Dynamic clustering

Ernst, Anja Franziska

DOI:
10.33612/diss.196176258

Publication date:
2022

Chapter 4

Mixture multilevel vector-autoregressive modelling

With the growing availability of intensive longitudinal data, the modelling approaches that handle such data become increasingly popular in the social sciences. These approaches reflect a shift in psychology towards an interest in the dynamic interplay between variables over time rather than their concurrent relationships. Particularly vector-autoregressive (VAR) models have been used to analyse emotion dynamics (e.g., Van der Krieke et al., 2017). In a VAR(1) model a vector of observations at a given time-point is regressed on the vector of observations at the previous time-point (Lütkepohl, 2005). VAR models provide insight into within-individual fluctuations, thus how one system state relates to a subsequent state. VAR models provided insight into the propensity of affective states to persist over time. This persistence is coined emotional inertia, and is often quantified as the autoregression of an emotion (Kuppens & Verduyn, 2017). Emotional inertia has been linked to psychological well-being and mental disorders (Brose et al., 2015a; Houben et al., 2015; Koval & Kuppens, 2012; Van Roekel et al., 2018).

In psychology, VAR models are typically used as single-subject model, thereby focusing on the dynamics of a single individual. Such single-subject models hold much promise for psychology, because they offer the possibility to tailor interventions and therapy, and to offer personalised feedback (e.g., Fisher & Boswell, 2016). However, individualized models still face issues that need to be overcome. That is, VAR models are often overfitted, thereby inadequately characterising the individual the model was uniquely fitted to (Bulteel et al., 2018). Further, they have been shown to be under-powered with the number of measurement occasions that are feasible in clinical practice (i.e., 75 – 100) (Mansueto et al., 2020). Also, VAR models have been found to differ markedly between individuals (Brose et al., 2015b; Hamaker et al., 2016), rendering comparisons between individuals difficult. Commonly, researchers want to generalise results to a population of individuals, and/or pool data of several individuals to assess how an individual compares to others. To this end a multisubject model is needed that integrates information across several subjects, while accounting for inter-individual differences in within-individual processes. So far two of such extensions have been made to the VAR model: multilevel VAR and latent class VAR models.

4.0.1 Multilevel VAR — accounting for continuous inter-individual differences

In the multilevel VAR model, inter-individual differences in the VAR coefficients are expressed through random effects. Here we use the term multilevel VAR (Bringmann et al., 2013), but different variants of the model are also known as mixed-effects VAR (Gorrostieta et al., 2012), hierarchical dynamic model (Driver & Voelkle, 2018), and dynamic multilevel model (Jongerling et al., 2015). Illustrations of the multilevel VAR model can be found

This chapter is currently submitted for publication as Ernst, A. F., Timmerman, M. E., Ji, F., Jeronimus, B. F., & Albers, C. J. (Submitted-b). Mixture multilevel vector-autoregressive modelling.

in Rovine and Walls (2006) and Schuurman et al. (2016). Kuppens et al. (2010) used this model to show that there is a relation between inertia and individual difference variables of psychological maladjustment, such as depression and neuroticism.

In the multilevel VAR, *random coefficients* (i.e., of means, autoregression coefficients, and cross-lagged regression coefficients) account for quantitative inter-individual differences. These random coefficients are specific for every individual and consist of a *fixed effect* and a *random effect* (e.g., Snijders & Bosker, 2011b). Such a fixed effect is the population average (mean or regression) coefficient across individuals, while a random effect represents the individual deviation from this average coefficient. Random coefficients are predicted as a weighted average of the population average coefficient and the individual's ideographic coefficients, meaning the individuals' random coefficients, are pulled (shrunken) towards the population average (see Section 1.6 in De Leeuw & Meijer, 2008). Estimates from the population are thus pooled, whilst the model accounts for quantitative inter-individual differences. The latter is often done based on the assumption of multivariate normally distributed random effects.

While random coefficients can capture some of the unobserved heterogeneity among dynamic processes of various individuals, some of the heterogeneity among individuals seems to be caused by discrete and qualitatively different processes (i.e., Carstensen et al., 2000; Hamaker et al., 2016; Hay & Diehl, 2011; Houben et al., 2015; Lane et al., 2019). In such cases, generalizability can be achieved only by identifying the subgroups of individuals that exhibit qualitative differences in dynamic processes (Voelkle et al., 2014). Random effects cannot account for such discrete unobserved heterogeneity. Discrete unobserved heterogeneity is accounted for by inferring latent subgroups from the data using exploratory methods, this stratifies the heterogeneous sample into homogeneous subgroups (Lubke & Muthén, 2005).

4.0.2 Latent class VAR — accounting for discrete inter-individual differences

Recently, latent class VAR models have been proposed to account for discrete unobserved heterogeneity in intensive longitudinal data (e.g., Anderlucci & Viroli, 2015; Ernst et al., 2020); for a review see Ernst et al. (2021). These models infer latent subgroups from the data in an exploratory fashion by incorporating a mixture model into the VAR model. The aim is to uncover subgroups with distinct patterns in means, autoregressions, and cross-regressions. In the context of a mixture model, subgroups are denoted components (McLachlan & Peel, 2004). The components themselves are unknown, and both the components and which component an individual belongs to needs to be inferred from the data. The latent class VAR models that have been proposed so far for psychological data, estimate only fixed-coefficients for all individuals in a component (e.g., Anderlucci & Viroli, 2015; Ernst et al., 2020). Thus, these models rely on the overly restrictive assumption that all individuals belonging to a single component can be modelled through identical parameters.

4.0.3 Mixture multilevel VAR — accounting for continuous and discrete inter-individual differences

So far, multilevel VAR models and latent class VAR models have been used to analyse multisubject intensive longitudinal data. In this paper, we propose a model that combines these two models, to account for individual differences while deriving generalisable results. We refer here to the proposed methodology as mixture multilevel vector-autoregressive modelling (MMVAR). By expressing the random coefficients for the individual's parameters as a mixture, the fixed effects correspond to the average random coefficients in a component (i.e., across individuals belonging to a component, per parameter), and the random effects represent the individual's deviation from their component mean.

4.1 Model specification

To present the MMVAR model, we consider an example with a single time lag (i.e., a VAR(1)), and two variables: positive affect (PA) and negative affect (NA). Model extensions to higher time lags (i.e., VAR(p), with p the number of lags), and to higher numbers of variables are straightforward to specify, yet can be cumbersome to estimate (see section MMVAR model specification and estimation considerations). The MMVAR has a within-person part (at level 1), and a between-person part (at level 2), with the mixture defined at level 2.

The model at level 1 is as follows:

Level 1:

$$\begin{pmatrix} y_{PA\,i\,t} \\ y_{NA\,i\,t} \end{pmatrix} = \begin{pmatrix} \mu_{PA\,i} \\ \mu_{NA\,i} \end{pmatrix} + \begin{pmatrix} w_{PA\,i\,t} \\ w_{NA\,i\,t} \end{pmatrix} \tag{4.1}$$

$$\begin{pmatrix} w_{PA\,i\,t} \\ w_{NA\,i\,t} \end{pmatrix} = \begin{pmatrix} \phi_{11\,i} & \phi_{21\,i} \\ \phi_{12\,i} & \phi_{22\,i} \end{pmatrix} \begin{pmatrix} w_{PA\,i,\,t-1} \\ w_{NA\,i,\,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{PA\,i\,t} \\ \varepsilon_{NA\,i\,t} \end{pmatrix}. \tag{4.2}$$

In Equation 4.1, $y_{PA\,i\,t}$ and $y_{NA\,i\,t}$ represent the observed scores of PA and NA respectively, for individual i ($i = 1,\ldots,N$) at time t ($t = 1,\ldots,T_i$);[1] $\mu_{PA\,i}$ and $\mu_{NA\,i}$ represent the within-person means for individual i, thereby expressing the latent trait scores; $w_{PA\,i\,t}$ and $w_{NA\,i\,t}$ represent the deviations from the within-person means at time t, thereby expressing the individual dynamics. In Equation 4.2, these dynamics are modelled further, where $\phi_{11\,i}$ and $\phi_{22\,i}$ represent the autoregression coefficients that reflect emotional inertia of PA and NA for individual i; $\phi_{12\,i}$ and $\phi_{21\,i}$ represent cross-regression coefficients that reflect the associations of PA and NA between the current time-point t and the previous time-point for individual i; these auto-and cross-regressions are collected in the VAR matrix Φ_i; $\varepsilon_{PA\,i,\,t}$ and $\varepsilon_{NA\,i,\,t}$ represent innovations, also called white noise error.

The model at level 2 includes a mixture of components that governs inter-individual differences in trait scores and dynamics. Every individual, i, is assumed to belong to exactly one of K mixture components ($k = 1,\ldots,K$); components memberships are given by a latent indicator variable of component membership $z_{i\,k} = 1$ if i belongs to component k and 0 otherwise. Components memberships are thus unknown a priori. That is, at level 2, we include a mixture of K components, including fixed and random effects, as follows

$$\begin{pmatrix} \mu_{PA\,i} \\ \mu_{NA\,i} \\ \phi_{11\,i} \\ \phi_{21\,i} \\ \phi_{12\,i} \\ \phi_{22\,i} \end{pmatrix} \sim MVN\left(\begin{pmatrix} \gamma_{PA\,k} \\ \gamma_{NA\,k} \\ \gamma_{11\,k} \\ \gamma_{21\,k} \\ \gamma_{12\,k} \\ \gamma_{22\,k} \end{pmatrix}, \begin{pmatrix} \tau_{11\,k} & & & & & \\ \tau_{12\,k} & \tau_{22\,k} & & & & \\ \tau_{13\,k} & \cdots & \tau_{33\,k} & & & \\ \tau_{14\,k} & \cdots & \cdots & \tau_{44\,k} & & \\ \tau_{15\,k} & \cdots & \cdots & \cdots & \tau_{55\,k} & \\ \tau_{16\,k} & \cdots & \cdots & \cdots & \cdots & \tau_{66\,k} \end{pmatrix}\right), \tag{4.3}$$

$$\begin{pmatrix} \varepsilon_{PA\,i\,t} \\ \varepsilon_{NA\,i\,t} \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11\,k} & \\ \sigma_{12\,k} & \sigma_{22\,k} \end{pmatrix}\right), \tag{4.4}$$

where $\gamma_{.\,k}$ represents a fixed effect for component k, and the random coefficients of individuals in this component are multivariate normally distributed with mean zero, and covariance matrix Ψ_k; the elements of Ψ_k, $\tau_{.\,k}$, represent the variances (diagonal elements) and covariances (off-diagonal elements) between random effects within component k; further, components have mixing proportions π_k where $\sum_{k=1}^{K} \pi_k = 1$. In an unconstrained model the covariances

[1] We assume a pre-sample containing the previous observations is available, as common in VAR(1) notation.

are estimated for every component; useful constraints for this covariance matrix, to facilitate estimation and interpretation, are considered in the next section.

Importantly, all expected values of innovations and random effects stated in the assumptions are conditional on the predictor. Additionally, the model relies on the following assumptions: Innovations have a mean of zero, within and between individuals, and for individuals with membership in component k they are multivariate normally distributed with covariance matrix Σ_k that contains the variances $\sigma_{11\,k}$ and $\sigma_{22\,k}$ and the covariance $\sigma_{12\,k}$. Other specifications and constraints for this covariance matrix are discussed in the section below. Innovations and random effects are assumed to be independent, of each other, and of the predictor. The model further relies on equidistant time intervals for all individuals. The means and covariance structure of the outcome variables are assumed constant over time for each individual (stationarity, i.e., the reverse characteristic polynomial of the matrix containing the individual's autoregressive and cross-lagged regression parameters has no roots in or on the complex unit circle (Lütkepohl, 2005)). Thus, we assume that there are no trends or cycles present in these data over time.

Component memberships, indicated by z_i, are unknown and need to be predicted via posterior probabilities, yielding a probabilistic clustering of the N individuals into the K components. For details on mixture models see McLachlan and Peel (2004), for details on multilevel mixture models see Asparouhov and Muthén (2008), for details on their estimation procedure see Muthén and Shedden (1999).

4.2 MMVAR model specification and estimation considerations

We highlight here three main issues that should be considered in the model specification and the estimation of a MMVAR model:

(a) Constraints can aid estimation: A model with a large number of parameters may be difficult to estimate on finite sample sizes; a solution may be to increase the sample size, and/or to reduce the number of parameters to estimate. If the latter approach is used, then the parameters to reduce should be chosen such that the resulting model still fits reasonably well. In an unconstrained MMVAR model the number of possible model parameters increases exponentially with the number of outcome variables. This can lead to convergence issues. Which parameters to estimate must depend on the research question and structure of the data.

(b) Equality constraints can aid interpretation: A model with many parameters may fit the data well and may be possible to estimate; yet, it might be that a simpler representation of the parameters can be achieved that still fits the data well, and is better to interpret. For instance, below we consider the possibility to employ equality constraints to create a fully crossed grouping of components that differ in trait scores and of components that differ in dynamics.

(c) Model specification influences interpretation: There may be different models that fit the data approximately equally well, yet have a different representation, and thus interpretation. Then it comes down to selecting the one that fits a priori knowledge best and is best to interpret. For example, below we contrast different model specifications to account for inter-individual differences in within-person means and to standardise coefficients in a multilevel model.

4.2.1 (a) Constraints can aid estimation

We consider constraints pertaining to the covariance matrix of the innovations, Σ_k, and to the covariance matrix of the random effects, Ψ_k. In standard multilevel VAR models (e.g., Bringmann et al., 2013; Schuurman et al., 2016), the innovation covariance matrix is fixed to be equal across all individuals in the population, whereas in a MMVAR model it may depend on the mixture (i.e., Σ_k). The innovations represent the influences that all factors that were not directly measured have on our two outcome variables. Such factors could be within-measurement effects of NA on PA (and vice versa), but they could also be external unmeasured factors, such as the weather, stress at the work place or at home. The exposure to variability in external factors and the response to external events is known to differ markedly across individuals (see e.g., Wichers et al., 2009). Therefore, an extension to the standard multilevel VAR model has been proposed where all elements of the innovation covariance matrix are modelled through random coefficients, thereby accounting for continuous inter-individual differences in innovations (Hamaker et al., 2018; Jongerling et al., 2015). These papers have shown that ignoring these individual differences might result in biased estimates of the innovation covariance matrix and the other model parameters.

In the MMVAR model specification in Equation 4.4, we account for inter-individual differences in innovation covariances via discrete coefficients. That is, we estimate several such covariance matrices (i.e., Σ_k, with $k = 1 \dots K$), one for every component of the mixture. While our model could be extended even further to account for continuous inter-individual differences within every component, we prefer the more parsimonious MMVAR specification here. Such an extension would increase the number of parameters dramatically, thereby likely yielding estimation issues in empirical practice.[2] We do not see this issue as hampering, as we expect that our discretely modelled inter-individual differences in innovations will typically cover the essential differences exhibited in the innovation parameters. Therefore, we expect the additional insights that would be gained through adding random effects for the innovation parameters to be minor.[3]

Further constraints on the innovations covariance, Σ_k, and the random effects covariance structure, Ψ_k, are also possible. For instance, the covariance between innovations of different outcome variables, $\sigma_{12\,k}$, and the covariance between the various random effects, $\tau_{12\,k}$, can be constrained to zero, when it seems reasonable to assume that these are linearly unrelated. Further, the covariance matrices Σ_k and Ψ_k might be constrained to be equal across components.

4.2.2 (b) Equality constraints can aid interpretation

Using equality constraints to create a fully crossed grouping of trait scores and dynamics

In our model specification, we consider mixtures that capture inter-individual differences in trait scores and dynamics. In an empirical application, the resulting grouping may be governed by the trait scores, the dynamics, or both. To advance the interpretability of a model, it may be useful to impose a distinct grouping with regard to the trait scores, and a distinct grouping with regard to the dynamics, where any relationship between the two is modelled via the mixing proportions (i.e., prior probabilities of component membership), π_k. This can be achieved by imposing equality constraints among a fully crossed combination of these groupings.

For instance, when assuming two distinct dynamic processes and two distinct trait scores in the population, a fully crossed grouping can be obtained as follows. One could estimate a mixture of four components, C1, C2, C3, and C4; herewith constrain the estimates for dynamics to be equal in C1 and C2, and in C3 and C4, further, constrain the estimates for the trait scores to be equal in C1 and C3, and in C2 and C4. Herewith, it is necessary to

[2] In our two variable example estimating these three additional random coefficients would increase the number of unique elements of only a single random effect covariance matrix, Ψ_k, from 21 to 45.

[3] In the software Mplus (see p. 757 in Muthén & Muthén, 2013), thus far random innovation variances can only be calculated for multilevel VAR models, and cannot be be calculated for multilevel mixture models.

constrain the covariances between the random effects for dynamics and the random effects for trait scores to be zero, because they lack a clear interpretation (i.e., $\tau_{13\,k}$, $\tau_{14\,k}$, $\tau_{15\,k}$, $\tau_{16\,k}$, $\tau_{23\,k}$, $\tau_{24\,k}$, $\tau_{25\,k}$, and $\tau_{26\,k}$; as defined in Equation 4.3). The estimated mixing proportions allow for an easy interpretation with regard to how often certain dynamic patterns co-occur with certain trait scores. An illustration of these constraints on empirical data is given in Appendix B.

4.2.3 (c) Model specification influences interpretation

The interpretation of the model parameters depends on the model specification that was employed during the model estimation. Here we discuss how to estimate the MMVAR model.

Accounting for inter-individual differences in within-person means

In Equations 4.1 and 4.2, the predictors are the deviations from the within-person means at time t, $w_{PA\,i,t-1}$ and $w_{NA\,i,t-1}$. In most multilevel software (e.g., lme; Bates et al., 2015), however, one cannot estimate the within-person mean simultaneously, while estimating a multilevel model. To tackle this issue, we consider two work-arounds to allow for inter-individual differences in within-person means nonetheless: (1) No centering (NC): fitting the level 1 model on the raw data, and (2) within-person centering (WPC): applying a two-step approach, firstly within-person center all lagged predictors, and secondly estimate the multilevel model on the centered data. These two specifications are simply reparametrizations of the identical underlying model, meaning that the two parameterizations describe exactly the same model.

(1) For NC, we use the lagged-1 outcomes, $y_{PA\,i,t-1}$ and $y_{PA\,i,t-1}$, as predictors. Thus $y_{PA\,i,t-1}$ and $y_{NA\,i,t-1}$ replace $w_{PA\,i,t-1}$ and $w_{NA\,i,t-1}$ in Equation 4.2:

$$y_{PA\,i\,t} = \mu^{NC}_{PA\,i} + \phi^{NC}_{11\,i} y_{PA\,i,t-1} + \phi^{NC}_{21\,i} y_{NA\,i,t-1} + \varepsilon^{NC}_{PA\,i\,t} \tag{4.5}$$
$$y_{NA\,i\,t} = \mu^{NC}_{NA\,i} + \phi^{NC}_{12\,i} y_{PA\,i,t-1} + \phi^{NC}_{22\,i} y_{NA\,i,t-1} + \varepsilon^{NC}_{NA\,i\,t}. \tag{4.6}$$

Then, $\mu^{NC}_{PA\,i}$ and $\mu^{NC}_{NA\,i}$ represent VAR intercepts rather than mean values. VAR intercepts lack a directly meaningful interpretation. This contrasts to the within-person means, $\mu_{PA\,i}$ and $\mu_{NA\,i}$, specified in Equation 4.1, which represents an individual's trait score. The lack of interpretability of the VAR intercepts poses problems beyond the interpretation of the coefficient. While the NC model and the WPC model are structurally equivalent (as shown in Appendix 1 of Hamaker & Grasman, 2015), they rely on different assumptions; the former assumes the random effects of the intercept to be normally distributed, while the latter assumes the individual trait scores to be normally distributed. If we wanted to include predictors at level 2, for instance, we want them to explain variance in the trait scores rather than variance in an arbitrary value, like the VAR intercept. Thus, it is advantageous to account for individual differences in trait scores through a coefficient that is interpretable, and for which we are comfortable forming particular distributional expectations. These properties hold for our second solution.

(2) For WPC, we first within-person center all lagged predictors based on the sample within-person mean, and subsequently use the within-person mean centered lagged-1 outcomes, $w_{PA\,i,t-1}$ and $w_{NA\,i,t-1}$, as predictors:

$$y_{PA\,i\,t} = \mu^{WPC}_{PA\,i} + \phi^{WPC}_{11\,i} w_{PA\,i,t-1} + \phi^{WPC}_{21\,i} w_{NA\,i,t-1} + \varepsilon^{WPC}_{PA\,i\,t} \tag{4.7}$$
$$y_{NA\,i\,t} = \mu^{WPC}_{NA\,i} + \phi^{WPC}_{12\,i} w_{PA\,i,t-1} + \phi^{WPC}_{22\,i} w_{NA\,i,t-1} + \varepsilon^{WPC}_{NA\,i\,t}. \tag{4.8}$$

Now the intercepts, $\mu_{PA\,i}^{WPC}$ and $\mu_{NA\,i}^{WPC}$, will be interpretable as an individual's trait score (see Hamaker & Grasman, 2015). In the following discussion on WPC it is important to distinguish between (A) standard multilevel models (i.e., a multilevel model that does not include a lagged version of the outcome variable as predictor) and (B) multilevel VAR models (i.e., a multilevel model that includes at least one predictor that is a lagged version of the outcome variable, see Hamaker & Grasman, 2015).

(A) In standard multilevel analysis, WPC[4] of level 1 predictors is generally advantageous whenever the within-person and the between-person slopes differ (i.e., when the relationship between outcome and predictor differs across the levels of the analysis). In these cases estimates of the coefficients based on NC predictors will be biased, because they will represent a mix of the within-person and the between-person slopes.[5] Because the WPC predictor contains only within-person variance, estimates of coefficients based on WPC predictors represent the within-person slope. In the case of standard multilevel models, WPC of the predictor variable thus solves the problem of disentangling within-and between-person slopes, and, because the WPC predictor does not contain any stable between-person differences, it also solves the endogeneity problem at level 2 (i.e., the within-person centered predictor cannot be correlated with the random effects (see Hamaker & Muthén, 2020)).

(B) In contrast to standard multilevel analysis, the favourable properties of WPC level 1 predictors do not hold for multilevel VAR models. Hamaker and Grasman (2015) showed that in the special case of a multilevel VAR model, WPC causes downward bias in the random slope for the autoregressive predictor. This bias does not occur for WPC in standard multilevel models.

In multilevel models, one assumption for obtaining consistent regression estimates is zero correlation between the random effects and the covariates (i.e., no level 2 endogeneity). In the case of a multilevel VAR model, including lagged outcomes (i.e., $y_{i,\,t-1}$) as predictors will violate this assumption due to the so-called initial condition problem (see e.g., Skrondal & Rabe-Hesketh, 2014; Wooldridge, 2005), where the first response (i.e., $y_{i,\,1}$) is endogenous by design when modelling $y_{i,\,2}$ due to the common unobserved heterogeneity (i.e., random effects).

In our current model setup, we took the WPC approach to avoid level 2 endogenity. However, this introduces a new level 1 endogeneity issue (i.e., the correlation between covariates and innovations), because the within-person mean and within-person error terms will be correlated due to the common error terms. This problem causes downward bias in the estimated autoregressive coefficients when the true autoregressive coefficient is positive; this is known as 'Nickell's bias' (Asparouhov & Muthén, 2019; Nickell, 1981). The original paper (Nickell, 1981), and multiple following papers (e.g., Bun & Carree, 2005) investigated the Nickell's bias analytically, and in Monte-Carlo studies (e.g., Hamaker et al., 2015a; Maddala, 1971; Nerlove, 1967, 1971). Many solutions have been proposed, for instance instrumental-variable based generalised method of moments estimators (Ahn & Schmidt, 1995; Anderson & Hsiao, 1982; Arellano & Bond, 1991; Bhargava & Sargan, 1983), although these estimators often do not perform well in finite samples. Other solutions to this problem have been proposed within the maximum-likelihood structural equation models (ML-SEM) framework in a series of recent papers (Allison et al., 2017; Moral-Benito, 2013; Moral-Benito et al., 2019); these ML-SEM estimators demonstrate improved finite-sample performance over the generalised method of moments estimators. Their practical use is limited, to the best of our knowledge ML-SEM estimators are currently only implemented in Stata and are also computationally more intensive (Williams et al., 2018). In psychology, Asparouhov et al. (2018)

[4]Often referred to as group-mean centering or cluster-mean centering in traditional multilevel models.

[5]These coefficient will only be biased if the within-person and between-person slopes differ. In multilevel VAR models these slopes will always differ (see, Hamaker & Grasman, 2015).

proposed a dynamic structural equation model that solves Nickell's bias by utilising latent mean centring.

In the current study we deem the WPC estimator acceptable to use, because the bias of the dependent lagged outcome tends to vanish quickly in the intensive longitudinal setting where the number of time-points is high, as has been shown by Asparouhov et al. (2018), Bun and Carree (2005), and Nickell (1981), and the same result is also seen in Simulation 1 below.

Standardisation in a multilevel model

The considerations on centering variables in a multilevel model, that were discussed in the section above, apply similarly to the standardisation in a multilevel model. For instance, to calculate coefficients that are standardised within-persons one could either standardise the regression parameters during the calculation of the coefficients, or one could standardise the individual data before the calculation. These two procedures are associated with different model assumptions about the distributions of the standardised and unstandardised parameters, and can consequently lead to different results (Schuurman et al., 2016). When standardising the regression parameters during the calculation of the coefficients, calculations are carried out on the unstandardised data, thus the individual unstandardised parameters are assumed to be normally distributed. Then, the standardised individual regression parameters will usually not be normally distributed (i.e., unless all individual data are standardised with the same constant). In contrast, when standardising the data before the calculation, one assumes that the standardised regression parameters, rather than the unstandardised regression parameters, are normally distributed.

In a multivariate VAR model, the model contains cross-lagged associations that quantify the lagged influence of one variable on another variable. Often researchers would like to compare the relative strength of associations of the various variables via the cross-lagged associations. To account for differences in the variable's measurement scales and variances, standardised coefficients should be employed for such comparisons. Even if all predictors are assessed on the same measurement scale, coefficients should be standardised for comparisons to account for differences in the predictors' variances.

Standardised and unstandardised coefficients represent proportions of *uniquely* explained variance. Because the predictors are dependent in VAR models, they will explain parts of the same variance in the outcome variable, and therefore it is impossible to determine which predictor has the strongest association with the outcome variable overall. Standardised coefficients can, however, indicate which predictor explains the largest unique variance in the outcome. In multilevel VAR models, various standardisation techniques are possible (for a comparison on within-person, between-person and grand standardisation techniques in multilevel VAR models see, Schuurman et al., 2016).

4.3 Simulations

All the scripts used to generate data, carry out the simulations and analyse the data are provided on the project's OSF page.[6]

[6] https://osf.io/cq248/?view_only=1a42ed6895524ec29e5656fd523d2d03

4.3.1 Simulation 1: The potential downward bias of auto-and cross-regressive slopes

To assess the issue of potential downward bias of the autoregressive and cross-regressive coefficients we carry out a simulation study. With this simulation we will establish the magnitude of the downward bias that can be caused by within-person centring or no centering. We thereby extend the study of Hamaker and Grasman (2015) to the multivariate case of the multilevel VAR model. We also investigate the relation between the downward bias and the number of time points.

We consider the effects of no centring (NC) and within-person centring (WPC) in a multilevel VAR model, both with fixed and random specifications of the autoregressive and cross-regressive coefficients. We are mainly interested to find the extent to which the bias that has been found by Hamaker and Grasman (2015) for WPC autoregressive slopes, will also pertain to WPC cross-regressive slopes. We thus add onto the results by Hamaker and Grasman (2015) in two ways, namely (1) to replicate their results for WPC autoregressive slopes, and (2) to extend their results to multilevel VAR models that include more than a single outcome and a single lagged predictor. To avoid needless complications, we neglect the mixture framework in this small simulation, and consider a standard multilevel VAR. We consider the effects of the following factors: (a) fixed vs. random coefficients, (b) the size of the cross-regressive slopes, and (c) the number of time-points. Our simulation was performed in R (R Core Team, 2020), multilevel VAR models were estimated with the lmer function from the R-package lme4 (Bates et al., 2015) in the fashion described by Bringmann et al. (2013).[7]

Design and procedure

To generate the data, the following factors were kept constant. We generate multilevel VAR(1) models each with two outcome variables and two lagged predictors. The fixed effect within-person slopes, γ_{11}, and γ_{22}, were equal to .3 for autoregressions. The level 2 variances of the within-person mean, τ_{11} and τ_{22}, equalled 3, the grand means of both variables equalled 0. The innovation variances were set to 3. Covariances of random effects and innovations were set to zero. Each generated data set contains 100 individuals. The following three factors were varied in a completely crossed design with 100 replications each: (a) the standard deviations of the individual within-individual slopes, τ_{33}, τ_{44}, τ_{55}, and τ_{66}, equalled 0 or .2; (b) the fixed effect within-individual slopes for the cross-regressions, γ_{21}, and γ_{12}, equalled 0 or .2; (c) the number of time-points equalled 20, 50 or 100.

All ($2 \times 2 \times 3 \times 100 =$) 1,200 resulting data sets were analysed with WPC and NC. For WPC, we fit two pairs of single outcome multilevel models using both within-person mean centered lagged-1 outcomes, $w_{PA\,i,\,t-1}$ and $w_{NA\,i,\,t-1}$, as predictors. Following the notation in Equations 4.1 and 4.2, the model to estimate the simulated data was:

$$y_{PA\,i\,t} = \mu^{WPC}_{PA\,i} + \phi^{WPC}_{11\,i} w_{PA\,i,\,t-1} + \phi^{WPC}_{21\,i} w_{NA\,i,\,t-1} + \varepsilon^{WPC}_{PA\,i\,t} \tag{4.9}$$
$$y_{NA\,i\,t} = \mu^{WPC}_{NA\,i} + \phi^{WPC}_{12\,i} w_{PA\,i,\,t-1} + \phi^{WPC}_{22\,i} w_{NA\,i,\,t-1} + \varepsilon^{WPC}_{NA\,i\,t}. \tag{4.10}$$

For NC, we fit two pairs of single outcome multilevel models using both lagged-1 outcomes on the original scales, $y_{PA\,i,\,t-1}$ and $y_{PA\,i,\,t-1}$, as predictors, estimating the following model:

$$y_{PA\,i\,t} = \mu^{NC}_{PA\,i} + \phi^{NC}_{11\,i} y_{PA\,i,\,t-1} + \phi^{NC}_{21\,i} y_{NA\,i,\,t-1} + \varepsilon^{NC}_{PA\,i\,t} \tag{4.11}$$
$$y_{NA\,i\,t} = \mu^{NC}_{NA\,i} + \phi^{NC}_{12\,i} y_{PA\,i,\,t-1} + \phi^{NC}_{22\,i} y_{NA\,i,\,t-1} + \varepsilon^{NC}_{NA\,i\,t} \tag{4.12}$$

[7]That is, multiple outcome variables are accounted for by estimating a separate multilevel equation for each outcome variable.

Subsequently, per condition, the bias was estimated for the autoregressive and cross-regressive slopes separately. In the following we take WPC as example, but the following applies equally to the NC estimates. The parameters of interest are the fixed effects $\gamma_{11}^{WPC}, \gamma_{12}^{WPC}, \gamma_{21}^{WPC}, \gamma_{22}^{WPC}$ that correspond to the random slopes $\phi_{11\,i}^{WPC}, \phi_{12\,i}^{WPC}, \phi_{21\,i}^{WPC}, \phi_{22\,i}^{WPC}$. Bias was estimated by averaging across the 100 replications within a simulation condition (e.g., $\widehat{\text{Bias}}(\gamma_{11}^{WPC}) = \frac{1}{100}\sum_{r=1}^{100} \hat{\gamma}_{11r}^{WPC} - \gamma_{11}$). Because the two single outcome multilevel models represented in Equations 4.9 and 4.10 are symmetric, we denote the average of $\widehat{\text{Bias}}(\gamma_{11}^{WPC})$ and $\widehat{\text{Bias}}(\gamma_{22}^{WPC})$ as the bias estimate for the *autoregressive* coefficients. Similarly, we denote the average of $\widehat{\text{Bias}}(\gamma_{12}^{WPC})$ and $\widehat{\text{Bias}}(\gamma_{21}^{WPC})$ as the bias estimate for the *cross-regressive* coefficients. To quantify the uncertainty of the bias estimates due to the limited number of replications, we calculated Monte Carlo errors using equation 7 in Koehler et al. (2009). Just like the autoregressive and cross-regressive estimates, the Monte Carlo errors are displayed as averages across the two variables *PA* and *NA*. All results are multiplied by 1,000 and included in Table 4.1.

Results

The following conclusions can be drawn from Table 4.1.

1. Similar to Hamaker and Grasman (2015), WPC yields autoregressive coefficients that are underestimated across different conditions; cross-regressive coefficients seem biased for WPC only when there is variation for the random slopes.
2. For NC, autoregressive coefficients seem unbiased; in contrast, cross-regressive coefficients are underestimated when there is variation for the random slopes.
3. The bias of cross-regressive coefficients are in general more severe for NC than for WPC; the bias tends to increase with the variance of the random slopes for NC and WPC.
4. For WPC and NC, the biases discussed above for autoregressive and cross-regressive coefficients reduce when the number of time-points increases.

Table 4.1: Simulation results.

	Bias (Monte Carlo Error) × 1000					
	Autoregressive			Cross-regressive		
Estimator	$T = 20$	$T = 50$	$T = 100$	$T = 20$	$T = 50$	$T = 100$
$\tau_{33} = \tau_{44} = \tau_{55} = \tau_{66} = 0$, $\gamma_{21} = \gamma_{12} = 0$						
NC	−0.025(2.335)	−0.969(1.385)	−0.667(0.865)	0.502(2.246)	−2.347(1.320)	1.265(0.932)
WPC	−75.997(2.172)	−28.412(1.355)	−13.966(0.855)	0.104(2.423)	−2.440(1.379)	1.210(0.940)
$\tau_{33} = \tau_{44} = \tau_{55} = \tau_{66} = 0$, $\gamma_{21} = \gamma_{12} = .2$						
NC	7.591(3.335)	−2.331(2.481)	−2.444(2.073)	0.040(2.886)	0.032(2.164)	−0.497(2.121)
WPC	−70.022(3.007)	−28.485(2.407)	−14.528(2.052)	0.308(3.121)	0.039(2.229)	−0.510(2.153)
$\tau_{33} = \tau_{44} = \tau_{55} = \tau_{66} = .2$, $\gamma_{21} = \gamma_{12} = 0$						
NC	−6.131(2.370)	−2.984(1.555)	−1.939(1.059)	−32.912(2.096)	−15.129(1.292)	−6.999(0.873)
WPC	−77.226(2.200)	−28.481(1.510)	−14.229(1.048)	−15.436(2.226)	−6.436(1.316)	−2.488(0.881)
$\tau_{33} = \tau_{44} = \tau_{55} = \tau_{66} = .2$, $\gamma_{21} = \gamma_{12} = .2$						
NC	−6.267(3.168)	−4.586(2.441)	−3.381(2.102)	−29.804(2.645)	−13.820(2.297)	−7.925(2.185)
WPC	−78.731(2.927)	−29.158(2.394)	−14.603(2.084)	−14.420(2.822)	−6.430(2.340)	−4.240(2.225)

4.3.2 Simulation 2: Performance of MMVAR model estimation

We will assess the performance of MMVAR in case of correct model specification under various relevant conditions through the following simulation. Performance will be evaluated with regard to recovery of the model parameters, and the component memberships. We will explore to what extent the recovery depends on two factors. One factor pertains to the partitioning of individuals with respect to their VAR coefficients: (a) the distance of VAR coefficients between components relative to the distance within components. The other factor pertains to the efficiency: (b) the number of observations per person. The simulation was carried out in R (R Core Team, 2020), the data was analysed in Mplus version 8 (Muthén & Muthén, 2013). For details on the estimation procedure of the multilevel mixture see Muthén and Shedden (1999).

Design and procedure

Data were generated according to a MMVAR model, as defined in Equations 4.1, 4.2, 4.3 and 4.4. The following data characteristics were kept constant. We used two variables, one time lag, a total number of individuals of $N = 100$, and four mixture components. Individuals were equally distributed over these components; every component contained thus 25 individuals. In line with the equality constraints that were discussed in the MMVAR model specification and estimation considerations section, each of the four components had identical fixed effects and random effect variances on within-person means to one other component (i.e, $\gamma_{PA\,k}$, $\gamma_{NA\,k}$, $\tau_{11\,k}$, and $\tau_{22\,k}$ were equal for C1 and C2, and were equal for C3 and C4). Also each component shared identical fixed effects and random effect variances on VAR coefficients with yet another component (i.e., $\gamma_{11\,k}$, $\gamma_{21\,k}$, $\gamma_{12\,k}$, $\gamma_{22\,k}$, $\tau_{33\,k}$, $\tau_{44\,k}$, $\tau_{55\,k}$, and $\tau_{66\,k}$ were equal for C1 and C3, and were equal for C2 and C4).[8] This created a fully crossed grouping of two distinct trait score patterns, and two distinct dynamic patterns across the four components. These equality constraints were imposed on the parameters during the data generation, as well as during the model estimation. The fixed effects of the within-person means equalled $\left(\begin{smallmatrix} 10 & 25 \end{smallmatrix}\right)^\top$ and $\left(\begin{smallmatrix} 12 & 23 \end{smallmatrix}\right)^\top$. The variances of the random effects for within-person means (i.e., τ_{11}, τ_{22}) equalled 0 in every component. With regard to within-person means, components were thus very well separated. In all components the variance of the random effects for individual VAR coefficients (i.e., τ_{33}, τ_{44}, τ_{55}, τ_{66}) equalled .01. All covariances between random effects equalled zero. The covariance matrix of the white noise series, Σ_k, equalled $\boldsymbol{I}_2 + .5$ in every component.

Two factors, distance and observations, were varied in a completely crossed design.

(a) Within component distance — expressed through the variances of the VAR coefficients — was kept fixed; between component distance — expressed through the mean absolute difference between fixed effect VAR coefficients of the different components — was varied at two levels:

- ◇ *small distance*, where the mean absolute difference between the two fixed VAR matrices, $\Phi_{k=1}$ and $\Phi_{k=2}$, equalled .23, resulting in an average misclassification probability of .02 (misclassification probability is determined based on the overlap between components on individuals' true VAR coefficients; for details see Melnykov, 2012);
- ◇ *large distance*, where the mean absolute difference between the two fixed VAR matrices, $\Phi_{k_2=1}$ and $\Phi_{k_2=2}$, equalled .28, resulting in an average misclassification probability of .005;

In every condition the fixed coefficients for one VAR matrix equalled $\begin{pmatrix} .5 & -.2 \\ -.2 & .5 \end{pmatrix}$. Four equal deviations were added to this matrix to create the fixed VAR coefficients

[8] By generating data according to these equality constraints we tested whether the estimation procedure detects component differences in both within-person means and VAR coefficients.

for the other dynamic component, according to the required distance in that condition. Random coefficient VAR matrices for all individuals in the condition were then generated by drawing random samples from the normal distribution with mean equal to the fixed coefficients, and variance of the random effects of VAR coefficients. The resulting random VAR matrices were then used to generate individuals' time-series. All random VAR matrices were checked to be stationary with their reverse characteristic polynomial having no roots in or on the complex circle, and were re-generated until they met this condition.

(b) The number of observations per person: 75 or 125.

For each of the resulting 4 conditions we simulated 10 data sets. The estimation of the MMVAR model with the equality constraints was initialised with 10 random starts. The best-fitting solution across these starts was indicated by the log likelihood.

Outcome measures

To evaluate the performance of our MMVAR estimation we investigated (1) estimation issues, (2) the sensitivity to local minima, and (3) computation time. To evaluate the quality of the parameter estimates we investigated the recovery of (1) the fixed effect within-person means, (2) the fixed effect VAR coefficients, (3) the random effect variances of within-person means, (4) the random effect variances of VAR coefficients, (5) the mixing proportions, and (6) the component membership by the crisp component predictions.

Parameter recovery was quantified through mean absolute differences (MAD) between true and estimated effects for the within-person means, the VAR coefficients, the variance of the random effects, and the mixing proportions. For within-person means and VAR coefficients, the recovery of the fixed effects is presented below; the recovery of the variance of the random effects and the mixing proportions is presented in Appendix A. Recovery of component membership was assessed with the adjusted Rand index (ARI; Hubert & Arabie, 1985) between the true partitioning of individuals and the modal estimated crisp prediction into four mixture components. The ARI takes on a value of 1 in case of perfect classification recovery and 0 when the classification recovery could be expected by chance; according to Steinley (2004), ARIs above .80 indicate good recovery, ARIs between .65 and .80 indicate moderate recovery. Effect sizes of the simulation factors on the recovery of parameters were determined with an ANOVA including both simulation factors and their two-way interaction as predictors, and MAD or ARI values as outcome variable. To quantify the uncertainty of our simulation study due to a finite number of replications we calculated Monte Carlo errors of the fixed effect within-person means and the fixed effect VAR coefficients for each simulation condition. Monte Carlo errors provide insight into between-simulation variability (Koehler et al., 2009).

Results

Estimation issues, sensitivity to local minima and computation time In 22 of the 40 data sets (55%) the likelihood value of the best fitting solution was not replicated among the 9 remaining starts. This suggests that many of the final solutions may represent local minima rather than global minima. This could be alleviated by increasing the number of random starts. We have selected such a limited number of random starts to keep the substantial computation time within feasible limits. In the current simulation, Mplus took on average 21.10 hours ($SD = 7.27$) per data set, running in parallel on a 64x based computer with 8 GB RAM. In 7 of the 40 data sets (17%), the likelihood value of the best fitting solution was indicated as a saddle point. For two data sets the final solution included empty components (both in conditions with 75 observations per person) which can be indicative of estimation problems (see e.g. pp. 433-435 in Bishop, 2006); the results from these two data sets were excluded in the following analyses.

Recovery of within-person means Table 4.2 lists the mean MADs per factor level, Figure 4.1a shows the MADs per simulation condition. Table 4.3 displays an ANOVA comparing the within-person mean MADs across simulation factors. The effects were all non-significant suggesting there are none to only small differences in recovery across simulation conditions. Given that fixed effect within-person means had a distance of 2 between components with random effect variance of 0, the MADs ($Median = .092$, $IQR = [.051, .205]$) indicate they were recovered well across simulation conditions.

Recovery of VAR coefficients Table 4.2 shows the mean MADs per factor level, Figure 4.1b displays the MADs per simulation condition. Effect sizes in Table 4.3 highlight particularly the interaction between distance and observations ($\hat{\eta}_p^2 = .114$); this indicates that a high number of observations compensates for a small distance between components, and vice versa. Fixed effect VAR coefficients were recovered well with small differences in MADs across simulation conditions ($Median = .020$, $IQR = [.016, .024]$).

Recovery of random effect variances The recovery of random effect variances of the within-person means and the VAR coefficients is presented in Appendix A. The estimated random effect variances are shown in Figure 4.7, and their recovery in Figure 4.6. The recovery of the random effect variances mirrors the recovery of their corresponding fixed effects, as can be seen by comparing Figure 4.7 and Figure 4.1. These figures suggest that the estimates for the variances are very close to their true data generating value, they are, however, consistently underestimated.

Recovery of component memberships In contrast to the recovery of the specific parameters that were discussed thus far, the recovery of component membership reflects the recovery across all parameters. Table 4.2 shows the mean ARIs per factor level, Figure 4.2 displays the ARIs per simulation condition. The effect sizes in Table 4.4 suggest that a high number of observations ($\hat{\eta}_p^2 = .321$) and large distances between components ($\hat{\eta}_p^2 = .446$) facilitated membership recovery. On average, ARIs indicate component membership were retrieved well ($M = .838$, $SD = .087$), with medium differences across simulation conditions, as indicated by the ANOVA.

Recovery of mixing proportions The recovery of mixing proportions is shown in Figure 4.8 in Appendix A. Comparing Figure 4.8 to Figure 4.2 shows that the recovery of mixing proportions closely mirrors the recovery of component memberships.[9]

Monte Carlo error of fixed effect within-person means and fixed effect VAR coefficients Table 4.5 shows the Monte Carlo errors (MCEs) associated with the parameter estimates for the fixed effect within-person means and the fixed effect VAR coefficients. In the table we present the average MCEs over all MCE values within a replication (i.e., averaged across all variables and across all components within a replication). The MCEs give an indication

[9]When overlap between components is substantial the recovery of component memberships does not give a good indication of the estimation performance anymore, in those cases the estimates of the mixing proportions should be preferred. Because the components in our simulation were well separated, we focus here on the component memberships.

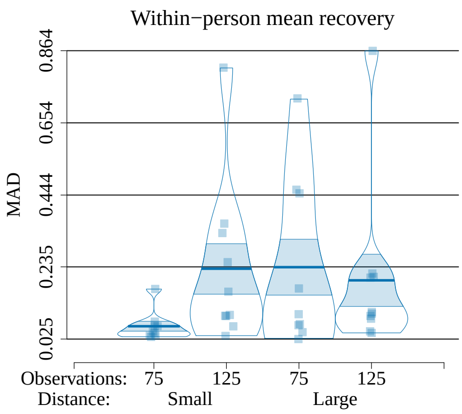
of the between-simulation variability of the parameter estimates that were obtained with our simulation design.

Table 4.2: MAD and ARI means (*SD*) across simulation factors. WPM-MAD denotes the MADs between estimated and true fixed effect within-person means, VAR-MAD the MADs between estimated and true fixed effect VAR coefficients, ARI the adjusted Rand index between predicted and true component membership. For each factor and outcome, the level with the best recovery is highlighted in bold.

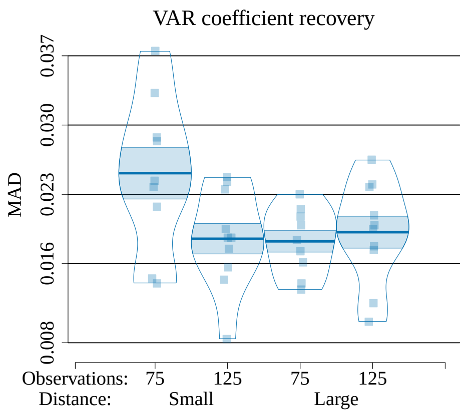
Factor	Level	WPM-MAD	VAR-MAD	ARI
Observations	75	**.148** (.195)	.022 (.007)	.796 (.088)
	125	.213 (.233)	**.019** (.005)	**.875** (.069)
Distance	Small	**.150** (.189)	.022 (.007)	.787 (.075)
	Large	.214 (.240)	**.019** (.004)	**.889** (.067)

Table 4.3: ANOVA results on MADs between estimated and true within-person means (WPM) and on MADs between estimated and true VAR coefficients. Any effect size above .1 is highlighted in bold.

	WPM				VAR coefficients			
	df	*F*	*p*	$\hat{\eta}_p^2$	*df*	*F*	*p*	$\hat{\eta}_p^2$
Observations	1	.871	.357	.025	1	2.490	.124	.068
Distance	1	.847	.364	.024	1	2.610	.115	.071
Observations × Distance	1	2.210	.146	.061	1	4.380	.044	**.114**



(a)



(b)

Figure 4.1: Parameter recovery for the fixed parameters across simulation conditions. A line indicates the mean MAD, a coloured band the region within one standard error of the mean.

Table 4.4: ANOVA results on the ARIs. Any effect size above .1 is highlighted in bold.

	df	*F*	*p*	$\hat{\eta}_p^2$
Observations	1	16.100	$<.001$	**.321**
Distance	1	27.400	$<.001$	**.446**
Observations $\times$ Distance	1	.075	.785	.002

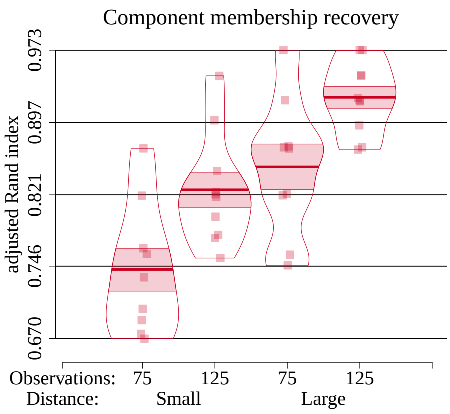


Figure 4.2: Component membership recovery across simulation conditions. A line indicates the mean ARI, a coloured band the region within one standard error of the mean.

Table 4.5: Average Monte Carlo error of the fixed effect within-person means (WPM) and the fixed effect VAR coefficients, averaged across replications and all fixed effect estimates within a replication.

	WPM		VAR coefficients	
	Small	Large	Small	Large
75	.030	.130	.009	.007
125	.100	.120	.007	.006

4.4 Application: COGITO study

4.4.1 Methods: The model to be estimated

To demonstrate the applicability of our proposed model, we will analyse the data from the COGITO study (for details see Schmiedek et al., 2010). This study tracked the emotions of 204 individuals across about 100 daily measurements. We analyse the data of a young sample (N = 101, age range: 20–31) and an older sample (N = 103, age range: 64–80) jointly, using the MMVAR model. The age differences in this sample result in inter-individual differences in emotion dynamics and trait scores (Hamaker et al., 2018). We will not include age in the model estimation — in this way, we demonstrate the advantage of accounting for discrete inter-individual heterogeneity by identifying components in an exploratory fashion. Measurements were made once a day. If the gap between measurements was longer than a day, missing values were inserted and filled with multiple imputation using the Amelia R package (Honaker et al., 2011), prior to estimating a MMVAR model. Here we assume data to be missing at random, even though measurements were not made on Sundays and holidays. Because we are only interested in illustrating the MMVAR model, and we expect the possible biasing effects to be rather small, we deem this missing data procedure as sufficient. The number of non-missing observations lies between 87 and 109 (M = 100.91, SD = 2.64). These data have been analysed with a multilevel VAR model by Hamaker et al. (2018).

As outcome variables in our MMVAR model, we used positive affect (PA) and negative affect (NA), each measured as the average score of 10 items from the Positive Affect Negative Affect Schedule (PANAS, Watson et al., 1988); PA and NA were assessed on a continuous scale from 0 (does not apply at all) to 7 (applies very well) (Brose et al., 2015a). In our analysis of this data, all lagged predictors (i.e., the lagged version of PA and NA) are within-person centred. This implies that the intercepts can be interpreted as individuals' trait scores. In total, we estimated four MMVAR models, as defined in Equations 4.1, 4.2, 4.3 and 4.4. Three models were specified without equality constraints, estimating two to four components. One model was specified with equality constraints as described in the MMVAR model specification and estimation considerations section; consequently, each of four components was constrained to have identical fixed effects and random effect variances on within-person means to one other component, and each component shared identical fixed effects and random effect variances on VAR coefficients with yet another component (henceforth, the 2×2 model).

For each model we used 50 starts — one based on an initialisation and 49 based on random perturbations of this initialisation.[10] Initialisations for the models were achieved in the following ways. For the constrained model, we used the multilevel VAR results on the COGITO data that were reported in Hamaker et al. (2018), where age group was used as a grouping variable instead of component memberships. For the unconstrained models, we employed results achieved by fitting a latent class VAR model (Ernst et al., 2020) with the corresponding number of latent classes to the data; for this we used the LCVAR R-function.[11] In all models the covariances between random effects were constrained to zero (i.e., the off-diagonal elements in Ψ_k defined in Equation 4.3).

All Mplus analysis scripts, including initialisations, model specifications, and results are supplied on the project's OSF page.[12]

4.4.2 Results

For each model, the final solution across all starts was selected based on the likelihood value. The log likelihood value, and Akaike's and Bayesian information criterion (AIC and BIC respectively) for each final model are shown in Table 4.6. The model for the unconstrained 4 component model did not converge, possibly because of a high number of parameters. In the final solutions for the 3 component model and the 2×2 model, the likelihood value of the best fitting solution was indicated as a saddle point. We encountered similar problems in Simulation 2, where for 17% of the data sets the final solution that was indicated as a saddle point. We might have encountered these estimation problems because in our empirical data set many participants displayed only very small variation in their NA scores.

Because the information criteria in Table 4.6 do not unanimously favour one model we selected to present the 3 component solution in the section below, because this solution gives us the most parameters to interpret. The 2×2 component solution is presented and interpreted in detail in Appendix B.

[10]For the unconstrained three component model a high proportion of the starts resulted in local minima/convergence problems, because of the high number of parameters. To increase the number of converged starts we used the best solution after 35 starts as a second initialisation for the following 21 starts with different convergence criteria. For details see the corresponding Mplus input files on the project's OSF directory.

[11]https://github.com/AnieBee/LCVAR

[12]https://osf.io/cq248/?view_only=1a42ed6895524ec29e5656fd523d2d03

Table 4.6: The log likelihood value of the final solution for each MMVAR model, with the associated information criteria values, and the number of free parameters that were estimated in the model.

Number of Components	Free parameters	Log likelihood	AIC	BIC
2 unconstrained	31	$-43,278$	86,618	86,878
2×2 constrained	39	$-41,884$	83,847	84,173
3 unconstrained	47	$-41,865$	83,824	84,218
4 unconstrained		no convergence		

4.4.3 The 3 component solution

The estimated component-specific MMVAR parameters of the model with three components are presented in Table 4.7. The fixed effect within-person means shown in this table, which we interpret as average trait scores, are graphically presented in Figure 4.3a, and the fixed effect VAR coefficients in Figure 4.3b. The confidence intervals (CIs) for the fixed effects are shown in Table 4.7. All these CIs exclude zero, except for the CIs for both cross-regressive coefficients in Component 1, and the cross-regressive coefficient from NA to PA in Component 3. Across components cross-regressive coefficients are negligibly small in both the 3 component and in the 2×2 component solution. Similar results were found with latent class VAR models on a comparable data set (Ernst et al., 2020). We interpret this in keeping with the notion that cross-regressive effects that are reported in single-individual VAR models are often the result of overfitting (Bulteel et al., 2018). The average trait scores in Figure 4.3a all differ significantly between components, with no overlap between any of the CIs. The average trait scores for PA are smallest for Component 1, followed by Component 2, and highest for Component 3. For NA, the reverse can be seen. The average trait scores also show that there is a floor effect for NA. Individuals in Components 2 and 3 predominantly report little to no NA. The small random effects associated with trait score NA in Components 2 and 3 show that there is little within-component variation in NA trait scores, particularly in Component 3.

We compared the resulting components on external variables that have not been included in the model estimation. In the following we assign individuals into components based on their modal component membership probabilities. Table 4.8 lists age and gender of the components' members, and also the mixing proportions of the components. Figure 4.4 shows the distribution of age within each of the three components, as based on the predicted modal component membership. Figure 4.4 shows that members of Component 1 are predominantly young, of Component 2 a bit more often young than older, and of Component 3 are predominantly older.

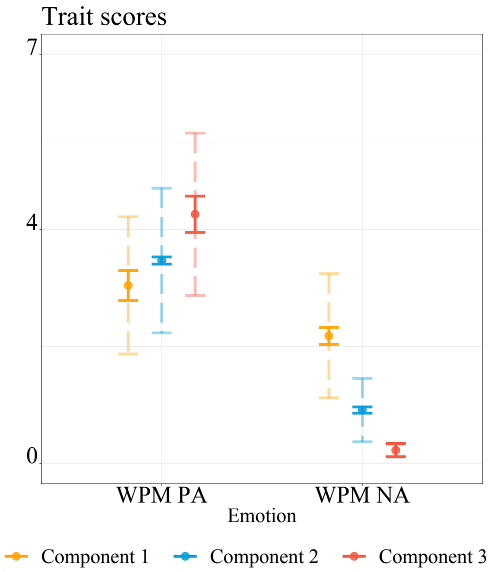
Figure 4.5 compares the members of the three components on their PANAS scores (Watson et al., 1988) before the ESM study took place (left) and two years after the ESM study took place (right); the corresponding values are listed in Table 4.9. Table 4.9 also shows the Neuroticism, Openness, and Extraversion scores of the components' members before the ESM study.

We conducted a MANOVA to investigate the component differences that are displayed in Table 4.8 and Table 4.9. All continuous variables that were assessed before the ESM study were used as outcome variables (i.e., PA, NA, Neuroticism, Extraversion, Openness and age); Modal component membership was used as independent variable. The MANOVA was significant ($p < .001$). To test subsequently which components differed and on which variables, we performed a paired t-test for each of the six variables. Each of these six post-hoc tests was Bonferroni-corrected for multiple testing across the combinations of components.

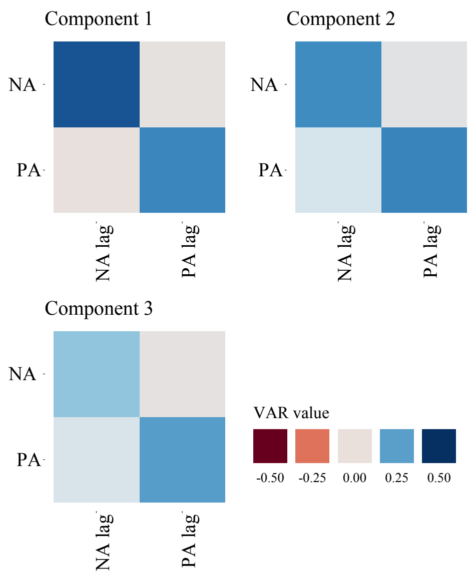
Because we have to account also for multiple testing with regard to the number of post-hoc tests, we present in Table 4.10 only those differences for which the adjusted p-values are below $\alpha = .05/6$. Table 4.10 shows that for age, Neuroticism and NA there are significant differences between components.

Table 4.7: Final MMVAR model parameters for the model with three unconstrained components: fixed effects (random effect variances) [95% confidence intervals for the fixed effect].

	Component 1	Component 2	Component 3
$\hat{\gamma}_{PA\,k}$ (WPM PA)	3.05 (1.38) [2.80; 3.30]	3.47 (1.54) [3.41; 3.53]	4.26 (1.93) [3.95; 4.57]
$\hat{\gamma}_{NA\,k}$ (WPM NA)	2.18 (1.13) [2.03; 2.33]	.91 (.3) [.86; .96]	.22 (.01) [.11; .33]
$\hat{\gamma}_{11\,k}$ (PA to PA)	.31 (.01) [.23; .39]	.31 (.01) [.28; .34]	.26 (.02) [.22; .30]
$\hat{\gamma}_{21\,k}$ (NA to PA)	$<.01$ ($<.01$) [$-.04$; .04]	.04 ($<.01$) [.03; .05]	.04 (.01) [$-.01$; .09]
$\hat{\gamma}_{12\,k}$ (PA to NA)	.01 ($<.01$) [$-.05$; .07]	.02 ($<.01$) [.02; .02]	.01 ($<.01$) [.004; .02]
$\hat{\gamma}_{22\,k}$ (NA to NA)	.42 (.01) [.36; .48]	.29 (.02) [.26; .32]	.17 ($<.01$) [.11; .23]



(a) The fixed effect within-person means. Solid error bars represent the 95% CIs. Dashed oblique error bars show one standard deviation estimates of the corresponding random effects (i.e., $\sqrt{\hat{\tau}_{11\,k}}$ or $\sqrt{\hat{\tau}_{22\,k}}$) above and below the fixed effect.



(b) The fixed effect VAR coefficients.

Figure 4.3

4.4.4 Conclusions

The COGITO sample comprised 101 young adults (aged 20–31) and 103 older adults (aged 65-80), as outlined in the Method section, and these distinct age groups are evident in Figure 4.4. Intriguingly, the three component solution seems to resemble largely, but not fully, two common age groups and a component that is a blend of both age groups (see Table 4.8, from Component 1 to 3; ‘Young adults’ (M_{Age}= 29, 21%), ‘Blended’ (M_{Age}= 42, 42%), and ‘Older adults’ (M_{Age}= 66, 37%)), and the results of the ‘Young adults’ and ‘Older adults’

Table 4.8: Means (SDs) on age, and gender proportions and mixing proportions for each component. Individuals are assigned into components based on their modal component membership probabilities.

	Age	Proportion female	Mixing proportion
Component 1	29.19 (14.18)	.57	.21
Component 2	41.69 (21.95)	.51	.42
Component 3	66.01 (14.93)	.46	.37

Table 4.9: Means (SDs) of positive affect (PA) and negative affect (NA) either before the start of the intensive longitudinal data collection, or 2 years after the intensive longitudinal data collection, and Means (SDs) on the NEO before the start of the intensive longitudinal data collection. PA and NA were measured with the PANAS on a continuous scale from 0 to 7. The NEO was assessed on a continuous scale from 0 to 4. Individuals were assigned into components based on modal component membership probabilities.

	PANAS				NEO		
	Before study PA	Before study NA	Follow-up PA	Follow-up NA	Neuroticism	Extraversion	Openness
Component 1	4.46 (1.01)	3.49 (.96)	4.27 (.88)	3.01 (1.31)	2.08 (.47)	2.39 (.45)	2.64 (.34)
Component 2	4.58 (.98)	2.91 (1.24)	4.44 (1.07)	2.41 (1.12)	1.84 (.38)	2.16 (.40)	2.48 (.35)
Component 3	4.76 (.82)	1.94 (.87)	4.68 (.98)	1.43 (.87)	1.47 (.34)	2.19 (.37)	2.45 (.29)

Table 4.10: All significant comparisons of the Bonferroni-adjusted post-hoc paired t-tests with modal component memberships as the independent variable. Different components were compared on their mean values of age, PA, NA, Neuroticism, Extraversion, and Openness. Because six of these paired t-tests were conducted, we show only the results where the adjusted p-values are below $.05/6$.

	Components	t	df	adjusted p
Age	1 vs. 2	-3.88	116.29	.001
Age	1 vs. 3	-13.25	88.45	$< .001$
Age	2 vs. 3	-8.33	150.52	$< .001$
Before study NA	1 vs. 3	8.65	77.53	$< .001$
Before study NA	2 vs. 3	5.80	152.01	$< .001$
Neuroticism	1 vs. 3	7.45	66.22	$< .001$
Neuroticism	2 vs. 3	6.38	153.81	$< .001$

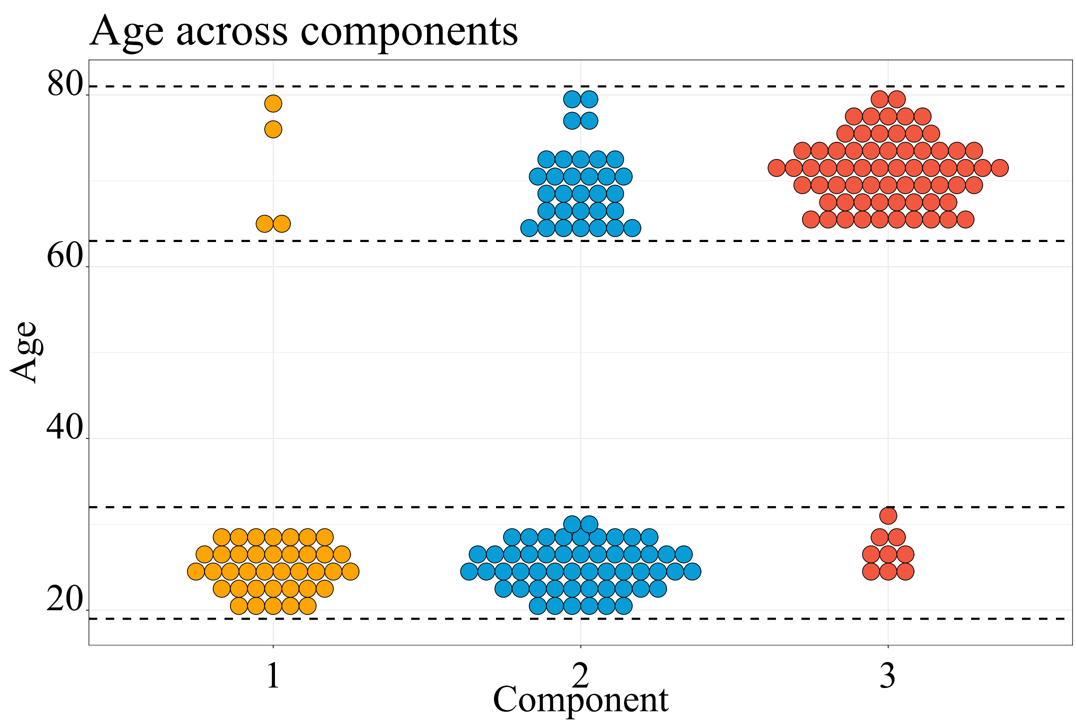


Figure 4.4: Age distributions within components. Dashed lines indicate the boundaries of the age groups in this sample. Individuals are assigned into components based on their modal component membership probabilities.

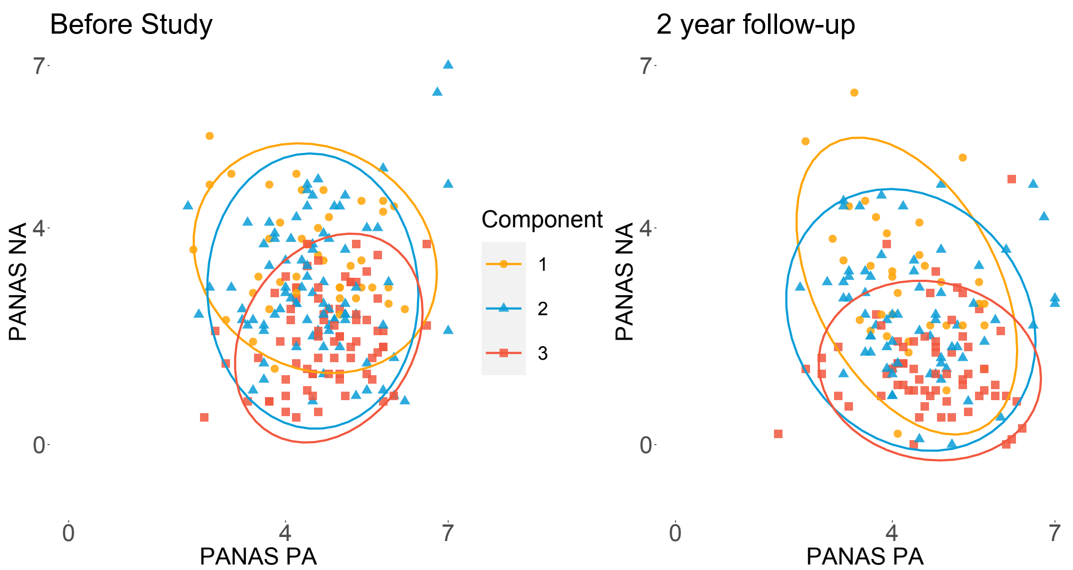


Figure 4.5: PANAS scores before the ESM study took place (left) and two years after the ESM study took place (right), split by modal component membership of the individual. Ellipses show the multivariate t-distribution at the 95% confidence level.

components align with expected lifespan differences. Most prominently, the higher within-person mean (WPM) of PA and lower NA (also smaller NA variance, see Figure 4.3a and Table 4.7) for the 'Older adults' compared to the 'Young adults' over 100 days are in keeping with a lifespan trajectory of increasing well-being that is known as the "positivity effect" (see Carstensen & DeLiema, 2018; Carstensen et al., 2011; Reed et al., 2014). The socioemotional selectivity theory explains this positivity effect via changes in how older adults structure their emotion goals to maximize positive interactions and minimize negative social encounters and learn to cognitively "dampen" the impact of NA (Carstensen & DeLiema, 2018; Carstensen et al., 2011). This is evident in our model via the lower fixed effect autoregressive coefficients of NA and not PA for the 'Older adults' (see Figure 4.3b and Table 4.7), and average PA and NA levels (Figure 4.5 and Table 4.9). These differences in emotion dynamics are supported by the significantly lower average neuroticism (see Tables 4.9 and 4.10) in line with typical lifespan trajectories (see Roberts et al., 2006). Although we can only assess between-subject differences of those within our two age ranges (as we included two age groups), the patterns described above fit the within-person lifespan trajectory that is known from the broader emotion and personality literature (Carstensen & DeLiema, 2018; Charles et al., 2001; McAdams et al., 2019), where the 'Blended' component emerges at its expected position with average age, in-between the 'Young adults' and 'Older adults', even though middle-aged adults have not been sampled. There must be a host of factors, including biological ageing, social indicators, personality differences, and contexts that guide our preferences, decisions, and that what we see, hear, and remember, from which this pattern of emotion dynamics emerges. Interestingly, Figure 4.4 shows that a few of the old participants were classified into the 'Young adults' component and vice versa, a few young participants were classified into the 'Older adults' component. In future research it might be interesting to find out more about these individuals who present an atypical 'mindset' for their age.

One way in which the model expands upon conventional knowledge is the notion of significant component differences in the WPM or trait of PA (Figure 4.3a), while these components did not significantly differ in extraversion (see Tables 4.9 and 4.10). Whole Trait Theory predicts that personality traits should roughly equal the mean of the density distribution of relevant states (Fleeson & Jayawickreme, 2015), but our results rather support the position that traditional personality trait self-descriptors (in general) and aggregated momentary assessment of daily scores do not converge neatly (Augustine & Larsen, 2012; Rauthmann et al., 2019); As we observed convergence of fixed effect WPM of NA and neuroticism, but not for fixed effect WPM of PA and extraversion. Overall our results clearly hold promise for the MMVAR approach to deepen our understanding of the link between processes within- and between-individuals across different temporal measurement levels, which is the new frontier in psychology (Collins, 2006; Hopwood et al., 2021).

Finally, Component 2 constitutes a blend between young and older adults, and it may not surprise that in the 2×2 component solution this group breaks into two while the two age group components reappear with almost identical classifications (see Figure 4.10 in Appendix B). We take this as evidence for the robustness of our classification. The 2×2 component solution is shown in Appendix B.

4.5 Discussion

4.5.1 Data requirements

In this paper we introduced mixture multilevel vector-autoregressive modelling (MMVAR) for identifying similar trait levels and dynamic processes across individuals in intensive longitudinal data. A simulation study (Simulation 2) evaluated our MMVAR estimation procedure implemented in Mplus (Muthén & Muthén, 2013). This simulation showed good recovery of component memberships and parameters under a range of empirically relevant conditions. However, the estimation was time-intensive, and a substantial proportion of the algorithm's starts ended in local minima. The estimation procedure achieved satisfactory recovery for the lowest number of observations considered (i.e., 75), and 100 individuals. It might be that even with a smaller number of observations and/or individuals proper recovery would have been achieved; yet, the recovery can be expected to depend also on other model parameters, including the distance between the component parameters (i.e., fixed effect VAR coefficients and within-person means).

In Simulation 2, distance between components, as expressed in distance between fixed effect VAR coefficients, appeared the most important factor in recovery. The distance between component parameters plays a crucial role especially regarding the recovery of the modal component membership. Modal component memberships are achieved by assigning each individual to the component for which they have the highest posterior probability. Consequently, the recovery of the modal component membership cannot be perfect when substantial overlap exists between components. In contrast, other parameters, for instance the mixing proportions, can be recovered perfectly despite overlap in components because they are based on the posterior probabilities.

Increasing the number of time-points caused little improvement in parameter recovery in Simulation 2. This could be because we investigated the recovery of level 2 parameters. In general a higher number of time-points will likely improve the estimation of level 1 parameters, while a higher number of individuals will likely improve the estimation of level 2 parameters. We conjecture that the random effect variances for VAR coefficients were consistently underestimated in Simulation 2 because the random effect variances for the within-person means were overestimated, because their true value was zero.

4.5.2 Model evaluation

We discussed several considerations that should be made when employing MMVAR. Amongst others, empirical researchers should evaluate which model constraints they can employ to aid model estimation. The MMVAR model relies on numerous parameters, and overparameterization can quickly lead to estimation/convergence problems. Further, empirical researchers should consider whether to within-person center (WPC) variables. In multilevel VAR models, WPC causes bias in the parameter estimates of autoregressive coefficients, as we showed in this paper through a simulation study (Simulation 1). Work by Hamaker and Grasman (2015) shows, however, that WPC in multilevel VAR models prevents biased estimates of the coefficients of level 2 predictors;[13] this bias does occur when variables are not centered (NC). Simulation 1 showed also that cross-regressive coefficients are underestimated, for WPC and NC, when there is variation for the random slopes. This bias of cross-regressive coefficients is in general more severe for NC than for WPC. Luckily, all the biases discussed here reduce when the number of time-points increases, for both autoregressive and cross-regressive coefficients — as is shown in Simulation 1.

The exploratory nature of MMVAR is what makes it particularly useful. In our empirical data, MMVAR could identify in an exploratory manner three components of individuals that exhibit distinct emotion dynamics and trait scores. MMVAR therefore did not rely on an observed grouping variable, but grouped individuals based on the characteristics of interest: the emotion dynamics and trait scores. This exploratory grouping is advantageous, firstly because the model does not need a good grouping predictor to be known or observed. MMVAR thus lends itself nicely to data sets where the causes of/associations with the inter-individual differences are unknown/unobserved. Secondly, the exploratory grouping of MMVAR gives a more direct description of the differences in emotion dynamics and trait scores than a grouping based on observed predictors. This was nicely illustrated in our empirical example. That is, in our empirical example participants differed markedly in age. Age differences are associated with differences in emotion dynamics and trait scores (Carstensen et al., 2011). The

[13]Meaning here coefficients that represent the influence that level 2 predictors have on the level 2 parameters.

estimated MMVAR model showed that differences in emotion dynamics and trait scores were substantially related to age, as shown in Figure 4.4: One component was uncovered (Component 1) that consisted of almost entirely young participants, but a few of the old participants were classified into this component as well. Vice versa was found for a component of predominately older adults (Component 3). These individuals who present an atypical 'mindset' for their age are done more justice in an exploratory grouping where they are grouped directly based on the characteristics of interest, than in a model that explicitly employs a grouping based on observed predictors. MMVAR thus properly accounts for such atypical cases, while retaining the ability to generalize to the population of individuals. This is unlike a model that is entirely focused on the individual, as is the case in single-individual models.

Using post-hoc tests we could test whether differences in certain observed variables are associated with the MMVAR components, which were identified based on differences in emotion dynamics and trait scores. In our empirical example, we found significant differences between the components with regards to age, neuroticism and negative affect. Such differences can account for between-component inter-individual differences. Because MMVAR is a multilevel model, observed variables like age could also be used to explain within-component inter-individual differences, by employing them as level 2 predictors. Similarly, level 1 predictors could be employed to account for trends or cycles over time.

In sum, MMVAR proves to be a promising strategy to identify *latent* subgroups that exhibit distinct within-individual dynamics and trait scores in intensive longitudinal data. This was demonstrated in our empirical example of individuals of disparate age groups; three components were identified based on their distinct emotion dynamics and trait scores. MMVAR offers a summary of dynamic processes that accounts for inter-individual differences that are both, continuous (i.e., within components) and discrete (i.e., between components). By capturing continuous inter-individual variation within components through the (co)variances of random effects, MMVAR extends the latent class VAR models that have been proposed so far that account only for discrete inter-individual differences between components (e.g., Anderlucci & Viroli, 2015; Ernst et al., 2020). MMVAR can also be seen as an extension of multilevel VAR (e.g., Rovine & Walls, 2006) that adds the exploratory identification of latent classes to account for structural inter-individual differences in dynamics.

4.A Appendix for Simulation 2: Performance of MMVAR model estimation

Recovery of random effect variances of within-person means Figure 4.6a shows the mean absolute deviations (MAD) between estimated and true random effect variances of within-person means per simulation condition. Figure 4.7a displays the average estimates for the random effect variances per simulation condition; the value that was used to generate the data is represented by a dotted line. Figures 4.6a and 4.7a show identical values, because the true random effect variances of within-person means equalled zero. Therefore, the corresponding MADs can be directly interpreted as the average estimated random effect variances. The figures indicate that overall the random effect variances were recovered well across simulation conditions with the exception of a few outliers.

Table 4.11 displays the ANOVA, the effects were all non-significant suggesting there are none to only small differences in recovery across simulation conditions.

Recovery of random effect variances of VAR coefficients Figure 4.6b displays the MADs between estimated and true random effect variances of VAR coefficients per simulation condition. Figure 4.7b shows the average random effect variances of the VAR coefficients per simulation condition. These figures suggest that the variances are recovered well across all simulation conditions, with estimated variances very close to the true data generating value. The variances are, however, consistently underestimated, as can be seen in Figure 4.7b.

Table 4.11 shows the ANOVA results, the effects were all non-significant suggesting there are none to only small differences in recovery across simulation conditions.

Recovery of mixing proportions Figure 4.8 displays the MADs per simulation condition. The effect sizes in Table 4.12 suggest that a high number of observations ($\hat{\eta}_p^2 = .112$) and large distances between components ($\hat{\eta}_p^2 = .253$) aided recovery of mixing proportions. This indicates there were medium differences across simulation conditions. On average, MADs indicate mixing proportions were retrieved well ($M = .017$, $SD = .010$).

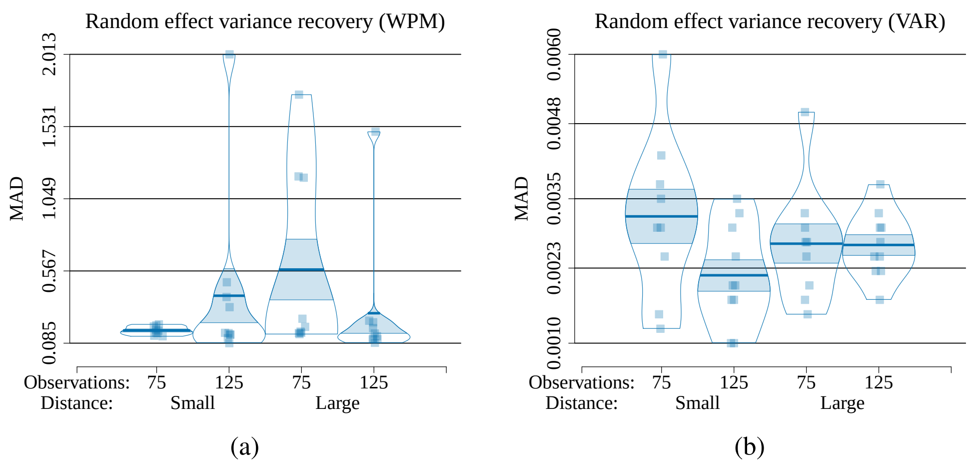


Figure 4.6: MADs between estimated and true random effect variances indicating the parameter recovery for the random effect variances across simulation conditions. A line indicates the mean MAD per condition, a coloured band the region within one standard error of the mean.

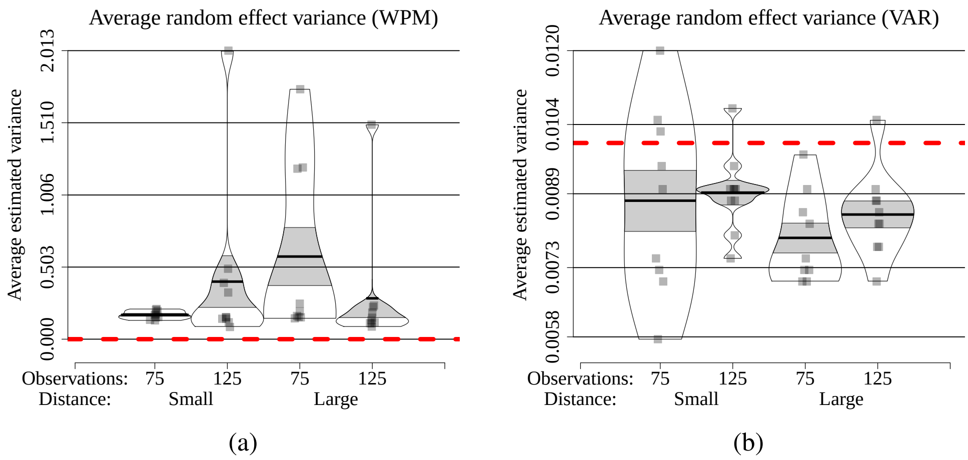


Figure 4.7: Average estimates for the random effect variances across simulation conditions. A line indicates the mean estimated variance, a coloured band the region within one standard error of the mean. A dotted line shows the true random variance value used in data generation.

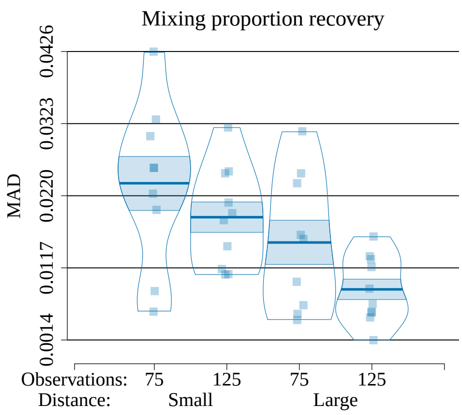


Figure 4.8: MADs between estimated and true mixing proportions indicating the recovery of mixing proportions across simulation conditions. A dark line indicates the mean MAD, a band the standard error.

Table 4.11: ANOVA results on MADs between estimated and true random effect variances of within-person means (WPM), and on MADs between estimated and true random effect variances of VAR coefficients. Any effect size above .1 is highlighted in bold.

	WPM				VAR coefficients			
	df	*F*	*p*	$\hat{\eta}_p^2$	*df*	*F*	*p*	$\hat{\eta}_p^2$
Observations	1	.037	.849	.001	1	2.500	.123	.068
Distance	1	.712	.405	.021	1	.026	.874	.001
Observations × Distance	1	2.830	.102	.077	1	2.290	.140	.063

Table 4.12: ANOVA results on MADs between estimated and true mixing proportions. Any effect size above .1 is highlighted in bold.

	df	*F*	*p*	$\hat{\eta}_p^2$
Observations	1	4.290	.046	**.112**
Distance	1	11.500	.002	**.253**
Observations × Distance	1	.110	.742	.003

4.B Appendix for Application: COGITO study

4.B.1 Results: The 2×2 component solution

The estimated MMVAR parameters of the 2×2 component model are presented in Table 4.13. The fixed effect within-person means shown in this table, which we interpret as average trait scores, are graphically presented in Figure 4.9a, the fixed effect VAR coefficients in Figure 4.9b. The confidence intervals (CIs) of the fixed effects, displayed in Table 4.13, all exclude zero, with the exception of the CI for the PA to NA cross-regressive coefficient for Components 3 and 4.

Average trait scores appear to be separated into ‘Unhappy components’ (Components 1 and 2) and ‘Happy components’ (Components 3 and 4). The ‘Unhappy components’ are characterised by low PA and high NA, as is evident from the fixed effect within-person means displayed in Figure 4.9a and in Table 4.13. The reversed pattern, high PA and low NA, can be seen for the ‘Happy components’. The ‘Happy’ and the ‘Unhappy’ components are further split into ‘High inertia’ and ‘Low inertia’ components (Components 1 and 3, and Components 2 and 4 respectively). Differences in inertia are particularly pronounced for NA as can be seen in Figure 4.9b and in Table 4.13.

We compared the resulting four components on external variables that have not been included in the model estimation. In the following, we assign individuals into components based on their modal component membership probabilities. Table 4.14 lists age and gender of the components’ members, and also the estimated mixing proportions of the components. The mixing proportions in Table 4.14 show that more individuals were classified into the ‘Happy’ rather than the ‘Unhappy’ components. Whether the ‘High inertia’ or ‘Low inertia’ components were more frequent depended on the component membership for the within-person means: Individuals in the ‘Happy’ components were more likely to be classified into the ‘Low inertia’ rather than the ‘High inertia’ component (i.e., Component 4 over Component 3). Individuals in the ‘Unhappy’ components were more likely to be classified into the ‘High inertia’ rather than the ‘Low inertia’ component (i.e., Component 1 over Component 2).

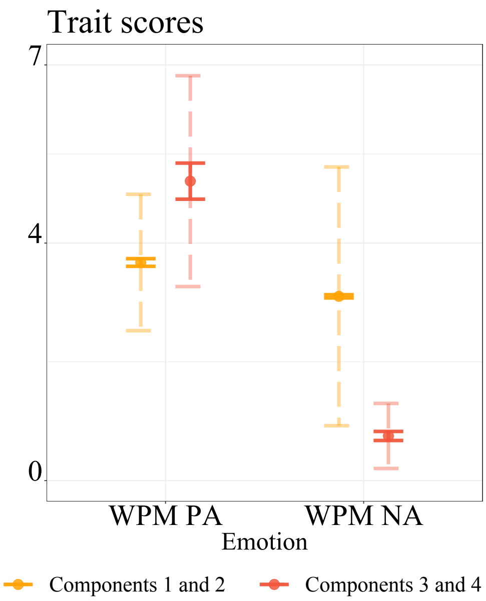
Components are clearly ordered by age, as can be seen in Table 4.14: Individuals in the ‘Happy’ components are older than the ‘Unhappy’ components, while the individuals in the ‘High inertia’ components have a lower mean age than the ‘Low inertia’ components. The differences in gender proportions across components, shown in Table 4.14, seem negligible. Figure 4.10 depicts the distribution of age within the four components. Figure 4.10 shows that, like in the 3 component solution, two of the MMVAR components correspond largely to the two different age groups. Figure 4.11 compares the members of the four components on their PANAS scores before the ESM study took place (left) and two years after the ESM study took place (right), the corresponding values are listed in Table 4.15. Table 4.15 shows also the Neuroticism, Openness, and Extraversion scores of the components’ members before the ESM study. Looking at external variables that were not included in the model estimation (see Table 4.14 and 4.15), components seem most different on values of NA. Differences in

NA are observable even 2 years after the study (see Table 4.15). Components seem to differ in Neuroticism but are comparable on Extraversion and Openness (see Table 4.15).

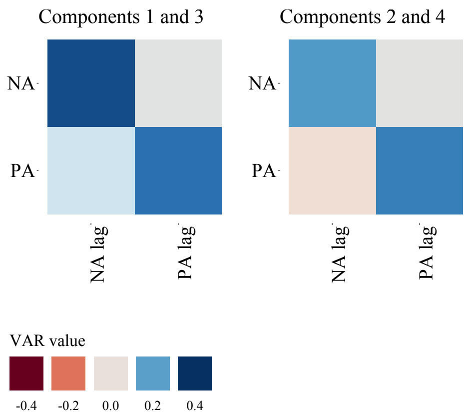
We conducted a MANOVA to investigate the component differences that are displayed in Table 4.14 and 4.15. All continuous variables that were assessed before the ESM study were used as outcome variables (i.e., PA, NA, Neuroticism, Extraversion, Openness and age); Modal component membership was used as independent variable. The MANOVA was significant ($p < .001$). To test subsequently which components differed and on which variables, we performed a paired t-test for each of the six variables. Each of these six post-hoc tests was Bonferroni-corrected for multiple testing across the combinations of components. Because we have to account also for multiple testing with regard to the number of post-hoc tests, we present in Table 4.16 only those differences for which the adjusted p-values are below $\alpha = .05/6$. Table 4.16 shows that for age, Neuroticism and NA there are significant differences between components.

Table 4.13: The final MMAVR model parameters: fixed effects (random effect variances) [95% confidence intervals for the fixed effect].

	Component 1	Component 2	Component 3	Component 4
$\hat{\gamma}_{PA\,k}$ (WPM PA)	3.67 (1.32) [3.61; 3.73]		5.04 (3.15) [4.74; 5.34]	
$\hat{\gamma}_{NA\,k}$ (WPM NA)	3.1 (4.74) [3.07; 3.13]		.75 (.3) [.67; .83]	
$\hat{\gamma}_{11\,k}$ (PA to PA)	.29 (.01) [.24; .34]	.26 (.01) [.23; .29]	.29 (.01) [.24; .34]	.26 (.01) [.23; .29]
$\hat{\gamma}_{21\,k}$ (NA to PA)	.05 (.01) [.02; .08]	$-.01$ ($<.01$) [$-.02; -.002$]	.05 (.01) [.02; .08]	$-.01$ ($<.01$) [$-.02; -.002$]
$\hat{\gamma}_{12\,k}$ (PA to NA)	.01 ($<.01$) [$-.01; .03$]	.01 ($<.01$) [$-.002; .02$]	.01 ($<.01$) [$-.01; .03$]	.01 ($<.01$) [$-.002; .02$]
$\hat{\gamma}_{22\,k}$ (NA to NA)	.35 (.01) [.30; .40]	.21 ($<.01$) [.18; .24]	.35 (.01) [.30; .40]	.21 ($<.01$) [.18; .24]



(a) The fixed effect within-person means. Solid error bars represent the 95% CIs. Dashed oblique error bars show one standard deviation estimates of the corresponding random effects (i.e., $\sqrt{\hat{\tau}_{11\,k}}$ or $\sqrt{\hat{\tau}_{22\,k}}$) above and below the fixed effect.



(b) The fixed effect VAR coefficients.

Figure 4.9

Table 4.14: Means (SDs) on age, and gender proportions and mixing proportions for each component. Individuals were assigned into components based on their modal component membership probabilities.

	Age	Proportion female	Mixing proportion
Component 1	29.55 (14.44)	.55	.20
Component 2	39.07 (20.67)	.62	.14
Component 3	42.39 (22.60)	.47	.29
Component 4	65.95 (15.00)	.46	.37

Table 4.15: Means (SDs) of positive affect (PA) and negative affect (NA) either before the start of the intensive longitudinal data collection, or 2 years after the intensive longitudinal data collection, and Means (SDs) on the NEO before the start of the intensive longitudinal data collection. PA and NA were measured with the PANAS on a continuous scale from 0 to 7. The NEO was assessed on a continuous scale from 0 to 4. Individuals were assigned into components based on modal component membership probabilities.

	PANAS				NEO		
	Before study PA	Before study NA	Follow-up PA	Follow-up NA	Neuroticism	Extraversion	Openness
Component 1	4.54 (.98)	3.43 (.90)	4.29 (.85)	2.88 (1.26)	2.04 (.45)	2.42 (.43)	2.63 (.35)
Component 2	4.92 (.98)	2.86 (1.36)	4.51 (1.26)	2.24 (1.34)	1.75 (.43)	2.21 (.42)	2.59 (.36)
Component 3	4.37 (.97)	2.91 (1.23)	4.40 (1.00)	2.44 (1.10)	1.88 (.40)	2.15 (.36)	2.45 (.33)
Component 4	4.75 (.81)	2.01 (.94)	4.67 (.98)	1.53 (.97)	1.50 (.37)	2.16 (.40)	2.44 (.29)

Table 4.16: All significant comparisons of the Bonferroni-adjusted post-hoc paired t-tests with modal component memberships as the independent variable. Different components were compared on their mean values of age, PA, NA, Neuroticism, Extraversion, and Openness. Because six of these paired t-tests were conducted, we show only the results where the adjusted p-values are below $.05/6$.

	Components	t	df	adjusted p
Age	1 vs. 3	−3.45	96.73	.005
Age	1 vs. 4	−12.73	82.10	< .001
Age	2 vs. 4	−6.39	39.79	< .001
Age	3 vs. 4	−6.91	95.80	< .001
Before study NA	1 vs. 4	7.97	82.79	< .001
Before study NA	3 vs. 4	4.67	106.34	< .001
Neuroticism	1 vs. 4	6.43	65.76	< .001
Neuroticism	3 vs. 4	5.66	117.89	< .001

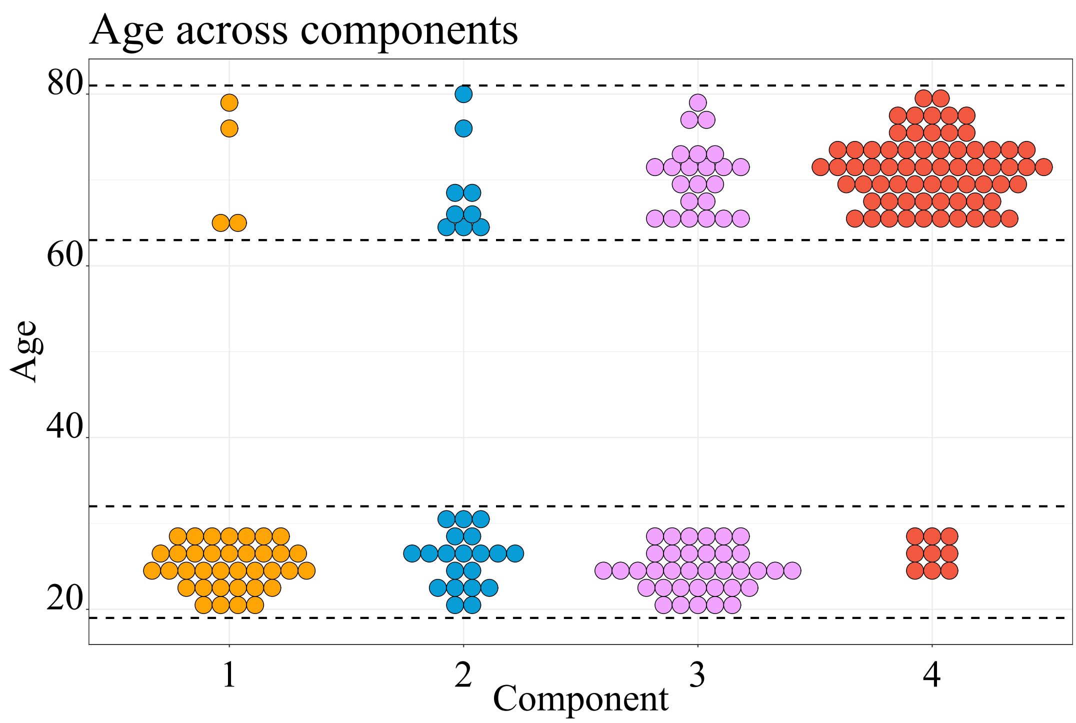


Figure 4.10: Age distributions within components. Dashed lines indicate the boundaries of the age groups in this sample. Individuals are assigned into components based on their modal component membership probability.

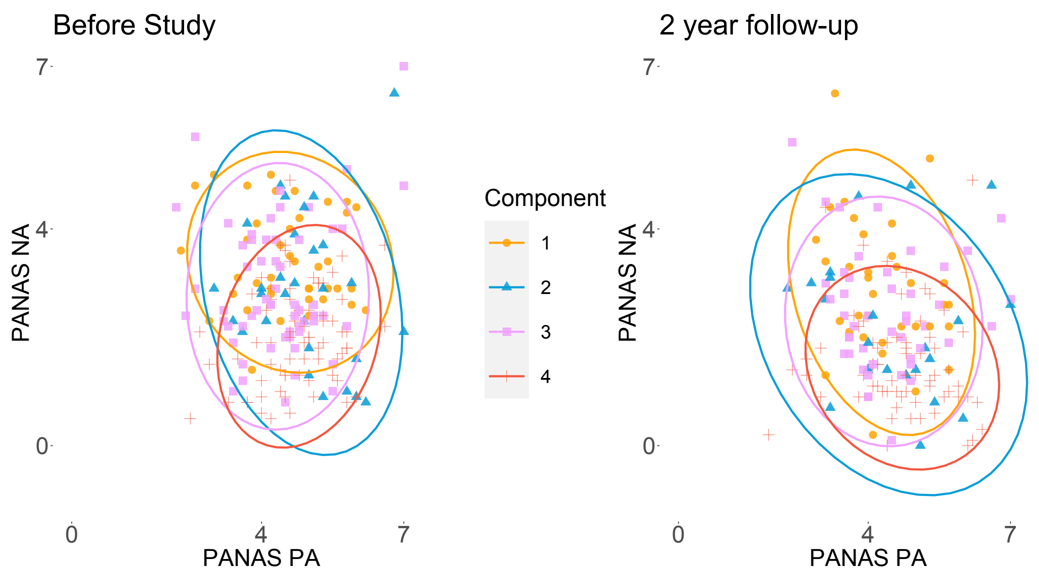


Figure 4.11: PANAS scores before the ESM study took place (left) and two years after the ESM study took place (right), split by modal component membership of the individual. Ellipses show the multivariate t-distribution at the 95% confidence level.