the number of days passed since the start of the study) may be added to deal with systematic trends over days (see also the application later on). When these systematic effects are added to the model, the between subject variance, the long-term between days variance, the short-term within day variance, the serial autocorrelation, and measurement error variance can be quantified after elimination of these trends.

Adaptations of Model (1) for different types of data

Until now, we have assumed that data are available with multiple measurements nested within signals nested within days nested within subjects. However, we realize that such data are not always available in EMA studies. In this section, we discuss

which models are available to analyze 3-level data (e.g., multiple signals nested in multiple days nested in multiple subjects) or 2-level data (multiple end-of-day reports nested in multiple subjects) and what the implications are for the estimation of the different variance components and serial correlation of the model. We start with some general remarks: First, remind that in Model (1) all concepts are defined as follows: The variance of random effects are called variance components (e.g., between subject variance, long-term between day variance, short-term within day variance). Conditionally on these random effects, the remaining residual variation may consist of errors that show a a serial correlation ($\epsilon_{(1)msdi}$ with a parameter θ, and a variance τ^2), and/or independent errors, called measurement error ($\epsilon_{(2)msdi}$ with variance $\sigma^2_{\epsilon_{(2)msdi}}$). From a mathematical-statistical point of view, the definition of these concepts is unambiguously. For example, in Model (1) and all the models we will discuss, the notion of *measurement* error for $\epsilon_{(2)msdi}$ and its variance ($\sigma^2_{\epsilon_{(2)msdi}}$) is unambiguous as this term prevents that observations at the same signal/time are perfectly correlated. In other words, when $\epsilon_{(2)msdi}$ is omitted from a model, this model assumes that observations at the same signal/time –whether collected or not– are perfectly correlated. As a result, one assumes no measurement error. However, from a substantive-psychological point of view, the interpretation of the different parameters of the model may vary depending on the data set at hand. For example, depending on the data, the serial correlation may pertain to successive signals for one data set but to successive days in another data set. Another example pertains to the variance of the measurement error ($\sigma^2_{\epsilon_{(2)msdi}}$); in Model (1) $\sigma^2_{\epsilon_{(2)msdi}}$ reflects pure 'measurement' error (using its psychometric meaning) but, depending on the data set at hand, it may also comprise 'true' within day variation in affect, measurement error, or both (see later). Therefore, we will label $\epsilon_{(2)msdi}$ in this section as independent errors (and $\sigma^2_{\epsilon_{(2)msdi}}$ as the variance of the independent errors) rather that measurement error (or the variance of the measurement error). Second, note that the estimation of $\sigma^2_{\epsilon_{(2)msdi}}$ does not necessarily require that multiple measurements at each signal/time are available. Third, in all the scenarios that will be discussed, the four-level Model (1) reduces to a three- or two-level model, which results in the inability

to estimate particular random effects and their variances. Note in this context that the representation of Model (1) with separate equations for the levels in Appendix A may be helpful to understand the different scenarios.

A first scenario is when only one measurement per signal is available with multiple signals nested within days nested within subjects. In this case, Model (1) reduces to the three-level LMM below:

$$Y_{sdi} = \beta_{000} + r_{00i} + r_{0di} + \epsilon_{(1)sdi} + \epsilon_{(2)sdi} \tag{5}$$

with all parameters being defined and having distributions as in Model (1). In this model, the between subject ($\sigma^2_{r_{00i}}$) and long-term between day variance ($\sigma^2_{r_{0di}}$) can still be estimated but the short-term within day variance cannot be estimated anymore. The reason is that it is impossible to estimate signal by day by subject specific random effects in this three-level model because the variation in valence scores from one signal to another signal within days may be due to 'true' short-term within day variance (as defined in Model (1)), measurement error or both. However, conditionally on the random effects, it is still possible to estimate the serial correlation ($\epsilon_{(1)sdi}$) and the variance of the independent errors ($\sigma^2_{\epsilon_{(2)sdi}}$). As there are different signals within the day, the serial correlation pertains to successive signals within days. Moreover, the independent errors $\epsilon_{(2)sdi}$ are necessary in this model because a model without this term would assume that measurements at the same signal (if they would be present) are perfectly correlated, which is not realistic for real data. However, given the fact that the signal by day by subject random effects and the independent errors can not be distinguished anymore, the variance of the independent errors ($\epsilon_{(2)sdi}$) may comprise both 'true' within day variation in affect from signal to signal (that is not serially correlated), measurement error or both. Therefore, we label $\sigma^2_{\epsilon_{(2)sdi}}$ as the variance of the independent errors rather than the measurement variance. In summary, in this model, the total within day variation (in the broad sense of the word) contains within day serial correlation, 'true' short-term within day variance and measurement error but the two last two sources of variance can not be distinguished anymore.

A second scenario is when multiple measurements are available at one signal nested within days within subjects (e.g., multiple measurements of valence at the end-of-the day). Note that this data-structure is exactly the same as in the previous scenario but with multiple measurements at one signal instead of one measurement at multiple signals. In this case, Model (1) reduces again to the three-level LMM below:

$$Y_{mdi} = \beta_{000} + r_{00i} + r_{0di} + \epsilon_{(1)mdi} + \epsilon_{(2)mdi} \tag{6}$$

with all parameters being defined and having distributions as in Model (1). Note that only the subscript for signal is replaced by a subscript for measurements. In this model, the between subject ($\sigma^2_{r_{00i}}$) and long-term between day variance ($\sigma^2_{r_{0di}}$) can be estimated but the short-term within day variance cannot be estimated anymore. The reason simply is that we only have one signal per day; as a result, there is no information on variation in valence scores between signals within days. Conditionally on these random effects, it is still possible to estimate the serial correlation ($\epsilon_{(1)mdi}$) and the independent errors ($\epsilon_{(2)mdi}$). However, in this case, the serial autocorrelation pertains to successive days as there is only one signal per day and successive signals consequently pertain to successive days; the serial autocorrelation does not pertain to successive measurements as measurements are nested within the same signal/day and cannot show a serial autocorrelation (because there is no variation in time for measurements). Finally, note that in this model, the variance of the independent errors ($\sigma^2_{\epsilon_{(2)mdi}}$) is measurement error. Finally, the last scenario is when only one measurement/signal per day is available like in EMA studies with end-of-the-day reports. In this case, Model (1) reduces to the two-level LMM with days nested within subjects below:

$$Y_{di} = \beta_{00} + r_{0i} + \epsilon_{(1)di} + \epsilon_{(2)di} \tag{7}$$

with all parameters being defined and having distributions as in Model (1). In this case, it is possible to estimate a subject-specific random effect ($\sigma^2_{r_{0i}}$) but no day by subject or signal by day by subject specific random effects anymore. As a result, it is not possible

anymore to estimate the long-term between day and the short-term within day variance. For these data, it is impossible to estimate day by subject random effects in this two-level model because the variation in valence scores from one day to another day may be due to 'true' long-term between day variance (as defined in Model (1)), measurement error or both. The signal by day by subject random effects cannot be estimated anymore because there is no data on signals varying within day (as there is only one signal per day). Conditionally, on these subject-specific random effects, it is still possible to estimate the serial correlation ($\sigma^2_{\epsilon_{(1)di}}$) and independent errors ($\sigma^2_{\epsilon_{(2)di}}$) of the model. In this case, the serial correlation pertains to successive days for the same reason as explained in the previous scenario. Note that, although there is only one signal and measurement per day, it is still possible to estimate the independent error variance ($\epsilon_{(2)di}$). Moreover, the inclusion of the independent errors in the model is necessary as the serial correlation assumes that observations at the same day (if they would be present), would be perfectly correlated which would be unrealistic for real data. However, given the fact that the day by subject random effects and the independent errors cannot be distinguished anymore, the variance of the independent errors ($\epsilon_{(2)di}$) may comprise both 'true' long-term between day variance in affect from day to day (that is not serially correlated) and/or measurement error but both cannot be distinguished anymore. As a result, in this model, the between subject variance, the serial correlation between successive days and the variance of the independent errors can be estimated but the 'true' long-term between day variance and measurement error cannot be distinguished anymore.

As a final remark, note that, depending on the data at hand, serial correlation structures for equally or unequally spaced data may be adopted (for more information on autocorrelation structures, see Littell et al., 2006). From a practical point of view, it is worth mentioning that Model (1) and all the models mentioned above can be estimated using SAS PROC MIXED (SAS Institute Inc., 2011); annotated SAS-code for Model (1) and the different scenarios is available in Appendix *B*.

Model to Estimate Affective Instability in Multiple Groups

Up to now, Model (1) allows the estimation of affective instability in one particular group of subjects. Next, to examine group differences in affective instability, we are mainly interested in the following three parameters: (1) the short-term within day variance ($\sigma^2_{r_{0sdi}}$), (2) the serial autocorrelation within days (with parameters τ^2 and θ), and (3) the long-term between day variance ($\sigma^2_{r_{00di}}$). The between subject variance ($\sigma^2_{r_{000i}}$) is not of direct interest for the analysis of affective instability but should be included in the model to deal with variability in affect that is due to differences between subjects. For example, random subject effects can take systematic differences in affect between subjects into account that may be due to (unknown) subject characteristics (e.g., biological factors). Data on available subject characteristics can be included as predictors in the model in a straightforward way but this does not change the methodology (See Appendix B to include covariates). Note that the variance and serial autocorrelation only model variability in the model that is not explained by such covariates.

To test the research question that different (diagnostic) groups differ from one another in affective instability, we propose four additional LMMs. These four models correspond to the fact that each of the variance parameters and the serial correlation (with its two parameters) mentioned above –short-term within day variance ($\sigma^2_{r_{0sdi}}$), serial autocorrelation (within days) ($\tau^2$and θ), between day variance ($\sigma^2_{r_{00di}}$), and between subject variance ($\sigma^2_{r_{000i}}$) – are made group-specific. For example, the model where the short-term within day variance $\sigma^2_{r_{0sdi}}$ is made group-specific, can be written as follows:

$$Y_{msdi} = \beta_{0000} + r_{000i} + r_{00di} + \sum_{g=1}^{G} r_{g0sdi} z_{gmsdi} + \epsilon_{(1)msdi} + \epsilon_{(2)msdi} \tag{8}$$

with $z_{gmsdi} = 1$ if Subject i belongs to Group g (g=1...g...G), and $z_{gmsdi} = 0$ otherwise. The random effects r_{g0sdi} of each group g are assumed to be normally distributed with variance parameters $\sigma^2_{r_{g0sdi}}$ and zero covariances. All other parameters have the same

distributions and meaning as in Model (1). The other models can be formalized in a similar way.

To test whether each of these models applies to a data-set at hand, the model with one group and the model with more groups are estimated using Restricted Maximum Likelihood and the model without and with group-specific variance components and/or serial correlation are compared using likelihood ratio tests (Verbeke & Molenberghs, 2000). With respect to model selection, it is theoretically possible to make all variance components and the serial correlation group-specific in one single model. When such a model would apply to the data, different groups would be characterized by (a) different short-term within day variance, (b) different serial autocorrelation within days, (c) different (long-term) between day variance, and (d) different between subject variance. It may be obvious that this model is very complex and probably too complex for many datasets. Therefore, we propose to estimate each of the four additional models separately and to evaluate them in terms of increase in likelihood in comparison to the model without group-specific component. Then, for the dataset at hand, the most appropriate model with a combination of group-specific variances and/or serial correlation can be selected.

From a practical viewpoint, it may be noted that these group-specific models can be estimated using the PROC MIXED procedure in SAS (SAS Institute Inc., 2011)(See SAS syntax in Appendix *B*). With respect to data dimensions, estimation procedures for LMMs can be considered stable and feasible for moderately sized data sets (in contrast with generalized LMMs which yield far more computational and/or convergence problems) (Verbeke & Molenberghs, 2000). However, it is impossible to give explicit guidelines in terms of data dimensions as so many factors play its role. In general, one may bear in mind that more data are necessary when the model complexity increases and when the measurements are unbalanced (no fixed time occasions). The reason is that by adding more random effects into the model, more parameters in the variance-covariance matrix have to be estimated. For example, in a two-level model (using fixed measurement occasions) with random intercepts and slopes, at least three

repeated measurements per subject are necessary since this model has a variance-covariance matrix with four parameters (variance for intercepts, variance for slopes, a covariance between both, and an error variance). Note that numerical problems are mostly related to the estimation of the variance-covariance structure and not to the estimation of fixed effects for which a closed form solution exist (weighted least squares). Finally, as missing data are rather the rule than the exception in EMA research, note that inference for the proposed models is valid under the assumption of missingness at random (MAR) (Little & Rubin, 2002; Verbeke & Molenberghs, 2000).

Results

To examine systematic trends within days and over days, Model (1) without serial autocorrelation was estimated with linear trends for passed hours (since the start of the day) and passed days (since the start of the study) as predictors. Results revealed neither a linear effect of passed hours, $F(1, 3234) = 0.18, p = 0.67$, nor a linear effect of passed days, $F(1, 3234) = 3.0, p = 0.08$, and no significant interactions of these variables with diagnostic group, $F(2, 3234) = 0.58, p = 0.56$, and, $F(2, 3234) = 0.17, p = 0.84$, respectively. However, when days were dummy coded in terms of day of the week (Monday, Tuesday, etc.), results indicated that subjects have significant higher valence scores, and feel consequently more positively, on weekend days compared to weekdays, $F(6, 3236) = 8.10, p < 0.0001$. More in particular, pairwise posthoc comparisons, revealed (a) that valence was highest on Saturday, which was significantly different from all weekdays, (b) that there were no significant differences among weekdays, and (c) that Sunday was significantly different from all weekdays with the exception of Thursday and Friday. There was no evidence that this effect was different for the diagnostic groups, $F(12, 3236) = 0.45, p = 0.94$. To further check for systematic trends of valence in the data, an Ordinary Least Squares (OLS) regression model with diagnosis and the dummy coded variable day of the week as predictors was estimated. Next, Locally Weighted Scatterplot Smoothing (LOWESS) was used to plot an average smoothed lowess curve through the scatterplot of time versus the residuals of this OLS

regression within days. These plots depicted almost similar horizontal lines with only very small fluctuations in the different diagnostic groups, which may be ignored given the large amount of variability present in the data. In a similar way, we inspected systematic effect of residuals over days and found no evidence to include (other) systematic effects of days in our model. These plots are available as a supplement. As a result, in all analyzes that follow, we have retained diagnosis and the dummy variables indicating day of the week and have dropped the other non-significant effects.

For the selection of the serial autocorrelation, we compared Model (1) with the effects mentioned above without serial autocorrelation to the same model with respectively an exponential and a Gaussian serial autocorrelation (resp. equations 2 and 3). Doing this, based on likelihood ratio tests, we opted for a model with an exponential serial autocorrelation (decrease in -2 log likelihood=192.7). Note that a formal test for the inclusion of the serial autocorrelation is not trivial because the test involves a 0-hypothesis on the boundaries of the parameter space (τ^2 being zero) implying that standard test procedures are invalid (for more information, see Discussion). In this particular case, however, the change in deviance in so large that the inclusion of a serial autocorrelation is justified without doubt.

The results of estimating this model revealed that all three diagnostic groups are, on average, in a neutral valence state with mean affective states of $0.36\,(SE = 0.73)$ for AN-BPT, $0.24\,(SE = 0.66)$ for AN-RT, and $-0.57\,(SE = 0.68)$ for BN-BPT. These means are not significantly different from one another, $F(2, 3236) = 0.53$, $p = 0.59$. Moreover, we found a significant effect for day of the week, $F(6, 3236) = 8.29, p < 0.0001$, with the same interpretation as mentioned above.

Of more interest for our research question, diagnostic group differences in affective instability, are the estimated variance components which are shown in Figure 4a. As can be seen in this figure, a large part of the total variability is related to between subject differences (52.90%) indicating that there are large differences in mean valence between subjects. Further, it can be seen that the long-term between day (6.36%) and short-term within day variance (10.03%) are similar in size but the largest part of the

affective instability is due to the serial autocorrelation within days (24.02%). As a result, the contribution of the serial autocorrelation to the within day instability is about twice the short-term within day variance. Finally, the error component accounts for 6.69% of the variance. In Figure 5a, it can be seen how the serial autocorrelation decreases as a function of hours between successive signals ($\tau^2_{\text{Model(1)}} = 4.02$ and $\theta_{\text{Model (1)}} = 2.0069$). For example, the serial correlation between two valence states with one hour time interval is 0.61.

Next, we estimated the four additional models by making (a) the short-term within day variance, (b) the within day serial autocorrelation, (c) the between day variance, and (d) the between subject variance components group-specific. In Table 1, likelihood ratio tests for each of these four models in comparison with Model (1) with exponential serial autocorrelation are summarized. It can be seen that there is evidence for group-specific short-term within day variance and group-specific within day serial autocorrelation and to a lesser extent for group-specific between day variance.

Next, we estimated a new model for valence with both group-specific short-term within day variance and group-specific within day serial autocorrelation. As expected, the three groups of eating disorders differed from one another in terms of the short-term within day variance and within day serial autocorrelation, $\chi^2(6) = 88.3, p < 0.0001$; the addition of a group-specific between day variance did not improve the model anymore, $\chi^2(2) = 2.7, p = 0.26$. In line with previous results, the three diagnostic groups did not differ from one another in mean valence, $F(2, 2782) = 0.54, p = 0.58$. There was also a significant effect of day of the week, $F(6, 3236) = 8.41, p < 0.0001$, with the same interpretation as mentioned above. In Figure 4b, the estimated variance components for this new model are depicted. When we compare Figures 4a and 4b, it can be seen that the between subject variance, the between day variance, and the error variance are very similar in both models. In addition, the mean of the group-specific variance components for both the short-term within day variance and the serial autocorrelation (τ^2) are very similar to the corresponding variance components in Model (1) ($\theta_{\text{Model (1)}} = 2.0069$ versus $\theta_{\text{AN-BPT}} = 1.6354$, $\theta_{\text{AN-RT}} = 3.6865$, and $\theta_{\text{BN-BPT}} = 2.8269$). In addition, in

Figure 4c, the within day instability for the three diagnostic groups is shown, which consists of the sum of the short-term within day variance and the part of the within day variability that is serial in nature ($\sigma^2_{r_{0sdi}} + \tau^2$). From this figure, it is clear that the group with AN-BPT shows more within day instability in comparison with the other two groups. Further, in Figure 5b, the exponential serial autocorrelations of the three groups are depicted. As can be seen in this figure, the serial autocorrelation is clearly lower for the group with AN-BPT in comparison with the other two groups. As a result, one could say that the groups with AN-RT and AN-BPT are characterized by more emotional inertia. The serial correlation between two valence states with one hour time interval are 0.54, 0.76, 0.70 for the groups with AN-BPT, AN-RT, and BN-BPT respectively. To test these findings more formally, we performed post-hoc pairwise comparison tests comparing Model (1) with respectively a model with group-specific within day variance and a model with group-specific within day serial autocorrelation for all pairs of groups. The results of these analyzes are depicted in Table 2. We can conclude that the group with AN-BPT shows indeed more within day affective instability, both in terms of the short-term within variance and serial dependency, than the other two groups. The latter two groups themselves are not significantly different from one another.

Discussion

A number of LMMs were proposed to examine diagnostic group differences in affective instability. The models we have proposed have several merits in the study of affective instability. First, a key motivation for developing these models was to have a tool for estimating both the variance and serial autocorrelation component of affective instability taking unequally spaced measures in time into account. By doing this, the serial correlation can be interpreted in terms of emotional inertia, reflecting the tendency to carry-over affective states from one moment to another (Brose et al., 2015; Koval & Kuppens, 2012; Koval et al., 2012; Kuppens et al., 2010; Suls et al., 1998). Moreover, the model yields the opportunity to evaluate whether the within day

instability is dominated by variance or serial autocorrelation. As a result, this model contributes to the framework of Jahng et al. (2008) who developed an index that is sensitive for both components but that does not allow simultaneous estimation of both components. Also Wang et al. (2012) have proposed a method to model individual differences in intra-individual variability in affect by separating both variance and temporal dependency but this model only applies to equally spaced measurements in time.

Note that we did not pay too much attention to the nature of the serial autocorrelation. The reason is that the serial correlation models only one aspect of the total variability present in the data with the largest part of the variability being captured by random effects at several levels (subjects, days, signals). Indeed, Chi and Reinsel (1989) and Verbeke and Molenberghs (2000) have argued that there is often strong competition between different stochastic sources like random effects and serial autocorrelation. For example, it is often the case in LMMs that the between subject variability severely dominates the within subject variability (as is also the case in our application), which implies that the exact parametric form of the serial correlation function can hardly be identified. Moreover, Chi and Reinsel (1989) have reported that a sufficient number of random effects in a model with white noise errors may be able to represent the serial correlations among the measurements taken on each individual because serial correlation can be replaced by very smooth subject-specific functions. Finally, Verbeke and Molenberghs (2000) have shown that the precise characterization of the serial correlation function g(.) is often extremely difficult in the presence of several random effects. These authors have illustrated (Verbeke & Molenberghs, 2000, see Section 10.3) that observed longitudinal profiles can often not distinguish between various serial autocorrelation functions, not even when many repeated measurements per subject are available. As such, including a serial autocorrelation, if present, is far more important than correctly specifying the serial autocorrelation function (e.g., exponential versus Gaussian). In practical applications in general, and in our application in particular, there is often a large change in likelihood when comparing a

model with and without serial autocorrelation but the change in likelihood is often similar for different types of serial autocorrelations. This can be explained by the fact that distinguishing exponential from Gaussian serial correlation functions requires measurements taken very closely in time, a feature that is not present in our data, nor in many other similar applications. As a consequence, the residual variability –after taking fixed and random effects into account– is often minimal and can be modeled with a rather simple serial autocorrelation. This also explains why the proposed serial correlation in the model may be conceived as rather simple from a time series point of view (Shumway & Stoffer, 2006). On the contrary, in time series models, no random effects can be included yielding a situation where almost all variability has to be modeled by the residual component requiring a serial autocorrelation structure that is far more complex both in terms of lag and/or nature. In this context, it is worth mentioning that when the nature of the serial autocorrelation, conditionally on a prespecified set of random effects, is of primary interest, one may adopt an approach using so-called fractional polynomials. This approach is flexible enough to allow various shapes to model the serial autocorrelation function (Lesaffre, Asefa, & Verbeke, 1999). However, it is not implemented in standard software yet and the optimization of random effects, serial autocorrelation, as well as a measurement error, is computationally very demanding and requires many repeated observations per subject. As a final remark, one may note that future research may further confirm whether the rather simple autocorrelation structures (e.g. exponential or Gaussian correlation structures) presented before capture adequately the serial correlations present in EMA data as more repeated observations are often present in such data compared to more traditional longitudinal data.

The proposed model may also be helpful to answer questions about the frequency of sampling data in future EMA protocols on affective instability. For example, in borderline personality disorders, affect is assumed to change very abrupt and quickly within hours during the day whereas in bipolar disorders, patients are already considered rapid cyclers when they show four or more different mood episodes during

the year (American Psychiatric Association, 2013). Failing to select the adequate time scale may result in inadequate results; for example, sampling very frequently over a short period of time may incorrectly lead to the conclusion that subjects with bipolar disorder do not show cycles in mood because the cycling occurs at a much slower pace and is not detected in the short time frame that is selected. Although there is general consensus that research questions should govern the sampling scheme of an EMA design (for example, weekly measures, end-of-the-day measurement or multiple measurements in 90 minutes blocks during the day), such decisions are often made on intuitive grounds. In our application, the total variance is decomposed in the measurement error variance, the short-term within day variance, serial autocorrelation, between day variance, and between subject variance. Recently, Shiyko and Ram (2011) have demonstrated that such a variance decomposition approach using LMMs in EMA-type data is useful to identify and quantify the relative speed of change processes, which, in turn may help decisions with respect to sampling frequency in EMA protocols. This is an important issue that has important implications for both the participants' burden and for researchers' ability to capture and study dynamic processes (in affect). Briefly stated, when the total variance is dominated by within day variance and/or serial autocorrelation, one should sample frequently within days, when it is dominated by between day variance, one should sample less within days but at multiple days, when it is equally divided one should sample multiple times within days for multiple days (For more information, see Shiyko & Ram, 2011). In this context, it is noteworthy that LMMs are very flexible in dealing with different time scales because levels –signals nested in days nested in weeks nested in months, and so forth– can easily be added or omitted in modeling the data as discussed in the theoretical section.

However, this research is not without limitations and further model developments are still necessary. In this paper, we have focused on the concept of affective instability and the examination of diagnostic group differences herein. A first limitation is that our models are restricted to the examination of groups (or categorical) differences in affective instability. As such, the variance and serial correlation components are

assumed to be the same within groups. However, one may also be interested in individual differences or within person dynamics in both the variability and serial autocorrelation component of affective instability. One way to examine individual differences (e.g., subjects' score on a dimension, like, for example depression) in the variability component of affective instability is to include time-invariant variables (e.g., a subject's depression score) as random subject, random day by subject, and/or random signal by day by subject effects in the model. In such a model, it is possible to adjust the estimated between subject variance, between day variance and/or within day variance for the continuous variables of interest. However, testing hypotheses that part of the between subject, between day and/or within day variance is related to such a time invariant predictor is not trivial because such hypotheses test for zero-variance components, which are on the boundary of the parameter space. In this case, classical Wald and likelihood ratio tests are invalid and specific adapted tests are needed. For example, it has been shown for some very specific null-hypotheses on the boundaries that the correct asymptotic null-distributions are often a mixture of chi-square distributions rather than a single chi-square distribution (See Stram & Lee, 1994, 1995; Verbeke & Molenberghs, 2000). A full discussion of this issue is beyond the scope of this article but more information can be obtained in Ke and Wang (2014); Morrell (1998); Stoel, Garre, Dolan, and van den Wittenboer (2006); Stram and Lee (1994, 1995); Verbeke and Molenberghs (2000).

In addition, as discussed in the introduction, other models have been proposed to model individual differences (instead of group differences) in affective instability. First, heterogeneous mixed models have been proposed to examine the effect of categorical and continuous covariates on between and within subject variance (but not between day variance) (Hedeker et al., 2008, 2012; Hoffman, 2007). In addition, Wang et al. (2012) have proposed a Bayesian estimation approach for equally spaced data that models interindividual differences in both components of affective instability. In this model, the AR parameters as well as the error/innovation variances may be subject-specific and predicted by covariates. Finally, the differential equation model for affective dynamics

(Oravecz et al., 2011) has great appeal as intra-individual variability and autoregression parameters can be estimated for unequally spaced data. Moreover, all stochastic parameters of the model –the home base, variance, and regulatory processes including the serial correlation– can be made subject-specific and related to covariates. Unfortunately, this model is difficult to implement, can not be estimated using standard software, and comes at a considerable computational cost (Oravecz & Tuerlinckx, 2011). For example, these authors mention that the computation time for an EMA dataset with 80 subjects who report at 63 signals takes about 75 minutes using parallel computing on a computing node. For a similar dataset, estimation time for the models we have proposed is a matter of minutes on a standard computer.

Another limitation of the proposed models is that they do not focus on cycles like circadian rhythms in affect, affect associated with menstrual cycles, and seasonal changes in mood which may be present in the data (Ram et al., 2005). In theory, systematic cycles that are uniformly present in all subject may be dealt with in the fixed part of the model and individual differences may be modeled by the inclusion of random effects. In our application, we found that subjects experience more pleasant feelings during the weekend in line with literature (Armeli, Carney, Tennen, G., & O'Neil, 2000; Vansteelandt, Rijmen, Pieters, Probst, & Vanderlinden, 2007). This effect was modeled by including day of the week as dummy variables. As further analyzes did not reveal evidence for cycles, we have opted to model these effects –as far as they were present– by the serial autocorrelation. However, when the focus of research question is explicitly on cycles, one may adopt models that are especially developed to model them such as frequency-domain time series methods including spectral densitry, Fourier analysis, parametric sinusoidal curve fitting, or ARMA time series models (Baehr et al., 2000; Browne & Nesselroade, 2005; Ram et al., 2005). However, such cycles may be hard to model as they are typically not synchronized across persons and may be characterized by subject-specific frequency, amplitude (minimum and maximum), and phase. In addition, in their research on weekly cycles in affect, Ram et al. (2005) concluded that affect is more likely to be an amalgamation of responses to a multitude

of internal and external factors, and with all of these factors shifting in both systematic and unsystematic ways over time, underlying cycles (if present) will likely be obscured.

Finally, affective instability is the dependent variable in our model and our model is not developed to predict outcome on the basis of affective instability in contrast with other models. For example, Wang et al. (2012) have developed a model in which variability and temporal dependency are included as random effects and that allows one to predict outcome variables using both components. The results of this study showed that both components have differential predictability of health outcomes, and consequently, should be modeled separately. Also in research on emotional inertia (Brose et al., 2015; Koval & Kuppens, 2012; Koval et al., 2012; Kuppens et al., 2010), it has been shown that persons with high neuroticism, low self-esteem and depression are characterized by higher levels of emotional inertia (or higher serial dependency) in both positive and negative emotions. However, in some studies (see, for example, Kuppens et al., 2010), the unequally spaced nature of the EMA data is simply ignored, which may result in inadequate results as discussed before (see Littell et al., 2006; Oravecz & Tuerlinckx, 2011; Voelkle & Oud, 2013; Voelkle et al., 2012).

In general, EMA provides large and complex data and there may be a strong competition between systematic effects and different stochastic sources like random effects, serial autocorrelation and measurement error. As a result, variability modeled in one part of the model may disappear in another part of the model and vice versa. It may be clear that different models focus on different aspects of the data and the selection of a model will strongly depend on the research question. We hope that the proposed models are useful to adequately model group differences in affective instability using EMA data and may further stimulate the development of new intensive longitudinal models for EMA data in general and for the study of affective instability in particular.

References

American Psychiatric Association. (2013). *Diagnosis and statistical manual of mental disorders (5th ed.)*. Washington, DC: American Psychiatric Association.

Armeli, S., Carney, M., Tennen, H., G., A., & O'Neil, T. (2000). Stress and alcohol use: A daily process examination of the stressor-vulnerability model. *Journal of Personality and Social Psychology*, *78*, 979–994. doi: 10.1037/0022-3514.78.5.979

Baehr, E., Revelle, W., & Eastman, C. (2000). Individual differences in the phase and amplitude of the human circadian temperature rhythm with an emphasis on morningness–eveningness. *Journal of Sleep Research*, *9*, 117–127. doi: 10.1046/j.1365-2869.2000.00196.x

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. New York: Guilford.

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

Browne, M., & Nesselroade, J. (2005). Representing psychological processes with dynamic factor models: Some promising uses and extensions of arma times series models. In A. Maydeu-Olivares & J. McArdle (Eds.), *Psychometrics: A festschrift to roderick p. mcdonald*. Mahwah, NJ: Lawrence Earlbaum.

Chi, E., & Reinsel, G. (1989). Models for longitudinal data with random effects and ar(1) errors. *Journal of the American Statistical Association*, *84*, 452–459. doi: 10.2307/2289929

Corstorphine, E., Waller, G., Ohanian, V., & Baker, M. (2006). Changes in internal states across the binge-vomit cycle in bulimia nervosa. *The journal of nervous and mental disorders*, *194*, 446–449. doi: 10.1097/01.nmd.0000221303.64098.23

Crayen, C., Eid, M., Lischetzke, T., Courvoisier, D., & Vermunt, J. (2012). Exploring dynamics in mood regulation–mixture latent markov modeling of ambulatory assessment data. *Psychosomatic Medicine*, *74*, 366–376. doi: 10.1097/PSY.0b013e31825474cb

Deboeck, P. (2012). Modeling nonlinear dynamics in intraindividual variability. In M. Mehl & T. Conner (Eds.), *Handbook of research methods for studying dailiy life* (pp. 407–458). New York: The Guilford Press.

Diener, E., & Emmons, R. A. (1984). The independence of positive and negative affect. *Journal of Personality and Social Psychology*, *47*, 1105–1107. doi: 10.1037/0022-3514.47.5.1105

Diggle, P., Heagerty, P., Liang, K.-Y., & Zeger, S. (2002). Parametric models for covariance structure. In P. Diggle, P. Heagerty, K.-Y. LIang, & S. Zeger (Eds.), *Analysis of longitudinal data (2nd ed.)* (pp. 81–113). New York: Oxford University Press.

Ebner-Priemer, U., Kuo, J., Kleindienst, N., Welsch, S., Reisch, T., Reinhard, I., ... Bohus, M. (2007). State affective instability in borderline personality assessed by ambulatory monitoring. *Psychological Medicine*, *37*, 961–970. doi: 10.1017/S0033291706009706

Eid, M., & Diener, E. (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. *Journal of Personality and Social Psychology*, *76*, 622–676. doi: 10.1037//0022-3514.76.4.662

Farmer, R., Nash, H., & Dance, D. (2004). Mood patterns and variations associated with personality disorder pathology. *Comprehensive Psychiatry*, *45*, 289–303. doi: 10.1016/j.comppsych.2004.03.009

Feldman Barrett, L., & Russell, J. (1998). Independence and bipolarity in the structure of affect. *Journal of Personality and Social Psychology*, *74*, 967–984. doi: 10.1037/0022-3514.74.4.967

Hedeker, D., Berbaum, M., & Mermelstein, R. (2006). Location-scale models for multilevel ordinal data: Between- and within-subjects variance modeling. *Journal of Probability and Statistical Science*, *4*, 1–20. doi: 10.1002/sim.5338

Hedeker, D., Demirtas, H., & Mermelstein, R. (2009). A mixed ordinal location scale model for analysis of ecological momentary assessment (ema) data. *Statistics and Its Interface*, *2*, 391–401. doi: 10.1111/j.1541-0420.2007.00924.x

Hedeker, D., Mermelstein, R., & Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (ema) data. *Biometrics*, *64*, 627–634. doi: 10.1111/j.1541-0420.2007.00924.x

Hedeker, D., Mermelstein, R., & Demirtas, H. (2012). Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. *Statistics in Medicine*, *31*, 3328–3336. doi: 10.1002/sim.5338

Hoffman, L. (2007). Multilevel models for examining individual differences in within-person variation and covariation over time. *Multivariate Behavioral Research*, *42*, 609–629. doi: 10.1080/00273170701710072

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

Ke, Z., & Wang, L. (2014). Detecting individual differences in change: methods and comparisons. *Structural Equation Modeling*. doi: 10.1080/10705511.2014.936096

Koval, P., & 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., & Sheeber, L. (2012). Getting stuck in depression: The roles of rumination and emotional inertia. *Cognition and Emotion*, *26*, 1412–1427. doi: 10.1080/02699931.2012.667392

Kring, A. M., Feldman Barrett, L., & Gard, D. (2003). On the broad applicability of the affective circumplex: Representations of affective knowledge among schizophrenia patients. *Psychological Science*, *14*, 207–214. doi: 10.1111/1467-9280.02433

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

Larsen, R. (1987). The stability of mood variability: A spectral analytic approach to

daily mood assessments. *Journal of Personality and Social Psychology*, *52*, 1195–1204. doi: 10.1037/0022-3514.52.6.1195

Lesaffre, E., Asefa, M., & Verbeke, G. (1999). Assessing the goodness-of-fit of the laird and ware model: An example: The jimma infant survival differential longitudinal study. *Statistics in Medicine*, *18*, 835–854. doi: 10.1002/(SICI)1097-0258(19990415)18:7<835::AID-SIM75>3.0.CO;2-7

Linehan, M. (1993). *Cognitive-behavioral treatment of borderline personality disorder.* New York: Guilford press.

Littell, R., Miliken, G., Stroup, W., Wolfinger, R., & Schabenberger, O. (2006). *Sas for mixed models (2nd ed.).* Cary NC: SAS Intitute Inc.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data (2nd ed.).* New York: Wiley.

Mayer, C., Waller, G., & Walters, A. (1998). Emotional states and bulimic psychopathology. In H. Hoeck, J. Treasure, & M. Katzman (Eds.), *Neurobiology in the treatment of eating disorders* (pp. 271–289). Chistester, UK: Wiley.

Mehl, M., & Conner, T. (2012). *Handbook of research methods for studying daily life.* New York: Guilford Press.

Morrell, C. (1998). Likelihood ratio testing of variance components in the linear mixed-effects model using restricted maximum likelihood. *Biometrics*, *54*, 1560–1568. doi: 10.2307/2533680

Oravecz, Z., & Tuerlinckx, F. (2011). The linear mixed model and the hierarchical ornstein-uhlenbeck model: Some equivalences and differences. *British Journal of methematical and Statistical Psychology*, *64*, 134–160. doi: 10.1348/000711010X498621

Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2011). A hierarchical latent stochastic differential equation model for affective dynamics. *Psychological Methods*, *16*, 468–490. doi: 10.1037/a0024375

Oud, J., & Jansen, R. (2000). Continuous time state space modeling of panel data by means of sem. *Psychometrika*, *65*, 199–215. doi: 10.1007/BF02294374

Ram, N., Chow, S.-M., Bowles, R., Wang, L., Grimm, K., Fujita, F., & Nesselroade, J. (2005). Examining interindividual differences in cyclicity of pleasant and unpleasant affects using spectral analysis and item response modeling. *Psychometrika*, *70*, 773–790. doi: 10.1007/s11336-001-1270-5

Rijmen, F., Vansteelandt, K., & De Boeck, P. (2008). Latent class models for diary method data: parameter estimation by local computations. *Psychometrika*, *73*, 167–182. doi: 10.1007/s11336-007-9001-8

Russell, J. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*, 145–172. doi: 10.1037/0033-295X.110.1.145

Russell, J., & Barrett, L. (1999). Core affect, prototypical episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*, 805–819. doi: 10.1037//0022-3514.76.5.805

SAS Institute Inc. (2011). Introduction to mixed modeling procedures (chapter 6). In S. I. Inc. (Ed.), *Sas/stat 9.3 user's guide* (pp. 125–139). Cary, NC: SAS Institute Inc.

Shiyko, M., & Ram, N. (2011). Conceptualizing and estimating process speed in studies emloying ecologcail momentary assessment designs: A multilevel variance decompostion approach. *Multivariate Behavioral Research*, *46*, 875–899. doi: 10.1080/00273171.2011.625310

Shumway, R. H., & Stoffer, D. S. (2006). *Time series analysis and its applications (2nd ed.)*. New York: Springer.

Stoel, R., Garre, F., Dolan, C., & van den Wittenboer, G. (2006). On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints. *Psychological Methods*, *4*, 439–455. doi: 10.1037/1082-989X.11.4.439

Stram, D., & Lee, J. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, *50*, 1171–1177. doi: 10.1111/sjos.12044

Stram, D., & Lee, J. (1995). Correction to: Variance components testing in the longitudinal mixed effects model. *Biometrics*, *51*, 1196. doi: 10.2307/ 2533038

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

Tennen, H., Affleck, G., & Armeli, S. (2005). Personality and daily experience revisited. *Journal of Personality*, *73*, 1465–1483. doi: 10.1111/j.1467-6494.2005.00355.x

Vansteelandt, K., Claes, L., Muehlenkamp, J., De Cuyper, K., Lemmens, J., Probst, M., ... Pieters, G. (2012). Variability in affective activation predicts non-suicidal self-injury in eating disorders. *European Eating Disorder Review*, *21*, 143–147. doi: 10.1002/erv.2220

Vansteelandt, K., Rijmen, F., Pieters, G., Probst, M., & Vanderlinden, J. (2007). Drive for thinness, affect regulation and physical activity in eating disorders: A daily life study. *Behaviour Research and Therapy*, *45*, 1717–1734. doi: 10.1016/j.brat.2006.12.005

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York: Springer.

Voelkle, M., & Oud, J. (2013). Continuous time modelling with individually varying time intervals for oscillating and non-oscillating processes. *British journal of mathematical and Stastistical Psychology*, *66*, 103–126. doi: 10.1111/j.2044-8317.2012.02043.x

Voelkle, M., Oud, J., Davidov, E., & Schmidt, P. (2012). An sem approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological Methods*, *17*, 176–192. doi: 10.1037/a0027543

Waller, G., Kennerley, H., & Ohanian, V. (2007). Schema focused cognitive-behavior therapy with eating disorders. In L. Riso, P. du Poit, & J. Young (Eds.), *Cognitive schemas and core beliefs in psychiatric disorders: A scientist-practitioner guide* (pp. 139–175). New York: American Psychiatric Association.

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

West, S. G., & Hepworth, J. T. (1991). Statistical issues in the study of temporal data: Daily experiences. *Journal of Personality*, *59*, 609–662. doi: 10.1111/j.1467-6494.1991.tb00261.x

Zeigler-Hill, B., & Abraham, J. (2006). Borderline personality features: Instability of self-esteem and affect. *Journal of Social and Clinical Psychology*, *25*, 668–687. doi: 10.1521/jscp.2006.25.6.668

Table 1
*Model (*1*) and models with group-specific components: Valence*

Model	**−2 log likelihood**	**statistical test**	**p -value**
Model (1) with exponential serial correlation	27022.4		
Group-specific between group variance	27019.2	$\chi^2(2) = 3.2$	0.2018
Group-specific between day variance	27016.1	$\chi^2(2) = 6.3$	0.0428
Group-specific within day variance	26958.4	$\chi^2(2) = 64.0$	< 0.0001
Group-specific serial autocorrelation	26937.2	$\chi^2(4) = 85.2$	< 0.0001

Table 2
Post-hoc Pairwise Comparisons for Affective Instability: Valence

Model	**−2 log likelihood**	**statistical test**	**p-value**
AN-RT vs. AN-BPT			
Model (1) with exponential serial correlation	18148.2		
Group-specific within day variance (WDV)	18102.6	$\chi^2(1) = 45.6$	< 0.0001
Group-specific serial autocorrelation (SAC)	18080.3	$\chi^2(2) = 67.9$	< 0.0001
Group-specific WDV + SAC	18078.8	$\chi^2(3) = 69.4$	< 0.0001
AN-RT vs. BN-BPT			
Model (1) with exponential serial correlation	18333.5		
Group-specific within day variance (WDV)	18333.5	$\chi^2(1) = 0.0$	1
Group-specific serial autocorrelation (SAC)	18330.5	$\chi^2(2) = 3.0$	0.22
Group-specific WDV + SAC	18327.7	$\chi^2(3) = 5.8$	0.12
AN-BPT vs. BN-BPT			
Model (1) with exponential serial correlation	17500.0		
Group-specific within day variance (WDV)	17457.1	$\chi^2(1) = 42.9$	< 0.0001
Group-specific serial autocorrelation (SAC)	17447.9	$\chi^2(2) = 52.1$	< 0.0001
Group-specific WDV + SAC	17447.8	$\chi^2(3) = 52.2$	< 0.0001

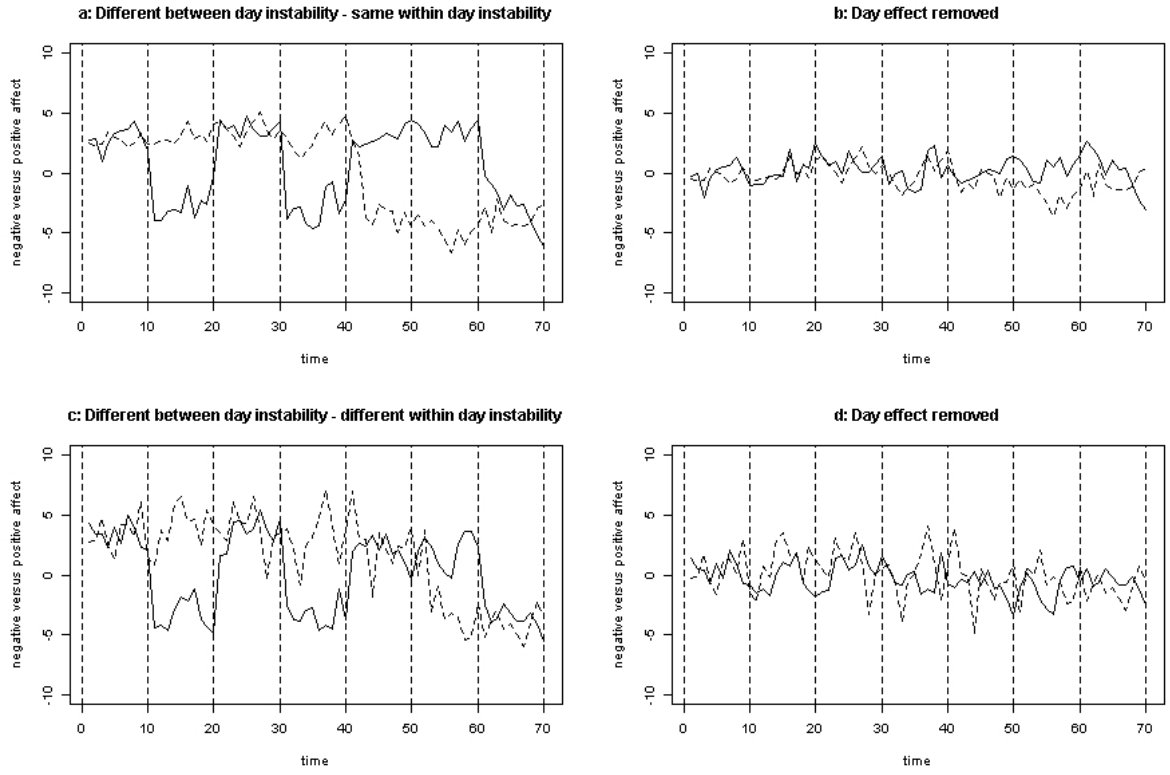

Figure 1. Between day and within day instability.

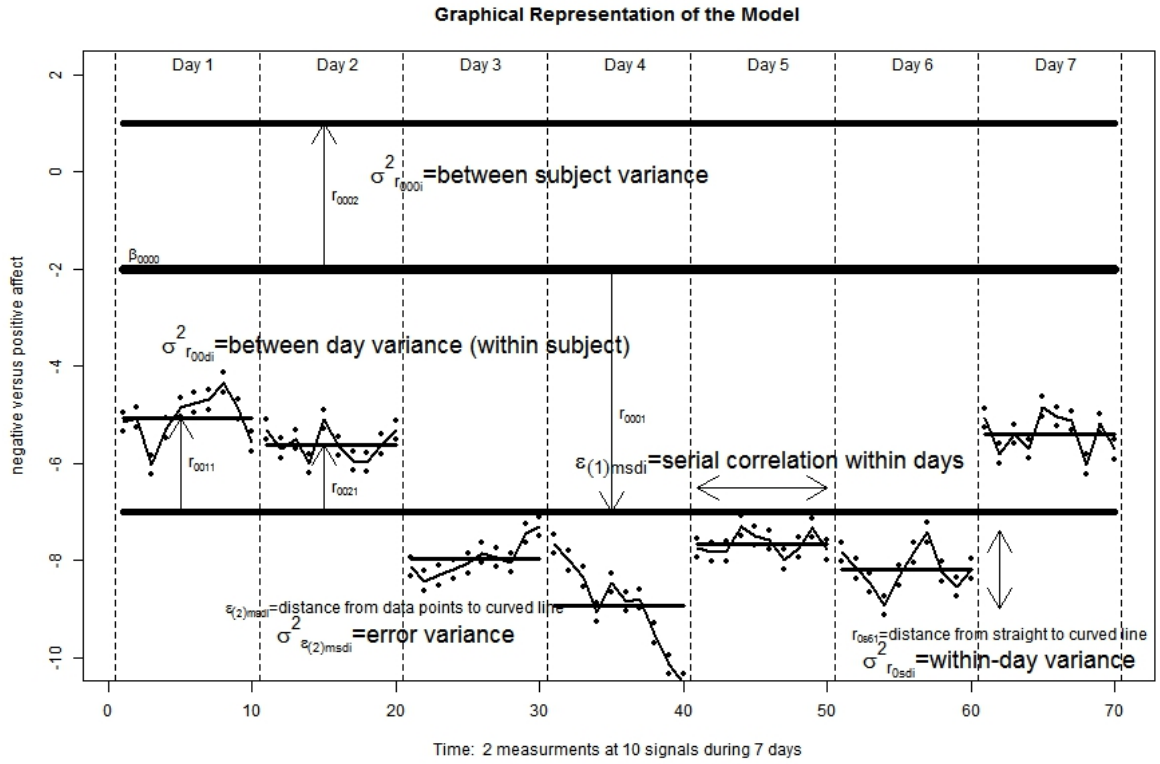

Figure 2. Graphical Representation of Model (1).

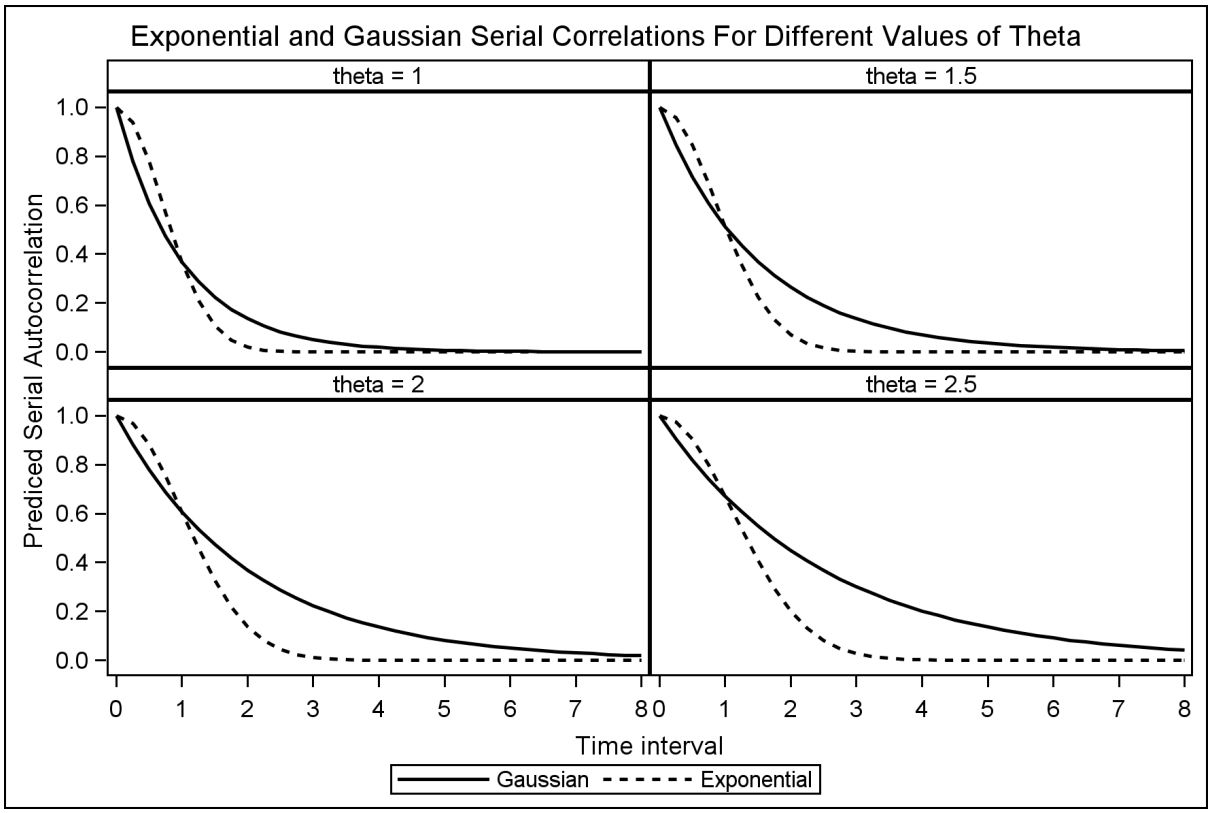

Figure 3. Exponential and Gaussian Serial Autocorrelation for different values of θ.

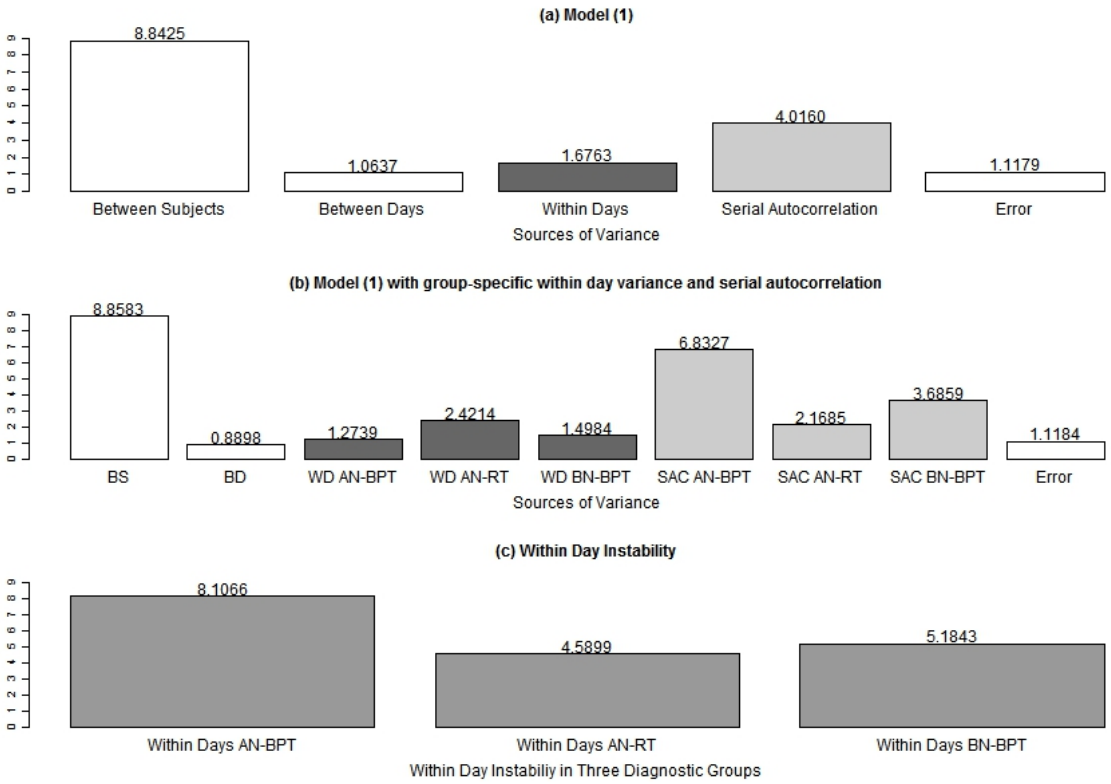

Figure 4. Valence: (a) Variance components in Model (1) (τ^2 for serial autocorrelation). (b) Model with group-specific short-term within day variance and serial autocorrelation, and (c) Group-specific within day affective instability in three groups of eating disorders.

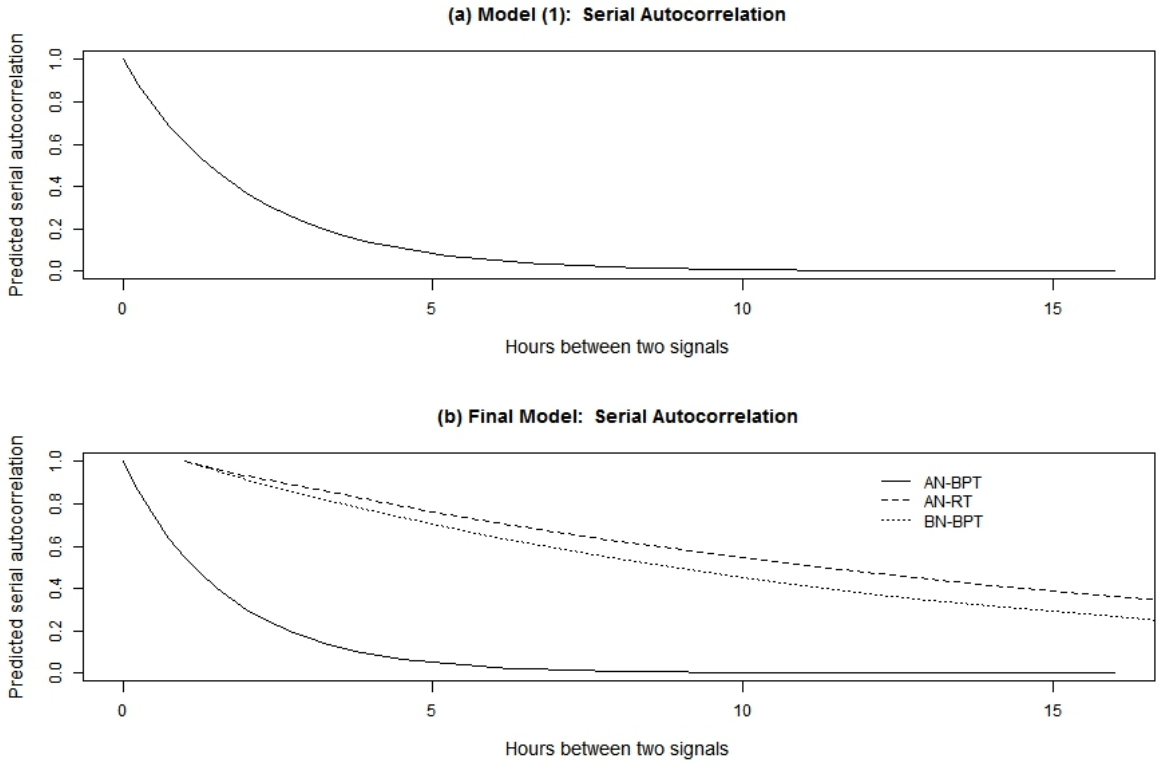

Figure 5. Exponential serial autocorrelation in (a) Model (1), and (b) the model with group-specific short-term within day variance and group-specific serial autocorrelation for valence.

Appendix *A*: Model (1) with separate equations for each level

To further clarify the model, Model (1) can be rewritten by expressing its four levels explicitly:

(a) Measurement level (Level 1):

$$Y_{msdi} = \beta_{0sdi} + \epsilon_{(1)msdi} + \epsilon_{(2)msdi} \tag{9}$$

with β_{0sdi} being the average affect at Signal s, on Day d for Subject i. Furthermore, $\epsilon_{(1)msdi}$ represents the serial autocorrelation, and $\epsilon_{(2)msdi}$ the measurement error.

(b) Short-term within-day (or signal) level (Level 2):

$$\beta_{0sdi} = \beta_{00di} + r_{0sdi} \tag{10}$$

with β_{00di} being the average affect on Day d for Subject i. r_{0sdi} is a random effect for Signal s at Day d in Subject i reflecting variation within days that is normally distributed with $r_{0sdi} \sim N(0, \sigma^2_{r_{0sdi}})$. As a result, $\sigma^2_{r_{0sdi}}$ is the short-term within day variance.

(c) Between day level (Level 3)

$$\beta_{00di} = \beta_{000i} + r_{00di} \tag{11}$$

with β_{000i} the average affect of Subject i. r_{00di} is a day by subject-specific random effect reflecting variation between days that is normally distributed with $r_{00di} \sim N(0, \sigma^2_{r_{00di}})$. At this level, $\sigma^2_{r_{00di}}$is the long-term between day variance.

(d) Between subject level (Level 4)

$$\beta_{000i} = \beta_{0000} + r_{000i} \tag{12}$$

with β_{0000} the overall intercept and r_{000i} a random subject-specific effect that is normally distributed with $r_{000i} \sim N(0, \sigma^2_{r_{000i}})$. Note that $\sigma^2_{r_{000i}}$ is the between subject variance.

Appendix *B*: Annotated SAS Code for different scenarios

Below, one can find the annotated SAS-code for Model (1), the scenarios described in the paper and the model with group-specific components. We will use numbers between parentheses (e.g., (1)) in the SAS-code and annotate the code below. In this code, 'time' is a variable indicating the time passed since the start of the day (e.g., 7:30 a.m.) expressed in hours and 'passeddays' is the number of days passed since the start of the study. The SAS REPEATED statement requires an unique order of all observations within the blocks denoted by the SUBJECT-statement. The variables 'phd' and 'dayorder' indicate the order of all observations in this statement but have, beside their ordering role, no effect on the analysis (see the file with SAS-code provided as supplementary material for instructions how to obtain these variables). Note that, depending on the data at hand, serial autocorrelations for equally (e.g., AR(1)) or unequally spaced observations can be used and that many types of serial autocorrelation structures are available in SAS. An exemplary dataset and file with SAS-code is provided as supplementary material.

(a) Data: different measurements nested within signals nested within days nested within subjects (4-level model). This is Model (1).

```
PROC MIXED DATA=data;
CLASS id day signal phd diagnosis;
(1) MODEL affect = diagnosis /SOLUTION;
(2)(7)RANDOM intercept /SUBJECT=id TYPE=UN  GROUP=diagnosis;
(3)(7)RANDOM intercept /SUBJECT=day(id) TYPE=UN GROUP=diagnosis;
(4)(7)RANDOM intercept / SUBJECT=signal(day*id) TYPE=UN GROUP=diagnosis;
(5)(6)(7) REPEATED phd / SUBJECT=day(id) TYPE=SP(EXP or GAU) (time) LOCAL GROUP=diagnosis;
RUN;
```

Annotations:

-Fixed Effects

(1) Inclusion of fixed effects in the model (e.g., diagnosis, time, other covariates)

-Random Effects

(2) Subject-level: between subject variance ($\sigma^2_{r_{000i}}$) SUBJECT=id implies that observations may be dependent within subjects; observations of different subjects are independent

(3) Day-level: long-term between day variance ($\sigma^2_{r_{00di}}$) SUBJECT=day(id) implies that observations may be dependent within days (or day by subject combinations); observations of different days or subjects are independent

(4) Signal-level: short-term within day variance ($\sigma^2_{r_{0sdi}}$). SUBJECT=signal(day*id) is used because SAS does not allow TYPE=signal(day(id)). Observations may be dependent within signals (or signal by day by subject combinations); observations of different signals or days or subjects are independent.

-Errors

(5) SUBJECT=day(id) TYPE=SP(EXP or GAU) (time): serial autocorrelation component ($\epsilon_{(1)msdi}$ with θ and τ^2)

SUBJECT=day(id) implies that observations may be dependent within days (or day by subject combination) with the (Exponential or Gaussian) serial autocorrelation being a function of the time interval between observations (time = passed hours since the start of the day). The serial correlation pertains to successive signals. Observations of different days and/or subjects are independent.

(6) The LOCAL option adds an observational error to a time series structure: Variance of measurement error ($\sigma^2_{\epsilon_{(2)msdi}}$)

(7) The GROUP-statement makes each of the involved variance or serial autocorrelation component group-specific. Model (1) is estimated using this SAS code without the GROUP statement.

(b) Data: different signals nested within days nested within subjects (3-level model).

```
PROC MIXED DATA=data;
CLASS id day signal phd diagnosis;
(1)MODEL affect = diagnosis /SOLUTION;
(2)RANDOM intercept /SUBJECT=id TYPE=UN ;
(3)RANDOM intercept /SUBJECT=day(id) TYPE=UN;
(5)(6) REPEATED phd / SUBJECT=day(id) TYPE=SP(EXP or GAU) (time) LOCAL;
RUN;
```

Annotations:

-Fixed Effects

(1) Inclusion of fixed effects in the model (e.g., diagnosis, time, other covariates)

-Random Effects

(2) Subject-level: between subject variance ($\sigma^2_{r_{00i}}$) SUBJECT=id implies that observations may be dependent within subjects; observations of different subjects are independent

(3) Day-level: long-term between day variance ($\sigma^2_{r_{0di}}$) SUBJECT=day(id) implies that observations may be dependent within days (or day by subject combinations); observations of different days or subjects are independent

-Errors

(5) SUBJECT=day(id) TYPE=SP(EXP or GAU) (time): serial autocorrelation component ($\epsilon_{(1)sdi}$ with θ and τ^2) SUBJECT=day(id) implies that observations may be dependent within days (or day by subject combination) with the (Exponential or Gaussian) serial autocorrelation being a function of the time interval between observations (time = passed hours since the start of the day). The serial correlation pertains to successive signals. Observations of different days and/or subjects are independent.

(6) The LOCAL option adds an observational error to a time series structure: Variance of measurement error ($\sigma^2_{\epsilon_{(2)sdi}}$)

(c) Data: different measurements nested within days nested within subjects (3-level model).

```
PROC MIXED DATA=data;
CLASS id day signal dayorder diagnosis;
(1)MODEL affect = diagnosis /SOLUTION;
(2)RANDOM intercept /SUBJECT=id TYPE=UN ;
(3)RANDOM intercept /SUBJECT=day(id) TYPE=UN;
(5)(6) REPEATED dayorder / SUBJECT= id TYPE=SP(EXP or GAU) (passeddays) LOCAL;
RUN;
```

Annotations:

-Fixed Effects

(1) Inclusion of fixed effects in the model (e.g., diagnosis, time, other covariates).

-Random Effects

(2) Subject-level: between subject variance ($\sigma^2_{r_{00i}}$) SUBJECT=id implies that observations may be dependent within subjects; observations of different subjects are independent

(3) Day-level: long-term between day variance ($\sigma^2_{r_{0di}}$) SUBJECT=day(id) implies that observations may be dependent within days (or day by subject combinations); observations of different days or subjects are independent

-Errors

(5) SUBJECT=id TYPE=SP(EXP or GAU) (passeddays): serial autocorrelation component ($\epsilon_{(1)mdi}$ with θ and τ^2). Note that this SUBJECT statement is different compared with the previous scenario! SUBJECT=id implies that observations may be dependent within subjects with the (Exponential or Gaussian) serial autocorrelation being a function of the time interval between observations (passeddays = passed days since the start of the study). The serial correlation pertains to successive days (measurements are taken at the same time within each day). Observations of different subjects are independent.

(6) The LOCAL option adds an observational error to a time series structure: Variance of measurement error ($\sigma^2_{\epsilon_{(2)mdi}}$)

(d) Data: different days nested within subjects (2-level model).

```
PROC MIXED DATA=data;
CLASS id day signal dayorder diagnosis;
(1)MODEL affect = diagnosis /SOLUTION;
(2)RANDOM intercept /SUBJECT=id TYPE=UN ;
(5)(6) REPEATED dayorder / SUBJECT=id TYPE=SP(EXP or GAU) (passeddays) LOCAL;
RUN;
```

Annotations:

-Fixed Effects

(1) Inclusion of fixed effects in the model (e.g., diagnosis, time, other covariates)

-Random Effects

(2) Subject-level: between subject variance ($\sigma^2_{r_{0i}}$) SUBJECT=id implies that observations may be dependent within subjects; observations of different subjects are independent

-Errors

(5) SUBJECT=id TYPE=SP(EXP or GAU) (time): serial autocorrelation component ($\epsilon_{(1)di}$ with θ and τ^2) SUBJECT= id implies that observations may be dependent within subjects with the (Exponential or Gaussian) serial autocorrelation being a function of the time interval between observations (passeddays = passed days since the start of the study). The serial correlation pertains to successive days. Observations of different subjects are independent.

(6) The LOCAL option adds an observational error to a time series structure: Variance of measurement error ($\sigma^2_{\epsilon_{(2)di}}$)