Running head: DSEM OF COGITO DATA

At the frontiers of modeling intensive longitudinal data: Dynamic structural equation models for the affective measurements from the COGITO study

E. L. Hamaker,[1] T. Asparouhov,[2], A. Brose,[3,4,5] F. Schmiedek,[5,6] and B. Muthén[2]

1. Methodology and Statistics, Faculty of Social and Behavioural Sciences, Utrecht University, The Netherlands; 2. Muthén and Muthén, Los Angeles, CA, USA; 3. Humboldt University Berlin, Germany; 4. KU Leuven, Belgium; 5. Max Planck Institute for Human Development, Center for Lifespan Psychology, Berlin, Germany; 6. German Institute for International Educational Research (DIPF), Center for Education and Lifespan Development, Frankfurt am Main, Germany

This manuscript is submitted to *Multivariate Behavior Research* (special section on COGITO study).

Author Note: This study was supported by the Netherlands Organization for Scientific Research (NWO; VIDI Grant 452-10-007). Correspondence concerning this paper should be addressed to E. L. Hamaker, Methods and Statistics, Faculty of Social and Behavioural Sciences, Utrecht University, P.O. Box 80140,3508 TC, Utrecht, The Netherlands. Email: e.l.hamaker@uu.nl.

Abstract

Dynamic structural equation modeling (DSEM) is a newly emerging class of techniques by which we can model the dynamic patterns in intensive longitudinal data. When using DSEM to analyse the time series of multiple individuals, we specify a time series model at the within-person level and allow for individual differences at the between-person level in the parameters that describe the dynamics. We use DSEM to analyze affective data from the COGITO study, which consists of two samples of over one hundred individuals each who were measured for one hundred days each. We use composite scores of positive and negative affect and apply a multilevel vector autoregressive model to investigate individual differences in means, autoregressions and cross-lagged effects. Then we extend the model with random residual variances, and finally we investigate whether the random effects mediate the effect of prior depression on later depression scores. We point out some additional options, and discuss several unresolved issues.

At the frontiers of modeling intensive longitudinal data: Dynamic structural equation models for the affective measurements from the COGITO study

The COGITO study, which took place a decade ago, is an impressive project that reflects an admirable determination on part of its researchers as well as the participants. It consists of a younger and an older sample of over 100 individuals each, who visited the laboratory on about 100 days to perform a battery of cognitive tests and fill out self-report questionnaires. This endeavor is all the more impressive if one realizes that at the time, gathering intensive longitudinal data was not only cumbersome, but it was also considered unnecessary by many, because short-term, within-person fluctuations were assumed to reflect mere noise.

Since then, however, the tide has turned. First, technological developments like smart phones and wearable devices have led to new forms of data collection—such as ambulatory assessments, experience sampling, ecological momentary assessments, and electronic daily diaries (Conner, Tennen, Fleeson, & Feldman Barret, 2009; Trull & Ebner-Priemer, 2013)—that have placed intensive longitudinal data within reach of mainstream psychology. Second, the growing body of research made possible by these innovations now forms a convincing testimony of the meaningfulness of short-term, within-person fluctuations. Moreover, researchers are coming to realize that intensive longitudinal data offer a unique opportunity to study psychological processes as they unfold over time (Hamaker & Wichers, 2017). Related to these developments, new statistical modeling techniques are being developed that focus on the within-person dynamic patterns in intensive longitudinal data (for an overview, see Hamaker, Ceulemans, Grasman, & Tuerlinckx, 2015).

A recent and particularly promising innovation in this respect is *dynamic structural equation modeling* (DSEM; cf. Asparouhov, Hamaker, & Muthén, 2017, in preparation). DSEM consists of a general, latent variable, multilevel time series model, and focusses on the autocorrelation structure of longitudinal data. Specifically, it is a combination of two time series techniques: a) *state-space modeling* (e.g., Durbin & Koopman, 2001; Harvey, 1989); and b) *dynamic factor analysis* (e.g., Molenaar, 1985; Zhang, Hamaker, & Nesselroade, 2008). But where time series analysis is traditionally concerned with $N = 1$ data (i.e., data from a single case; cf. Box & Jenkins, 1970; Chatfield, 2004; Hamilton, 1994; Kim & Nelson, 1999; Shumway & Stoffer, 2006), DSEM encompasses both $N = 1$ and $N > 1$ data. The latter is based on a multilevel extension of the time series model: It accounts for individual difference in dynamics by allowing the parameters of the time series model at the lower, within-person level to have a distribution at the higher, between-person level. DSEM also includes extensions with categorical latent variables and cross-level classification across persons and time, which will not be covered here (for details see Asparouhov et al. (2017) and Asparouhov and Muthén (2015) respectively).

In this paper, we apply DSEM, as it is implemented in Mplus version 8, to the daily affective ratings from the COGITO study. Our goal is to illustrate several different approaches that can be taken to such data with DSEM, and to show the research questions we may tackle with this. Because DSEM is still in its infancy, it is important to realize that there are many unanswered — and often unasked — questions; we will elaborate on some of these in the current paper.

This paper is organized as follows. First, we begin this paper by explaining how intensive longitudinal data can be decomposed into different sources of variance, as this is fundamental to DSEM. We use the affective data from the COGITO study to illustrate this decomposition. Second, we discuss the basics of DSEM, including the general model structure, estimation, and the handling of unequally spaced observations. Third, we present a multilevel cross-lagged model that allows for individual differences in means, autoregressive relationships, and cross-lagged relationships, and apply this to the affective data from the COGITO study. Fourth, we extend this model with random innovation variances and covariance, to account for individual differences in these model features, and apply this to the COGITO data. Fifth, we specify a mediation model in which the random effects can mediate the effect of prior depressive symptoms on subsequent depressive symptoms, and apply this to the COGITO data. We end this paper with a discussion in which we highlight the strengths of the current modeling approach, and touch upon some urgent—yet unresolved—issues in modeling intensive longitudinal data.

Total, within-person, and between-person variability

There is quite a large body of literature regarding the actual meaningfulness of cross-sectional research for gaining knowledge about processes that operate within individuals (e.g., Borsboom, Mellenbergh, & Van Heerden, 2003; Cattell, 1952; Grice, 2004; Hamaker, 2012; Hamaker, Schuurman, & Zijlmans, 2017; Lamiell, 1998; Molenaar, 2004; Molenaar, Huizenga, & Nesselroade, 2003; Nesselroade, 2002; Schmitz & Skinner, 1993; Voelkle, Brose, Schmiedek, & Lindenberger, 2014). This topic is already complex in itself, but the discussion is further complicated by the fact that researchers sometimes use the same terms to refer to different phenomena. Therefore, we begin by briefly explaining our terminology and how this relates to the conventions in different strands of literature. Once this is in place, we discuss the decomposition of the variance of the affective measurements that were included in the COGITO study.

Terminology

While the terms *within-person* and *intraindividual* variability unambiguously refer to the variability *within a person over time and/or situations*, their linguistic counterparts *between-person* and *interindividual* variability have been used to refer to two distinct forms of variation: At times, they are used to refer to *stable, trait-like variation between persons* (Hamaker, Nesselroade, & Molenaar, 2007; Wang & Maxwell, 2015), and at other times they are used to refer to *cross-sectional variation*, that is, the variability between persons at a single occasion (Borsboom et al., 2003; Hamaker, Dolan, & Molenaar, 2005; Molenaar, 2004; Voelkle et al., 2014).

This ambiguous usage of the terms *between-person* and *interindividual* may have led some researchers to erroneously conclude that there is no difference between trait-like variation and cross-sectional variation, and as a result, cross-sectional variation may easily be mistaken as the counterpart of intraindividual or within-person variation. In the current paper, we will use these terms in accordance to the multilevel literature, such that between-person and interindividual are the complement of within-person and intraindividual. This implies that cross-sectional variability is the result of both within-person and between-person variability (cf. Hamaker et al., 2017; Hamaker et al., 2007; Schmitz, 2000).

Finally, we will use the term *total variability* to refer to the variability both across individuals and across time points. The latter is thus also the result of within-person and between-person variability but in contrast to cross-sectional variability, which is restricted to a single time point, the total variability can be thought of as an average of cross-sectional variability over time.

Decomposing the COGITO affective data

We consider the differences between these forms of variability using the composite scores obtained from the positive affect negative affect schedule (PANAS; Watson, Clark, & Tellegen, 1988), which the COGITO participants filled out at each measurement occasion (Brose, Voelkle, Lövdén, Lindenberger, & Schmiedek, 2015). The PANAS consists of ten items that measure positive affect (PA) and ten items that measure negative affect (NA). We analyze the data of the younger sample (aged 20-31), consisting of 101 individuals, and the older sample (aged 65-80), which consists of 103 individuals, separately.[1]

First, we estimate the total variances for PA and NA and the corresponding total correlation based on all the data in the sample, both across persons and time. Second, the within-person and between-person results are obtained using a basic multivariate multilevel model (sometimes referred to as the empty model, as it does not include any predictors): It decomposes the total variance into two parts, that is, the between part, which is formed by the within-person means (sometimes referred to as cluster or group means), and the within-part, which are formed by the within-person fluctuations over time around the within-person means. Let PA_{it} and NA_{it} be the PA and NA scores of individual i at time t. The decomposition into the within-person and between-person parts is

$$PA_{it} = \mu_{PA,i} + PA^*_{it} \quad \text{and} \quad NA_{it} = \mu_{NA,i} + NA^*_{it}$$

or in matrix notation

$$\begin{bmatrix} PA_{it} \\ NA_{it} \end{bmatrix} = \begin{bmatrix} \mu_{PA,i} \\ \mu_{NA,i} \end{bmatrix} + \begin{bmatrix} PA^*_{it} \\ NA^*_{it} \end{bmatrix}. \tag{1}$$

where the μ's are the within-person means that form the between-person part of the model, and PA^*_{it} and NA^*_{it} represent the temporal deviations of individual i at occasion t from these within-person means.

In Figure 1, the variance estimates are plotted along with their posterior standard deviations. The intraclass correlation (ICC) for each variable in each group is also reported: It is computed as the ratio of between-person variance to total variance, where the latter is the sum of within-person and between-person variance. Hence, the ICC indicates the proportion of the total variance that is accounted for by stable, trait-like between-person differences. Figure 1 shows that: a) the between-person variance is larger than the within-person variance for both PA and NA in both groups (ICCs lie between .64 and .89); b) the variance of PA is larger than that of NA in both groups and this is true for the total, the between and the within variances; and c) the older sample is characterized by larger ICCs, and also by larger differences between the variability in PA and NA than the younger sample.

Insert Figure 1 about here

Considering the correlations between PA and NA, we found that the total correlation in the younger sample was very close to zero ($\hat{\rho} = -.018$, with a 95%

credible interval: CI=[-.038, .001]), while in the older sample there was evidence for a weak negative correlation ($\hat{\rho} = -.100$, CI=[-.118, -.081]). At the between-person level there was no evidence for a correlation between PA and NA in either group ($\hat{\rho} = .077$, CI=[-.126, .273] in the younger sample; $\hat{\rho} = -.080$, CI=[-.272, .120]). In contrast, at the within-person level we found evidence for a weak negative correlation between PA and NA in both samples: $\hat{\rho} = -0.186$ with CI=[-.204, -.167] in the younger sample, and $\hat{\rho} = -.194$ with CI=[-.212, -.175].

These results can be understood in the following way (see also Brose, Voelkle, et al., 2015). The traits PA and NA, which are captured by the between-person part of the model, are not correlated in these populations, which implies that sampling an individual who is high on one of these traits and low on the other is about equally likely as sampling an individual who is high on both traits, or low on both traits. In contrast, it is not that common to experience both high levels of PA and NA at the same time, because repeated measures of PA and NA are typically negatively correlated within a person. In the current dataset, the within-person variability refers to day-to-day fluctuations, rather than moment-to-moment fluctuations; therefore, the correlation of -.19 in both samples suggests that individuals tend to experience good days with elevated PA and decreased NA, and bad days with elevated NA and decreased PA instead, although this pattern is not as strong as in momentary measurements (when individuals are asked how they are feeling right now).

In sum, these results show that there are marked differences between the within-person and the between-person relationships, and that neither is represented well by the results for the total. This also confirms the concern that has been expressed by many researchers before: Cross-sectional results are a mix of within-person and between-person relationships, and may not represent either very well (cf. Brose, Schmiedek, & Lindenberger, 2013; Schmitz, 2000).[2] This decomposition of the total variance into a within-person and a between-person part is fundamental to DSEM, as we will discuss next.

DSEM in a nutshell

Here we present the basis of DSEM as it is implemented in Mplus and applied in the current study. We begin by discussing the general model structure, consisting of a within-person and a between-person model part. There are extension that allow for time-varying parameters and regime-switching, which are not covered here (cf. Asparouhov et al., 2017, in preparation). Subsequently, we briefly discuss the estimation used in DSEM. Finally, we discuss how unequal intervals between observations are handled within DSEM in Mplus. This is a common problem, as intensive longitudinal typically are characterized by unequal intervals either by design, or because measurements were missed by the participants.

The model equations

When applying DSEM to multilevel data (meaning there is more than one person, dyad, or other unit, from which repeated measures have been obtained), the modeling approach is based on a decomposition. Let $\mathbf{y}_{it}$ be the vector with p observed variables for individual i at time t. This is decomposed into a within-person mean $\boldsymbol{\mu}_i$, and a time-specific within-person deviation from this mean

$\mathbf{y}^*_{it}$. Hence, the decomposition can be expressed as

$$\mathbf{y}_{it} = \boldsymbol{\mu}_i + \mathbf{y}^*_{it}. \tag{2}$$

The within-person mean is used in the between-person part of the model, to study the between-person differences, whereas the within-person deviations are modeled in the within-person part. Analogue to traditional structural equation modeling, both parts here consist of a *measurement equation*, which relates the scores from the decomposition to latent variables, and a *structural equation* that relates the latent variables to each other.

Within-person equations. The within-person deviations are modeled using the measurement equation

$$\mathbf{y}^*_{it} = \sum_{r=0}^{R} \boldsymbol{\Lambda}^*_{i,r} \boldsymbol{\eta}^*_{i,t-r} + \boldsymbol{\epsilon}^*_{it} \tag{3}$$

where the $\boldsymbol{\Lambda}$'s are $p \times q$ matrices with factor loadings, the $\boldsymbol{\eta}^*_{i,t-r}$'s are q-variate vectors with latent variables of individual i at time $t-r$, and $\boldsymbol{\epsilon}_{it}$ is a p-variate vector with residuals (i.e., random measurement errors). The latter come from a multivariate normal distribution with means zero, and covariance matrix $\boldsymbol{\Theta}^*$.

A specific feature of Equation 3 is that it allows for *lagged factor loadings*, that is, we can relate the current within-person centered observations $\mathbf{y}^*_{it}$ to concurrent latent variables $\boldsymbol{\eta}^*_{it}$, but also to preceding latent variables, that is $\boldsymbol{\eta}^*_{i,t-1}$ to $\boldsymbol{\eta}^*_{i,t-R}$, where R specifies the maximum lag. Such lagged factor loadings are characteristic of dynamic factor analysis as proposed by Molenaar (1985), and is a way to account for the potential autocorrelation in the data.

Note that the factor loading matrices have a subject index i, indicating that they can differ across individuals, in which case they will be modeled as random effects at the between-person level. Moreover, the residual variances (i.e., the variances of $\boldsymbol{\epsilon}^*_{it}$) can also be modeled as random across individuals: In this case, $\boldsymbol{\Theta}$ would also get a subject index. Random factor loadings will have a normal distribution at the between-person level, while random variances will have a log normal distribution at the between-person level.

The latent variables at the within-person level can be further modeled using the structural equation, that is,

$$\boldsymbol{\eta}^*_{it} = \boldsymbol{\alpha}^*_i + \sum_{s=0}^{S} \mathbf{B}^*_{i,s}\boldsymbol{\eta}^*_{i,t-s} + \boldsymbol{\Gamma}^*_i \mathbf{x}^*_{it} + \boldsymbol{\zeta}^*_{it} \tag{4}$$

where $\boldsymbol{\alpha}_i$ is a q-variate vector with random intercepts (which is typically fixed to zero for all i), the $\mathbf{B}$'s are $q \times q$ matrices with regression coefficients that allow us to regress the current latent variables on each other at the same occasion (when $s = 0$) or on preceding occasions (when $s = 1, \ldots, S$), $\boldsymbol{\Gamma}^*_i$ a $q \times m$ matrix with regression coefficients for the m exogenous variables $\mathbf{x}^*_{it}$ (which are the level 1 or time-varying predictors), and $\boldsymbol{\zeta}^*_{it}$ is a q-variate vector with residuals (i.e., dynamic error). The latter are assumed to come from a normal distribution with means equal to zero and covariance matrix $\boldsymbol{\Psi}^*$.

The structural equation is closely related to the transition or state equation in the state-space model (cf. Harvey, 1989; Durbin & Koopman, 2001). However, where the state equation in the state-space model regresses the current state solely on the state immediately preceding it, such that it can be thought of as a first-order vector autoregressive model, the current expression extends this to included lag zero regressions (when $s = 0$), and regressions at larger lags (i.e., up to lag S). While models with such higher lags are possible in the state-space model, this requires including latent variables in the state vector which have indices that do not correspond to the time index of the state vector itself.[3]

In addition to the random intercepts (in $\boldsymbol{\alpha}_i^*$) and random slopes (in the $\mathbf{B}^*$ matrices and $\boldsymbol{\Gamma}_i^*$), we can also include random residual variances. In this case $\boldsymbol{\Psi}^*$ also gets a subject index i, and the variances are assumed to come from a log normal distribution.

Between-person equations. At the between-person level we model the within-person means $\boldsymbol{\mu}_i$ from Equation 2, but also all the other random effects, which are defined in the within-person model. These include: a) random factor loadings in the $\boldsymbol{\Lambda}_{i,r}^*$ matrices at different lags; b) random latent intercepts in $\boldsymbol{\alpha}_i^*$; c) random variances in the covariance matrix $\boldsymbol{\Theta}_i^*$; d) random regression coefficients in the $\mathbf{B}_{i,s}^*$ matrices at different lags; e) random regression coefficients in the matrix $\boldsymbol{\Gamma}_i^*$ for the exogenous variables; and f) random variances in the covariance matrix $\boldsymbol{\Psi}_i^*$.

Specifically, the between-person level consists of a measurement equation that relates the within-person means $\boldsymbol{\mu}_i$ to the latent variables at the between-person level, denoted as $\boldsymbol{\eta}_i$. The latter is a v-variate vector that includes these same within-person means $\boldsymbol{\mu}_i$, and the other random effects in the model. Additionally, $\boldsymbol{\eta}_i$ may also include latent variables (e.g., a common factor) specified at this level to account for the covariance structure between all these random effects. Hence, the measurement equation at the between-person level is

$$\boldsymbol{\mu}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i \tag{5}$$

where $\boldsymbol{\nu}$ is a p-variate vector with intercepts, $\boldsymbol{\Lambda}$ is a $p \times v$ matrix with factor loadings, $\boldsymbol{\eta}_i$ is the v-variate vector with latent between level factor scores for individual i (including all the random effects and other latent variables), and $\boldsymbol{\epsilon}_i$ is the p-variate vector with residuals (i.e., systematic error, as it does not vary over time).

The latent variables can be further modeled using the structural equation

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\zeta}_i \tag{6}$$

where $\boldsymbol{\alpha}$ is a v-variate vector with random intercepts, $\mathbf{B}$ is a $v \times v$ matrix with regression coefficients that allow us to regress the latent variables on each other, $\boldsymbol{\Gamma}$ is a $v \times w$ matrix with regression coefficients for the w exogenous variables $\mathbf{x}_i$ (which are the level 2 or time-invariant predictors), and $\boldsymbol{\zeta}_i$ is a v-variate vector with residuals. The latter are assumed to come from a normal distribution with means equal to zero and covariance matrix $\boldsymbol{\Psi}$.

Estimation in DSEM

DSEM as it is implemented in Mplus version 8, is based on Bayesian estimation. This implies that the data are combined with prior distributions for the unknown model parameters, and estimation is based on sampling according to a Markov chain Monte Carlo (MCMC) procedure (Gelman, Carlin, Stern, & Rubin, 2013). While Mplus allows the user to specify the priors, it also provides default priors that are uninformative so that the data dominate the results (Asparouhov & Muthén, 2010; Gelman et al., 2013).

The MCMC algorithm is based on separating all the parameters in the model into blocks. Note that these parameters include the model parameters, but also the individual parameters (i.e., random effects), latent variables (both at the within-person level and the between-person level), and missing values. Some blocks, such as the ones that contain latent variables, only play a role in some models.

Within each block, the parameters are sampled from their conditional posterior distributions, which depend on the most recently sampled values for the other parameters, the data, and the prior distributions for the parameters. Below follows a brief description of the six steps that DSEM in Mplus uses.

The first step consists of sampling the *individual means* $\boldsymbol{\mu}_i$. Given these means, the within-person part can be computed as $\mathbf{y}^*_{it} = \mathbf{y}_{it} - \boldsymbol{\mu}_i$.

The second step consists of sampling the *remaining elements of* $\boldsymbol{\eta}_i$. These include all the other individual parameters, that is, the random (lagged) factor loading matrices $\boldsymbol{\Lambda}^*_{i,r}$, the random (lagged) regression coefficients in the structural matrices $\mathbf{B}^*_{i,s}$, and the random regression coefficients in $\boldsymbol{\Gamma}^*_i$ (i.e., all the random slopes in the model). Moreover, when the model includes additional latent variables (e.g., a common factors) at the between-person level, these are also sampled in this step.

The third step consists of sampling the *fixed effects*. This includes the means that define the distributions of the random effects, but also the within-person factor loadings and (lagged) regression coefficients that are not random across individuals (i.e., the within-person slopes that take on the same value for all individuals). Moreover, it also includes all the slopes at the between-person level, that is, the factor loadings in $\boldsymbol{\Lambda}$, and the regression coefficients in $\mathbf{B}$ and $\boldsymbol{\Gamma}$.

The fourth step consists of sampling the *(residual) variances and covariances at the between-person level* in $\boldsymbol{\Theta}$ and $\boldsymbol{\Psi}$. These can be of the random effects, but also of a common factor that is specified at the between-person level.

The fifth step consists of sampling the *residual variances and covariances at the within-person level* in $\boldsymbol{\Theta}$.

Finally, the sixth step consists of *latent variables at the within-person level* $\boldsymbol{\eta}^*_{it}$ (e.g., a common factor, or a moving average term). These have to be sampled sequentially within a person, because their conditional posterior distribution depends on the latent variables at at preceding and subsequent occasions.

There are additional complexities, most notably the handling of missing data, the handling of the beginning and ending of a series (as there are no prior latent vectors, and no subsequent later observations, which are used to sample $\boldsymbol{\eta}^*_{it}$), and sampling random variances (which is done using a Metropolis-Hastings algorithm). These issues are discussed in more detail in Asparouhov et al. (in preparation).

When using an MCMC algorithm to estimate parameters, we have to decide on the number of iterations: This either has to be specified by the user, or we can allow Mplus to terminate when a default criterion is reached. Mplus uses the potential scale reduction (PSR), which is computed for each model parameter separately, by dividing the *total variability across two chains* of the MCMC algorithm, by the *variance within a chain* (Asparouhov & Muthén, 2010; Gelman et al., 2013). This quantity should be very close to one, as the between-chain variance should be close to zero, such that the total variance across two chains becomes (almost) identical to the within-chain variance if enough iterations are used.

Unequal time intervals between observations

A particular challenge when dealing with time series data and intensive longitudinal data are unequal intervals between observations. These may result from missing observations, for instance, when we have daily measurements but participants sometimes forget to fill out the end-of-day questionnaire. However, non-equidistant observations may also be a part of the research design: In experience sampling and ecological momentary assessments, participants are prompted at random time points during the day to ensure they are caught in their daily life rather than that they prepare for the next measurement. In the COGITO data, the majority of observations were obtained with an interval of one day, but there were also larger intervals, for instance because people missed a day, or because there were no measurements taken on the weekend or during holidays.[4]

Unequal intervals are a concern when the interest is in lagged relationships, because the strength of a lagged effect depends on the size of the time interval between the measurements (cf. Gollob & Reichardt, 1987). An elegant way of dealing with this issue is the use of continuous time models (Deboeck & Preacher, 2015; Voelkle, Oud, Davidov, & Schmidt, 2012). In DSEM in Mplus we can account for different intervals using the tinterval command and choosing a particular time grid. Heuristically, it implies that we may have a continuous time dimension with observations located on this dimension at unequal intervals. By choosing a grid, this continuous time dimension is divided into blocks of the length we specified. These blocks are numbered 1, 2, 3, et cetera, and they either contain an observed value, or not. If there is no observed value in a particular block, this block is assigned a missing value. Note that the actual observations could have been taken at any time within the block they fall into, which is why the approach results in *approximately* equally spaced observations. Choosing a finer grid will result in the insertion of more missing values, but also in a better approximation of equidistant measurement occasions. In the current example, a logical choice is a grid based on intervals of one day, which implies that if there is an interval between two consecutive measurements of more than one day, Mplus will add missing values (i.e., as many as there are missing days). More details on the exact procedure are given in (Asparouhov et al., in preparation).

Model 1: A multilevel VAR(1) model for PA and NA

We will begin by considering a multilevel model based on a first-order vector autoregressive (VAR(1)) model, which is represented in Figure 2. The latter is one of the most basic time series models (for N=1 data), and consists of regressing a vector with observations at occasion t on the vector of observations at the preceding occasion. Multilevel extensions of this basic time series model have emerged in the psychological literature, allowing for the analysis of multivariate time series of multiple individuals simultaneously (cf. Bringmann et al., 2013; Schuurman, Ferrer, de Boer-Sonnenschein, & Hamaker, 2016). Another way to view this model is as a multilevel extension of a cross-lagged panel model, allowing for individual differences in means and in lagged effects.

As indicated above, DSEM is based on the decomposition into a within-person and a between-person component, as shown in Equation 1. This decomposition is also shown in the left part of Figure 2. The within-person means $\mu_{PA,i}$ and $\mu_{NA,i}$ can be thought of as a person's long-run equilibria for the variables PA and NA respectively. Such equilibria have also been interpreted as a person's trait scores, or — if we are taking a dynamic system's perspective — as the attractor of the system to which an individual returns in the absence of external perturbations. The temporal deviations from these within-person means, PA^*_{it} and NA^*_{it}, can be thought of as the affective states of a person at occasion t. Both parts are estimated simultaneously in DSEM; we will discuss them successively below.

Insert Figure 2 about here

Within-person model

The within-person deviations PA^*_{it} and NA^*_{it} are regressed on themselves and each other at the preceding occasion, PA^*_{it-1} and NA^*_{it-1}, as represented in the top right part of Figure 2. Hence, the level 1 or within-person equations can be expressed as

$$PA^*_{it} = \phi_{PP,i}PA^*_{it-1} + \phi_{PN,i}NA^*_{it-1} + \zeta_{PA,it}$$
$$NA^*_{it} = \phi_{NN,i}NA^*_{it-1} + \phi_{NP,i}PA^*_{it-1} + \zeta_{NA,it}$$

or (using the general matrix notation $\boldsymbol{y}_{it} = \boldsymbol{\Phi}_i\boldsymbol{y}_{it-1} + \boldsymbol{\zeta}_{it}$) as

$$\begin{bmatrix} PA^*_{it} \\ NA^*_{it} \end{bmatrix} = \begin{bmatrix} \phi_{PP,i} & \phi_{PN,i} \\ \phi_{NP,i} & \phi_{NN,i} \end{bmatrix} \begin{bmatrix} PA^*_{it-1} \\ NA^*_{it-1} \end{bmatrix} + \begin{bmatrix} \zeta_{PA,it} \\ \zeta_{NA,it} \end{bmatrix}, \tag{7}$$

where $\phi_{PP,i}$ and $\phi_{NN,i}$ are the autoregressive parameters for PA and NA respectively; $\phi_{PN,i}$ is the cross-lagged regression effect from NA to PA at the next occasion; and $\phi_{NP,i}$ is the cross-lagged regression coefficient from PA to NA at the next occasion. The ζ's represent the residuals, often referred to as *innovations*, or alternatively as *random shocks*, *disturbances*, *system noise* or *dynamic errors*. The latter term is to distinguish it from measurement error, which tends to affect the measurements only at one occasion; in contrast, dynamic error tends to change the actual course of the system, because it also affects future measurements through the lagged relationships (Schuurman, Houtveen, & Hamaker, 2015). By definition, these innovations have a mean of zero both within and between individuals. They are assumed to come from a multivariate normal distribution. Hence, at the within-person level, we estimate two variances and a covariances for the innovations.

All the lagged parameters have a subject index i to indicate that individuals may differ with respect to the magnitude of these parameters. Individual differences in autoregressive parameters have been an important topic in affect research, as this parameter indicates how quickly a person restores equilibrium after being perturbed, and thus, how good a person may regulate emotions. For this reason, this parameter has also been referred to as *inertia*,[5] *carryover*, and *regulatory weakness*. Typically, its value lies between 0 and 1, and the closer it is to 1, the longer it takes a person to return to his/her equilibrium. Empirical studies, including a study that uses data from the COGITO study (Brose, Schmiedek, Koval, & Kuppens, 2015), have shown that the autoregressive parameter is positively related to current and future depression, being female, and neuroticism (cf., Koval, Kuppens, Allen, & Sheeber, 2012; Kuppens, Allen, & Sheeber, 2010; Suls, Green, & Hillis, 1998).

Individual differences in the cross-lagged parameters reflect differences in predictive relationships, and are especially interesting because they potentially represent causal mechanisms. Although the existence of time-varying omitted variables that serve as a cause of both variables cannot be ruled out, the current decomposition into a within-person part and a between-person part ensures that the cross-lagged relationships that are estimated are not biased by stable, time-invariant omitted variables (Halaby, 2004; Hamaker, Kuiper, & Grasman, 2015; Ousey, Wilcox, & Fisher, 2011). The within-person structure that is obtained with this model, has therefore stirred the interest of researchers in the context of psychopathology, as a way to investigate the dynamic network of symptoms and how they affect each other over time (e.g., Bringmann et al., 2013). The cross-lagged effects are sometimes referred to as *spill-over*, as they may represent the cascade effect of functioning or behavior in one domain into another domain (Almeida, Wethington, & Chandler, 1999; Bolger, DeLongis, Kessler, & Wethington, 1989; Masten & Cichetti, 2010).

Between-person model

While the initial decomposition in Equation 1 suggests that at the between level we have two variables, that is, the within-person means for PA and NA (i.e., μ_{PA} and μ_{NA}), closer inspection of Equation 7 shows that we have four additional random effects, that is: two random autoregressive parameter ($\phi_{PP,i}$ and $\phi_{NN,i}$), and two random cross-lagged regression parameters ($\phi_{PN,i}$ and $\phi_{NP,i}$). If we simply consider these to be related to each other without any particular constraint, we have

$$\begin{bmatrix} \mu_{PA,i} \\ \mu_{NA,i} \\ \phi_{PP,i} \\ \phi_{PN,i} \\ \phi_{NP,i} \\ \phi_{NN,i} \end{bmatrix} = \begin{bmatrix} \gamma_{P} \\ \gamma_{N} \\ \gamma_{PP} \\ \gamma_{PN} \\ \gamma_{NP} \\ \gamma_{NN} \end{bmatrix} + \begin{bmatrix} u_{P,i} \\ u_{N,i} \\ u_{PP,i} \\ u_{PN,i} \\ u_{NP,i} \\ u_{NN,i} \end{bmatrix}, \tag{8}$$

where the six γ's represent the fixed or averaged effects, and the six u's represent the individual deviations from these means. These individual deviations are assumed to come from a multivariate normal distribution, that is,

$$\boldsymbol{u}_i \sim MN(\mathbf{0}, \boldsymbol{\Omega}), \tag{9}$$

where $\mathbf{\Omega}$ is a 6×6 covariance matrix, which implies there are 6 variances (i.e., the random effects), and 15 covariances. Combining this with the 6 fixed effects, we now have a model with 27 parameters at the between-person level, which gives a total of 30 parameters when combined with the 3 parameters estimated at the within-person level. Below, we begin with estimating this basic dynamic multilevel model.

Results

We estimate this model with DSEM in Mplus version 8, using 50,000 iterations and a thinning of 10 iterations (meaning our results are based on 5,000 iterations; cf. Gelman et al., 2013), and two chains. Model convergence is checked using the proportional scale reduction Mplus reports (which has to be close to 1), and by checking the trace plots for an absence of trends, spikes, or other irregularities. The parameter estimates for the fixed and random parameters, and their 95% credible intervals (CI) are presented in the left part of Table 1.

Only the CIs for the fixed effect $\phi_{NP,i}$ in the older sample contains zero, indicating that there is no evidence that there is an effect of PA on NA over time on average in this group. This does not necessarily imply that this is not an important parameter in the model: As it is included as a random parameter, there may be meaningful individual differences that are centered around zero. In this case, however, the variance for this parameter also seems rather small in this group, implying that the individual $\phi_{NP,i}$'s are all close to zero.[6]

To interpret the magnitude of the cross-lagged effects, we consider the standardized results that Mplus provides, which are based on standardizing the parameters per person first, and then taking the average of these (cf. Schuurman et al., 2016). These prove to be all rather small: The average standardized effect from

NA to PA is .034 (CI=[.008, .059]) in the younger sample and .089 (CI=[.059, .119] in the older sample; the average standardized effect from PA to NA is .038 (CI=[.017, .059] in the younger sample, and .026 (CI=[.001, .051] in the older sample. Note that for the latter effect, the corresponding fixed effect had a CI that contained zero. This seemingly contradictory result stems from the fact that the averaged cluster specific parameter always has a smaller CI than the corresponding model parameter, which reflects the kind of inferences that can be based on them: The averaged cluster specific parameter allows for inferences conditional on these clusters (i.e., other observations within the same clusters), whereas the conventional model parameters allow for inferences to other clusters from the population. Note that in this case, the lower bound of the CI of the average standardized effect is very close to zero, so we do not have to consider this an important contradiction. Mplus also provides the averaged within-person proportion of explained variance, which is .18 (CI=[.17, .20]) for PA and .21 (CI=[.19, .23]) for NA in the younger sample, and .26 (CI=[.24, .28]) for PA and .16 (CI=[.14, .17]) for NA in the older sample. Furthermore, the standardized results also reveal that the average within-person correlation between the innovations is -.19 (CI=[-.21, -.17]) in the younger sample, and -.27 (CI=[-.29, -.25]) in the older sample.

The standardized results also include the correlations between the six random effects included in the models. To represent these for both groups in a convenient and concise way, we make use of the function qgraph in R (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012). The resulting network representations presented in the upper part of Figure 3 show the correlation estimates that have credible intervals that do not include zero. Red connections represent positive (warm) correlations, and blue connections represent negative (cold) correlations, and the strength of these correlations is represented by the thickness of the connections. The correlations in the younger sample ranged from .22 to .46 (in absolute values), and in the older sample they ranged from .25 to .49 (in absolute values).

Insert Figure 3 about here

These networks show that there are marked differences between the younger and older samples: While the networks are equally dense (i.e., they have about the same number of connections with equal strengths), the patterns are very different. The two negative relationships in the network of the younger sample indicate that individuals with a higher autoregressive parameter for PA tend to have lower cross-lagged parameters from PA to NA, and individuals with a higher autoregressive parameter for NA tend to have lower cross-lagged parameters from NA to PA. The positive correlation between the autoregressive parameter for NA and the mean of NA indicates that individuals with a higher trait level for NA also tend to have more carry-over in their NA from one day to the next. Finally, there is a positive connection between the two means of PA and NA, which runs counter to the results we reported before in the decomposition, where the between-person correlations were close to zero. The currently positive correlation implies that individuals who have a higher mean on PA also tend to have a higher mean on NA; note however that this relationship is quite weak, and further inspection of the corresponding CI indicated that the lower bound was quite close to zero.

In contrast, in the older sample we see the network consists of two separate parts: One part connecting the parameters associated with PA, including the mean $\mu_{PA,i}$, and the lagged regression coefficients $\phi_{PP,i}$ and $\phi_{PN,i}$ (on the left side), and another part connecting the parameters associated with NA, including the mean $\mu_{NA,i}$, and the lagged regression coefficients $\phi_{NN,i}$ and $\phi_{NP,i}$ (on the right side). Higher mean PA is associated with *less* carry-over in PA, and a smaller effect from NA on PA. This pattern seems to suggest that high levels of PA are not the result from the maintenance of PA, but that they may emerge from the ability to shield NA to spill-over into PA. In contrast, higher mean NA is associated with *more* carry-over in NA, and a stronger effect from PA to NA. Finally, the autoregressive parameter is positively correlated to the cross-lagged parameter in the same part, meaning that carry-over and spill-over are positively related.

Model 2: Extending the multilevel VAR(1) model with random innovation variance

In the model presented above we considered random means and random lagged parameters. In contrast, the innovation variances and their covariance were fixed to be equal across individuals and estimated at the within-person level. While typically within-person (or level 1) residuals are not given much thought in multilevel modeling, they are actually of great interest here: They represent everything that was not measured explicitly, but that affects the course of the variables that were observed, including social interactions, weather conditions, demands at work or at home, and physiological factors such as sugar or caffeine intake, hormones, and health.

Rationale for individual differences in innovation variances and covariance

As Jongerling, Laurenceau, and Hamaker (2015) argued, we may expect individual differences in the variance of this collective term, due to two kinds of differences between individuals. First, there may be individual differences in reactivity to unobserved influences. For instance, there are a number of studies that show that individuals who are depressed a characterized by larger increases in negative affect in response to external negative events than non-depressed individuals: This phenomenon is referred to as *stress sensitivity* Wichters et al., 2009. Similarly, non-depressed individuals are characterized by a stronger increase in PA in response to a positive event in comparison to depressed individuals, which is referred to as the *reward experience* Wichters et al., 2009. Second, there may be individual differences in variability of unobserved influences. That is, some people lead more constant lives than others, and this difference in exposure to variability in external factors may also be captured by individual differences in the innovation variance.

While Jongerling et al. (2015) considered a univariate model and thus only needed a random innovation variance to account for individual differences in this part of the model, the current model is based on two variables, such that we also have a covariance between the innovations to consider. Individual differences in the covariance between innovations can be understood as resulting from three possible sources: a) within-day effects from PA to NA; b) within day effects from NA to PA; and c) unobserved (or “third”) variables that vary over time and that affect both variables (for instance, the weather, social interactions, external demands, physiological states, etc.). In most instances, the covariance will be the result of a mix of these three sources. If we want to actually distinguish between these different explanations, we should study the dynamics at the within-day timescale using multiple measurements within a day, for instance through the use of experience sampling (cf., Conner et al., 2009). Here we suffice with assuming that correlations between innovations are likely to result from all three sources, that is, from unobserved causes as well as within-day reciprocal effects between PA and NA. Moreover, because we believe that all these sources are likely to be characterized by individual differences, we want to extend the model presented above with a random covariance between the innovations as well as random innovation variances.

Random innovation variances and covariance

In its basic form, the innovations are said to come from a multivariate normal distribution, that is,

$$\boldsymbol{\zeta}_i \sim MN(\mathbf{0}, \boldsymbol{\Psi}), \tag{10}$$

which implies that in the bivariate case we have two variances and one covariance for the innovations. Here we want to extend the model such that the covariance matrix of the innovations can differ across individuals, meaning $\boldsymbol{\Psi}$ gets a subject index i.

In DSEM we can account for such individual differences using a normal distribution for the log of the innovation variance, which is the same as saying that the innovation variance comes from a log normal distribution. This way, we can allow the random log of an innovation variance to be correlated with the other random effects, using a multivariate normal distribution for all the random effects.

Extending the model with a random covariance is more complicated though, as it requires us to model this using an additional latent variable. Here, we specify a latent variable η_{it} that represents what the two innovations have in common, while the unique parts are modeled as residuals. Specifically, the innovations are modeled as

$$\begin{bmatrix} \zeta_{PA,it} \\ \zeta_{NA,it} \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} \eta_{it} \end{bmatrix} + \begin{bmatrix} e_{PA,it} \\ e_{NA,it} \end{bmatrix} \tag{11}$$

where the latter two residuals e are by definition uncorrelated; these form the unique parts of the innovations. The covariance matrix of the innovations for individual i can now be expressed as

$$\mathbf{\Psi}_i = E\left[\boldsymbol{\zeta}_{it}\boldsymbol{\zeta}_{it}^T\right] = \begin{bmatrix} E[e^2_{PA,it}] + E[\eta^2_{it}] & \\ -E[\eta^2_{it}] & E[e^2_{NA,it}] + E[\eta^2_{it}] \end{bmatrix}, \tag{12}$$

which shows that the variances of the innovations are the sum of the unique variances and the shared variance (i.e., the covariance term), while the covariance is determined by the negative variance of the common factor. The latter is the result of choosing the factor loadings of this common factor to be 1 and -1: This implies that for each individual, the covariance between the innovations will be negative.[7]

The model defined in Equations 1, 7, and 12, is visually represented in Figure 4. It shows that we do not estimate any parameters at level 1, but that instead, we have 9 random effects, that are estimated at the between level, which we discuss below.

Insert Figure 4 about here

Between-person model

In comparison to Model 1, the current model has three parameters less at the within-person level; instead there are six additional parameters at the between-person level. These include three fixed effects (i.e., $\gamma_{log(\sigma^2_{e_P})}$ and $\gamma_{log(\sigma^2_{e_N})}$ for the unique parts of the innovation variances, and $\gamma_{log(-\sigma)}$ for the covariance between the innovations), and three random effects (i.e., two random variances for the unique parts of the innovations ($E[e^2_{PA,it}] = \sigma^2_{eP,i}$ and $E[e^2_{NA,it}] = \sigma^2_{eN,i}$)), and one random covariance between the innovations ($-E[\eta^2_{it}] = \sigma_i$)). If we simply consider all nine random effects to be related to each other without any particular constraints, the between-person model is

$$\begin{bmatrix} \mu_{PA,i} \\ \mu_{NA,i} \\ \phi_{PP,i} \\ \phi_{PN,i} \\ \phi_{NP,i} \\ \phi_{NN,i} \\ \log(\sigma^2_{e_P})_i \\ \log(\sigma^2_{e_N})_i \\ \log(-\sigma)_i \end{bmatrix} = \begin{bmatrix} \gamma_P \\ \gamma_N \\ \gamma_{PP} \\ \gamma_{PN} \\ \gamma_{NP} \\ \gamma_{NN} \\ \gamma_{log(\sigma^2_{e_P})} \\ \gamma_{log(\sigma^2_{e_N})} \\ \gamma_{log(-\sigma)} \end{bmatrix} + \begin{bmatrix} u_{P,i} \\ u_{N,i} \\ u_{PP,i} \\ u_{PN,i} \\ u_{NP,i} \\ u_{NN,i} \\ u_{log(\sigma^2_{e_P}),i} \\ u_{log(\sigma^2_{e_N}),i} \\ u_{log(-\sigma),i} \end{bmatrix}, \tag{13}$$

where the nine γ's represent the fixed or averaged effects, and the nine u's represent the individual deviations from these means. These individual deviations are assumed to come from a multivariate normal distribution, that is,

$$\boldsymbol{u}_i \sim MN(\mathbf{0}, \boldsymbol{\Omega}), \tag{14}$$

where $\boldsymbol{\Omega}$ is a 9×9 covariance matrix, which implies there are 9 variances (i.e., the random effects), and 36 covariances. Combining this with the 9 fixed effects, we now have a model with 54 parameters at the between-person level. Below, we begin with estimating this basic model, and subsequently discuss an extension that can be used to further model the covariance structure of the random effects.

Results

Again, we used 50,000 iterations and a thinning of 10. The parameter estimates and their 95% credible intervals are presented in the right part of Table 1. Note that none of these intervals includes zero, indicating that there is ample evidence that these fixed effects differ from zero.

We also represented the correlations between the 9 random effects in each group as a network in the lower part of Figure 3. When comparing these networks to the networks based on Model 1, it is striking that the relationships between the random effects that are included in both models (i.e., means and lagged parameters) change depending on whether or not the variances and covariance of the innovations are included as random effects. This emphasizes the need to consider potential randomness in these residual variances, because ignoring these additional sources of individual differences may result in biased estimates of the other model parameters (cf. Jongerling et al., 2015).

When comparing the networks of the two samples based on Model 2, it is clear that the younger sample is characterized by a much denser network than the older sample. However, there are also some similarities across the two groups. First, what is interesting to see is that both groups are characterized by strong positive connections between the individual's mean on NA (i.e., $\mu_{NA,i}$), the unique part of the innovation variance of NA (i.e., $\log(\sigma^2_{eN,i})$), and the (negative) covariance between the innovations of PA and NA (i.e., $\log(-\sigma_i)$). This implies that individuals with a relatively high mean on NA also tend to be characterized by a larger innovation variance (as this is the sum of $\sigma^2_{eN,i}$ and σ_i), and a more negative covariance between the innovations.

Second, in both groups there are also relatively strong positive relationships between the autoregressive coefficient for NA (i.e., $\phi_{NN,i}$), the mean of NA (i.e., $\mu_{NA,i}$), and the unique part of the innovation variance of NA (i.e., $\log(\sigma^2_{eN,i})$). This implies that individuals who are relatively high on NA on average, tend to be characterized by a stronger carry-over effect for their NA, and they have larger (unique) innovation variances.

Third, in both groups there are positive relationships between the unique part of the innovations of PA and NA and the covariance between the innovations of PA and NA. This implies that individuals who are more reactive to external influences and/or who experience more variability in external influences, tend to do so for both PA and NA.

Fourth, there is a negative correlation between the mean of PA (i.e., $\mu_{PA,i}$) and the lagged effect from NA to PA (i.e., $\phi_{PN,i}$) in both samples. This implies that individuals who are relatively high on PA on average, tend to be characterized by a lower cross-lagged effect from NA to PA.

Finally, there is a negative relationship in both groups between the autoregressive parameter of PA (i.e., $\phi_{PP,i}$), and the negative covariance between the innovations. This implies that when individuals are characterized by more carry-over in their PA, they also tend to have a smaller value for the log of the negative covariance, which corresponds to larger negative covariance between the innovations.

The current analysis is closely related to studies that have propagated a dynamic network view of symptoms of psychopathology (e.g. Bringmann et al., 2013). However, the between-person network that was presented in previous studies typically consists of the within-person means only. This may be explained by the fact that these analyses are often based on running a multilevel regression model for each variable separately, such that the covariance between all the random effects are not estimated as part of the model. In contrast, the DSEM approach taken here makes the study of all the random effects and their interrelatedness self-evident. In this regard it is also interesting to note that the strongest connections are actually not between the two means in these samples: In the younger sample there is a weak positive correlation, whereas in the older sample there is no evidence that the means are related. The latter is in agreement with the idea that the trait-like dimensions of PA and NA are not related to each other in the population.

While the current approach makes it easy to investigate the relationships between all the random effects in the model, this also leads to an increased complexity: Where we started with just two observed variables in the original decomposition, we now have nine random effects and 36 correlations between them. Alternatively however, one may also argue that where we started out with a dataset of $2 \times 100 \times 100$ (i.e., variables by persons by time points), we have now reduced this to a an 9 by 9 correlation matrix.[8] Either way, this is still a large number of relationships, which we may want to model further, for instance by using predictors or latent variables to account for their interrelatedness. We consider one particular approach next.

Model 3: Mediation through the random effects

In the COGITO dataset there are diverse person characteristics that were measured prior and/or after the intensive longitudinal measurement period. Here we consider depressive symptoms measured by the Center for Epidemiologic Studies Depression (CESD) scale both before and after the daily measurement phase. Prior research, including analysis of this issue with the data from the COGITO study (cf. Brose, Schmiedek, et al., 2015), has shown that the autoregressive parameter of affective measurements is positively related to current levels of depression , but also that it is predictive of depression two and a half years later, suggesting that it flags an important malfunctioning regulation mechanism (Houben, Van den Noortgate, & Kuppens, 2015; Koval et al., 2012; Kuppens et al., 2010; Kuppens et al., 2012). To see whether we can replicate these two findings in a single model, we include pre-CESD as a predictor of the nine random effects of the dynamic multilevel model described above, and post-CESD as an outcome variable of these random effects and of pre-CESD. We compute additional parameters in Mplus, based on a product of the two parameters that form a mediated effect through a particular random effect. This allows us to investigate these nine mediators, using the posterior distribution for these indirect effects.

In Table 2, we have included the point estimates and the 95% CIs for the nine indirect effects and the direct effect. It shows that in both samples there is a positive *direct effect* of pre-CESD on post-CESD, with the effect in the older sample being about twice as large as that in the younger sample. Furthermore, in the younger sample there is evidence that two of the nine *indirect effects* are different from zero. The first of these is through the autoregressive parameter of NA (i.e., mediated by ϕ_{NN}), which confirms previous findings (Brose, Schmiedek, et al., 2015; Kuppens et al., 2012): People with more depressive symptoms at the start (i.e., higher pre-CESD), tend to have more carryover of negative affective states (i.e., higher autoregression in NA), which in turn is associated with more depressive symptoms at the post-assessment (i.e., higher post-CESD). This effect exists after controlling for previous depression (i.e., the direct effect). The second indirect effect is through the covariance of the innovations, and illustrates the importance of looking past the usual suspects of means and autoregression: It implies that higher pre-CESD precedes a stronger negative covariance between the innovations (either because of within day reciprocal effects or because of an unmeasured common cause), which is followed by higher post-CESD. This latter result is in agreement with the idea that people who differentiate less in their affective responses (i.e., individuals with a larger negative covariance between innovations), run more risk of becoming depressed (van Borkulo, Boschloo, Borsboom, Penninx, & Schoevers, 2015).

The two indirect effects added together are 0.199, which is about two-thirds of the direct effect, which is 0.290. Moreover, note that the individual means of PA and NA did not mediate the effect of prior depressive symptoms on later depressive symptoms. In the younger sample, only the mean of NA was related to pre-CESD (point estimate and 95% CI: 0.78 [0.34,1.25]), whereas in the older sample the mean of PA was related to pre-CESD (point estimate and 95% CI: -2.00 [-2.94,-1.02]), but none of the means were predictive of post-CESD. Although highly speculative, the largely absent association between depression scores and average PA and NA from the daily measurements may reflect a form of adaption in scale use, where depressed individuals grow used to feeling more NA and less PA, and thus change their reference point when filling out daily diaries.

Discussion

We have introduced DSEM and showcased some of the exciting new possibilities it has to offer for analyzing intensive longitudinal data, using the daily affective measurements from the COGITO study. Our multilevel models can be thought of as forming a bottom-up extension of a time series model, in which we use the latter at the within-person level, and allow for individual differences in the parameters at the between-person level. Alternatively, we can think of it as a top-down extension of the well-known cross-lagged panel model, in that instead of assuming all the parameters to be fixed (i.e., equal) across all individuals, we allow them to be random. We considered random means, lagged parameters, innovation variances and innovation covariance, and used these random effects as mediators between previous and subsequent depressive symptoms.

Surpluses of DSEM in these applications

These applications of DSEM as implemented in Mplus version 8 highlight a number of advantages over other software packages that could be used for dynamic multilevel analyses. Most importantly, in comparison to regular multilevel regression software, the current approach allows for: a) multiple outcome variables without the need to add an additional level for this; b) random residual variances; c) (random) covariance between residuals; d) correlations between all random effects in the model; e) outcome variables at the between-person level, and hence also mediation of between-level predictors through the random effects to between-level outcomes. While some of the limitations of standard multilevel software may be circumvented by experienced users (e.g., including multiple outcome variables can be done by stacking the variables, and using dummies to indicate whether it is the first or second variable etc.), the ease with which these issues are handled in Mplus is unparalleled.

Furthermore, when using standard multilevel software for the estimation of models with lagged effects, missing observations are troublesome. For instance, if we consider a first-order autoregressive model and we have a missing value on the outcome variable at occasion t, this implies that at $t + 1$ we will have a missing value on the predictor. Both cases would thus be deleted. In contrast, DSEM handles missing data in a similar vein as is done in the Kalman filter (cf., Harvey, 1989): This implies that when there is a missing observation, the filter cannot update a prediction using the observed score, and instead uses the predicted rather than the updated score when it proceeds to the next occasion. In so doing, all data are used.

Related to this, if we have unequally spaced observations by design (such as in experience sampling, which is by definition based on (quasi-)random measurement times), we can add missing values to make our cases (approximately) equally spaced. In theory, this approach could also be used in standard multilevel software, but it easily leads to so many missing observations that there would be very few cases left for which there is an observed value for the outcome as well as the lagged predictor. In contrast, DSEM handles missing data using an algorithm based on the Kalman filter (Harvey, 1989), and it can easily handle additional missings. Moreover, this approach is built into DSEM, and only requires one extra line of code in which the user specifies the variable that should be used to determine the distances (i.e., time intervals) between the observations.

Another important advantage of DSEM is that the fundamental decomposition into the within-person and between-person components is done simultaneously as the estimation of the model parameters. As a result, DSEM is not haunted by the well-known problems associated with centering variables using sample means, which is particularly notorious in the context of autoregressive relationships (Hamaker & Grasman, 2014; Judson & Owen, 1999; Nickell, 1981).

To be fair, most (if not all) of these advantages of DSEM also apply to other Bayesian software packages, such as WinBUGS, jags, and stan: These are flexible programs that allow researchers to estimate a wide variety of models. However, the advantages of DSEM over these other Bayesian programs is that it uses tailor-made code, which makes it relatively easy to specify these complex multilevel models. Moreover, DSEM is based on efficient programming, making it relatively fast and stable.

Unresolved issues

Since DSEM is a newly emerging class of analyses techniques, there are very few guidelines and rules of thumb that the user can turn to. It seems that with each new modeling possibility, there are new questions that arise. Here we provide a brief and non-exhaustive list of what we consider the most urgent matters that need to be resolved within the next few years.

First, we need to know more about the trade-off between the number of people, the number of time points, the number of variables, and the number of random effects and other parameters in the model. This will clearly depend on the circumstances, and we should not expect a one-size-fits-all rule of thumb. However, we need to develop a sense of what are reasonable ratios between these quantities, for instance through simulation studies that investigate the accuracy and reliability of parameter estimates, coverage rates, and efficiency for basic models, such as the bivariate VAR(1) model presented in this paper.

Second, we need to find appropriate ways to compare nested and non-nested models. The DIC is not functioning well for these models, due to the many parameters that need to be sampled: All the individual parameters are also treated as unknown parameters in the Bayesian context, which have to be sampled at each iteration. This makes the DIC very unstable, even when very large numbers of iterations are used. Moreover, because the DIC is based on a likelihood in which the random effects are included, the DIC is not always comparable across models due to the parametrization (cf. Celeux, Forbes, Robert, & Titterington, 2006). For instance, when latent variables are added to a model (as in our application when going from Model 1 to Model 2), this model is no longer based on the same parametrization, and as a result, the DIC of the model with is not comparable to the DIC of the model without this latent variable. At this point, there is no obvious solution to this problem. Alternatively, we can check whether there is evidence that a parameter differs from zero based on its CI, but for variance terms, these will never include zero by definition. Hence, at this point there is no easy to use tool to determining whether a random effect is necessary, or that the variance across individuals can be fixed to zero.

Third, related to model comparison, we need to develop measures to evaluate model fit. This may prove rather challenging, as even in N=1 time series analysis, there is not a strong tradition of summarizing model fit. This may be because the focus in time series analysis is not specifically on the proportion of explained variance (as it tends to be in regression analysis), but also on the autocovariance structure of the data. On the other hand, there is no saturated model that could serve as a basis for comparison as is typically done in structural equation modeling. Evaluating the appropriateness of a time series model has often relied on investigating whether the residuals (i.e., innovations) behave like white noise, meaning there are no autocorrelations in these series (Box & Jenkins, 1970): This illustrates the strong focus on trying to capture the autocovariance structure in the observed data. In multilevel multivariate extensions, this implies we have to consider the autocorrelation and partial autocorrelation functions for each variable and the cross-lagged correlations functions for each pair of variables for each person: Clearly, it will not be easy to keep track of the many functions, and even more so, it would also require the development of new rules of thumb regarding—for instance—the number of acceptable violation of white noise (both within and across individuals).

Fourth, we need to develop additional ways to further deal with the covariance structure of the random effects in some of these models. As shown in Models 1 and 2, even with a relatively simple bivariate dataset, we can easily obtain a rather complex covariance structure at the between-person leve, that begs for further modeling. For instance, one could factor analyse these between-person differences in parameters, to see whether there particular random effects that are more strongly related due to a potential common cause. Obviously, such between-person models would benefit from substantive theory, and it is our hope that researchers will be inspired to consider theoretical accounts in view of the new modeling opportunities and to investigate whether existing theories can be translated into hypotheses about the interrelatedness of these random effects at the between-person level.

Fifth, we need more insight on the distribution of the data, and the effects this has on parameter estimation. While measures of positive affect are typically characterized by a distribution that could be well described by a normal distribution (both at the within-person and the between-person level), measures of negative affect tend to behave quite differently. That is, some individuals have distributions that look normal, but many others have a strong floor effect in that they indicate at a large number of occasions that they experience no negative affect at all (cf. (Brose, De Roover, Ceulemans, & Kuppens, 2015)). How to deal with such distributions continues to be a challenging question. Potential solutions could be to handle such variables as categorical (i.e., ordinal) rather than as continuous, to use factor models rather than sum scores, and a combination of these two (i.e., use ordinal indicators of a continuous latent variable). Another possible approach would be to use a regime-switching model, which will be possible with Mplus version 8.1 (see Asparouhov et al., 2017).

Sixth, we need to figure out how we should handle potential trends in the data. In the time series literature, the convention is to de-trend data before considering further relationships between variables (e.g., Chatfield, 2004), and this approach has also been advocated by some in panel research and intensive longitudinal studies (e.g., Berry & Willoughby, 2016; Wang, Hamaker, & Bergman, 2012). Recently, however, Wang and Maxwell (2015) questioned this approach in the context of longitudinal multilevel modeling. Some have considered the possibility that the trend results from the same underlying dynamics as the short-term fluctuations, a view that is common in the dynamic systems literature (Bisconti, Bergeman, & Boker, 2004; Van der Maas et al., 2006) and continuous time modeling (Voelkle et al., 2012). Hence, there are various ways to handle trends. The cross-classification option offered by DSEM (which we did not consider here), allows for the modeling of an average (across individuals) slope over time of an unknown shape (Asparouhov et al., in preparation). Alternatively, we can add time as a predictor and allow for a random slope so that individuals can be characterized by different trends. And finally, we can assume the trend results from the underlying dynamics, and that it does not require explicit modeling, which is the approach we have taken here. How we handle potential trends in the data, can have serious consequences for the parameter estimates that describe the dynamics of the process: It should therefore reflect our conviction of the underlying mechanism, and whether we believe trends are something that happens separately from the dynamics, or as an intrinsic part of it.

To conclude

With the many novel options offered with the creation of the DSEM toolbox, we also see many new challenges and unresolved issues that require a solution. But we should not let this discourage us at this point. Rather, it is our hope that the fast technological and statistical developments that are taking place now will inspire many of us to pursue a research line in this innovative field: We need psychometricians, applied statisticians, quantitative psychologists and substantive researchers to explore this exciting new frontier, so that ten years from now we can look back and smile at how little was known today.

References

Almeida, D. M., Wethington, E., & Chandler, A. L. (1999). Daily transmission of tension between marital dyads and parent-child dyads. *Journal of Mariage and the Family*, *61*, 49-61.

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2017). Dynamic latent class analysis. *Structural Equation Modeling*, *24*, 257-269.

Asparouhov, T., Hamaker, E. L., & Muthén, B. (in preparation). Dynamic structural equation modeling.

Asparouhov, T., & Muthén, B. (2010). *Bayesian analysis of latent variable models using Mplus.* Technical Report: Version 3. Retrieved from: http://statmodel.com/download/Bayes3.pdf.

Asparouhov, T., & Muthén, B. (2015). General random effect latent variable modeling: Random subjects, items, context, and parameters. In J. R. Harring, L. M. Stapleton, & S. N. Beretvas (Eds.), *Advnaces in multilevel modeling for educational research: Addressing practical issues found in real-world applications.* Charlotte, NC: Information Age Publishing.

Berry, D., & Willoughby, M. T. (2016). On the practical interpretability of cross-lagged panel models: Rethinking a developmental workhorse. *Child Development*, 1-24. doi: 10.1111/cdev.12660

Bisconti, T., Bergeman, C. S., & Boker, S. M. (2004). Emotional well-being in recently bereaved widows: A dynamical system approach. *Journal of Gerontology, Series B: Psychological Sciences and Social Sciences*, *59*, 158-167.

Bolger, N., DeLongis, A., Kessler, R. C., & Wethington, E. (1989).

Journal of Marriage and the Family, *51*, 175-183.

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*(2), 203-219. doi: 10.1037/0033-295X.110.2.203

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. San Francisco, CA: Holden-Day.

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, ... Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLoS ONE*, *8*, e60188, 1-13.

Brose, A., De Roover, K., Ceulemans, E., & Kuppens, P. (2015). Older adults' affective experiences across 100 days are less variable and less complex than younger adults'. *Psychology and Aging*, *30*, 194-208. doi: 10.1037/a0038690

Brose, A., Schmiedek, F., Koval, 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

Brose, A., Schmiedek, F., & Lindenberger, U. (2013). Affective states contribute to trait reports of affective well-being. *Emotion*, *13*, 940-948. doi: 10.1037/a0032401

Brose, A., Voelkle, M. C., Lövdén, M., Lindenberger, U., & Schmiedek, F. (2015). Differences in the between-person and the within-person structures of affect are a matter of degree. *European Journal of Personality*, *29*, 55-71. doi: 10.1002/per.1961

Cattell, R. B. (1952). The three basic factor-analytical research designs: Their interrelations and derivatives. *Psychological Bulletin*, *49*(5), 499–520. doi:

10.1037/h0054245

Celeux, G., Forbes, F., Robert, C. P., & Titterington, D. M. (2006). Deviance informaiton criteria for missing data models. *Bayesian Analysis*, *1*, 651-674.

Chatfield, C. (2004). *The analysis of time series: An introduction* (6th ed.). London: Chapman and Hall.

Conner, T. S., Tennen, H., Fleeson, W., & Feldman Barret, L. (2009). Experience sampling methods: A modern idiographic approach to personality research. *Social and Personality Psychology Compass*, *3*, 292-313.

Deboeck, P., & Preacher, K. (2015). No need to be discrete: A method for continuous time mediation analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*, 1-15. doi: 10.1080/10705511.2014.973960

Durbin, J., & Koopman, S. J. (2001). *Time series analysis by state space methods.* New York, NY: Oxford University Press.

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48*(4), 1-18.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. (2013). *Bayesian data analysis* (3rd ed.). Boca Raton, FL: CRC Press, Taylor & Francis Group.

Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development*, *58*, 80–92.

Gottman, J. M., Murray, J. D., Swanson, C. C., Tyson, R., & Swanson, K. R. (2002). *The mathematics of marriage: Dynamic nonlinear models.* Cambridge, MA: MIT Press.

Grice, J. W. (2004). Bridging the idiographic-nomothetic divide in ratings of self and others. *Journal of Personality*, *72*(2), 203-241.

Halaby, C. N. (2004). Panel models in sociological research: Theory intro practice. *Annual Review of Socioloy*, *30*, 507-544. doi: 10.1146/annurev.soc.30.012703.110629

Hamaker, E. L. (2012). Why researchers should think "within-person" a paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (p. 43-61). New York, NY: Guilford Publications.

Hamaker, E. L., Ceulemans, E., Grasman, R. P. P. P., & Tuerlinckx, F. (2015). Modeling affect dynamics : State-of-the-art and future challenges. *Emotion Review*, *7*, 316-322.

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2005). Statistical modeling of the individual: Rationale and application of multivariate time series analysis. *Multivariate Behavioral Research*, *40*(2), 207-233. doi: 10.1207/s15327906mbr4002_3

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

Hamaker, E. L., Kuiper, R., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102-116. doi: 10.1037/a0038889

Hamaker, E. L., Nesselroade, J. R., & Molenaar, P. C. M. (2007). The integrated trait-state model. *Journal of Research in Personality*, *41*(2), 295-315. doi: 10.1016/j.jrp.2006.04.003

Hamaker, E. L., Schuurman, N. K., & Zijlmans, E. A. O. (2017). Using a few snapshots to distinguish mountains from waves: Weak factorial invariance in the context of trait-state research. *Multivariate Behavioral Research*, *52*(1), 47-60. doi: 10.1080/00273171.2016.1251299

Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science*(26), 10-15. doi: 10.1177/0963721416666518

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

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter.* Cambridge: University Press.

Houben, M., Van den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, *141*, 901930. doi: 10.1037/a0038822

Jongerling, J., Laurenceau, J.-P., & Hamaker, E. L. (2015). A multilevel ar(1) model: Allowing for inter-individual differences in trait-scores, inertia, and innovation variance. *Multivariate Behavioral Research*, *184*, 334-349.

Judson, R. A., & Owen, A. L. (1999). Estimating dynamic panel data models: A guide for macroeconomists. *Economics Letters*, *65*, 9-15. doi: http://dx.doi.org/10.1016/S0165-1765(99)00130-5

Kim, C.-J., & Nelson, C. R. (1999). *State-space models with regime switching: Classical and Gibbs-sampling approaches with applications.* Cambridge, MA: The MIT Press.

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

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

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

Lamiell, J. T. (1998). 'Nomothetic' and 'idiograhic': Contrasting Windelband's understanding with contemporary usage. *Theory and Psychology*, *8*, 23-38.

Masten, A. S., & Cichetti, D. (2010). Devlopmetnal cascades. *Development and Psychopathology*, *22*, 491-495. doi: 10.1017/S0954579410000222

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, *50*(2), 181–202. doi: 10.1007/bf02294246

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology - this time forever. *Measurement: Interdisciplinary Research and Perspectives*, *2*(4), 201-218. doi: 10.1207/s15366359mea0204_1

Molenaar, P. C. M., Huizenga, H. M., & Nesselroade, J. R. (2003). The relationship between the structure of interindividual and intraindividual variability: A theoretical and empirical vindication of developmental systems theory. In U. M. Staudinger & U. Lindenberger (Eds.), *Understanding human development: Dialogues with lifespan psychology* (p. 339-360). Norwell, MA: Kluwer Academic Publishers. doi: 10.1007/978-1-4615-0357-6_15

Nesselroade, J. R. (2002). Elaborating on the differential in differential psychology.

Multivariate Behavioral Research, *37*, 543-561.

Nickell, S. (1981). Biases in dyanmic models with fixed effects. *Econometrica*, *49*(6), 1417-1426.

Ousey, G. C., Wilcox, P., & Fisher, B. S. (2011). Something old, something new: Revisiting competing hypothesis of the victimization-offending relaitonship among adolescents. *Journal of Quantitative Criminology*, *27*, 53-84. doi: 10.1007/s10940-010-9099-1

Schmitz, B. (2000). Auf der Suche nach dem verlorenen Individuum: Vier Theoreme zur Aggregation von Prozessen. *Psychologische Rundschau*, *51*, 83-92.

Schmitz, B., & Skinner, E. (1993). Perceived control, effort, and academic performance: Interindividual, intrainidividual, and multivariate time-series analyses. *Journal of Personality and Social Psychology*, *64*, 1010–1028.

Schuurman, N. K., Ferrer, E., de Boer-Sonnenschein, M., & Hamaker, E. L. (2016). How to compare cross-lagged associations in a multilevel autoregressive model. *Psychological Mehtods*, *21*, 206-221.

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

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

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

Trull, T. J., & Ebner-Priemer, U. (2013). Ambulatory assessment. *Annual Review* *of Clinical Psychology*, *9*, 151-176.

Van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*, 842-861.

van Borkulo, C. D., Boschloo, L., Borsboom, D., Penninx, B. W. J. H., & Schoevers, R. A. (2015). Association of symptom network structure and the course of depression. *JAMA Psychiatry*, *72*, 1219-1226.

Voelkle, M., Brose, A., Schmiedek, F., & Lindenberger, U. (2014). Toward a unified framework for the study of between-person and within-person strcutures: Building a bridge between two research paradigms. *Multivariate Behavioral Research*, *49*(3), 193-213. doi: 10.1080/00273171.2014.889593

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

Wang, L., Hamaker, E. L., & Bergman, C. S. (2012). Investigating inter-individual difference in short-term intra-individual variability. *Psychological Methods*, *17*, 567-581.

Wang, L., & Maxwell, S. E. (2015). On disaggregating between-person and witin-person effects with longitudinal data using multilevel models. *Multivariate Behavioral Research*, *20*(1), 63-83. doi: 10.1037/met0000030

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063-1070.

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

Zhang, Z., Hamaker, E. L., & Nesselroade, J. R. (2008). Comparisons of four methods for estimating a dynamic factor model. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*(3), 377-402. doi: 10.1080/10705510802154281

Table 1: Point estimates (posterior means) and 95% credible intervals for fixed effects (i.e., means) and random effects (i.e., variances) in both groups.

Variable	Model 1				Model 2			
	Fixed Effects (means)		Random Effects (variances)		Fixed Effects (means)		Random Effects (variances)	
	Young	Older	Young	Older	Young	Older	Young	Older
μ_{PA}	3.09 [2.88,3.31]	4.56 [4.28,4.83]	1.18 [0.87,1.62]	1.86 [1.39,2.54]	3.10 [2.88, 3.32]	4.58 [4.29, 4.85]	1.27 [0.94,1.74]	1.96 [1.45,2.72]
μ_{NA}	0.98 [0.83,1.13]	0.32 [0.21,0.42]	0.60 [0.44,0.83]	0.30 [0.22,0.41]	0.97 [0.82, 1.13]	0.31 [0.21, 0.43]	0.60 [0.44,0.85]	0.31 [0.23,0.43]
ϕ_{PP}	0.33 [0.28,0.39]	0.42 [0.36,0.47]	0.06 [0.04,0.08]	0.06 [0.05,0.09]	0.37 [0.32, 0.42]	0.47 [0.42, 0.53]	0.06 [0.04,0.08]	0.07 [0.05,0.10]
ϕ_{PN}	0.05 [0.01,0.09]	0.17 [0.09,0.25]	0.02 [0.01,0.04]	0.10 [0.07,0.16]	0.08 [0.04, 0.13]	0.15 [0.07, 0.24]	0.03 [0.02,0.05]	0.07 [0.04,0.13]
ϕ_{NP}	0.04 [0.01,0.07]	0.02 [-0.02,0.06]	0.01 [0.01,0.02]	0.03 [0.02,0.04]	0.03 [0.00, 0.06]	0.02 [0.00, 0.03]	0.01 [0.00,0.01]	0.00 [0.00,0.00]
ϕ_{NN}	0.37 [0.32,0.42]	0.23 [0.17,0.29]	0.06 [0.04,0.09]	0.06 [0.05,0.09]	0.40 [0.34, 0.45]	0.27 [0.22, 0.33]	0.07 [0.05,0.09]	0.06 [0.05,0.09]
$log(\sigma^2_{eP})$					-1.33 [-1.51,-1.17]	-2.53 [-2.75,-2.32]	0.69 [0.50,0.97]	1.18 [0.88,1.62]
$log(\sigma^2_{eN})$					-2.04 [-2.33,-1.75]	-4.37 [-4.73,-4.01]	1.92 [1.39,2.71]	3.30 [2.44,4.55]
$log(-\sigma)$					-3.29 [-3.64,-2.98]	-4.78 [-5.09,-4.43]	2.00 [1.40,2.95]	2.60 [1.92,3.60]

Model 1 is the multilevel VAR(1) model with random means for daily positive and negative affect (μ_{PA} and μ_{NA}), and random autoregressive and cross-lagged regression coefficients (ϕ_{PP} to ϕ_{NN}). Model 2 also includes random log of the variance of the unique parts of the innovations ($log(\sigma^2_{eP})$ and $log(\sigma^2_{eN})$); and random log of the negative covariance between the innovations ($log(\sigma)$).

Table 2: Point estimates (posterior means) and 95% credible intervals for direct and indirect effects of pre-CESD on post-CESD in both groups.

Effect	Young	Older
direct	0.290 [0.062,0.522]	0.585 [0.076,1.206]
mediated by μ_{PA}	*0.058 [-0.011,0.154]*	*0.054 [-0.018,0.147]*
mediated by μ_{NA}	*0.024 [-0.062,0.130]*	*0.011 [-0.022,0.070]*
mediated by ϕ_{PP}	*0.003 [-0.032,0.050]*	*0.003 [-0.020,0.043]*
mediated by ϕ_{PN}	*0.000 [-0.053,0.061]*	*-0.003 [-0.106,0.097]*
mediated by ϕ_{NP}	*-0.019 [-0.178,0.087]*	*-0.048 [-0.691,0.470]*
mediated by ϕ_{NN}	0.127 [0.036,0.258]	*-0.011 [-0.069,0.020]*
mediated by $log(\sigma^2_{eP})$	*0.000 [-0.059,0.055]*	*-0.046 [-0.127,0.007]*
mediated by $log(\sigma^2_{eN})$	*-0.009 [-0.103,0.076]*	*0.079 [-0.015,0.212]*
mediated by $log(-\sigma)$	0.072 [0.004,0.185]	*0.029 [-0.035,0.122]*

The nine model parameters are: the means of daily positive and negative affect (μ_{PA} and μ_{NA}); the autoregressive and cross-lagged regression coefficients (ϕ_{PP} to ϕ_{NN}); the log of the variance of the unique parts of the innovations ($log(\sigma^2_{eP})$ and $log(\sigma^2_{eN})$); and the log of the negative covariance between the innovations ($log(\sigma)$).

Footnotes

[1]At this point, multiple group analyses are not part of DSEM in Mplus version 8 yet. Alternatively, we could treat the data as one sample, and have a grouping variable that is used at the between-person level to see whether there are differences between the groups. However, as the two samples were sampled from different age populations, we decided to treat them as two separate samples here.

[2]Brose et al. (2013) actually investigated the effect of current affective state on the way individuals answer questions with respect to what they are like in general (i.e., when they are given a trait instruction), using data from the COGITO study. The results showed that the current affective state a person is in contributed substantially to the way a person evaluates their general affect. It is likely that the relative contributions of trait and state depend on the instruction that is given, but studies like the one by ? (?) show that using a particular instruction does not eliminate the potential contamination of other sources of variability.

[3]For instance, if we want to estimate a second-order autoregressive process, that is $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \zeta_t$, we can write the state vector of the state-space equation as

$$\begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \zeta_t \\ 0 \end{bmatrix},$$

which shows that time index of the elements of the state vector at for instance time t (left-hand side) is t and $t - 1$.

[4]In the younger sample, for instance, the intervals were 1 for 10,066 cases, 2 for 7,321 cases, 3 for 1,580 cases, 4 for 636 cases, 5 for 168 cases, 6 for 89 cases, 7 for 54 cases, and so on, with a maximum interval of 54 days.

[5]While inertia is a term that has been used extensively to refer to autoregression and autocorrelation in affective research, starting with Suls et al. (1998) and Gottman, Murray, Swanson, Tyson, and Swanson (2002), this should not be confused with the usage of this term in the context of differential equations and physics.

[6]Note that the CI of a variance term will never include zero, as the prior is specified to be larger than zero. Hence, using the CI of a variance estimate to investigate whether there is evidence that a parameter differs from zero, is not as straight forward as it is for other parameters.

[7]Note that in the current setup, we have to choose the innovation covariance to be either positive or negative for all individuals. This may seem needlessly restrictive or even an undesirable limitation. We have therefore also considered an alternative model, in which we fixed the variance of this common factor to 1 and allowed one of the factor loadings to be estimated freely and to vary randomly across individuals: This would allow for both positive and negative covariances between the innovations. However, this alternative setup does not guarantee that every covariance matrix can be fitted well, and in our case it caused convergence problems. An alternative setup allows all covariance matrices to be fitted well would involve the Cholesky decomposition.

[8]The authors thank Nilam Ram for pointing out this alternative perspective.

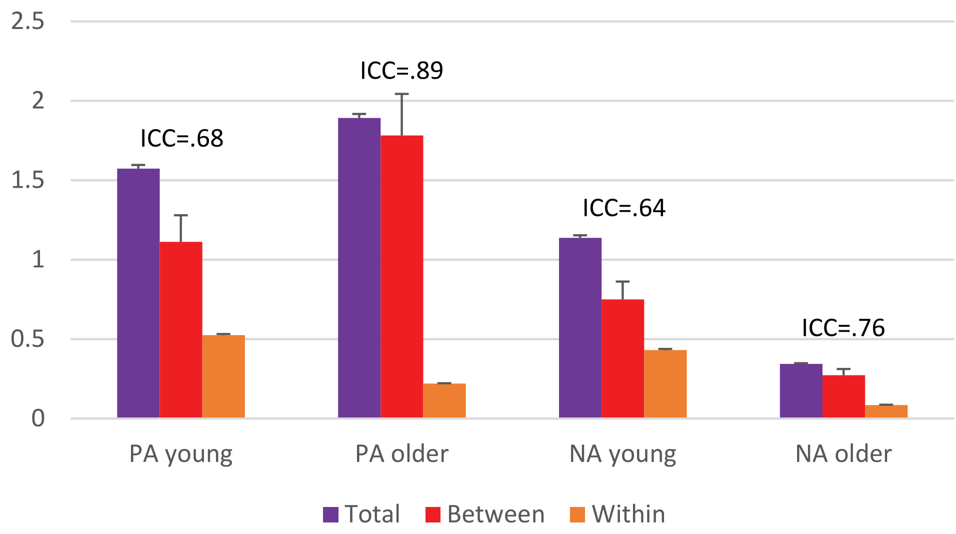
Figure Captions

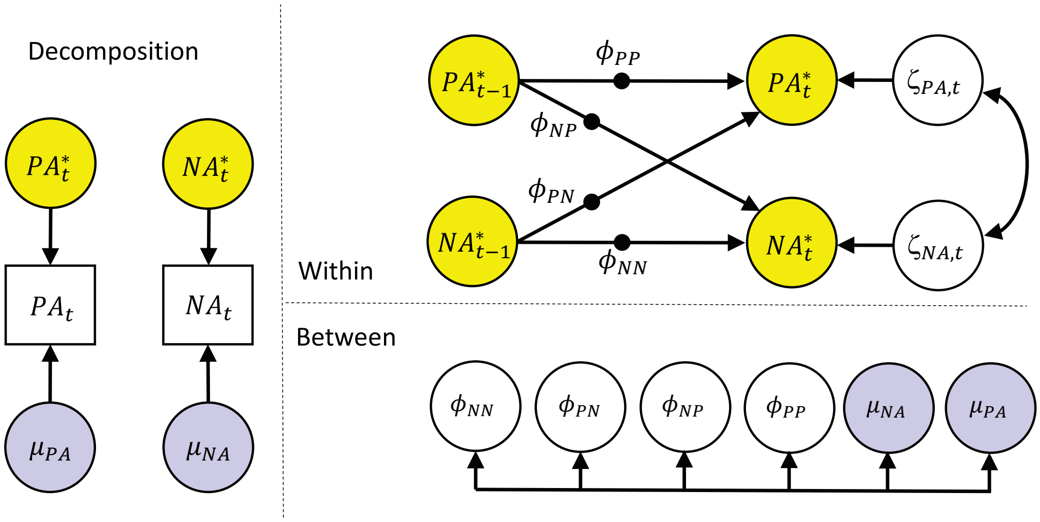
Figure 1. Variance estimates for the younger and older samples, with the posterior standard deviations, including: 1) total variance (across time and persons); 2) between-person variance (i.e., variance in within-person means across people); and 3) within-person variance (i.e., variance within people across time). For each variable in each group, we also included the intraclass correlation (ICC), which represents the proportion of total variance that is account for by the stable between-person differences.

Figure 2. Representation of the multilevel VAR(1) model. Left part contains the decomposition into within-person (time-varying) and between-person (time-invariant) components. Top right contains the within-person model, which is a VAR(1) model. Bottom right contains the between-person model, which includes the between-person components from the decomposition, as well as all the random effects of the model, corresponding to the solid black circles in the within-person model.

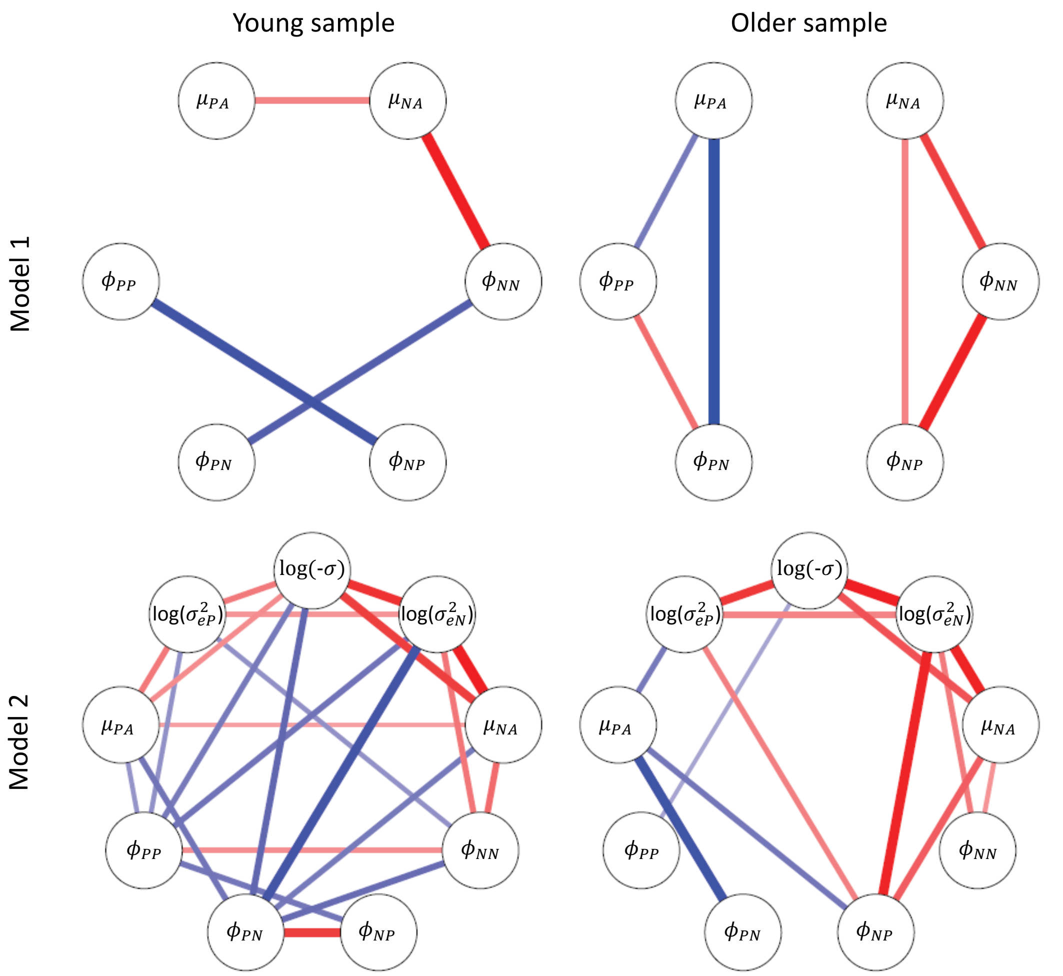
Figure 3. Network representations of the correlation structure of the random effects of the multilevel VAR(1) models in the two samples. Upper part contains the networks for model 1 based on six random effects, while the bottom part contains the networks based on model 2, which has nine random effects. Only correlations whose 95% credible interval did not include zero are included here. Red connections represent positive correlations, blue connections represent negative correlations. Strength of the correlations is represented by the thickness of the connection.

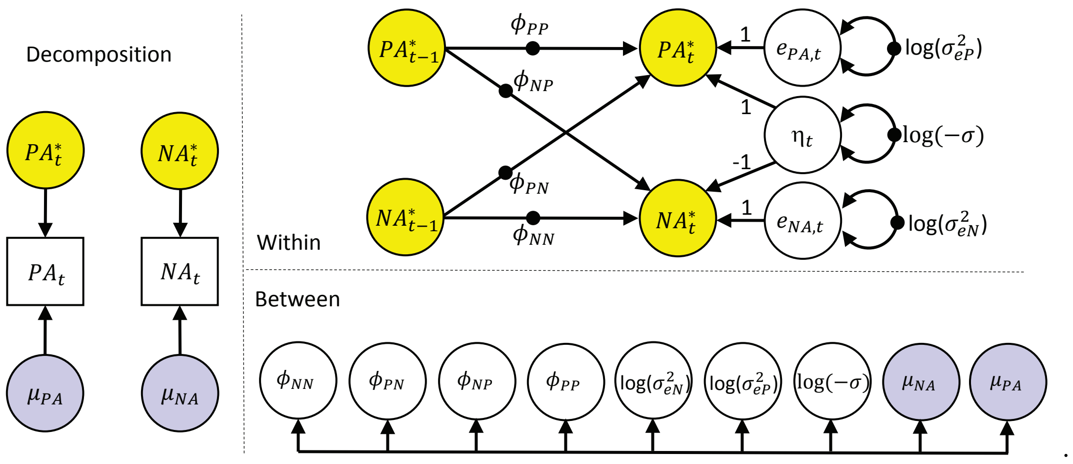
Figure 4. Representation of the multilevel VAR(1) model with random innovation variances and covariance. Left part contains the decomposition into within-person (time-varying) and between-person (time-invariant) components. Top right contains the within-person model, which is a VAR(1) model. Bottom right contains the between-person model, which includes the between-person components of the observed data as well as all the random effects of the model, corresponding to the solid black circles in the within-person model.


2.5
2
1.5
1
0.5
0
ICC=.68
ICC=.89
ICC=.64
ICC=.76
PA young
PA older
NA young
NA older
Total
Between
Within



Decomposition
PA_t^*
NA_t^*
PA_t
NA_t
μ_{PA}
μ_{NA}
Within
PA_{t-1}^*
NA_{t-1}^*
ϕ_{PP}
ϕ_{NP}
ϕ_{PN}
ϕ_{NN}
PA_t^*
NA_t^*
$\zeta_{PA,t}$
$\zeta_{NA,t}$
Between
ϕ_{NN}
ϕ_{PN}
ϕ_{NP}
ϕ_{PP}
μ_{NA}
μ_{PA}


Young sample
Older sample
Model 1
Model 2
μ_{PA}
μ_{NA}
ϕ_{PP}
ϕ_{NN}
ϕ_{PN}
ϕ_{NP}
log(-σ)
log(σ^2_{eP})
log(σ^2_{eN})


Decomposition
PA_t^*
NA_t^*
PA_t
NA_t
μ_{PA}
μ_{NA}
Within
PA_{t-1}^*
NA_{t-1}^*
ϕ_{PP}
ϕ_{NP}
ϕ_{PN}
ϕ_{NN}
PA_t^*
NA_t^*
1
1
-1
1
$e_{PA,t}$
η_t
$e_{NA,t}$
$\log(\sigma_{eP}^2)$
$\log(-\sigma)$
$\log(\sigma_{eN}^2)$
Between
ϕ_{NN}
ϕ_{PN}
ϕ_{NP}
ϕ_{PP}
$\log(\sigma_{eN}^2)$
$\log(\sigma_{eP}^2)$
$\log(-\sigma)$
μ_{NA}
μ_{PA}