

University of Groningen

Dynamic clustering

Ernst, Anja Franziska

DOI:
[10.33612/diss.196176258](https://doi.org/10.33612/diss.196176258)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Ernst, A. F. (2022). *Dynamic clustering: classifying people through ecological momentary assessment*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.
<https://doi.org/10.33612/diss.196176258>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chapter 3

Inter-individual differences in multivariate time-series: Latent class vector-autoregressive modelling

In recent years, research has expanded into the dynamic processes of psychological phenomena. Smartphones and tablets are providing an infrastructure that allows for the collection of time-intensive, multivariate data using ambulatory assessment (AA). Intensive longitudinal AA data can provide insight into emotion dynamics. These dynamics reflect how individuals react to their environment and regulate their emotions. For instance, emotional inertia, the propensity of affective states to persist over time and resist to change, is often quantified as the autoregression of an emotion (Kuppens & Verduyn, 2017). High emotional inertia may be indicative of an inability to recover from negative emotions as a result of impaired emotion regulation. High inertia may also be a sign of increased preoccupation, resulting in emotional insensitivity and decreased involvement with the environment (Brose et al., 2015a). Inertia is negatively associated with psychological well-being (Houben et al., 2015; Koval & Kuppens, 2012) and has been linked to mental disorders. For instance, depression and manic episodes are characterised by high levels of inertia of positive or negative affect. This expresses general under-reactivity and inflexibility (American Psychiatric Association, 2013). While single-subject time-series models can provide insight into emotion dynamics, these models are unwieldy in a large sample and prevent a proper characterisation of the population of individuals. To arrive at a proper generalisation, one needs to account for the presence of inter-individual differences in emotion dynamics in multi-subject AA data.

Theories surrounding emotion regulation suggest that individuals perceive and predict situations differently and construct qualitatively different emotional processes (Barrett, 2018) that differ in their persistence (Verduyn & Lavrijsen, 2015) and therefore display qualitatively distinct dynamic affect patterns (Houben et al., 2015). Good emotion regulation may manifest itself, for instance, by rarely moving out of positive affect states and by moving out of negative affect states quickly (Hay & Diehl, 2011). Such individual differences can be captured with multi-subject time-series models that identify distinct subgroups of people with similar emotion dynamics. Before we introduce our novel latent class vector-autoregressive modelling (LCVAR) strategy to identify groups of people with similar emotion dynamics we first outline the benefits of such time-series clustering models to popular alternatives.

When using time-series models the inter-individual differences that are allowed follow a continuum from single-subject models to narrow panel models that estimate dynamic patterns with one set of *fixed*-coefficients.¹ Such fixed-coefficients can become substantially biased

This chapter has been published as Ernst, A. F., Albers, C. J., Jeronimus, B. F., & Timmerman, M. E. (2020). Inter-individual differences in multivariate time-series: Latent class vector-autoregressive modeling. *European Journal of Psychological Assessment*, 36(3), 482–491. <https://doi.org/10.1027/1015-5759/a000578>.

¹To circumvent the ambiguity surrounding the term “random effects” we refer to a coefficient as *fixed* when it is identical across individuals and *random* when it may vary across individuals.

in the presence of inter-individual differences (Hamaker et al., 2015b). Inter-individual differences arise as observed or unobserved heterogeneity (Lubke & Muthén, 2005). Observed time-series heterogeneity can be explained by an observed variable, for instance gender. Unobserved heterogeneity is caused by variables that remain unknown, which occurs rather often.

Unobserved heterogeneity may emerge from a distribution of quantitative deviations around a uniform effect or reflect discrete and qualitatively different processes. Such quantitative inter-individual differences can be modelled through random-coefficients. Hierarchical Bayesian models use random-coefficients to allow for time-series models where all coefficients may vary across individuals and Bayesian estimation offers the distribution of the coefficients great flexibility (e.g., Driver & Voelkle, 2018). Thereby, the multilevel vector autoregressive (VAR) model captures within-individual dynamics (level 1) and continuous inter-individual differences (level 2) (Schuurman et al., 2016). Random-coefficients allow for quantitative differences (i.e., differences in coefficient values) but are unable to capture qualitative differences (i.e., differences in time-series models) and cannot explain why certain coefficients are positive for some and negative for others.

In contrast, discrete unobserved heterogeneity is accounted for by inferring latent subgroups from the data using exploratory methods. This can uncover subgroups with distinct patterns in affect dynamics. In the context of unobserved heterogeneity, subgroups are denoted latent classes (Lubke & Muthén, 2005). The latent classes themselves are unknown, and which latent class an individual belongs to needs to be inferred from the data.

3.0.1 Time-series models for discrete unobserved heterogeneity

Latent Markov models (LMMs) have been proposed to discover latent classes in AA data. These models estimate latent class membership at time point t based on the latent class membership at $t - 1$ and the current observed variable(s). Latent classes represent different mood states and the stability of mood states is expressed by transition probabilities between latent classes. To allow for inter-individual differences in mood fluctuation LMMs may nest latent classes within latent classes (Crayen et al., 2017) or model transition probabilities through random-coefficients (Asparouhov et al., 2017). However, the transition probabilities in LMMs give no insight into the dynamics between all observed emotions.

In this article we propose a time-series clustering model that will identify latent classes (henceforth called clusters) of individuals based on similarity in emotion dynamics between all observed emotions. Individuals belong to exactly one cluster throughout. These models allow researchers to assess various theories on emotion regulation and pathology that imply groups of individuals who display different emotion dynamic patterns.

Time-series clustering

We focus here exclusively on time-series clustering methods which cluster on the temporal persistence (inertia) of emotions and their dynamic interplay. Of the methods introduced so far, some are non-probabilistic (i.e., they operate by minimisation of a criterion, e.g., Bulteel et al., 2016). However, model-based clustering methods (i.e., based on a mixture model, McLachlan & Peel, 2004) are generally preferred because they offer a framework for formal inference and have theoretically favourable properties; For example, if the model is correct and the sample is large enough, the clustering of each individual is optimal (McLachlan, 2011).

In line with Jacques and Preda (2014) we distinguish between *filtering* and *adaptive* time-series clustering (sometimes called multi-step and simultaneous clustering, respectively), see Ernst et al. (2021) for an overview of these methods. Filtering methods firstly summarise

the dynamics of every individual (e.g., by summarising the time-series of every individual into ideographic VAR coefficients), and subsequently form clusters based on the individual summaries. Conversely, adaptive methods combine summarising and clustering individuals' dynamics (i.e., the summary into VAR coefficients depends on the clusters and vice versa). Because filtering methods summarize individual time-series before they cluster, the data summary may provide a suboptimal representation of the time-series for the clustering approach. This occurs if the data resulting from the compression is not discriminant for the different clusters (Bouveyron & Brunet-Saumard, 2014), or when the individual time-series are over-fitted. Adaptive methods avoid this problem, because the description of dynamic processes adapts during the clustering, which prevents premature compression of the data. Adaptive methods thus enable a cluster-specific description of the time-series. Consequently, different clusters can be modelled by qualitatively different time-series models.

Existing adaptive time-series clustering methods (e.g. Anderlucci & Viroli, 2015; Vermunt, 2007) are suboptimal for time-series with more than 10 measurement occasions and/or varying lengths, which are both common in AA data. Our adaptive, model-based LCVAR model combines: (1) VAR time-series models, describing intra-individual dynamic processes and (2) a latent class model, clustering individuals with similar dynamic characteristics via fixed-coefficients. This combination provides some crucial advantages: (a) characteristics of AA emotion data are accounted for, including multiple variables, a high number of time points, and possible differences in length of individual's time-series; (b) the model is ideally suited to identify inter-individual differences in emotion dynamics because the clustering is based on emotion dynamic patterns, emotion intensity levels and influences of external variables; and (c) through the adaptive estimation of the model, clusters of individuals exhibiting qualitatively different emotion dynamics can be distilled. Consequently, (d) accurate generalisations can be made both at the individual and the population level. Before illustrating these advantages on an empirical example of emotion dynamics, we define the LCVAR model formally.

3.0.2 Model specification

The LCVAR model can be seen as a multilevel model for occasions (level 1) nested within the individuals (level 2). At level 1, the model for the individual time-series includes a) covariate coefficients for trends and exogenous variables (Equation 3.1) and b) a VAR model of lag order p (VAR(p); cf. Chapters 2 – 3, Lütkepohl, 2005) for emotion dynamics (Equation 3.2). At level 2, we model the individual parameters via K latent classes – that is, with fixed-coefficients for each of the K clusters (Equations 3.3 – 3.5). Specifically, the observed time-series \mathbf{y}_i of individual i ($i = 1, 2, \dots, N$) is represented as follows

$$\text{Level 1 : } \mathbf{y}_{it} = \mathbf{w}_{it} + \mathbf{B}_i \mathbf{x}_{it}, \quad (3.1)$$

$$\mathbf{w}_{it} = \left(\sum_{a=1}^p \mathbf{A}_{ia} \mathbf{w}_{i,t-a} \right) + \mathbf{u}_{it} \quad \mathbf{u}_{it} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i) \quad (3.2)$$

$$\text{Level 2 : } \mathbf{B}_i = \sum_{k=1}^K z_{ik} \mathbf{B}_k \quad (3.3)$$

$$\mathbf{A}_{ia} = \sum_{k=1}^K z_{ik} \mathbf{A}_{ka} \quad (3.4)$$

$$\boldsymbol{\Sigma}_i = \sum_{k=1}^K z_{ik} \boldsymbol{\Sigma}_k, \quad (3.5)$$

where \mathbf{y}_{it} represents the $m \times 1$ vector of m observed endogenous variables at time point t ($t = 1, 2, \dots, T_i$);² \mathbf{B}_k a $m \times q$ covariate matrix expressing the influence of q exogenous variables, captured in $q \times 1$ vector \mathbf{x}_{it} , on the endogenous variables of individual i ; \mathbf{w}_{it} a $m \times 1$ vector containing a zero-mean VAR(p) time-series of the m endogenous variables; \mathbf{A}_{ka} a $m \times m$ matrix containing the VAR coefficients at lag a ; \mathbf{u}_{it} white noise error with $m \times m$ covariance $\mathbf{\Sigma}_k$; Every individual, i , is assumed to belong to exactly one of K clusters ($k = 1, 2, \dots, K$), with an indicator variable of cluster membership $z_{ik} = 1$ if $i \in k$ and 0 otherwise.

The inclusion of time-varying variables \mathbf{x}_{it} in the model allows to account for the influence of observed exogenous variables. This includes static variables such as gender and treatment condition, as well as time-varying variables, such as the weather. Time itself may be included, to capture trends in the time-series. For time-series with non-zero mean we capture the mean values on the observed endogenous variables, \mathbf{y}_i , through intercepts included in \mathbf{B}_k . This property will be illustrated in our empirical example where \mathbf{B}_k will capture mean values on all variables at various times of day for each cluster. We prefer this formulation over including a VAR-intercept in Equation 3.2, because VAR-intercepts lack a clear interpretation. In our formulation, the exogenous variables influence only a single measurement and these influences do not carry over to any future time point.

The actual dynamics are expressed via the VAR coefficients contained in \mathbf{A}_{ka} , describing the dynamic influence of the endogenous variables at time lag a on themselves at time point t . The matrix contains autoregressive coefficients which quantify emotional inertia, and cross-regressive coefficients which capture emotion augmentation and blunting (Kuppens & Verduyn, 2017). Our LCVAR model allows for comparisons with regard to emotional intensity and the influence of exogenous variables (via \mathbf{B}_k) as well as emotion dynamic patterns (via \mathbf{A}_{ka}). To model an equal influence of exogenous variables or equal covariance across the population, \mathbf{B}_k or $\mathbf{\Sigma}_k$ might be constrained to be equal across clusters. The VAR(p) model assumes equidistant time points between measurements of all individuals.

3.0.3 Estimation

We propose a maximum likelihood estimation procedure based on the conditional representation of the time-series. The big advantage of this representation is that it enables us to model both long time-series and time-series of unequal length between individuals. This is so, because the dimension of the within-cluster covariance matrices does not depend on the length of the time-series. We thereby extend the work of Michael and Melnykov (2016) to the multivariate case. Using Equation 3.2 we condition on the p previous observations, modelling the distribution of a time-series as a sample of size T_i from an m -variate conditional normal distribution. The log-likelihood of observing data $\mathbf{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$ with respect to the model parameters contained in vector $\Theta = \{\tau_1, \dots, \tau_K, \mathbf{\Sigma}_1, \dots, \mathbf{\Sigma}_K, \mathbf{A}_1, \dots, \mathbf{A}_K, \mathbf{B}_1, \dots, \mathbf{B}_K\}$ is thus

$$\log[\mathcal{L}(\Theta; \mathbf{Y})] = \sum_{i=1}^N \log \left[\sum_{k=1}^K \tau_k \prod_{t=1}^{T_{ik}} \Phi(\mathbf{y}_{it} | \mathbf{y}_{i,t-1}, \dots, \mathbf{y}_{i,t-p_k}, \mathbf{x}_{it}, \dots, \mathbf{x}_{i,t-p_k}; \boldsymbol{\mu}_{kt}(\mathbf{A}_k, \mathbf{B}_k), \mathbf{\Sigma}_k) \right], \quad (3.6)$$

where τ_k represents the prior probabilities of cluster membership (i.e., the mixing proportions); Φ is the multivariate Gaussian density corresponding to component k parameterised

²As common in VAR(p) models, we assume an additional pre-sample, containing p previous observations, is available.

by Σ_k , as defined in Equation 3.5 and a conditional mean, $\mu_{kt}(\mathbf{A}_k, \mathbf{B}_k)$, depending on $\mathbf{y}_{i,t-1}, \dots, \mathbf{y}_{i,t-p_k}, \mathbf{x}_{it}, \dots, \mathbf{x}_{i,t-p_k}$. That every component can be modelled by a Gaussian parameterised by Σ_k becomes obvious when realising that the deviation between \mathbf{y}_{it} and the conditional mean vector $\mu_{kt}(\mathbf{A}_k, \mathbf{B}_k)$ corresponds to \mathbf{u}_{it} . Due to the adaptive estimation of the model, the lag order, p_k , may differ between clusters with $\mathbf{A}_k = (\mathbf{A}_{k1} \dots \mathbf{A}_{kp_k})$.

Because the maximisation of Equation 3.6 is not straightforward it is maximised indirectly via the complete data log likelihood under the posterior distribution of cluster membership (\mathcal{Q} -function). Cluster membership is unknown and estimated via posterior probabilities of cluster membership. We estimate probabilistic cluster membership and cluster-wise time-series parameters through the Expectation Maximization (EM) algorithm (Dempster et al., 1977). The \mathcal{Q} -function, the EM algorithm, parameter estimates and their proof are given in Appendix A. The estimation assessments for the computer implementation of the algorithm are listed in Appendix B.

LCVAR can be estimated with the `LCVARc1ust` R function that is downloadable from github.³ Given a number of clusters, K , and a range of lag orders, the algorithm considers every combination⁴ of lag orders within this range. Every combination of lag orders corresponds to a different model. To accommodate the selection of the lag order, we extended the HQ information criterion (Quinn, 1980) which weighs model fit against time-series length and the number of time-series parameters (see Equation 3.10 in Appendix A). For a fixed number of clusters, K , the HQ_{LCVAR} indicates the model with the ideal number of lags across all starts and lag combinations. Researchers are recommended to choose among the final models, indicated by HQ_{LCVAR} , for different numbers of clusters based on interpretability and estimation convergence across the various EM starts.

3.0.4 Simulation

The quality of the estimation method to recover cluster parameters and cluster memberships is evaluated in a simulation study described in Appendix C. The results indicate excellent recovery performance under various simulated conditions. The R scripts we used to generate data, carry out the simulation and analyse the results are provided on the project's OSF page.⁵

3.1 Empirical Example

3.1.1 Methods

We showcase our model using data from the ‘How Nuts are the Dutch’ study described in detail by Van der Krieke et al. (2016). For up to 31 days, participants were assessed every morning, midday and evening on eight emotions that covered the commonly recognised dimensions of affective valence and arousal (Yik et al., 2011). This yielded scores on Positive Affect Activation (PAA; enthusiastic and cheerful), Positive Affect Deactivation (PAD; relaxed and content), Negative Affect Activation (NAA; irritable and anxious) and Negative Affect Deactivation (NAD; tired and gloomy). Each emotion was assessed using one item scored on a continuous scale from 0 to 100.

To fulfil the VAR assumption of equidistant time-points, missing data and values for each night measurement were imputed. We have imputed these observations to account for the longer time-period between successive measurements of different days that might otherwise downward-bias the VAR-coefficients. We deem this imputation procedure as sufficient for

³<https://github.com/AnieBee/LCVAR>.

⁴A combination denotes a selection where the order of the selection is not considered.

⁵https://osf.io/j3p6h/?view_only=f85d373c6be9447bbf814ac7d6db432d.

showcasing our model, in the future it could be interesting to account for these breaks by extending the hierarchical structure of the model to nest measurements within days. Data were imputed using Amelia II (Honaker et al., 2011) taking time-of-day, lags of previous and leads of following measurements into account. We consider participants who completed at least 71 measurements ($N = 410$); hence, the number of individual measurements ranged from 71 to 91 ($M = 78.97$, $SD = 4.96$).

LCVAR was estimated including time-of-day as exogenous dummy variable. Both the coefficients for exogenous variables and the VAR coefficients were constrained to be cluster-specific. The partitioning of individuals is thus based on emotion dynamic patterns and their time-of-day averages. We considered models of three to six clusters with one to three lags. For every combination of lag orders, 17 models were estimated; one based on a rational start, 15 based on pseudo-random starts and one based on the best solution achieved so far for other lag order combinations for the same number of clusters. The maximum number of EM iterations was set to 50. Tolerance for relative convergence of the estimated log likelihood was $\text{Conv} = 1 \times 10^{-7}$. The R-scripts we used to analyse the data are provided on the project's OSF page.

3.1.2 Results

From the resulting four models (i.e., three to six clusters) we chose the six cluster solution as final solution because it offered the most comprehensive summary of our data. This solution is interpreted below. The optimal solution, as indicated by the HQ_{LCVAR} , summarised time-series in each cluster via a VAR(3) model. In contrast, using ideographic analysis the HQ criterion selected a VAR(1) model for every individual time-series in our sample. Support for higher-order lag coefficients is found at the cluster-level partly because statistical power increases via the aggregation of time-series, and because accuracy of the cluster-level time-series model decreases for individual time-series through the inclusion of fixed-intercepts per cluster. This decrease in accuracy could be circumvented by centering individual's time-series at zero. We refrained here from centering because we are interested in a clustering that includes emotional intensity. A future solution could be the inclusion of random-intercepts per cluster.

All clusters showed positive autocorrelations over three lags. Given that previous work showed median duration for episodes of anxiety, irritation, relaxation, and contentment of four hours or less (Verduyn & Lavrijsen, 2015) we understand the relative stabilities we observed as emotion sequences throughout the day. The negative emotions (anxiety, gloomy, irritable) showed slightly stronger autocorrelations than the positive emotions, in line with the literature (Verduyn & Lavrijsen, 2015). In most clusters positive emotions weakly augmented the intensity of subsequent positive emotions and thus gave rise to stronger experiences; positive emotions also blunted the intensity of subsequent negative emotions (i.e., opposite valences), while the inverse was observed for negative emotions (cf. Pe & Kuppens, 2012).

Cluster 1 was coined 'average Jane and Joe' given that this cluster captured the largest proportion of participants and showed the most intermediate positive and negative emotion intensities and mood symptom levels (Figure 3.1 and Table 3.1). Henceforth cluster 1 serves as our reference category.

Cluster 2 captures an 'extended average' cluster of participants. Next to small differences in emotional intensity (e.g., feeling more tired and irritable) cluster 2 people showed virtually identical day trends and emotion dynamics compared to cluster 1. Nonetheless, slightly more pronounced increases in deactivated positive emotions in the evening (Figure 3.1) and somewhat more persistent positive emotions and blunting effects (Figure 3.2) differentiate cluster 2 from 'average Jane and Joe', such as the lagged relationship between contentment and

feeling gloomy. Positive deactivated emotions blunted subsequent positive activated emotions over three lags, which we understood as the result of a sequence of emotion-eliciting events throughout the day, and such emotional blunting may reflect other appraisal tendencies in cluster 2 than 1. Also the larger proportion of women in cluster 2 (Table 3.2) may have driven some differences between cluster 1 and 2, as women tend to experience slightly more persistent emotions than men (Verduyn & Lavrijsen, 2015), among others, and cluster 2 is also marked by more depressive symptoms (Table 3.1).

Cluster 3 captures the ‘unwinding evening people’ who feel slightly more relaxed and content in the evening relative to ‘average Jane and Joe’ and seem less irritable despite being tired (Figure 3.1). This suggests that the unwinding evening people may experience a more relaxing home environment in which they decompress from the day. Moreover, compared to people in clusters 1 and 2, people in this cluster experience less anxiety, depression, and NA (Table 3.1 and Figure 3.1), which are closely connected to irritability (Vidal-Ribas et al., 2016). Finally, the unwinding people also show slightly stronger autocorrelations for positive emotions (Figure 3.2), which suggests slower dynamics, and a stronger valence than arousal focus (see Barrett, 1998, for an elaboration).

Cluster 4 combines ‘unhappy people’ who report markedly lower intensities of positive emotions and more intense negative emotions (Figure 3.1) and heightened symptom levels of anxiety and depression (Table 3.1). The relative proportion of this unhappy cluster (9%) converges with estimates of the proportion of people with heightened mood symptom levels in the general Dutch population (De Graaf et al., 2012) and although cluster 4 is similar to the previous clusters in terms of age it comprises a slightly larger proportion of men (Table 3.2). Cluster 4 shows the highest autocorrelations for all negative emotions (Figure 3.2) and while feelings of relaxation also seem fairly stable over three lags, their experience of being cheerful was most unstable of all clusters. Emotional blunting processes seem slightly pronounced over lag 2 and 3 (Figure 3.2). Given these results cluster 4 people are likely to be characterised by high neuroticism scores (Jeronimus, 2019).

Cluster 5 captures relatively ‘happy people’ who are more often middle-aged (Table 3.2) and tend to report less intense anxiety, gloom, and irritability than people in the previous clusters (1 to 4, see Figure 3.1) and lowest symptom levels of depression combined with the highest PA and lowest NA scale scores (Table 3.1). The happy people seem to show slightly weaker emotion autocorrelations than cluster 1–4, with the exception of feeling tired over one and three lags, whereas PA augmentation processes seems slightly stronger (Figure 3.2). Overall the happy people feel least tired during the day and show the lowest intensities for gloomy, anxiety, and irritability (Figure 3.1).

Cluster 6 captures ‘morning types’ (or ‘larks’) who feel relatively relaxed and content (PAD) during the day but exhibit a characteristic drop in enthusiasm and cheerfulness (PAA) during the evening when they become tired (Figure 3.1). High positive affect combined with stable low negative emotion intensities during the day is in line with previous reports on morning people (Smith et al., 2002). The proportion ‘morning people’ who rise early and perform mentally and physically best in the morning and retire early in the evening (7%) is identical to previous estimates in population samples (Biss & Hasher, 2012), and the relatively high mean age for this cluster (Table 3.2) is also in line with this literature (Smith et al., 2002). The morning people reported similar levels of PA and NA as the happy cluster but slightly lower anxiety levels yet more symptoms of depression (Table 3.1). Daily cycles in tiredness and enthusiasm (Figure 3.1) distinguish cluster 6 from 5.

Altogether, this empirical example demonstrates the promise of LCVAR as we discerned clusters based on identifiable differences in emotional intensity levels, inertia and diurnal dynamics. Whereas cluster 1 and 2 seem to capture average participants, cluster 3 and 6 distinguish evening and morning people whereas cluster 4 and 5 contrast happy versus unhappy people. The observed proportions of some of these cluster typologies align with those

reported in the psychology literature which can be seen as a tentative method validation.

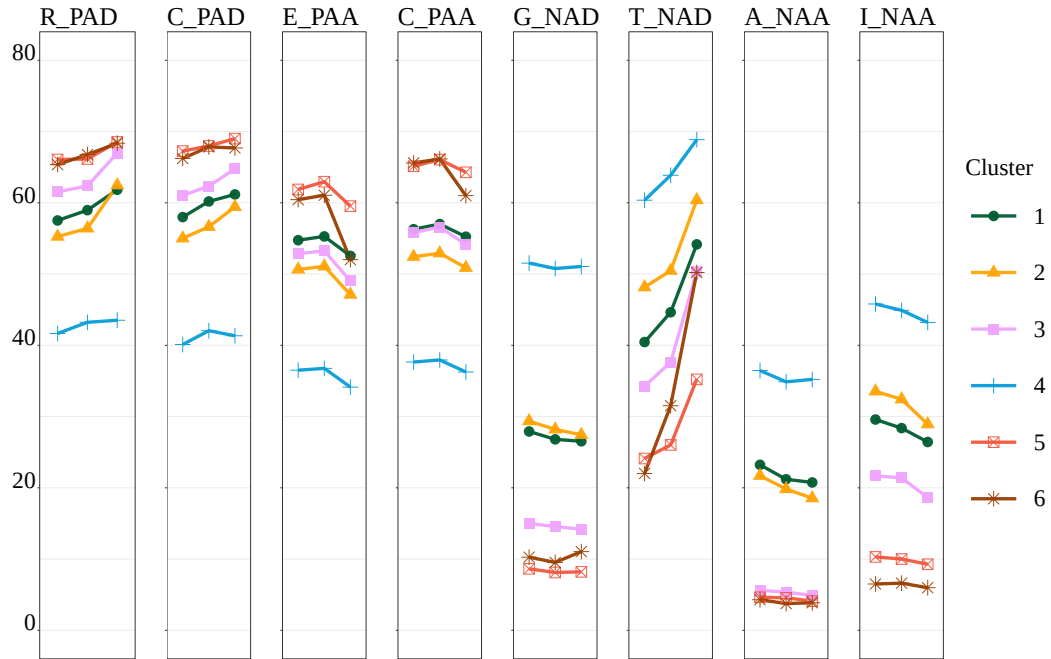


Figure 3.1: Mean values on the endogenous variables within the clusters (y-axis) for morning, midday and evening measurements (x-axis) as indicated by the cluster-wise covariate coefficients. Abbreviations denote the endogenous variables: relaxed (R_PAD), content (C_PAD), enthusiastic (E_PAA), cheerful (C_PAA), gloomy (G_NAD), tired (T_NAD), anxious (A_NAA) and irritable (I_NAA). All assessed on a scale from 0–100.

Table 3.1: Mean (SD) of the clusters on cross-sectional measures, clusters are ordered based on mean PA scores. Cluster averages are calculated using crisp cluster membership.

Cluster	Depression (0 - 27)	Anxiety/Panic (0 - 3)	PANAS PA (10 - 50)	PANAS NA (10 - 50)
5 ‘happy’	3.78 (3.06)	.38 (.49)	36.54 (4.82)	15.28 (4.09)
6 ‘morning people’	4.62 (3.37)	.24 (.51)	35.69 (4.58)	15.34 (5.04)
3 ‘unwinding evening’	4.97 (3.26)	.49 (.57)	33.94 (6.30)	17.84 (4.73)
1 ‘average Jane and Joe’	6.73 (3.80)	.97 (.82)	31.34 (6.56)	22.46 (6.54)
2 ‘extended average’	8.23 (4.70)	.96 (.79)	31.05 (6.98)	22.99 (7.01)
4 ‘unhappy’	12.14 (4.91)	1.64 (1.05)	25.89 (7.04)	28.61 (8.09)

3.2 Discussion

In this paper we introduced latent class vector-autoregressive modelling (LCVAR) for identifying similar processes across individuals in time-intensive AA data. LCVAR captures the inter-individual differences at the sub-group level. A simulation study to evaluate the algorithm to estimate LCVAR showed excellent recovery of cluster membership and cluster parameters under a range of empirically relevant conditions. A satisfactory performance required a substantial number of observations (i.e., over 50); further performance was ameliorated by large differences between cluster parameters.

Table 3.2: Mean (SD) of the clusters on demographic variables and cluster proportions, clusters are ordered based on mean PA scores. Cluster averages on demographic variables are calculated using crisp cluster membership.

Cluster	% Male	Age	Proportion ($\hat{\tau}_k$)
5 ‘happy’	.32 (.47)	48.90 (13.46)	.12
6 ‘morning people’	.31 (.47)	52.72 (13.09)	.07
3 ‘unwinding evening’	.15 (.36)	39.06 (12.34)	.26
1 ‘average Jane and Joe’	.17 (.38)	38.90 (12.81)	.28
2 ‘extended average’	.10 (.30)	33.75 (11.76)	.18
4 ‘unhappy’	.31 (.47)	42.08 (12.56)	.09

In our empirical data, LCVAR distilled six groups of individuals exhibiting distinct emotion dynamic patterns. The adaptive nature of LCVAR was useful, in that higher-order lag coefficients were identified for some clusters. Had time-series been summarised in an ideographic manner, they would have been presented with a lower lag order. Because support for the higher lag orders was only found after aggregating the data, this result could not have been achieved had the time-series been summarised independently of the clustering process (i.e., using a filtering method).

In our LCVAR solution, it appeared that in various clusters, many cross-regressive effects were close to zero. Though this might be indicating a lack of a clear cluster pattern, we interpret them in keeping with the notion that cross-regressive effects found through ideographic time-series analysis are often the result of overfitting (Bulteel et al., 2018). In this paper we have only considered point estimates of LCVAR parameters. Because LCVAR is a model-based time-series clustering method (i.e., based on a mixture model) it provides a framework for formal inference, and estimates of parameter standard errors (and thereby confidence intervals) can be calculated via derivatives of the \mathcal{Q}^* function given in Equation 3.8 in Appendix A (McLachlan & Peel, 2004).

Latent class models should be applied only when distinct subgroups are to be expected in the data, for instance due to qualitatively different subtypes. If this or other assumptions of the model are not met, LCVAR can retrieve illusory subgroups. For instance, because non-normality is used to infer the presence of latent classes, spurious latent classes may arise due to non-normality in the data (Bauer, 2007).

In sum, LCVAR proves a promising strategy to cluster intensive time-series data on within-person dynamics as demonstrated in our empirical study of 410 participants who each provided over 70 measurements in which six clusters of emotion dynamics were identified. Although LCVAR offers a summary of time-series that accounts for inter-individual differences, the model does not yet capture variation within clusters beyond the covariance. We are currently working on an extension of LCVAR that includes random cluster coefficients. In the future, it will be interesting to go beyond the cluster-level perspective by accounting for individual differences and within-cluster variation to a greater extent.

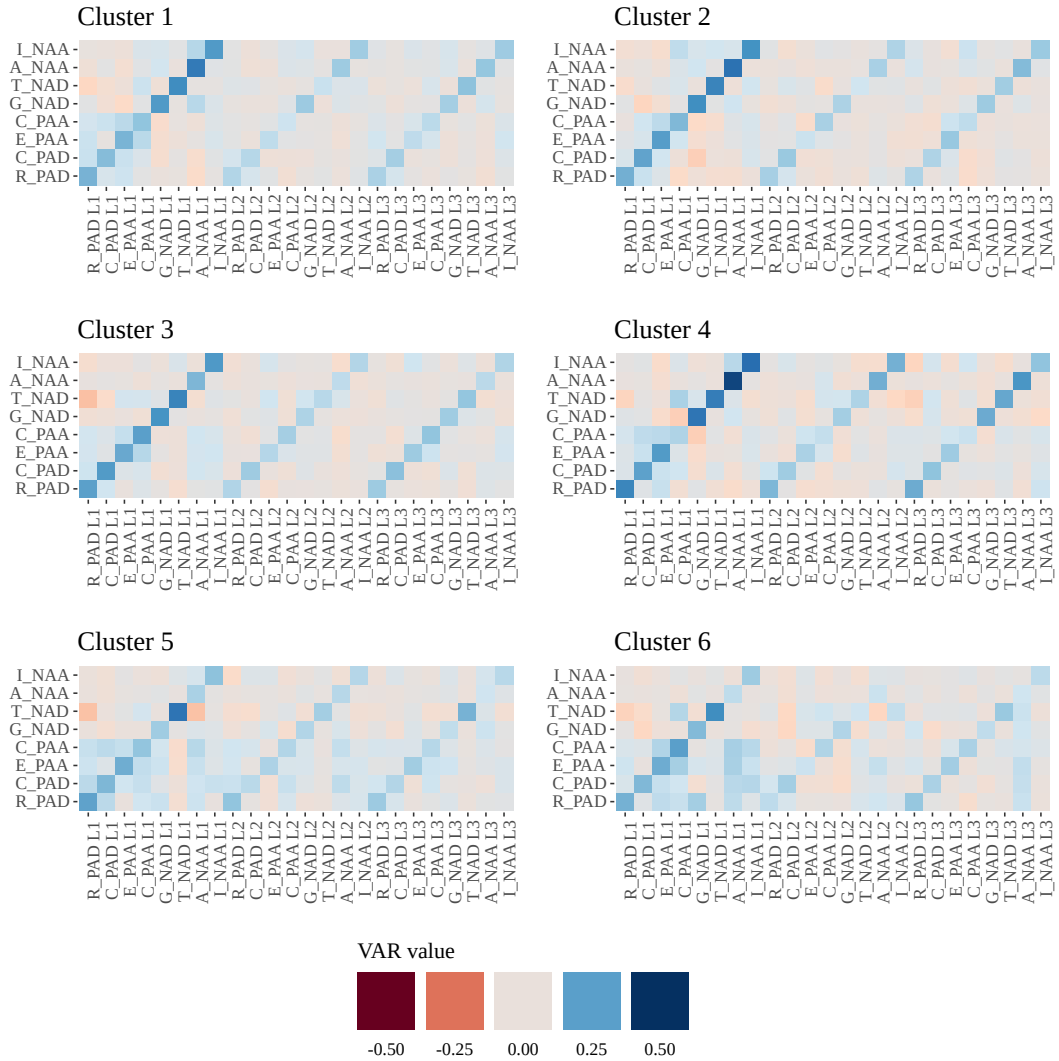


Figure 3.2: Emotion dynamics within the clusters as indicated by the cluster-wise VAR(3) coefficients. Abbreviations denote the endogenous variables: relaxed (R_PAD), content (C_PAD), enthusiastic (E_PAA), cheerful (C_PAA), gloomy (G_NAD), tired (T_NAD), anxious (A_NAA) and irritable (I_NAA). Lag 1 coefficients are denoted by L1, lag 2 by L2 and lag 3 by L3. Lag coefficients quantify the influence the endogenous variables (columns) have on the endogenous variables (rows) at 1, 2 or 3 time points in the future.

3.A Appendix: EM algorithm

To make inferences about the component distributions we aim to maximise the log likelihood given in Equation 3.6. Because the maximisation of Equation 3.6 is not straightforward, the log likelihood of the complete data (i.e., of observed data \mathbf{Y} and unobserved cluster membership \mathbf{Z}) is utilised. Cluster membership is unobserved, thus the expected value of the complete data log likelihood under the posterior distribution of cluster membership is maximised. Using the time-series representation in Equation 3.2 we condition on the p previous observations,

$$\begin{aligned} \mathcal{Q}(\Theta^{(b)}; \Theta^{(b-1)}) &= \mathbb{E}_{\mathbf{Z}|\mathbf{Y}, \Theta^{(b-1)}} [\log(\mathcal{L}(\Theta^{(b)}; \mathbf{Y}, \mathbf{Z}))] = \\ & \sum_{i=1}^N \sum_{k=1}^K \pi_{ik}^{(b)} \left(\log(\tau_k) + \sum_{t=1}^{T_{ik}} \log \left(\Phi(\mathbf{y}_{it}; \boldsymbol{\mu}_{kt}(\mathbf{A}_k, \mathbf{B}_k), \boldsymbol{\Sigma}_k) \right) \right), \end{aligned} \quad (3.7)$$

with (b) indicating the iteration of the EM algorithm; $\boldsymbol{\mu}_{kt}(\mathbf{A}_k, \mathbf{B}_k)$ representing the conditional mean vector; $\boldsymbol{\Sigma}_k$ as defined in Equation 3.5; $\Phi(\mathbf{y}_{it}; \boldsymbol{\mu}_{kt}(\mathbf{A}_k, \mathbf{B}_k), \boldsymbol{\Sigma}_k)$ denotes the conditional density function $\Phi(\mathbf{y}_{it} | \mathbf{y}_{i,t-1}, \dots, \mathbf{y}_{i,t-p_k}, \mathbf{x}_{it}, \dots, \mathbf{x}_{i,t-p_k}; \boldsymbol{\mu}_{kt}(\mathbf{A}_k, \mathbf{B}_k), \boldsymbol{\Sigma}_k)$, in line with the notation employed by Michael and Melnykov (2016).

The likelihood of observing a VAR time-series \mathbf{Y} of individual i belonging to cluster k of length T_i with m endogenous variables equals

$\mathcal{L}(\mathbf{A}_k, \mathbf{B}_k, \boldsymbol{\Sigma}_k; \mathbf{Y}) = 2\pi^{-T_i m/2} \det(\boldsymbol{\Sigma}_k)^{-T_i/2} \exp\{-\sum_{t=1}^{T_i} \frac{1}{2}(\mathbf{u}_{kt}^\top \boldsymbol{\Sigma}_k^{-1} \mathbf{u}_{kt})\}$ with \mathbf{u}_{kt} indicating the deviation between \mathbf{y}_t and the conditional mean vector $\boldsymbol{\mu}_{kt}(\mathbf{A}_k, \mathbf{B}_k)$, $\mathbf{u}_{kt} = \mathbf{y}_t - \mathbf{B}_k \mathbf{x}_t - (\sum_{a=1}^p \mathbf{A}_{ka}(\mathbf{y}_{t-a} - \mathbf{B}_k \mathbf{x}_{t-a})) = \mathbf{w}_{kt} - (\sum_{a=1}^p \mathbf{A}_{ka} \mathbf{w}_{k,t-a})$, as defined in Equations 3.2 to 3.5 (for a general proof of this likelihood for a VAR model that does not include clusters and exogenous variables, see Equation 3.4.5 in Lütkepohl, 2005).⁶ Employing Equation 3.7 and the likelihood of observing a VAR time-series, the resulting \mathcal{Q}^* function is

$$\begin{aligned} \mathcal{Q}^*(\Theta^{(b)}; \Theta^{(b-1)}) &= \sum_{i=1}^N \sum_{k=1}^K \pi_{ik}^{(b)} \left(\log(\tau_k) - \frac{m T_{ik}}{2} \log(2\pi) + \frac{T_{ik}}{2} \log(\det(\boldsymbol{\Sigma}_k^{-1})) \right. \\ & \left. - \frac{1}{2} \left(\sum_{t=1}^{T_{ik}} \left(\mathbf{w}_{ikt} - \left(\sum_{a=1}^{p_k} \mathbf{A}_{ka} \mathbf{w}_{i,k,t-a} \right) \right)^\top \boldsymbol{\Sigma}_k^{-1} \left(\mathbf{w}_{ikt} - \left(\sum_{a=1}^{p_k} \mathbf{A}_{ka} \mathbf{w}_{i,k,t-a} \right) \right) \right) \right), \end{aligned} \quad (3.8)$$

with $\boldsymbol{\Sigma}_k$ of size $m \times m$, $\mathbf{w}_{ikt} = \mathbf{y}_{it} - \mathbf{B}_k \mathbf{x}_{it}$ as defined in Equation 3.1 and

$\Theta = \{\tau_1, \dots, \tau_K, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_K, \mathbf{A}_1, \dots, \mathbf{A}_K, \mathbf{B}_1, \dots, \mathbf{B}_K\}$. Here * indicates that the probability of a pre-sample, containing the first p_k observations ($t = -p_k + 1, \dots, 0$), is neglected. By maximising Equation 3.8 for a given number of clusters and lag orders, parameter estimates can be obtained. These parameter estimates are given in Equations 3.11 – 3.14, and their proofs are given in Equations 3.15 – 3.18. Posterior probabilities of cluster membership, $\boldsymbol{\pi}$, can be found with Bayes' rule given \mathbf{Y} and an initialisation $\Theta^{(0)}$

$$\pi_{ik}^{(b)} = \frac{\tau_k^{(b-1)} \prod_{t=1}^{T_{ik}} \Phi(\mathbf{y}_{it}; \boldsymbol{\mu}_{kt}(\mathbf{A}_k^{(b-1)}, \mathbf{B}_k^{(b-1)}), \boldsymbol{\Sigma}_k^{(b-1)})}{\sum_{k'=1}^K \tau_{k'}^{(b-1)} \prod_{t=1}^{T_{ik'}} \Phi(\mathbf{y}_{it}; \boldsymbol{\mu}_{k't}(\mathbf{A}_{k'}^{(b-1)}, \mathbf{B}_{k'}^{(b-1)}), \boldsymbol{\Sigma}_{k'}^{(b-1)})}. \quad (3.9)$$

⁶This corresponds to the likelihood of observing a sample \mathbf{Y} of size T_i from an m -variate normal distribution, $\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{Y}) = 2\pi^{-T_i m/2} \det(\boldsymbol{\Sigma})^{-T_i/2} \exp\{-\sum_{t=1}^{T_i} \frac{1}{2}(\mathbf{y}_t - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_t - \boldsymbol{\mu})\}$.

The EM algorithm proceeds as follows. In the b th iteration of the algorithm posterior probabilities based on $\Theta^{(b-1)}$ are determined via Equation 3.9 (E-step). Subsequently, parameter estimates in $\Theta^{(b)}$ are updated in such a way as to maximise the \mathcal{Q}^* -function in Equation 3.8 (M-step) using Equations 3.11–3.14 in succession. Thus, the K matrices $\mathbf{A}_k^{(b)}$ are updated based on $\mathbf{w}_{i k t}^{(b)} = \mathbf{y}_{i t} - \mathbf{B}_k^{(b-1)} \mathbf{x}_{i t}$; then the K matrices $\Sigma_k^{(b)}$ are updated based on $\mathbf{u}_{i k t}^{(b)} = \mathbf{w}_{i k t}^{(b)} - \mathbf{A}_k^{(b-1)} \tilde{\mathbf{w}}_{i, k, t-1}^{(b)}$; subsequently the K matrices $\mathbf{B}_k^{(b)}$ are updated based on $\mathbf{A}_k^{(b)}$ and $\Sigma_k^{(b)}$; lastly $\boldsymbol{\tau}^{(b)}$ is updated based on $\boldsymbol{\pi}^{(b)}$ determined in the E-step. E-steps and M-steps are repeated sequentially until convergence. Relative convergence is determined when $\frac{\log[\mathcal{L}(\Theta^{(b)}; \mathbf{Y})] - \log[\mathcal{L}(\Theta^{(b-1)}; \mathbf{Y})]}{|\log[\mathcal{L}(\Theta^{(b-1)}; \mathbf{Y})]|} < \text{Conv}$, with $\mathcal{L}(\Theta^{(b)}; \mathbf{Y})$ as defined in Equation 3.6.

3.A.1 Initialisations for the EM algorithm

The computer implementation of our EM algorithm uses three distinct start strategies to achieve initialisations $\Theta^{(0)}$. Rational starts are based on the k -means partitioning of individuals' ideographic VAR and ideographic covariate coefficients. For pseudo-random starts K individuals are randomly selected as cluster centres. Then individuals are partitioned into the cluster to which their ideographic VAR and ideographic covariate coefficients are closest. This crisp partitioning of individuals into clusters serves as posterior probabilities in the calculation of initial parameter estimates. Estimates for prior probabilities, τ_k , are determined using Equation 3.14. The covariate coefficients, \mathbf{B}_k , are initially approximated by the least squares regression coefficients of the exogenous variables. Based on these estimates of \mathbf{B}_k , initial estimates for the VAR matrices, \mathbf{A}_k and the within-cluster covariance matrices, Σ_k , are estimated using Equation 3.11 and Equation 3.12. Finally, estimates of the covariate coefficients, \mathbf{B}_k , can be made based on Equation 3.13. These estimates for τ_k , \mathbf{A}_k , Σ_k and \mathbf{B}_k are used to initialise the EM algorithm. Equations 3.14, 3.11, 3.12 and 3.13 are given below. The computer implementation of the algorithm includes safeguards to prevent loss of precision, components collapsing onto a single individual or singular covariance matrices. These assessments are listed in Appendix B.

To estimate models efficiently a third type of start is implemented based on the best solution achieved so far for other lag order combinations for the same number of clusters (as indicated by the HQ_{LCVAR} defined in Equation 3.10). This start uses the classification reached by the best solution as posterior probabilities in the calculation of initial parameter estimates. The estimates for all initial \mathbf{A}_k are based on the highest lag order considered in the current EM estimation. Estimation is accelerated by assigning clusters a lag order depending on their initial estimates for \mathbf{A}_k before the first EM iteration.⁷ All subsequent estimates are based on the assigned lag orders fixed for the clusters, while individuals' probabilistic memberships switch between clusters.

3.A.2 Model selection

LCVAR models can be compared with regard to their log likelihood as given in Equation 3.6. We do not use likelihood based information criteria such as the BIC for model selection, however, because various time-series models for a fixed number of clusters can differ in their effective number of parameters and the effective number of time points used to calculate these parameters. Rather, to accommodate selection of the lag order, we extended the HQ information criterion which weighs forecasting precision against time-series length and the number of time-series parameters (Quinn, 1980). For a fixed number of clusters, K , the

⁷For instance, if a combination of two lag 1 a lag 2 and a lag 4 cluster are considered, the cluster with the highest lag 3 and lag 4 coefficients are assigned 4 lags, the cluster with the highest lag 2 coefficients of the remaining clusters is assigned 2 lags, the remaining clusters are assigned 1 lag.

HQ_{LCVAR} indicates the optimal model across all starts and lag combinations, selecting thus the model with the ideal number of lags,

$$HQ_{LCVAR} = \sum_{k=1}^K \tau_k \left(\log(\det(\boldsymbol{\Sigma}_k)) + \frac{2p_k(m^2) \log(\log(\sum_{i=1}^N \pi_{ik} T_{ik}))}{\sum_{i=1}^N \pi_{ik} T_{ik}} \right). \quad (3.10)$$

To consider a collection of cluster numbers, the algorithm estimates models for every number of clusters, K , within a specified selection of cluster numbers. An optimal model for each value of K , based on the HQ_{LCVAR} , is determined across several starts and lag combinations. Researchers are recommended to choose among these models based on interpretability and estimation convergence across the various EM starts.

3.A.3 Estimation of the model

The estimators of vector-autoregressive coefficients are

$$\mathbf{A}_k^{(b)} = \sum_{i=1}^N \pi_{ik}^{(b)} \left(\sum_{t=1}^{T_{ik}} \mathbf{w}_{ikt}^{(b)} \tilde{\mathbf{w}}_{i,k,t-1}^{(b)\top} \right) \left(\sum_{i=1}^N \pi_{ik}^{(b)} \left(\sum_{t=1}^{T_{ik}} \tilde{\mathbf{w}}_{i,k,t-1}^{(b)} \tilde{\mathbf{w}}_{i,k,t-1}^{(b)\top} \right) \right)^{-1}, \quad (3.11)$$

(see Equation 3.15) with $m \times mp_k$ matrix $\mathbf{A}_k^{(b)} = \begin{pmatrix} \mathbf{A}_{k1}^{(b)} & \dots & \mathbf{A}_{k p_k}^{(b)} \end{pmatrix}$ and $mp_k \times 1$ vector $\tilde{\mathbf{w}}_{ikt}^{(b)} = \begin{pmatrix} \mathbf{w}_{ikt}^{(b)} & \dots & \mathbf{w}_{i,k,t-p_k+1}^{(b)} \end{pmatrix}^\top$. It follows from Equation 3.1 that $\mathbf{w}_{ikt}^{(b)} = \mathbf{y}_{it} - \mathbf{B}_k^{(b-1)} \mathbf{x}_{it}$ and $\mathbf{u}_{ikt}^{(b)} = \mathbf{w}_{ikt}^{(b)} - \mathbf{A}_k^{(b-1)} \tilde{\mathbf{w}}_{i,k,t-1}^{(b)}$ where (b) indicates the iteration of the EM algorithm. The estimates of the covariance matrices can be obtained by

$$\boldsymbol{\Sigma}_k^{(b)} = \sum_{i=1}^N \pi_{ik}^{(b)} \left(\sum_{t=1}^{T_{ik}} \mathbf{u}_{ikt}^{(b)} \mathbf{u}_{ikt}^{(b)\top} \right) \left(\sum_{i=1}^N \pi_{ik}^{(b)} T_{ik} \right)^{-1}, \quad (3.12)$$

(see Equation 3.16). The derivative of Equation 3.8 with respect to the coefficient matrices \mathbf{B}_k is found by defining a $q \times (p_k + 1)$ matrix $\tilde{\mathbf{x}}_{ikt} = \begin{pmatrix} \mathbf{x}_{it} & \dots & \mathbf{x}_{i,t-p_k} \end{pmatrix}$, a $(mp_k + m) \times 1$ vector $\tilde{\mathbf{y}}_{ikt} = \begin{pmatrix} \mathbf{y}_{it} & \dots & \mathbf{y}_{i,t-p_k} \end{pmatrix}^\top$ and a $m \times (mp_k + m)$ matrix $\boldsymbol{\delta}_k^{(b)} = \begin{pmatrix} \mathbf{I}_m & -\mathbf{A}_k^{(b)} \end{pmatrix}$ where \mathbf{I}_m indicates the identity matrix of size m . The partial derivative leads to

$$\text{vec}(\mathbf{B}_k)^{(b)} = \left(\sum_{i=1}^N \pi_{ik}^{(b)} \left(\sum_{t=1}^{T_{ik}} \left(\tilde{\mathbf{x}}_{ikt}^\top \otimes \mathbf{I}_m \right)^\top \boldsymbol{\delta}_k^{(b)\top} \boldsymbol{\Sigma}_k^{-1(b)} \boldsymbol{\delta}_k^{(b)} \left(\tilde{\mathbf{x}}_{ikt}^\top \otimes \mathbf{I}_m \right) \right) \right)^{-1} \sum_{i=1}^N \pi_{ik}^{(b)} \left(\sum_{t=1}^{T_{ik}} \left(\tilde{\mathbf{x}}_{ikt}^\top \otimes \mathbf{I}_m \right)^\top \boldsymbol{\delta}_k^{(b)\top} \boldsymbol{\Sigma}_k^{-1(b)} \boldsymbol{\delta}_k^{(b)} \tilde{\mathbf{y}}_{ikt} \right), \quad (3.13)$$

(see Equation 3.17) where \otimes denotes the Kronecker product and $\text{vec}()$ represents the vec operator of vectorisation. With reference to the prior probabilities, constraining all probabilities to sum to 1, the partial derivative of Equation 3.8 leads to the following estimator

$$\tau_k^{(b)} = \frac{\sum_{i=1}^N \pi_{ik}^{(b)}}{N}, \quad (3.14)$$

(see Equation 3.18).

$$\begin{aligned}
\frac{\delta \mathcal{Q}^*(\Theta^{(b)}; \Theta^{(b-1)})}{\delta \mathbf{A}_k} &= \sum_{i=1}^N \pi_{ik} \left(\sum_{t=1}^{T_{ik}} \Sigma_k^{-1} (\mathbf{w}_{ik t} - \mathbf{A}_k \tilde{\mathbf{w}}_{i,k,t-1}) \right. \\
&\quad \left. \tilde{\mathbf{w}}_{i,k,t-1}^\top \right) \quad (3.15) \\
0 &= \sum_{i=1}^N \pi_{ik} \left(\sum_{t=1}^{T_{ik}} \Sigma_k^{-1} (\mathbf{w}_{ik t} - \mathbf{A}_k \tilde{\mathbf{w}}_{i,k,t-1}) \tilde{\mathbf{w}}_{i,k,t-1}^\top \right) \\
\sum_{i=1}^N \pi_{ik} \left(\sum_{t=1}^{T_{ik}} \mathbf{A}_k \tilde{\mathbf{w}}_{i,k,t-1} \tilde{\mathbf{w}}_{i,k,t-1}^\top \right) &= \sum_{i=1}^N \pi_{ik} \left(\sum_{t=1}^{T_{ik}} \mathbf{w}_{ik t} \tilde{\mathbf{w}}_{i,k,t-1}^\top \right) \\
\mathbf{A}_k &= \sum_{i=1}^N \pi_{ik} \left(\sum_{t=1}^{T_{ik}} \mathbf{w}_{ik t} \tilde{\mathbf{w}}_{i,k,t-1}^\top \right) \\
&\quad \left(\sum_{i=1}^N \pi_{ik} \left(\sum_{t=1}^{T_{ik}} \tilde{\mathbf{w}}_{i,k,t-1} \tilde{\mathbf{w}}_{i,k,t-1}^\top \right) \right)^{-1}.
\end{aligned}$$

$$\begin{aligned}
\mathcal{Q}^*(\Theta^{(b)}; \Theta^{(b-1)}) &= \sum_{i=1}^N \sum_{k=1}^K \pi_{ik} \left(\log(\tau_k) - \frac{m T_{ik}}{2} \log(2\pi) + \frac{T_{ik}}{2} \log(\det(\Sigma_k^{-1})) \right. \\
&\quad \left. - \frac{1}{2} \sum_{t=1}^{T_{ik}} \mathbf{u}_{ik t}^\top \Sigma_k^{-1} \mathbf{u}_{ik t} \right) \quad (3.16) \\
&= \sum_{i=1}^N \sum_{k=1}^K \pi_{ik} \left(\log(\tau_k) - \frac{m T_{ik}}{2} \log(2\pi) + \frac{T_{ik}}{2} \log(\det(\Sigma_k^{-1})) \right. \\
&\quad \left. - \frac{1}{2} \sum_{t=1}^{T_{ik}} \text{tr} \left\{ \mathbf{u}_{ik t} \mathbf{u}_{ik t}^\top \Sigma_k^{-1} \right\} \right) \\
\frac{\delta \mathcal{Q}^*(\Theta^{(b)}; \Theta^{(b-1)})}{\delta \Sigma_k^{-1}} &= \sum_{i=1}^N \pi_{ik} \left(\frac{T_{ik}}{2} \Sigma_k^\top - \frac{1}{2} \sum_{t=1}^{T_{ik}} \mathbf{u}_{ik t} \mathbf{u}_{ik t}^\top \right) \\
0 &= \sum_{i=1}^N \pi_{ik} \left(\frac{T_{ik}}{2} \Sigma_k^\top - \frac{1}{2} \sum_{t=1}^{T_{ik}} \mathbf{u}_{ik t} \mathbf{u}_{ik t}^\top \right) \\
\Sigma_k^\top = \Sigma_k &= \sum_{i=1}^N \pi_{ik} \left(\sum_{t=1}^{T_{ik}} \mathbf{u}_{ik t} \mathbf{u}_{ik t}^\top \right) \left(\sum_{i=1}^N \pi_{ik} T_{ik} \right)^{-1}.
\end{aligned}$$

$$\begin{aligned}
\frac{\delta \mathcal{Q}^*(\Theta^{(b)}; \Theta^{(b-1)})}{\delta \text{vec}(\mathbf{B}_k)} &= \sum_{i=1}^N \pi_{ik} \left(\sum_{t=1}^{T_{ik}} \left(\tilde{\mathbf{x}}_{ikt}^\top \otimes \mathbf{I}_m \right)^\top \boldsymbol{\delta}_k^\top \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\delta}_k \right. \\
&\quad \left. \left(\tilde{\mathbf{y}}_{ikt} - \left(\tilde{\mathbf{x}}_{ikt}^\top \otimes \mathbf{I}_m \right) \text{vec}(\mathbf{B}_k) \right) \right) \\
0 &= \sum_{i=1}^N \pi_{ik} \left(\sum_{t=1}^{T_{ik}} \left(\tilde{\mathbf{x}}_{ikt}^\top \otimes \mathbf{I}_m \right)^\top \boldsymbol{\delta}_k^\top \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\delta}_k \right. \\
&\quad \left. \left(\tilde{\mathbf{y}}_{ikt} - \left(\tilde{\mathbf{x}}_{ikt}^\top \otimes \mathbf{I}_m \right) \text{vec}(\mathbf{B}_k) \right) \right) \\
\sum_{i=1}^N \pi_{ik} \left(\sum_{t=1}^{T_{ik}} \left(\tilde{\mathbf{x}}_{ikt}^\top \otimes \mathbf{I}_m \right)^\top \boldsymbol{\delta}_k^\top \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\delta}_k \tilde{\mathbf{y}}_{ikt} \right) &= \sum_{i=1}^N \pi_{ik} \left(\sum_{t=1}^{T_{ik}} \left(\tilde{\mathbf{x}}_{ikt}^\top \otimes \mathbf{I}_m \right)^\top \boldsymbol{\delta}_k^\top \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\delta}_k \right. \\
&\quad \left. \left(\tilde{\mathbf{x}}_{ikt}^\top \otimes \mathbf{I}_m \right) \text{vec}(\mathbf{B}_k) \right) \\
\text{vec}(\mathbf{B}_k) &= \left(\sum_{i=1}^N \pi_{ik} \left(\sum_{t=1}^{T_{ik}} \left(\tilde{\mathbf{x}}_{ikt}^\top \otimes \mathbf{I}_m \right)^\top \boldsymbol{\delta}_k^\top \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\delta}_k \left(\tilde{\mathbf{x}}_{ikt}^\top \otimes \mathbf{I}_m \right) \right) \right)^{-1} \\
&\quad \sum_{i=1}^N \pi_{ik} \left(\sum_{t=1}^{T_{ik}} \left(\tilde{\mathbf{x}}_{ikt}^\top \otimes \mathbf{I}_m \right)^\top \boldsymbol{\delta}_k^\top \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\delta}_k \tilde{\mathbf{y}}_{ikt} \right).
\end{aligned} \tag{3.17}$$

Using a Lagrange multiplier λ to impose: $\sum_{k=1}^K \tau_k = 1$,

$$\begin{aligned}
\frac{\delta \mathcal{Q}^*(\Theta^{(b)}; \Theta^{(b-1)})}{\delta \tau_k} &= \frac{\sum_{i=1}^N \pi_{ik}}{\tau_k} + \lambda \\
0 &= \frac{\sum_{i=1}^N \pi_{ik}}{\tau_k} - N \\
\tau_k &= \frac{\sum_{i=1}^N \pi_{ik}}{N}.
\end{aligned} \tag{3.18}$$

3.B Appendix: Assessments during the EM algorithm

To prevent singular covariance matrices and single-individual clusters, the cluster memberships are reset in the E-step when components appear to collapse onto a single individual. That is, when crisp cluster membership indicates a cluster of less than three individuals, the posterior probabilities of three random individuals are set to 1.01 for this cluster and the posterior probabilities are subsequently scaled to sum to 1 for every individual. Also, the covariance matrix of the cluster is reset by increasing each element by 10. To combat loss of precision during the calculation of posterior probabilities and log likelihoods – at the end of the E-step and the M-step, respectively – the centre of the exponential sum is shifted during these calculations, forcing the largest value to zero. The computer implementation of the algorithm includes several additional checks to account for loss of precision, though the two checks mentioned above rendered these additional assessments unnecessary; no estimation problems requiring the additional checks were encountered during the simulation study and the empirical analysis. The following additional examinations were performed. If during the E-step the likelihood of a time-series $\left(\prod_{t=1}^{T_{ik}} \Phi(\mathbf{y}_{it}; \boldsymbol{\mu}_{kt}(\mathbf{A}_k^{(b-1)}, \mathbf{B}_k^{(b-1)}), \boldsymbol{\Sigma}_k^{(b-1)})\right)$ is indeterminate due to underflow, the likelihood is replaced with the mean likelihood of all time-series. After the E-step, if any posterior probabilities of cluster membership are indeterminate or tend to infinite values, posterior probabilities are set to $1/K$. Whenever posterior probabilities are reset, they are subsequently scaled to sum to 1 across every individual. After the M-step, if any of the data log likelihoods $\left(\log \left[\sum_{k=1}^K \tau_k^{(b)} \prod_{t=1}^{T_{ik}} \Phi(\mathbf{y}_{it}; \boldsymbol{\mu}_{kt}(\mathbf{A}_k^{(b)}, \mathbf{B}_k^{(b)}), \boldsymbol{\Sigma}_k^{(b)}) \right]\right)$ is effected by underflow, it is reset with a value of -9000 . Because the EM algorithm for multivariate normal mixtures breaks down for singular covariance matrices, the diagonal of the estimate of the covariance matrix is increased by a value of .01 when the covariance of a component density appears to be singular by a threshold value of 1×10^{-200} . Whenever posterior probabilities, likelihoods, or covariance matrices had to be reset, a warning is issued and convergence within the next two EM iterations is prohibited.

3.C Appendix: Simulation

We will assess the performance of LCVAR under perfect model specification through the following simulation. Performance will be evaluated with regard to recovery, that is, the degree to which the cluster membership and the parameters of the VAR matrices (\mathbf{A}_k) can be retrieved. We will investigate to what extent the recovery depends on five data characteristics. Three of these factors pertain to the partition of individuals: (1) the number of clusters, (2) the relative cluster sizes and (3) the distance between clusters relative to the distance within clusters. Two factors reflect efficiency: (4) the number of lags, and (5) the number of time points. The simulation was carried out in R (R Core Team, 2018). The R scripts we used to generate data, carry out the simulation and analyse the results are provided on the project's OSF page.⁸

3.C.1 Design and procedure

The following data characteristics were kept constant. The covariance matrix of the white noise series, $\boldsymbol{\Sigma}_k$, equalled $\mathbf{I}_m + .5$ in every cluster. The number of variables was set to $m = 4$, the total number of individuals was set to $N = 120$, the number of exogenous variables was set to $q = 4$, consisting of an intercept, one categorical variable with three levels and one

⁸https://osf.io/j3p6h/?view_only=f85d373c6be9447bbf814ac7d6db432d

continuous variable. The influence of exogenous variables was equal across clusters with

$$\mathbf{B}_k = \begin{pmatrix} 0 & 2 & 3 & .2 \\ 0 & 2 & 3 & .4 \\ 0 & 2 & 3 & .6 \\ 0 & 2 & 3 & .8 \end{pmatrix},$$

$k = 1, \dots, K$, where the first column indicates the intercept, the second and third column represent the influence of the categorical variable and the fourth column shows the influence of the continuous variable. To generate data the five factors introduced above were varied in a completely crossed design.

1. The number of clusters: two or four.
2. The relative size of the clusters:
 - ◇ *equal distribution* of participants across clusters;
 - ◇ *majority condition*, with one cluster containing 60% of individuals and the remaining individuals evenly distributed over the other cluster(s).
3. The distance between the different clusters relative to the distance within clusters. Cluster distance was expressed through the Euclidean distance between VAR matrices of the different clusters, within cluster variation was expressed through the white noise covariance matrix Σ_k :
 - ◇ *small distance*, where the Euclidean distance between any two VAR matrices, \mathbf{A}_k , equalled $\sqrt{72/625} \approx .339$;⁹
 - ◇ *large distance*, where the Euclidean distance between any two VAR matrices, \mathbf{A}_k , equalled $\sqrt{.32} \approx .566$.¹⁰

In every condition the coefficients for one VAR matrix were randomly sampled from $U[.5, .7]$ (autoregressive coefficients) or $U[-.4, .4]$ (cross-regressive coefficients). If a lag(2) process was generated, lag 2 coefficients were randomly sampled from $U[-.2, .2]$ (auto- and cross-regressive coefficients). Eight equal deviations were added to this matrix to create VAR matrices for all other clusters in the condition. Coefficients were sampled until all VAR matrices in the condition were stationary with their reverse characteristic polynomial having no roots in or on the complex circle.

4. The number of lags p : one or two.
5. The number of time points per person T_i : 50 or 150.

For each of the resulting 32 conditions we simulated 15 data sets. Time series were generated according to Equations 3.1 and 3.2. The value of the continuous exogenous variable was randomly drawn from $N(20, 20)$ for every individual at every time point. The value of the categorical exogenous variable switched to the next category at every measurement. For each data set the appropriate time-series model (VAR(1) or VAR(2)) with cluster-specific covariate coefficients was estimated for the correct number of clusters.

⁹Resulting in a mean absolute difference between parameters of .06 for VAR(1) time-series, and a mean absolute difference of .03 for VAR(2) time-series.

¹⁰Indicating a mean absolute difference between parameters of .1 in a VAR(1) time-series, and a mean absolute difference of .05 for VAR(2) time-series.

Models were estimated using the EM algorithm as described in Appendix A. The following thresholds were set during the simulation. In contrast to the procedure employed during the model estimation of empirical data, cluster memberships were reset in the E-step when components appeared to collapse onto a single individual by a threshold of 1×10^{-15} . In this case posterior probabilities below 1×10^{-15} were set to $1/K$, but covariance matrices were not reset after a collapse. Tolerance for relative convergence of the estimated log likelihood equalled 1×10^{-7} . The maximum number of EM iterations was set to 25. The algorithm was initialised with 1 rational start and 10 pseudo-random starts. The best-fitting solution was indicated by the likelihood.¹¹ For each resulting model the crisp cluster membership and the cluster VAR slopes were considered.

3.C.2 Simulation results

To determine the performance of LCVAR we investigated the recovery of (1) the VAR coefficients and (2) the cluster membership.

Recovery of VAR coefficients The recovery of VAR coefficients was determined as the mean absolute differences (MAD) between estimated and true VAR coefficients. VAR coefficients were recovered well with an average MAD of .016 ($SD = .008$) and small differences across simulation conditions; the mean MAD per condition are given in Table 3.3. Effect sizes of the manipulated factors were determined with an ANOVA including all manipulated factors and all possible two-way interactions between factors. Effect sizes in Table 3.4 show the recovery of VAR coefficients improved particularly with a high number of observations ($\hat{\eta}_p^2 = .726$), a low number of clusters ($\hat{\eta}_p^2 = .476$) and a low number of lags ($\hat{\eta}_p^2 = .193$).

Table 3.3: MAD and ARI means (SD) across simulation conditions. MAD denotes the mean absolute differences between estimated and true VAR coefficients, ARI the adjusted Rand index between estimated and true cluster membership. For each factor the condition with the best recovery is highlighted in bold.

Factor	Levels	MAD	ARI
Lag	VAR(1)	.015 (.007)	.938 (.122)
	VAR(2)	.018 (.008)	.927 (.149)
Clusters	2	.013 (.005)	.951 (.098)
	4	.020 (.009)	.914 (.164)
Proportion	Equal	.015 (.007)	.941 (.113)
	Majority	.017 (.009)	.924 (.156)
Observations	50	.022 (.007)	.870 (.168)
	150	.011 (.004)	.995 (.030)
Distance	Small	.017 (.009)	.876 (.172)
	Large	.016 (.007)	.989 (.033)

Recovery of cluster membership To assess the recovery of cluster membership we calculated the adjusted Rand index (ARI; Hubert & Arabie, 1985) between the estimated classification of individuals into clusters and the true partitioning. The ARI takes on a value of 1 in case of perfect agreement between classifications and 0 when the agreement between classifications could have been expected by chance. The recovery of cluster membership was excellent with an average of .933 ($SD = .136$) and slight differences across simulation

¹¹Because the number of parameters was equal across starts, a model selection criterion was unnecessary.

Table 3.4: ANOVA Results on mean absolute differences (MAD) between estimated and true VAR coefficients. Effect sizes above .1 are highlighted in bold.

	<i>df</i>	<i>F</i>	<i>p</i>	$\hat{\eta}_p^2$
Lags	1	111.000	< .001	.193
Clusters	1	421.000	< .001	.476
Proportion	1	35.300	< .001	.071
Observations	1	1,228.000	< .001	.726
Distance	1	16.900	< .001	.035
Lags \times Clusters	1	21.000	< .001	.043
Lags \times Proportion	1	1.520	.218	.003
Lags \times Observations	1	8.740	.003	.018
Lags \times Distance	1	.825	.364	.002
Clusters \times Proportion	1	29.400	< .001	.060
Clusters \times Observations	1	50.400	< .001	.098
Clusters \times Distance	1	4.330	.038	.009
Proportion \times Observations	1	3.250	.072	.007
Proportion \times Distance	1	2.770	.096	.006
Observations \times Distance	1	22.400	< .001	.046

conditions. Mean ARIs for different conditions are shown in Table 3.3. An ANOVA on the ARI values including all manipulated factors and all possible two-way interactions between factors, in Table 3.5, shows especially a high number of observations ($\hat{\eta}_p^2 = .355$) and a large distance between clusters ($\hat{\eta}_p^2 = .314$) aided membership recovery. Also, the interaction between distance and observations ($\hat{\eta}_p^2 = .302$) indicates that a high number of observations compensates for a small distance between clusters and vice versa. The interaction between the distance and the number of observations is illustrated in Figure 3.3.

Table 3.5: ANOVA Results on the adjusted Rand index (ARI). Effect sizes above .1 are highlighted in bold.

	<i>df</i>	<i>F</i>	<i>p</i>	$\hat{\eta}_p^2$
Lags	1	2.100	.148	.004
Clusters	1	21.300	< .001	.044
Proportion	1	4.900	.027	.010
Observations	1	255.000	< .001	.355
Distance	1	212.000	< .001	.314
Lags \times Clusters	1	2.890	.090	.006
Lags \times Proportion	1	.866	.353	.002
Lags \times Observations	1	3.180	.075	.007
Lags \times Distance	1	1.360	.244	.003
Clusters \times Proportion	1	4.120	.043	.009
Clusters \times Observations	1	15.100	< .001	.032
Clusters \times Distance	1	12.100	.001	.025
Proportion \times Observations	1	3.660	.056	.008
Proportion \times Distance	1	4.420	.036	.009
Observations \times Distance	1	201.000	< .001	.302

In sum and as evident from Table 3.3, the recovery of VAR coefficients and cluster membership was best for a low number of parameters (i.e., conditions with few clusters or low lag orders), for a high number of time points to estimate cluster parameters (i.e., in conditions in which many observations or people were available for every cluster, as in the conditions with equal proportion of individuals) and for well separated clusters (i.e., conditions with large distances between clusters).

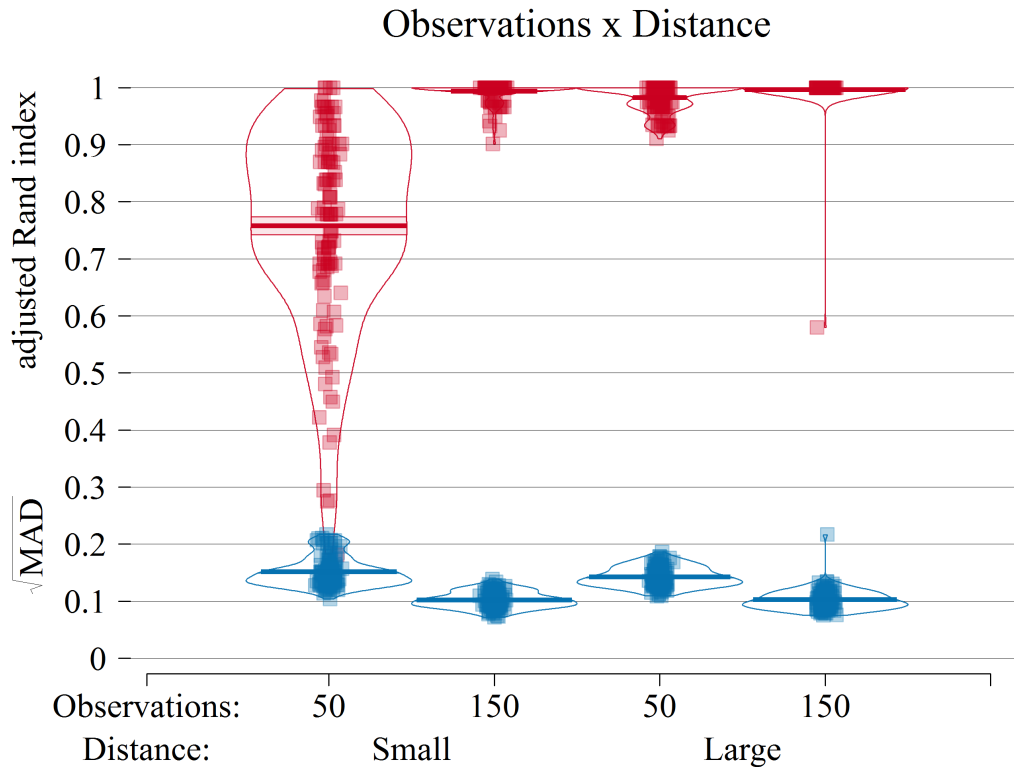


Figure 3.3: The square root mean absolute deviation (MAD) between estimated and true VAR coefficients and the adjusted Rand index across all simulation conditions. An opaque line shows the mean of the distribution within one of four observation \times distance conditions, a semi-opaque line shows the area within one standard error around the mean.