UNIVERSITY OF

OKLAHOMA GRADUATE

COLLEGE

# IMPACT OF TEMPORAL ORDER SELECTION ON CLUSTERING INTENSIVE LONGITUDINAL DATA BASED ON VAR MODELS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

Yaqi Li

Norman, Oklahoma
2023

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/337940

# IMPACT OF TEMPORAL ORDER SELECTION ON CLUSTERING INTENSIVE LONGITUDINAL DATA BASED ON VAR MODELS

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF PSYCHOLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Hairong Song, Chair

Dr. Dingjing Shi

Dr. Scott Gronlund

Dr. Kun Lu

© Copyright by Yaqi Li 2023
All Rights Reserved.

**Acknowledgments**

At this moment, I would like to extend my heartfelt appreciation and gratitude to the exceptional individuals who have stood by me throughout my rigorous five-year Ph.D. program. This dissertation would not have come to fruition without their unwavering support. Their unwavering love, intellect, and diligence enabled me to immerse myself in my research and ultimately complete this dissertation.

First and foremost, I am deeply indebted to my advisor, Dr. Hairong Song. She has been not only a mentor but also a friend and a profound inspiration in my life. I am immensely grateful to her for introducing me to the realm of quantitative psychology and nurturing me into a capable researcher. Throughout these years, she generously shared her valuable insights and guidance, granting me the freedom to explore my research interests. It is through her wholehearted dedication that this dissertation has reached its current form.

Additionally, I wish to express my sincere gratitude to my colleagues and scholars at the University of Oklahoma. I am particularly indebted to my lab mate Adon Rosen for sharing his invaluable suggestions regarding this work. Furthermore, I extend my thanks to Dr. David Bard and Dr. Yutian Thompson for providing me with a remarkable research opportunity at OUHSC. Collaborating with them and their team has been an enriching experience.

I am also immensely grateful to my committee members, Dr. Dingjing Shi, Dr. Scott Gronlund, and Dr. Kun Lu, for their insightful suggestions and constructive feedback during my General Exam and the development of this dissertation.

Furthermore, I would like to convey my deepest appreciation to my family. My parents, in particular, have been my unwavering pillars of support, especially during the most challenging

times. Without their unwavering encouragement and assistance, I would not have been able to complete this arduous journey towards attaining my degree.

To all those who have contributed to my academic and personal growth, I offer my heartfelt thanks. Your support, encouragement, and belief in me have made this accomplishment possible, and for that, I am eternally grateful.

## Table of Contents


Acknowledgments ..... iv

Abstract ..... xi

Introduction ..... 1

- Model-Based Approach for Clustering ILDs ..... 9
- VAR Models to Single-Individual ILDs ..... 15
- Clustering Multiple ILDs Based on VAR Models ..... 18
  - GMM Algorithm ..... 19
  - K-means Algorithm ..... 21
  - Comparison of GMM and k-means Algorithms ..... 21
  - Impact of Temporal Order Selection on Clustering ..... 23

Simulation Study ..... 25

- Method ..... 25
  - Design ..... 25
  - Data generation ..... 26
  - Clustering Procedure ..... 29
  - Hypothesis ..... 30
- Results ..... 31
  - The Recovery on Number of Clusters ..... 31
  - The Recovery on Cluster Membership ..... 32
  - The Recovery on Cluster-Specific Dynamics ..... 37

Empirical Study ..... 41

- Method ..... 41
  - Data ..... 41
  - Missing Data ..... 41
  - Data Analysis ..... 43
- Results ..... 45

Discussion and Conclusion ..... 51

- Recommendations and Suggestions ..... 60

Footnote ..... 76

Reference ..... 78

Appendix A ..... 87

Appendix B ..... 88

## List of Figures

Figure 1. Two-step procedure of model-based clustering method ..... 61

Figure 2. The accuracy of recovered number of clusters of under HO/LO and the GMM/k-means clustering algorithms..... 62

Figure 3. The accuracy of recovered cluster membership (ARI) under HO/LO and the GMM/k-means clustering algorithms.. ..... 63

Figure 4. The Euclidian distance between the true and estimated VAR coefficients under HO/LO and the GMM/k-means clustering algorithms.. ..... 64

Figure 5. Cluster-specific dynamiccal cofficients for Positive Affect (PA) and Negative Affect (NA). ..... 65

Figure 6. Score on personality measurements of three clusters obtained by GMM with LO strategy. ..... 66

Figure 7a. Network of correlations between trait affects, AR and CR coefficients, and big-five personality traits for Cluster 1 obtained by LO strategy with GMM. ..... 67

Figure 7b. Network of correlations between trait affects, AR and CR coefficients , and big-five personality traits for Cluster 2 obtained by LO strategy with GMM. ..... 67

Figure 7c. Network of correlations between trait affects, AR and CR coefficients , and big-five personality traits for Cluster 3 obtained by LO strategy with GMM..... 67

Figure B1. Score on personality measurements of three clusters obtained by HO strategy with GMM..... 88

Figure B2a. Network of correlations between trait affects, AR and CR coefficients , and big-five personality traits for Cluster 1 obtained by HO strategy with GMM. ..... 89

Figure B2b. Network of correlations between trait affects, AR and CR coefficients , and big-five personality traits for Cluster 2 obtained by HO strategy with GMM.. ..... 90
Figure B2c. Network of correlations between trait affects, AR and CR coefficients , and big-five personality traits for Cluster 3 obtained by HO strategy with GMM ..... 91

## List of Tables

Table 1.Average probabilities of correctly predicted number of clusters..... 70

Table 2. Average agreement between the true and the recovered cluster membership (ARIs) ... 71

Table 3.Average Euclidean distances between the true and estimated cluster-specific dynamics 72

Table 4. Cluster-specific AR and CR effects for the PA and NA of the three clusters obtained by LO strategy with GMM.. ..... 73

Table 5. Means of the external variables for the three clusters of PA and NA obtained by LO strategy with GMM..... 74

Table 6. Characteristics for the three clusters of PA and NA obtained by LO strategy with GMM ..... 75

Table B1a. Cluster-specific AR and CR effects of PA and NA for the Cluster 1 obtained by HO strategy with GMM..... 92

Table B1b. Cluster-specific AR and CR effects of PA and NA for the Cluster 2 obtained by HO strategy with GMM..... 92

Table B1c. Cluster-specific AR and CR effects of PA and NA for the Cluster 3 obtained by HO strategy with GMM..... 92

Table B2. Means of the external variables for three clusters of PA and NA obtained by HO strategy with GMM..... 93

Table B3. Characteristics for three clusters of PA and NA obtained by HO strategy with GMM. ..... 94

**Abstract**

In real-world research, intensive longitudinal data (ILDs) are typically collected from a group of individuals of interest, which enables researchers to model not only the within-individual dynamics of the studied processes but also the between-individual differences on the within-individual dynamics. Among the statistical techniques proposed for modeling ILDs of multiple individuals, clustering of intensive longitudinal data provides a meaningful way to quantify sample heterogeneity in dynamic processes, assuming that such heterogeneity reflects the distinct nature of the studied processes.

The aims of this dissertation are threefold: (a) to introduce a VAR-based clustering technique, (b) to examine the impact of temporal order selection on clustering accuracy and parameter estimation by a simulation study, and (c) to demonstrate the application of the clustering technique through an empirical analysis. Specially, I investigated the influence of two temporal order selection strategies: (1) using the most complex structure or highest order (HO) for all individual processes, and (2) using the most parsimonious structure or the lowest order (LO) for all individuals on the performance of two-step model-based clustering procedure. This procedure extracted dynamic coefficients from vector autoregressive (VAR) models and employed the Gaussian mixture model (GMM) and *K-means* clustering algorithms on the coefficients for cluster identification. Additionally, I also examined whether the influence varied across two clustering algorithms.

The simulation study showed that, regardless of the clustering algorithms used, LO strategy consistently outperformed HO strategy in terms of recovering the number of clusters, cluster membership, and cluster-specific AR and CR effects. GMM performed better than *K-means* when LO strategy was applied; however, the performance of GMM decreased while the

temporal orders increased. Additionally, GMM showed more vulnerability with smaller numbers of participants. The application of the two-step VAR-based method to affect data yielded a meaningful and informative clustering solution, which provided further insights of the uses of the model-based clustering approach Lastly, suggestions and recommendations were offered based on the results of the simulation and empirical analyses.

## Introduction

Intensive longitudinal data[1] (ILD) is characterized by multiple measurements collected over a short observation period, such as several times per day for a few days or weeks. The close spacing of ILD in time within individuals enables to capture the fluctuations of the outcome variables around their means over time, therefore, providing a more in-depth or fine-grained understanding of the dynamical mechanisms underlying time-ordered observations of psychological or behavioral processes, such as emotion (e.g., Pe & Kuppens, 2012), affect (e.g., Jahng, Wood, & Trull, 2008), stress (e.g., daSilva et al., 2021), interpersonal behaviors (e.g., Roche et al., 2014), and psychosomatic characteristics (e.g., Price et al., 2017; Von Leupoldt, Riedel, & Dahme, 2006).

Furthermore, ILD is typically collected in real-time rather than relying on retrospective reporting. This helps reduce the various types of cognitive bias (Schwarz & Sudman, 2012). The application of mobile or wearable devices to collect ILD, such as smartphone apps, eliminates the need for a laboratory setting. Thus, ILD has the capability to reflect an individual' s daily life as it is commonly lived, resulting in higher ecological validity (de Haan-Rietdijk, Voelkle, Keijsers, & Hamaker, 2017). Recent advancements in research designs and technologies, such as ecological momentary assessment (EMA) (Shiffman, Stone, & Hufford, 2008), experience sampling method (Chun, 2016), and daily diaries (Bolger, Davis & Rafaeli, 2003), enable repeated and intensive sampling of human behavior in everyday life. As a result, the research with ILDs has enjoyed a dramatic increase in use in social and behavioral research.

A common objective of the studies employing intensive longitudinal designs is to unfold how the preceding state of the system influences the subsequent state, which is defined as the temporal dependencies or lagged relationship (Hamaker, Asparouhov, Brose, Schmiedek, &

Muthen, 2018). These temporal dependencies are ubiquitous between psychological or behavioral processes (Hartmann et al., 1980; Hays, 1981). For instance, emotions and arousal at the individual level are often influenced by prior states (Wood and Brown, 1994). Previous research by Fuller et al. (2003) revealed a negative association between job stress on the current day and stress levels on the preceding day. Fairbairn and Sayette (2013) demonstrated that alcohol use diminishes the emotional impact of prior affective states on current emotions. Additionally, a review of 44 behavioral research studies with 248 independent sets of repeated measures data showed that autocorrelations ranged from 0.1 to 0.49, and 40% exceeded 0.25 (Busk and Marascuilo 1988). It is evident that this temporal dependence should not be neglected in research.

To capture those lagged relationships or temporal dependencies, various statistical models (i.e., time-series model), such as the autoregressive model (AR), vector autoregressive model(VAR), and autoregressive moving average model (ARMA) can be employed, which have found widespread use in fields such as physiology (e.g., Beltrame, et al., 2017; Nakamura, et al. 2016;), economics (e.g., Beck & Katz, 2011), neuroscience (Richiardi, 2013), and meteorology (e.g., Elangasinghe et al., 2014). These statistical models focus on studying a single unit dynamic process, which involves analyzing a large number of repeated measures from a single system (e.g., a person or a country). The single-unit ILD model, also known as the ideographic model, allows for capturing the unique dynamic process, guiding decision-making, or facilitating personalized intervention. However, single-unit models lack generalizability and struggle to assess differences between individuals. Furthermore, in social and behavioral research, ILD is often collected from a group of individuals with a shared characteristic or condition, such as individuals with depression suffer (e.g., Axelson et al., 2003; Whalen et al., 2008). In such cases,

extending ILD modeling to multiple units becomes essential for ILD's application in social and behavioral research. This extension involves modeling not only the within-individual dynamics but also the between-individual differences in dynamics.

In order to examine both between- and within-individual differences, researchers need to consider some degree of generation. Generalizing dynamic processes or obtaining a nomothetic model for multiple individuals requires making inferences at the inter-individual and intra-individual levels. This assumes that the processes are equal and interchangeable, which is known as the ergodicity assumption. However, this assumption is often violated for most behavioral and psychological processes (Molenaar, 2004). Research with ILD designs has shown that individuals vary widely in their behavioral and psychological processes across time, regardless of the specific domain being studied (e.g., Schenk, et al, 2017; Scott, Sliwinski, & Fields, 2013; Wright & Simms, 2016). Consequently, analyzing one level of the process is insufficient for generalizing to another level. Nonetheless, it is possible to generalize within-person dynamics to the between-person level by achieving conditional ergodicity (Molenaar, 2004). It requires control for unique factors of each structure, including autoregression, time trends, or distinct subgroups, to ensure equivalence between the within-person and between-person structures (Voelkle, Brose, Schmiedek, & Lindenberger, 2014).

A variety of statistical techniques have been proposed to analyze ILDs and generalize dynamic processes among multiple individuals to achieve conditional ergodicity. These approaches can be categorized into two classes based on how they handle between-individual differences.

One class of techniques, such as the dynamic multilevel model or mixed-effects model, focuses on quantifying between-individual differences as *quantitative*. These techniques assume

that the between-individual differences follow certain probability distributions, and these differences are reflected in the strength of dynamics. To model the within-individual dynamics, the quantitative techniques can utilize the statistics models from social sciences, such as dynamic factor models (e.g., Song & Ferrer, 2012), as well as methods commonly employed in engineering and econometrics, including state-space models (e.g., Chow, Ferrer, & Nesselroade, 2007) and time-series models (e.g., Babbin et al., 2015; Bringmann et al., 2013; Kuppens et al., 2012; Liu et al., 2013; Stefanovic et al., 2022). For example, one line of research utilizing this "*quantitative*" technique is through the dynamic multilevel model to investigate the dynamic relationship between negative affect and positive affect. In this model, the within-person dynamic process of affect is described through a time-series model at level-1, while between-person differences in dynamic relationships are accomplished with random effect at Level-2, such that the distribution of the coefficients is modeled with a fixed effect representing the average dynamic process across the population, and the variance terms representing the spread the person-specific dynamic coefficients value across the population. Studies have found a negative association between the carry-over effect of affect (a tendency to persist in a specific state for a considerable amount of time) and psychological well-being. Moreover, the carry-over effect has been found to have the capability to predict future psychological maladjustments, such as depression (e.g., Houben, Van den Noortgate, & Kuppens, 2015; Kuppens et al., 2012).

Additionally, between-individual differences in dynamics can be modeled using multilevel structural equation modeling (SEM) approaches, such as dynamic SEM, which offers a way to model multiple individuals' dynamic processes under SEM framework (e.g., Asparouhov, Hamaker, & Muthén, 2018; Hamaker, et al., 2018; McNeish & Hamaker, 2020; Song & Zhang, 2014). Dynamic SEM encompasses three distinct modeling approaches: time-

series analysis, such as AR and VAR models, to capture the dynamic processes within individual ILDs, multilevel modeling to capture the average dynamic structure across the population while considering individual differences in dynamic coefficients, and structural equation modeling (SEM) to incorporate multiple outcome variables, latent variables, and mediation effects. Within-person dynamics are modeled using a time-series model at the within-person level, while individual differences in dynamic coefficients (such as means, autoregression, and cross-lagged regressions), treated as latent variables, are modeled at the between-person level.

One concern of the "*quantitative*" approach is related to the assumption of a uniform model structure applied to the entire population, where the between-individual differences are quantified as the variation of the uniform model parameters. However, distinct dynamic models seem to exhibit among individuals, resulting from discrete and qualitatively different processes. (i.e., Carstensen et al., 2000; Hamaker et al., 2018; Hay & Diehl, 2011; Houben et al., 2015; Lane et al., 2019). For example, Hamaker et al. (2018) utilized the DSEM method to explore the dynamic relationship between positive emotion and negative emotion in young and old participants. They identified a distinct pattern in the dynamic relationship between the carry-over effect and the spill-over effect for two distinct samples of individuals. The carry-over effect refers to the impact of functioning or behavior form the previous moment to the current moment within on domain, while the spill-over effect refers to the impact of functioning or behavior in the previous moment within one domain on another domain in the current moment. Specifically, the study revealed that younger individuals exhibited a negative relationship between the carry-over effect and spill-over effect, indicating that those with a higher carry-over effect for positive emotion tend to have a lower spill-over effect from positive emotion to negative emotion. In contrast, older individuals showed a different pattern, where those with a high carry-over effect

for positive emotion tended to have a higher spill-over effect from negative emotion to positive emotion, but this was unrelated to the spill-over effect from positive to negative emotion. The existence of discrete and qualitatively different processes among individuals challenges the assumption of identical dynamic structures within the population. Ignoring this heterogeneity can introduce bias in model estimation (Bell, et al, 2010; Maas & Hox, 2004). Thus, neither random effects nor between-level structure can account for such discrete (un)observed heterogeneity. In these cases, generalizability can be achieved by identifying subclusters of individuals who exhibit qualitative differences in dynamic processes.

In contrast to quantitative techniques, clustering offers an alternative approach to capturing between-individual differences in dynamics by treating them as *qualitative*. These techniques assume that the between-individual differences reflect distinct nature in dynamic processes among individuals. The primary goal of this method is to assign individuals to different clusters that exhibit distinct differences in dynamic processes. It requires maximizing the within-clusters similarity in dynamic characteristics while minimizing between-clusters similarities. These techniques, known as model-based clustering, typically begin by fitting a time-series model, state-space model, or unified SEM (uSEM; Gates, et al, 2010; Kim, et al, 2007) to each individual's ILD. The estimated dynamic coefficients are then used as inputs for clustering algorithms to identify the clusters (e.g., Bulteel et al., 2016; Ernst et al., 2021; Gates et al., 2017).

Clustering of ILDs has been widely used as a data mining and data reduction technique in fields such as engineering, finance, health, and neuroscience (e.g., Aghabozorgi & Teh, 2014; Elangasinghe et al., 2014; Stetco, Zeng & Keane, 2013; Tran & Wagner, 2002; Yan et al., 2022). However, its application in social and behavioral research has been relatively limited. Recent

technological advancements, such as smartphones and sensors, have enabled the collection of large volumes of ILDs, creating new opportunities for exploring patterns within discrete subclusters of entire populations in psychological and behavioral processes. For example, clustering ILDs can be applied to identify distinct clusters of individuals who are smokers and exhibit similar dynamics in mood, affect, and smoking cravings during cessation periods. By identifying these group-specific patterns, tailored prevention and intervention strategies can be developed and implemented to achieve more effective outcomes. In this study, my aim was to contribute to the understanding of model-based clustering techniques by addressing a critical issue related to temporal order selection.

The issue of interindividual differences in temporal orders poses a challenge when using the model-based approach to cluster individual ILDs. Traditionally, identical temporal orders of the underlying dynamic processes, such as the number of lags for the autoregressive process, have been set for all individuals due to technical requirements. However, several studies(Bosley, Sandel, & Fisher, 2020; Chow, Hamagami, & Nesselroade, 2007; Gu, Preacher, & Ferrer, 2014; Rosmalen et al., 2012; Snippe et al., 2015) have highlighted that individuals exhibit different temporal orders in their psychological and behavioral processes, such as major depression (e.g. Price et al., 2017), autism (e.g. Volkmar, 2004), attention deficit (e.g. Karalunas & Nigg, 2020), as well as in neuroimaging studies involving healthy individuals (e.g. Finn et al., 2015; Price et al., 2017). For example, Ernst et al. (2021) found that the majority of participants' daily affect (approximately 94% of participants) could be best described by a lag-1 model, while for the rest of the individuals (about 6% of participants), lag-2 or lag-3 models were more appropriate. Similarly, Snippe et al. (2015) investigated the dynamic relationship between depression symptoms and repetitive thinking and discovered individual-specific patterns. They observed

that the daily decrease in repetitive thinking preceded a decrease in depressive symptoms the next day for one participant, while for another participant, depressive symptoms decreased two days after a decrease in repetitive thinking. No significant relationship was found between previous moments' repetitive thinking and current moment depressive symptoms for the remaining participants.

When individuals exhibit different temporal orders in their dynamic processes, as is very likely to occur, selecting an appropriate order for all individuals becomes a crucial issue before applying any clustering procedures. Two strategies have been proposed, up to date. One is to apply the most complex structure or highest order (HO) to all individual processes (Ernst et al., 2020; Kalpakis, Gada, & Puttagunta, 2001; Maharaj, 2000; Ren & Barnett, 2022; Xiong & Yeung, 2004). The other one is to choose the most parsimonious structure or the lowest order (LO) for all individuals (e.g., Bulteel et al., 2016; Gates et al., 2017; Zheng et al., 2013).

Utilizing HO strategy for temporal order selection likely leads to overfitting, whereas the LO practice likely leads to underfitting, at least for some of the individual processes in the data whose true optimal temporal order differs from the one used. The overfitting makes the model too complex and captures noise or idiosyncratic patterns specific to individual participants. The consequences of overfitting of a time-series model on the clustering include the identification of spurious clusters and inflated estimates of within-cluster variability, which can hinder accurate cluster identification and estimation of cluster-level dynamics. Underfitting, on the other hand, leads to the time-series model being too simplistic and an inability to capture meaningful variation within clusters, therefore masking the important individual differences in the data. To my best knowledge, the questions regarding the consequences of each method of temporal order selection: HO vs. LO (or overfitting vs. underfitting) when clustering individual processes with

the model-based approach have not been thoroughly answered yet, leaving the practices of HO and LO, as well as the model-based clustering techniques, unwarranted.

The goal of this study is two-fold. First, I aimed to introduce the model-based approach for clustering multiple ILDs to social and behavioral research. Second, I examined the effects of temporal order selection on ILD clustering by comparing the performance of HO and LO practices using simulated and empirical data.

In the following section, I first introduced the model-based approach for clustering ILDs. I then described the vector autoregressive (VAR) model (Hamilton, 1994; Lütkepohl, 2005), the one used as the " model" in the model-based approach in this study. Next, I presented a two-step procedure for clustering multiple ILDs in light of VAR models, with an emphasis on several commonly used clustering algorithms. I then reported the simulation study, where the effect of temporal order selection was examined systematically, and then an empirical analysis to further illustrate the effect of HO and LO strategy. Lastly, I discussed the findings and offered recommendations on the uses of clustering techniques.

**Model-Based Approach for Clustering ILDs**

The ILDs clustering approach can be classified into three categories based on whether they cluster directly with raw data, indirectly with features extracted from the raw data, or indirectly with generative models of the ILD (Aghabozorgi, Shirkhorshidi, & Wah, 2015; Liao, 2005).

The *shape-based clustering approach* (e.g., Berndt & Clifford, 1994; Cuturi & Blondel, 2017; Gravano, 2015), also known as the raw data clustering approach, attempted to utilize the similarity or distance measures (e.g., Euclidean distance) directly on the raw ILD (e.g., Euclidean distance) in combination with standard algorithms. This approach is straightforward

to implement. The primary objective of shape-based approaches is to match the given ILD from two individuals as well as possible based on the similarity of their fluctuations at each time point or identifying common trends occurring at different times by non-linear stretching and contracting of the time axes. For example, researchers and practitioners might be interested in identifying distinct clusters of individuals with depression based on the similarity of their patterns of change in negative emotions when exposed to specific situational cues. By tailoring interventions to a certain time within the specific clusters, researchers and practitioners can optimize the effectiveness and efficiency of treatments.

The performance of shape-based clustering approaches heavily relies on the choice of similarity or distance measures. Yahyaoui et al. (2016) and Wang et al. (2013) conducted a comprehensive review of similarity/distance measures for ILD, categorizing them into four main categories: lock-step measures (e.g., Euclidean distance), elastic measures (e.g., dynamic time warping (DTW)), pattern-based measures (e.g., spatial assembling distance), and threshold-based measures (e.g., threshold query-based similarity search). Various distance measures allow the shape-based clustering approach to achieve robust performance, even when partitioning multiple individuals that have multiple and/or unequal-length of ILDs. However, the accuracy of shape-based clustering is sensitive to noise, outliers, as well as high dimensionality in terms of the number of intensive longitudinal variables and length of the ILDs. Additionally, the efficiency of clustering is a concern in the shape-based approach, as computing distance measures can be computationally expensive, especially when dealing with high-dimensional ILDs. Moreover, this approach faces challenges in handling multiple ILDs with missing values and/or unequal intervals, as it lacks a straightforward method to address these issues.

The *feature-based clustering approach* (e.g., D' Urso & Maharaj 2009; Galeano & Peña, 2001; Wang, Smith, & Hyndman, 2006) conducts clustering analysis based on the statistical features of ILD (e.g., mean, autocorrelations, trend, and independent components). The approach typically consists of two stages. It first extracts a set of statistical features from each individual' s ILD, and then feeds the extracted features to cluster algorithms with the standard distance measures. One advantage of this approach is that it allows the use of classic algorithms like k-means, eliminating the need for developing novel clustering algorithms specifically for the extracted features. Additionally, feature-based methods extract low-dimensional features from raw ILDs, which can reduce the computational costs of the clustering and enhance the robustness to noise in the raw data. Moreover, utilizing a well-chosen set of features extracted from ILDs enables the maintenance of interpretability in certain clustering tasks. For instance, researchers can use entropy as a feature to cluster patients' heart rate intervals and discover that individuals with congestive heart failure exhibit lower entropy in their heart rate intervals. Understanding which features yield good performance for a specific task facilitates researchers to gain conceptual insights into the underlying properties of the data and conduct accurate decision-making (Fulcher,2017).

However, it is important to acknowledge that the features extracted from ILD data are not inherently linked to the specific clustering objective. Meanwhile, there is no established guideline indicating which features should be extracted for a given clustering task. In some cases, commonly used features like mean or variations of ILDs may not yield meaningful clustering results. For example, it happens when the goal is to identify subclusters based on the temporal relationship between two variables, such as the dynamic relationship between daily stress and negative affect. The process of extracting features from ILDs has the risk of focusing

solely on preserving the local structure of ILDs, resulting in the generation of nonrepresentative and meaningless features, which subsequently impairs the clustering performance (Ma, et al, 2021). Additionally, there is a risk of losing crucial information present in the original data during the feature extraction stage, potentially leading to suboptimal clustering outcomes. Although the feature-based approach offers advantages in terms of computational efficiency and robustness to noise in ILD clustering, compared to the shape-based approach, it is important to notice that isolating feature extraction from the clustering task overlooks the potential interaction between them. Neglecting this interaction can limit the ability of the feature-based approach to achieve meaningful partitioning results.

The *model-based clustering approach* (e.g., Bulteel et al. 2016; Coke and Tsao, 2010; Ernst et al. 2019; Piccolo, 1990; Xiong & Yeung, 2002) for ILDs is based on the dynamic coefficients of the chosen time-series models that are assumed to generate the observed ILDs. Such approaches partition individuals based on the similarity in how ILD varies over time. For example, a researcher may wish to cluster together individuals that tend to follow a rise in negative affect with an increase of negative affect the next day.

In comparison with the other two approaches, the model-based approach offers several advantages. Firstly, it allows for the numerical estimation of dynamic relationships within and between variables, thereby enabling statistical inferences to be made based on these relationships. This capability holds critical value in enhancing the understanding of the underlying dynamics of ILDs in social and behavioral research. Secondly, the model-based approach offers dimensionality (i.e., length of ILD) reduction by extracting meaningful dynamic coefficients from the raw ILD data. Furthermore, it facilitates the comparison of ILDs with varying lengths. This is achieved by partitioning individuals based on the similarity of the

extracted coefficients obtained from a given model, thus accommodating ILDs of different lengths and enabling the generation of meaningful clusters. In social and behavioral science research, the model-based method holds great promise. The coefficients derived from the statistical model capture the dynamic relationships among multiple behavioral or psychological processes. Consequently, clustering individuals based on these dynamic coefficients allows for the establishment of meaningful explanations for individual behavior or psychological phenomena, leading to the identification of discrete subclusters of people.

The model-based method can be implemented using different strategies, varying in the number of steps (i.e., one-step vs. two-step) and/or the specific clustering algorithms employed. In a typical one-step procedure, researchers fit a chosen time-series model (e.g., VAR model) to all individuals, which requires simultaneously partitioning the individuals and estimating the cluster-wise coefficients of the selected time-series model. It searches for the best partitioning of participants by minimizing the prediction errors of the selected time-series models that are separately fitted on each cluster (e.g., Bulteel et al., 2016; Jacques & Preda, 2014). For example, Bulteel et al. (2016) proposed an alternating least squares (ALS) algorithm for clustering multivariate ILDs based on the VAR(1) model, which is an extension of the *k-means* algorithm. This method mainly consists of three steps: first, initial cluster membership is tentatively assigned to each individual based on a specific criterion, often derived from other clustering methods such as hierarchical clustering; Second, a VAR (1) model is estimated for each cluster, considering the data from individuals within the same cluster. That is the ILDs from the individuals in the same cluster are aggregated to be fitted to a cluster-wise VAR(1) model, with dynamic coefficients that vary across clusters but remain identical within each cluster. Finally, cluster membership is updated based on prediction errors. Specifically, participants are

reassigned to the cluster where they have the minimum prediction error. This iterative process is repeated to search for the most appropriate cluster membership that minimizes the total prediction errors for all participants.

In contrast, a two-step procedure separates the model estimation and clustering processes into distinct steps. First, the dynamic coefficients of a chosen time-series model are extracted from each individual. Then, standard clustering algorithms are applied to these extracted coefficients to partition the individuals. Thus, this approach allows for within-cluster deviation in cluster-wise dynamics coefficients rather than treated as the error term. For example, Zheng et al. (2013) implemented a two-step procedure. In step 1, they fitted a VAR(1) model to each individual, estimating person-specific coefficients. In step 2, the similarities between pairs of individuals were calculated based on their coefficients, typically using a distance measure like Euclidean distance. These pairwise similarities were then used to identify clusters through a hierarchical clustering algorithm. In another example, Ernst et al. (2019) submitted the person-specific dynamic coefficients estimated through VAR model in step 1 to the Finite Gaussian mixture model (GMM) for cluster identification in step 2, under the assumption that the model coefficients follow a mixture of normal distributions.

The one-step method in model-based clustering has a major drawback in that it does not allow for within-cluster variation in dynamic coefficients. The two-step method offers more flexibility by allowing for within-cluster variation in dynamic coefficients, which can better capture the heterogeneity within clusters. Additionally, the one-step method partitions individuals into distinct clusters based on minimizing within-cluster predictive errors, while the two-step method forms clusters by quantifying the (dis)similarity of estimated dynamic coefficients. Furthermore, research has demonstrated that the two-step approach outperforms the

one-step approach in clustering multivariate ILDs across a wide range of conditions (Ernst et al., 2021). Besides, the iterative nature of the one-step approach can be computationally slow, especially when dealing with large datasets where various numbers of potential clusters and/or potential models with varied temporal order are considered. As the number of potential temporal orders and clusters increases, the number of comparisons that need to be made grows exponentially, leading to increased computational complexity. In this study on temporal order selection, I chose two-step procedures for clustering multivariate ILDs.

**VAR Models to Single-Individual ILDs**

Time-series models have been widely used in various fields to model dynamical mechanisms underlying time-ordered observations. When a dynamical system consists of multivariate processes, multivariate time-series models are needed. Among them, VAR models have gained popularity in social and behavioral research due to their capacity to model temporal dependence within and between variables through autoregressive (AR) and cross-regressive (CR), respectively. The former reflects the carryover effect of one behavior from the previous state to the current state. While the latter reflects the cascade effect of functioning or behavior in one domain into another domain from moment to moment (Almeida,Wethington, &Chandler, 1999; Bolger,DeLongis, Kessler, & Wethington, 1989; Masten & Cicchetti, 2010). For example, Rosmalen et al. (2012) used this technique to investigate the temporal relationship between depression and physical activity, and they found that the current depression level can be predicted by physical activity at previous occasions.

A VAR($p$) model indicates that the current state of a dynamic process is predicted by the state of itself or other variables in $p$ immediately preceding occasions after controlling for the influence of all other variables at the current occasions. The $p$ in VAR($p$) model is called lag or

temporal order of the process. The VAR(*p*) enables to provide insight into unique effects with a certain temporal order, thereby highlighting potential causal pathways between variables. Following is an example of a bivariate VAR (2) model:

$$Y_t = A + \Phi_{t-1}Y_{t-1} + \Phi_{t-2}Y_{t-2} + E_t$$

or

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}_t = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}_{t-1} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}_{t-1} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}_{t-2} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}_{t-2} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}_t \quad (1)$$

where $Y_t$ is a 2x1 vector of observed values at time *t*, the transition 2x2 matrices $\Phi_{t-1}$ and $\Phi_{t-2}$ contain the AR coefficients in the diagonals and the CR coefficients in the off-diagonal, $E_t$ represents a vector of residuals or innovations, and *A* is the intercept. The innovations are assumed to follow the multivariate normal distribution, with zero-mean and covariance matrix. If the off-diagonal elements of the innovation covariance matrix take on a non-zero value, reflecting a contemporaneous relationship between two processes.

Overall, the temporal dependencies are captured by the AR and CR coefficients. Those coefficients represent how quickly the process returns to the mean after being perturbed, while the innovation component represents a collection of internal or external events, such as various social interactions or stressful events, that push the process away from its mean at a given time. The stronger the effect size of AR and CR coefficients, the more effect carryover from the previous state, the longer it will take for the effect of innovation to disappear, and the slower the process will revert to its meaning. However, those coefficients are assumed not to be too strong (i.e., the value lies between 0 and 1, negative values may also occur), so the perturbations due to the innovation "die out" over time rather than accumulating (Hamilton, 1994). It should be noted that, like most of the time-series models, the VAR requires the time series being analyzed to be

stationary, that is, the statistical properties of the data (e.g., mean, variance, and autocovariance) need to be invariant over time, which require the eigenvalues of transition matrix Φ less than 1 in modulus (Lütkepohl, 2005).

The estimation of dynamic coefficients in a VAR(p) model can be performed using various estimation procedures, including least square estimation, Yule-Walker estimation, and maximum likelihood estimation (Lütkepohl, 2005). It is important to note that the least square estimation yields identical estimates to the maximum likelihood estimation. Although the Yule-Walker estimator shares the same asymptotic properties as the least square and maximum likelihood estimators, it is known to be less optimal in small samples (Lütkepohl, 2005). In this study, I employed the least square estimation method to derive the dynamic coefficients for each individual. The details of the VAR (1) model coefficients estimation procedure by least square estimation are shown in Appendix A.

A critical step when fitting a VAR(*p*) model to a single-individual ILD is to determine the temporal order *p*. Information criteria such as Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) are commonly used to assist in this selection of an optimal *p*. Specifically, by comparing the AIC or BIC values across a range of *p* values, one can identify the model with the lowest value as the optimal choice. However, incorrect decisions can still be made, leading to overfitting/underfitting. When the selected *p* is greater than the true order of the studied processes, overfitting occurs, while when it is smaller than the true order, underfitting occurs.

Research has shown that overfitting and underfitting are associated with different biases when fitting time series models to individual ILDs. For example, Lütkepohl (2015) showed that overfitting of a VAR model causes increased prediction errors while underfitting often generates

autocorrelated errors. Overall, an incorrect specification of *p* would lead to biased estimation (Braun & Mittnik, 2003), which consequently would impact the accuracy of clustering based on the estimated dynamic characteristics. Furthermore, including additional lags increases the number of estimated parameters, under which a large number of measurement occasions is required for reliable parameter estimation. This poses challenges for typical behavioral research where ILDs are collected from a relatively small number of time points. It is reasonably expected that overfitting could be particularly problematic for short ILDs.

**Clustering Multiple ILDs Based on VAR Models**

In the current study, I used a two-step clustering procedure in this study. In step 1, a VAR (*p*) model is fitted to each individual ILD, to extract person-specific AR and CR coefficients; in step 2, a specific clustering algorithm is performed on the person-specific coefficients derived in step 1 to partition individuals into different clusters. The performance of this two-step procedure depends on the choices of *p* in step 1 as well as the choice of clustering algorithm in step 2. While my major task is to examine the effects of HO and LO regarding the choices of *p*, I also investigate whether those effects depend on the specific algorithms chosen in step 2.

Clustering algorithms can be classified as probabilistic and non-probabilistic algorithms. Probabilistic algorithms impose certain probability distributions, such as a mixture of normal distributions over the parameters of a chosen model (e.g., AR and CR coefficients in a VAR model), and then identify the clusters of individuals whose model parameters, also known as features, follow the same distribution (Fraley & Raftery, 2002). Conversely, non-probabilistic algorithms typically rely on distance or similarity metrics, such as Euclidean distance, to partition individuals into different clusters, by which the within-cluster similarities of model parameters are maximized, and the between-cluster similarities are minimized. I chose the finite

Gaussian mixture model (GMM) and the k-means algorithms as the representatives of the probabilistic and non-probabilistic clustering algorithms, respectively, in the current study.

***GMM Algorithm***

The GMM algorithm (Banfield & Raftery,1993; Browne & McNicholas, 2014; Celeux & Govaert,1995) assumes that the AR and CR coefficients of the VAR (*p*) models come from a mixture of multivariate normal distributions, which included *k* unique distributions, and each unique distribution corresponds to a distinct cluster. The distribution of every participant is specified by a probability density function through a finite mixture model of *k* components, which takes the following form:

$$f(x_i;\alpha,\theta)=\sum_{k=1}^{K}\alpha_k f_k(x_i;\theta_k) \qquad (2)$$

where K is the number of clusters. $\alpha_k$ is the mixing weights or probability with constraints $\alpha_k \geq 0$ . The sum of all $\alpha_k$ is one (i.e. $\sum_{k=1}^{K}\alpha_k = 1$).$x_i = \{\phi_{i,1,1,1}, \phi_{i,1,2,1}, \ldots \phi_{i,m,m,p}\}$is the AR and CR coefficients for participant i which are derived from their personal-specific VAR(*p*) model. m represents the number of intensive longitudinal variables. *p* represents the temporal order of VAR(*p*) model. $f_k(x_i;\theta_k)$ is the kth cluster density function for a certain participant i' s AR and CR coefficients with the parameter set $\theta_k$. $f_k(x_i;\theta_k)$ are assumed to be from the same parametric distribution family for all clusters, which is a multivariate normal distribution in GMM, i.e. $f_k(x;\ \theta_k) \sim\ MVN(\mu_k,\Sigma_k)$. Thus, in GMM, Equation 2 transformed to the following form:

$$f(x;\alpha,\mu_k,\Sigma_k)=\sum_{k=1}^{K}\alpha_k \frac{1}{(2\pi)^{m/2}\,|\,\Sigma_k\,|^{1/2}} e^{-\frac{1}{2}(x_i-\mu_k)^T\Sigma_k^{-1}(x_i-\mu_k)} \qquad (3)$$

where $\mu_k$ are mean AR and CR coefficients across all the participants within kth cluster with $m^2 \times p$ dimensions. The matrix $\Sigma_k$ is variances/covariances of the individual dynamic coefficients within the kth cluster around the mean vector $\mu_k$ for kth cluster. The shape of the clusters is ellipsoidal, centered at the mean vector $\mu_k$ and with other geometric features, such as volume, shape, and orientation, determined by the covariance matrix $\Sigma_k$. The structure of the covariance matrix allows clusters with varied geometric features. The details about the structure of the covariance matrix $\Sigma_k$ of GMM are referred to Scrucca et al. (2016).

In GMM, the parameters $\alpha_k$ and $\theta_k$ for each cluster are usually unknown and require estimation. The estimation procedure for this model proceeds in a two-step, interactively processing between (a) estimating the posterior probability of cluster membership for each individual with a fixed set of parameters through Equation 4 (i.e., the probability of a given participant i with VAR(*p*) coefficients $x_i$ belongs to cluster k), and (b) updating the parameter estimates by fixing the probability of class membership for each individual via maximum likelihood. This estimation procedure usually uses the expectation–maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977). The technical details are referred to Bartholomew and Knott (1999), McLachlan and Basford (1988), and McLachlan and Peel (2000).

$$p(participant_i \in k \mid x_i; \alpha_k; \mu_k; \Sigma_k) = \frac{\alpha_k f_k(x_i; \mu_k; \Sigma_k)}{\sum_{k=1}^{K} \alpha_k f_k(x_i; \mu_k; \Sigma_k)} \quad (4)$$

where $C_k$ is the kth cluster.

In general, to estimate the mixture of clusters, the number of clusters, as well as the structure of the covariance matrix must be pre-specified. The selection of the number of clusters and covariance structure can be conceived of as a comparison between statistical models, where

the different statistical models represent different choices for the possible combination of the different numbers of clusters with different covariance structures. The optimal combination of the number of clusters with covariance structure can be determined by using information criteria, such as the BIC (Schwarz, 1978).

***K-means Algorithm***

On the other hand, the *k-means* algorithm (MacQueen, 1967), as one of the most widely used clustering algorithms, was designed to partition ILDs into *k* clusters based on the squared Euclidean distance of raw data or extracted features. The goal of the *k-means* algorithm is to minimize the within-cluster variance. This minimization is carried out through an iterative least square procedure in which individuals are placed in clusters such that the distance between each individual's VAR(*p*) coefficients and the mean vector $\mu_k$ of its cluster is as small as possible. The loss function for k-means is shown in Equation 5.

$$\sum_{k=1}^{K} \sum_{i \in C_k} (x_i - \mu_k)^T (x_i - \mu_k) \tag{5}$$

***Comparison of GMM and k-means Algorithms***

Compared to the loss function of GMM (Equation 3), the *k-means* (Equation 5) can be treated as the most parsimony form of GMM (Steinley &Brusco,2011). Firstly, the term $\sum_{i \in C_k}$ in Equation 5 is related to the $\alpha_k$ term in Equation 3. The difference is in the cluster membership. In GMM, cluster membership is probabilistic, such that every participant makes some degree of contribution to all *k* clusters with some non-negative probability, while cluster membership for k-means is either present or absent, such that one participant is either in or out of a cluster. Second, if the elements of vector $x_i$ are assumed to be independent within each cluster, meaning that the covariances (located in off-diagonal in the matrix $\Sigma_k$) between these elements are fixed to zero.

The covariance matrix $\Sigma_k$ is assumed to be identical across the diagonal elements and clusters, leading the denominator of Equation 3 becomes constant and can be ignored. Meanwhile, if applying a log operation to the GMM loss function, it becomes equivalent to the inner product of the loss function of *k-means* defined in Equation 5. Thus, the loss function of *k-means* corresponds to the GMM function when all the features are independent, all the elements $\Sigma_k$ are equal, and the covariance matrices for all clusters are identical. However, in reality, these constraints are often violated. Therefore, through comparison between GMM and k-means, the GMM tends to provide a greater degree of flexibility in finding the true, underlying cluster structure present in the data (Steinley & Brusco, 2011). Besides, the GMM algorithm assigns individuals to a cluster based on the highest posterior probability and generates a probabilistic group membership, which I transformed into a definitive value, such as 0 or 1. In contrast, the *k-means* algorithm assigns individuals to one and only one cluster, so no transformation step was needed.

GMM has a clear advantage over clustering algorithms that are not based on probabilistic models. However, it is important to consider the detrimental effect of high dimensionality on the performance of GMM. On the one hand, it may cause the feature space of time-series models to have a large number of random dimensions that do not contribute to cluster identification or are unrelated to the job of clustering (Scrucca & Raftery, 2016). On the other hand, the high dimensionality in terms of the number of AR and CR coefficients used in clustering stage brings a large number of parameters (e.g., the large number of elements of need to be estimated) that need to be estimated for GMM, which requires a large sample size to obtain the accurate clustering solution (Bouveyron & Brunet-Saumard, 2014). Additionally, the goodness of cluster recovery with the GMM cannot be guaranteed due to potential overfitting occurring with

exceedingly complex models (Steinley & Brusco, 2011). In contrast, the *k-means* clustering could avoid model overfitting and consequently produce a parsimonious cluster structure when high dimensionality occurs.

***Impact of Temporal Order Selection on Clustering***

When the two-step procedure based on a VAR(*p*) is used to cluster a sample of multivariate ILDs, overfitting will occur on those individual processes with true orders lower than the chosen *p*. This will yield some model coefficients approach zero, that do not vary across true clusters, the so-called non-effective coefficients. The non-effective coefficients are not informative for cluster identification; instead, they act more like noises in the clustering stage. Clustering accuracy could be attenuated if the procedure produces a large number of non-effective coefficients (Takano et al., 2020). In addition, as the *p* increases, the number of estimated coefficients of a VAR (*p*) model will increase. As a result, the feature space would become increasingly sparse, which will blur the similarities among clusters (Steinbach, Ertöz & Kumar, 2004). In this case, the clustering algorithms that partition data based on similarities would perform poorly. Research has shown that time-series models with high temporal order severely restricted clustering accuracy (Bouveyron & Brunet-Saumard, 2014; Scott and Thompson, 1983). Regarding the underfitting of VAR models in clustering ILDs, the unmodeled temporal relationships could manifest themselves as autocorrelation residuals that lead to biased estimates of the model coefficients, which in turn would impact the accuracy of clustering based on the estimated model coefficients.

It is important to acknowledge that previous studies on clustering ILDs mostly used lag-1 time series models, such as VAR (1) (e.g., Bulteel et al., 2016; Ernst et al., 2021). This choice of using lower temporal order increases the likelihood of underfitting, as higher temporal orders

that may exist in individual ILDs but are left unmodeled. Overall, current knowledge is very limited about the impact of temporal order selection or overfitting/underfitting on clustering ILDs using the model-based approach. Thus, I conducted a simulation study and empirical analyses to offer some insights into this imperative issue.

## Simulation Study

### Method

The purpose of the simulation study was to investigate the influence of temporal order selection, specifically HO vs. LO, on clustering multivariate ILDs across a wider range of conditions. In real-world scenarios, multivariate ILDs collected from a sample of participants typically exhibit diverse dynamic processes, such as some following a VAR (1) process while others adhering to a VAR (2) process. In situations where both processes coexist in the data, I sought to investigate the consequences of exclusively clustering individuals based on a VAR (1) dynamic coefficients (i.e., the LO strategy). In this case, the ILDs generated by a VAR (2) process would suffer from underfitting. Similarly, I also explore the results of exclusively clustering individuals based on a VAR (2) dynamic coefficients (i.e., the HO strategy), which would result in overfitting for the ILDs generated by a VAR (1) process. This pivotal question formed the core of the simulation study. Furthermore, I examined whether the impact of HO or LO on clustering performance is dependent on the specific algorithms employed for clustering individuals. I employed three criteria to evaluate the performance of the two-step model-based clustering approach: (a) recovery of the correct number of clusters, (b) recovery of cluster membership, and (c) recovery of cluster-specific model.

#### *Design*

Data were generated with both VAR (1) and VAR (2) processes with the manipulation of six factors: (1) the proportion of ILD generated by VAR(2): 0%, 10%, 40%, 60%, 90%, and 100%, (2) the number of time-ordered observation of the ILD: 50, 100, 500, and 1000, (3) the number of individuals: 60, 120, & 240, (4) the number of clusters: 2, 3, and 4, (5) the Euclidean distance between clusters: .05, .10, .15, .20, and .25, and (6) the number of effective coefficients

(i.e., the coefficients are varied across clusters) within a cluster: 2, 4, 6, 8, 10, and 12. In total, I generated 64,800 unique datasets by fully crossing all factors with 10 replications in each condition. The within-dataset factors included the temporal-order selection strategy (HO vs. LO) and clustering algorithms (GMM and k-means algorithms). All data were generated by R (R Core Team, 2021).

***Data generation***

The individual multivariate ILDs were generated based on the VAR(*p*) model, where the number of variables was fixed at 4, the temporal order *p* was set to be either 1 or 2 depending on the pre-determined percentages of each process, and the intercepts were set to zero. The individual ILD generated model was expressed as follows:

$$\underbrace{\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \\ y_{4,t} \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} \phi_{11,t-1} & \phi_{12,t-1} & \phi_{13,t-1} & \phi_{14,t-1} \\ \phi_{21,t-1} & \phi_{22,t-1} & \phi_{23,t-1} & \phi_{24,t-1} \\ \phi_{31,t-1} & \phi_{32,t-1} & \phi_{33,t-1} & \phi_{34,t-1} \\ \phi_{41,t-1} & \phi_{42,t-1} & \phi_{43,t-1} & \phi_{44,t-1} \end{bmatrix}}_{\Phi_{t-1}} \underbrace{\begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \\ y_{4,t-1} \end{bmatrix}}_{Y_{t-1}} + \underbrace{\begin{bmatrix} \phi_{11,t-2} & \phi_{12,t-2} & \phi_{13,t-2} & \phi_{14,t-2} \\ \phi_{21,t-2} & \phi_{22,t-2} & \phi_{23,t-2} & \phi_{24,t-2} \\ \phi_{31,t-2} & \phi_{32,t-2} & \phi_{33,t-2} & \phi_{34,t-2} \\ \phi_{41,t-2} & \phi_{42,t-2} & \phi_{43,t-2} & \phi_{44,t-2} \end{bmatrix}}_{\Phi_{t-2}} \underbrace{\begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \\ y_{3,t-2} \\ y_{4,t-2} \end{bmatrix}}_{Y_{t-2}} + \underbrace{\begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \\ \varepsilon_{4,t} \end{bmatrix}}_{E_t}$$

With

$$\begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \\ \varepsilon_{4,t} \end{bmatrix} \sim N\left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \underbrace{\begin{bmatrix} \sigma^2_{11} & \sigma^2_{12} & \sigma^2_{13} & \sigma^2_{14} \\ \sigma^2_{21} & \sigma^2_{22} & \sigma^2_{23} & \sigma^2_{24} \\ \sigma^2_{31} & \sigma^2_{32} & \sigma^2_{33} & \sigma^2_{34} \\ \sigma^2_{41} & \sigma^2_{42} & \sigma^2_{43} & \sigma^2_{44} \end{bmatrix}}_{\Sigma} \right) \quad (6)$$

where $Y_{t\ is}$ a 4 × 1 vector that represents the four intensive longitudinal variables for participant i at time occasion t; $\Phi_{t-1}$ and $\Phi_{t-2}$ are 4 × 4 matrix referring to the AR and CR coefficients for lag-1 and lag-2 respectively. $E_t$ is 4 × 1 innovation vector at time t. It follows a multivariate normal distribution, MVN (0, I+0.5), which is a zero mean vector and a covariance matrix Σ. $\sigma^2$ located

in Σ diagonal represents variances of the process noises, and $\sigma^2$ at Σ off-diagonal represents the covariance.

The distance between different clusters was captured through the distance between cluster-specific AR and CR coefficients in Φ matrix. Individuals within the same cluster shared the same VAR model to generate person-specific ILDs. In other words, individuals within the same cluster had identical AR and CR coefficients in their Φ matrices. The within-cluster variation only came from innovation terms.

Cluster-level Φ matrices were specified for the reference and the rest clusters in each condition. Initially, the AR and CR coefficients of the reference cluster were randomly sampled from uniform distributions. Specifically, the AR coefficients in lag 1, representing the diagonal elements of the $\Phi_{t-1}$ matrix, were sampled from a uniform distribution with upper and lower bounds [0.5,0.7], and the CR coefficients in lag-1, corresponding to the off-diagonal elements of $\Phi_{t-1}$ matrix, followed a uniform distribution ranging from -0.4 to 0.4. If a lag-2 process was present, both AR and CR coefficients in lag-2, elements of $\Phi_{t-2}$ matrix, were randomly generated from a uniform distribution between -0.2 and 0.2. Once the reference cluster was generated, the AR and CR coefficients for other clusters were generated by adding or subtracting the predetermined value of the distance (e.g., 0.1) from the effective coefficients of the reference cluster. For example, in the 100% VAR (1) for two clusters and 1 effective coefficient with 0.1 distance, the cluster-level Φ were:

$$\Phi_{reference} = \begin{bmatrix} 0.7 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.7 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.7 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.7 \end{bmatrix} \Phi_{cluster\,2} = \begin{bmatrix} 0.7 & 0.2 & 0.1 & 0.1 \\ 0.1 & 0.7 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.7 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.7 \end{bmatrix}$$

where $\Phi_{\text{reference}}$ and $\Phi_{\text{cluster 2}}$ are the AR and CR coefficients for the reference cluster and the second cluster, respectively.

An element existing difference between those two Φ matrices is $\phi_{12,t-1}$, with a value of 0.1 for reference cluster and 0.2 for cluster 2. This specific element is the effective coefficient. The effective coefficients were randomly allocated in the Φ matrix. If a lag-2 process was generated, the effective coefficients were evenly located in both lag-1 and lag-2 in the Φ matrix. The above processes of data generation were repeated until all cluster-specific VAR models met the two requirements: (a) assumption of stationarity, that is, the eigenvalues of the Φ matrix are less than 1 in modulus; (b) the total distance between any pair of cluster-level Φ is equal, such that $|\Phi_{\text{cluster1}}- \Phi_{\text{cluster2}}|=|\Phi_{\text{cluster2}}- \Phi_{\text{cluster3}}|=|\Phi_{\text{cluster1}}- \Phi_{\text{cluster3}}|^2$.

In each simulated dataset, the size of each cluster was determined by the proportion of ILDs generated by VAR (2) model. Specifically, when the proportions were set to 0% or 100%, the individuals were evenly distributed across clusters. For the proportions of 10% or 90%, one cluster contained 10% of individuals while the remaining individuals were evenly distributed across the other clusters. Similarly, for the proportions of 40% or 60%, one cluster contained 40% of individuals and the rest were evenly distributed across the remaining clusters. For instance, consider a simulated dataset with 10% VAR (2), 4 clusters, and a total of 120 individuals in this sample. One cluster would consist of 10% of individuals who followed the VAR (2) process, resulting in 12 individuals. The remaining 90% of individuals were evenly distributed across the other 3 clusters, with each cluster containing 30% of these individuals, resulting in 36 individuals for each cluster, who followed the same cluster-specific VAR (1) process.

***Clustering Procedure***

In each simulated dataset, person-specific AR and CR coefficients were estimated for all the individuals using both LO and HO strategy to all individuals in a dataset. For example, in a dataset with 10% VAR (2) and 90% VAR (1), the VAR (1) models were fitted to all individual ILDs under the LO method, whereas the VAR (2) models were fitted under the HO method. Subsequently, the GMM and *k-means* algorithms were utilized to cluster individuals within each dataset based on their person-specific AR and CR coefficients.

The number of clusters is often unknown and requires estimation in real-world scenarios. Thus, in the current simulation study, I relaxed the assumption of knowing the number of clusters for both *k-means* and GMM. Both algorithms were fitted with a variable number of clusters, ranging from $K$=1 to $K$=5 clusters. The GMM algorithm utilized BIC[3] to determine the optimal number of clusters as well as the optimal cluster covariance structures. The *k-means* algorithm used the gap index (see Tibshirani, Walther, & Hastie, 2001 for the details of the calculation of gap index) to identify the optimal number of clusters. To ensure the stability and accuracy of the gap index, 5000 bootstrapped samples were used for each *k-means* procedure.

Furthermore, neither GMM nor k-means guarantee a globally optimal solution (McLachlan & Peel, 2000). To mitigate the potential influence of starting positions, I used the effective initialization technique. Specifically, Banfield and Raftery (1993) and Fraley and Raftery (1998) suggested initializing the GMM algorithm with hierarchical agglomerative clustering of individual dynamic coefficients. This technique begins by creating clusters with one individual each and merges the two clusters that are associated with the lowest decrease in classification likelihood for Gaussian mixture model. Similarly, initializing *k-means* clustering with a hierarchical procedure, such as Ward's method (Ward, 1963), has been widely

recommended (Arabie & Hubert, 1992; Milligan, 1980; Milligan & Sokol, 1980; Waller et al., 1998). To ensure the quality of initial starting configurations, I initialized each algorithm with its respective hierarchical clustering counterpart and performed ten random starts. In the simulation study, I utilized the R package "mclust" (Scrucca et al., 2016) to select the optimal clustering results using the GMM algorithm, while "NbClust" (Charrad et al., 2015) was employed to calculate the gap index for determining the optimal number of clusters in k-means. Figure 1 depicts the two-step clustering procedure employed in the simulation study.

***Hypothesis***

Regarding designed factors, I have the following hypotheses: the performance of the clustering decrease with (a) fewer time-ordered observations, (b) fewer individuals, (c) a higher number of clusters, (d) smaller distances between clusters, and (e) fewer effective coefficients. Also, I expected when HO was applied, the performance of the cluster assumed to promote as the proportion of data generated by VAR (2) increased. Furthermore, the performance of GMM outperforms the k-means when LO was applied. However, GMM showed more vulnerability to high dimensionality in terms of a large number of dynamic coefficients used in clustering (i.e., application of HO strategy).

## Results

### *The Recovery on Number of Clusters*

To assess the recovery of the number of clusters, I computed the recovery rate which is the percentage of simulated datasets in which the correct number of clusters was accurately identified. Overall, the LO strategy demonstrated superior performance compared to the HO strategy, achieving an accuracy rate of 68% for LO and 40.8% for HO. This pattern held true across both the GMM and *k-means* algorithms. Furthermore, when comparing the two algorithms, GMM exhibited a higher recovery rate than k-means. Specifically, for the LO strategy, GMM achieved an accuracy rate of 75.3%, whereas *k-means* achieved 52.9%. In the case of the HO strategy, GMM had an accuracy rate of 42.1%, while k-means achieved 39.4%. These results indicate that GMM outperformed k-means in accurately identifying the correct number of clusters, particularly under the LO strategy.

The application of the HO strategy resulted in an unacceptable clustering performance, as it failed to accurately identify the correct number of clusters for more than half of the dataset, approximately 59.2%. The BIC and gap index, which are commonly used for model selection, tended to favor more parsimonious models, suggesting fewer clusters than the true value, particularly when the HO strategy was employed. Specifically, 6.2% of the datasets indicated a greater number of clusters, while 53% of the datasets indicated a fewer number of clusters than the true value under the HO strategy. Comparing the performance of k-means and GMM with the HO strategy, it was observed that *k-means* tended to yield fewer clusters. In approximately 55.8% of the datasets, *k-means* indicated a smaller number of clusters compared to the true value, while GMM did so in 51.3% of the datasets. Interestingly, even when the LO strategy was applied, *k-means* still had a tendency to prefer fewer clusters. Specifically, in 35.3% of the

datasets, *k-means* indicated a smaller number of clusters compared to the true value, while GMM did so in only 18.3% of the datasets.

Table 1 displayed the recovery rate for the number of clusters under each specific condition. The findings consistently demonstrated that the LO strategy consistently outperformed the HO across all conditions. However, the accurate rate patterns in terms of LO and HO strategies varied depending on specific data conditions. For example, higher accuracy in terms of recovery of the number of clusters, as my hypothesis, was observed in cases with a larger number of time-ordered observations, greater inter-cluster distances, fewer non-effective coefficients, and a lower number of clusters. This trend held true for both LO and HO. Interestingly, as the proportions of VAR (2) ILDs varied, Figure 2 revealed distinct patterns between LO and HO. Specifically, under the LO strategy, the recovery rates remained relatively stable varying proportions of ILDs generated by VAR (2). In contrast, under the HO strategy, the recovery rates increased as the proportions of ILDs generated by VAR (2) became higher.

The performance of the clustering algorithms was found to be influenced by the choice between the LO and HO strategies. Specifically, when the LO strategy was employed, the *k-means* algorithm consistently exhibited poorer performance compared to GMM. However, under the HO strategy, the performance of *k-means* was more comparable to that of GMM, especially when the number of time-ordered observations was large, such as $t$ = 1,000.

***The Recovery on Cluster Membership***

The adjusted Rand index (ARI), initially proposed by Hubert and Arabie (1985), is a commonly recommended method (Milligan & Cooper, 1986; Steinley, 2004) for evaluating the recovery of cluster membership. It provides a quantitative measure of the agreement between the predicted cluster memberships and the true cluster memberships, ranging from 0 to 1. A value of

0 signifies a random assignment of memberships, whereas a value of 1 indicates a perfect correspondence between the predicted and true cluster memberships.

The ARI is preferred over the misclassification rate, which measures the proportion of individuals assigned to the correct cluster, for evaluating clustering performance. The misclassification rate, traditionally used in evaluate the accuracy of membership identification, has several limitations. Firstly, it requires making arbitrary decisions when matching cluster structures between the predicted and true structures[4]. Secondly, it can only be used to compare cluster structures with the same number of clusters[4]. Lastly, the misclassification rate converges to a value of 1 - *1/K* (*K* is the number of clusters), while the ARI approaches zero when the cluster solutions resemble random assignments (Steinley & Brusco, 2011). In the current study, the ARI was utilized as the measure to evaluate clustering performance. According to Steinley (2004), an ARI value of $\geqslant 0.90$ indicates an excellent recovery, $\geqslant 0.80$ represents a good recovery, $\geqslant 0.65$ suggests a moderate recovery, and $< 0.65$ indicates a poor recovery.

Overall, the LO strategy demonstrated superior performance compared to the HO in terms of average recovery on cluster membership. The LO strategy achieved an average ARI of 0.71, indicating a moderate recovery, while the HO strategy had an average ARI of 0.46, indicating a poor recovery. Furthermore, when the LO strategy was employed, the GMM algorithm showed a good average recovery with an ARI of 0.79, indicating a strong agreement between predicted and true cluster memberships. On the other hand, the k-means algorithm yielded a relatively poor average recovery with an ARI of 0.63. Conversely, when the HO strategy was utilized, both GMM and the k-means algorithms exhibited poor recovery on cluster membership. The average ARI was 0.46 for GMM and 0.45 for the k-means algorithm, indicating a weak agreement between the predicted and true cluster memberships.

Table 2 reported the average ARIs for all data conditions and clustering algorithms. The results consistently showed that the average ARI increased systematically as the number of time-ordered observations, sample size, the distance of dynamics between clusters, and the number of effective coefficients increased. Notably, the number of time-ordered observations had the most significant impact on the recovery of cluster membership. The average ARI improved from a poor recovery, with an ARI of 0.24, to a good recovery, with an ARI of 0.83, as the number of time-ordered observations increased from 50 to 1000. This improvement can be attributed to the increased accuracy in parameter estimation for VAR models, which rely on a sufficient amount of time-ordered data.

Further inspection of the results of Table 2 revealed that the clustering with LO strategy generally yielded a higher average ARI than HO across all levels of all design factors. However, the extent of the LO strategy' s superiority over the HO method tended to decrease as the number of time-ordered observations, sample size, the number of clusters, and the distance of dynamics between clusters increased. Furthermore, the results indicated that the superiority of LO strategy in recovery was not dependent on the underlying data-generating model. Regardless of the temporal order of the underlying data-generating model (shown in Figure 3), the LO method consistently outperformed HO. Even when all the ILDs were generated by the VAR (2) model, the recovery of LO, with an ARI of 0.75, is still superior to HO, with an ARI of 0.5.

Based on the hypothesis, I anticipated to observing the differential clustering performance on HO and LO strategies across the levels of the proportion of ILDs generated by VAR (2). The proportion of ILDs generated by VAR (2) directly influenced the percentage of overfitted or underfitted person-specific models within a dataset. For instance, if all ILDs within a dataset were generated by VAR (1) models and the HO strategy (VAR (2)) was applied to all

individuals' ILDs, it would result in 100% of person-specific models being overfitted, and vice versa. Accordingly, I expected that the clustering recovery using the HO method would improve as the proportion of ILDs generated by VAR (2) increased. This is because a higher proportion of VAR (2) ILDs would lead to a decrease in the number of overfitted person-specific models within the dataset. On the other hand, I anticipated that the clustering performance for the LO strategy would worsen in such cases. However, the results partially support the hypothesis. The results revealed that the performance of clustering significantly improved when the HO method was applied, with increases of 0.19 for GMM and 0.16 for *k-means*, with the increment of the proportion of ILDs with higher-order VAR processes. In contrast, only a slight improvement in clustering performance was shown when the LO method was applied, with increases of 0.09 for GMM and 0.11 for *k-means*.

Additionally, the comparison of GMM and k-means algorithms in Table 2 showed that when LO was applied, GMM clustering was much better than the competing algorithm *k-means* for each level of each design factor. On the other hand, when HO was applied, the two cluster algorithms were competitive, and relative performance tended to be influenced by designed factors. Specifically, *k-means* yielded better performance for a large number of clusters (i.e., $k > 3$), a small number of sample sizes (i.e., $n < 60$), a larger number of time-ordered observations (i.e., $t > 500$), and a small number of effective coefficients (i.e., $EF < 4$). Among the design factors, the sample size and time-ordered observations were particularly noteworthy, as they led to the difference in the performance of the two clustering algorithms.

Relative to k-means, the GMM algorithm is more sensitive to the sample size, particularly when HO was applied. Specifically, the *k-mean* method yielded better recovery for $n=60$, with an ARI= 0.7 for k-means and ARI=0.46 for GMM, and $n=120$, with an ARI= 0.71 for

k-means and an ARI=0.53 for GMM. However, when *n*=240, GMM was superior, with an ARI=0.84, compared to k-means, with an ARI= 0.71. As sample size increased, GMM recovery also increased with improvements of 0.16 for LO and 0.35 for HO, which is particularly true when HO was applied. In contrast, *k-means* showed negligible improvements, with increases of only 0.01 for LO and 0.02 for HO.

The clustering performance varied with the number of time-ordered observations when the HO method was applied. The *k-means* algorithm exhibited better recovery for $t$ = 500, with an ARI of 0.70 for k-means and 0.61 for GMM, as well as for $t$ =1000, with an ARI of 0.79 for k-means and 0.73 for GMM. However, both GMM and k-means showed poor performance for clustering when the number of time-ordered observations was limited. GMM outperformed k-means for $t$ = 50, with an ARI of 0.06 for k-means and 0.18 for GMM, as well as for $t$=100, with an ARI of 0.25 for k-means and 0.32 for GMM. the recovery of k-means exhibited a more significant improvement, with a change of 0.73 in ARI, as the number of time-ordered observations increased from $t$=50 to $t$=1000. In contrast, the recovery of GMM was not as strongly influenced by the increase in the number of time-ordered observations, with a change of 0.55 in ARI. This suggests that the GMM algorithm is more vulnerable to the HO strategy. With the HO strategy, the number of features used in the clustering algorithm increases, leading to a larger number of parameters that need to be estimated for the GMM loss function (Equation 3). Therefore, a larger sample size is required to accurately estimate the parameters of GMM and properly partition individuals. On the other hand, the *k-means* algorithm does not involve a parameter estimation procedure and is not based on a probabilistic model. As a result, it is less likely to be influenced by the HO strategy, which involves a large number of features, and it is relatively insensitive to sample size.

In summary, when LO strategy was applied, in conjunction with adequate sample size (e.g., $n$ =240), the GMM algorithm demonstrated acceptable recovery (i.e., $ARI$ > 0.65) even with the small number of time-ordered observations (i.e., $t$ = 50). However, If the sample size is limited (e.g., $n$ =60), achieving acceptable clustering performance with GMM requires a moderate number of time-ordered observations (i.e., $t$ >100). In contrast, the k-means require a large number of time-ordered observations (i.e., $t$ >500) to obtain an acceptable recovery. On the other hand, when the HO method was applied, both a large number of time-ordered observations ($t$ > 500) and a large sample size (>240) are necessary for GMM to achieve acceptable recovery. Similarly, the k-means algorithm also requires a large number of time-ordered observations ($t$ > 500) to achieve acceptable recovery, regardless of the sample size.

***The Recovery on Cluster-Specific Dynamics***

To access the recovery of cluster-specific dynamics, the mean Euclidean distance between the estimated cluster-specific dynamics and the true dynamics from the underlying model-generating model was calculated. The following Equation calculated the mean Euclidean distance:

$$\frac{\sum_{k}^{K}\sqrt{\frac{1}{v^2}\sum_{j=1}^{v^2}(\hat{\phi}_{kj}-\phi_{kj})^2}}{K}$$

With

$$\phi_{kj}=\frac{1}{n}\left(\sum_{i}^{n}\phi_{participant_i,kj}\right)\ ,\ j=1,2...v^2 \quad \text{and} \quad participant_i \in cluster_k \tag{7}$$

where $v$ is the number of intensive longitudinal variables. $\hat{\phi}_{kj}$ and $\phi_{kj}$ are the cluster-level of estimated and true AR and CR effects coefficients, respectively. $k$ is the kth cluster. $n$ is the number of people within kth cluster.

The recovery of cluster-specific dynamics was conducted exclusively on datasets where the correct number of clusters was indicated. To obtain the cluster-specific dynamics, such as the AR and CR effects coefficients, the dynamics were averaged across individual-specific VAR models within each cluster. The quality of recovery was determined by the Euclidean distance, with a smaller distance indicating a better match to the true dynamics. It is important to note that clustering algorithms assign arbitrary numbers to label clusters. To ensure accurate assessment, the partitions were permuted prior to calculating the Euclidean distance, maximizing the alignment between the estimated and true clusters.

Across all simulated datasets, clustering with the LO strategy consistently outperformed the HO in terms of recovery on cluster-specific dynamics, with a mean value of 0.31 and 0.35, respectively. Notably, relative to the GMM, *k-means* demonstrated competitive performance compared to GMM, regardless of whether it was used in conjunction with the LO or HO strategy. This stands in contrast to the poor performance of k-means observed in the recovery of membership and the number of clusters.

Table 3 provided a comprehensive comparison of different combinations of temporal-order selection methods and clustering algorithms based on the average Euclidean distance between estimated cluster-specific dynamics and true values. Overall, the LO strategy consistently outperformed the HO strategy, as indicated by its lower mean Euclidean distance across all levels of the designed factors. Besides, the performance of the HO strategy in recovering cluster-specific dynamics was influenced by the underlying data-generating model (shown in Figure 4). Specifically, as the proportion of ILDs generated by VAR (2) increased, the recovery of cluster-specific dynamic coefficients improved. However, even when all ILDs were

generated by VAR (2) models, the LO strategy still exhibited better recovery than the HO method, with mean distance values of 0.33 and 0.36, respectively.

An interesting pattern was observed in the recovery of cluster-specific dynamics with respect to the distance of dynamics between clusters. Regardless of the chosen clustering algorithms or temporal order of the VAR model, as the distance of dynamics between clusters increased, there was a decrease in the recovery of cluster-specific dynamics. Specifically, the average Euclidean distance increased by 0.04 and 0.02 for LO and HO, respectively, when the distance of dynamics between clusters changed from 0.05 to 0.25. On the contrary, an increase in the distance of dynamics between clusters was associated with an improvement in both the recovery of membership (ARI increased by 0.41) and the accuracy of the number of clusters (accuracy increased by 38.4%). This indicates that when the distance between clusters' dynamics is large, the accuracy of partitioning individuals can be guaranteed, even though the accuracy of cluster-specific dynamics decreases.

One possible explanation for this observation is as the distance between the dynamic coefficients of two clusters increases, the magnitudes of dynamic coefficients (including AR and CR effect coefficients) in other clusters are more likely to approach 0, given the stationary assumption of the VAR model. To accurately estimate these small dynamic coefficients, a large number of time-ordered observations are required. Therefore, even if the biased estimation of dynamics does not significantly affect individual partitioning accuracy, the recovery of cluster-specific dynamics is susceptible to biased estimation.

Collectively, the results of recovery of the number of clusters, membership, and cluster-specific dynamics revealed that the use of LO in conjunction with the GMM algorithm yielded the highest level of precision in partitioning individuals across all the designed conditions.

Interestingly, it was observed that underfitting had a lesser negative impact on the model-based clustering performance compared to overfitting. Furthermore, the performance of the two clustering algorithms was influenced by the choice of temporal order strategy. When LO was employed, the GMM algorithm consistently outperformed k-means in terms of clustering performance. On the other hand, the superiority of GMM over k-means, when using HO, varied depending on the levels of the design factors. Specifically, *k-means* showed better performance when a large number of time-ordered observations (i.e., $t > 500$) were available. Additionally, both a large number of time-ordered observations and a larger sample size were found to significantly contribute to enhancing the performance of the clustering, especially for the GMM algorithm.

## Empirical Study

### Method

#### *Data*

The data of empirical study is from the "How Nuts are the Dutch" project (the details about the dataset see Van der Krieke et al., 2016). The dataset comprised 1393 participants from the general population of the Netherlands who were assessed three times per day (morning, midday, and night) over a period of 30 or 31 days on multiple mental health dimensions (e.g., Affect, welling-being, etc.). The variable used in the clustering analyses was momentary affect, measured by 12 items from the circumplex model of affect (Barrett & Russell, 1998; Yik, Russell, & Barrett, 1999). Positive affect (PA) was assessed by items of enthusiastic, energetic, cheerful, relaxed, content and calm, and negative affect (NA) was assessed by the items of irritable, nervous, anxious, tired, dull, and gloomy. Each item was scored on a continuous scale ranging from 0 to 100. The values of NA and PA were the averages of the corresponding six items, with a range of 0–100. To account for the missingness rate of data, I only included the participants who completed at least 90% of the diary measurements (> 81 of the 90 time-ordered observations), which resulted in a final sample size of 227 participants.

#### *Missing Data*

To meet the assumption of equidistant between two successive time points, an additional measurement point was added as a missing value after the previous day’ s night and before the next day’ s morning measurement event for each participant. To handle the missing data, I employed the Amelia II package implemented in R, which is a specialized multiple imputation package designed for addressing missingness in ILD (Honaker & King, 2010; Honaker et al., 2011). It provides robust methods for imputing missing values and allows us to account for the

temporal dependencies present in the data. Compared to the traditional multiple imputation methods, the Amelia II offers computational efficiency by replacing a full Markov Chain Monte Carlo (MCMC) algorithm with a faster bootstrap-based Expectation Maximization (BEM) algorithm, resulting in improved speed (Ji, et al., 2018). The BEM algorithm provides greater flexibility in handling various posterior distributions for the predictors, outcome variables and model parameters. This algorithm first conducts a bootstrap procedure on the original dataset to generate m samples, the sample size is *n* (*n* is equal to the size of the original dataset) with replacement. The *m* bootstrapped datasets are then used the EM algorithm to iteratively update the model parameters. After the parameter updates, missing observations in the original data set are imputed separately using each of the *m* sets of parameter estimates from the EM algorithm, resulting in m multiply imputed data sets. To apply the Amelia function, I took into account the dynamic relationship between previous measurements and current measurements. I also included the measurement event (i.e., morning event) as covariates. Each individual was treated as a different cross-section to account for between-individual variance.

Each individual PA and NA score were centered across the observed mean score of that individual to enable the partitioning of individuals based solely on their dynamics. By using the Amelia II, each of the *m* imputed data sets underwent the same model fitting procedures as if it were fully observed, resulting in *m* sets of parameter estimates from fitting the model of interest. To obtain a final set of parameter estimates, Rubin's pooling procedures (Rubin, 1996) were applied. The final point estimates were obtained by averaging the parameter estimates over the *m* replications.

In the current study, five imputations were created by the function of amelia() in the Amelia II R package , as increasing the number of imputations beyond five did not lead to

notable differences in the imputation and estimation results. For each individual, the dynamic coefficients were derived from the five imputed datasets and then pooled using Rubin's method. Subsequent clustering was performed on the pooled person-specific dynamic coefficients.

### *Data Analysis*

**Temporal Order.** To determine the specific value of p in the VAR(p) model for the HO and LO strategy under the empirical dataset, the study initially identified the optimal temporal order for each individual. The best VAR(p) model for each participant was determined using the AIC. Out of the 227 individuals in the sample, the analysis revealed that 144 individuals were best represented by a VAR (1) model, 46 individuals by a VAR (2) model, and 37 individuals by a VAR (3) model. Clustering was applied to both the most parsimonious model, VAR (1) (corresponding to the LO strategy), and the most complex model, VAR (3) (corresponding to the HO strategy).

**Clustering Procedures.** The BIC criterion was utilized to determine the optimal number of clusters, ranging from 1 to 8 clusters, and the membership of participants for the GMM algorithm. The Gap index was employed to determine the optimal number of clusters for the k-means algorithm, also ranging from 1 to 8 clusters. For both k-means and GMM, 100 random starting points were used to identify the globally optimal solution. And all the clustering models converged. According to the BIC results, a three-cluster solution was determined to be optimal for both the LO and HO strategies within the GMM algorithm. Conversely, based on the gap index, a single cluster was identified as the optimal number of clusters for both the LO and HO strategies within the k-means algorithm. Given the primary interest in identifying distinct dynamic patterns that deviate significantly from clusters. Furthermore, the simulation study indicated that the k-means algorithm has a tendency to yield a fewer number of clusters

compared to the true number of clusters. Thus, I ultimately chose the cluster solution generated by the GMM algorithm.

**Clustering Performance Evaluation.** To evaluate the cluster solutions of empirical data, I examined the differences between clusters using external variables that were measured in the "How Nuts are the Dutch" project, which employed a cross-sectional design. One of the external variables I considered was the assessment of the big-five personality traits, which were measured using the NEO-FFI-3 questionnaire (De Fruyt and Hoekstra,2014). This questionnaire consists of 96 items, each rated on a scale from 1 to 5, and encompasses five personality domains: extraversion, agreeableness, conscientiousness, neuroticism, and openness. The happiness index assesses the degree to which one judges the quality of one’ s life with a single item ranging from 1 to 10 (Fordyce, 1988; Priebe et al.,2010). The status of depression, anxiety, and stress of each individual was assessed through the DASS scale (Lovibond & Lovibond, 1995), which included 42 items with a range of 0-3. Additionally, I took into account demographic variables, including age and gender, to provide a descriptive analysis of each cluster's characteristics. Additionally, I calculated the mean values of PA and NA across 30/31 days to describe the trait levels of PA and NA for each individual.

**Results**

A comparison of cluster solutions for LO and HO strategies showed that when HO was applied, the coefficients of AR and CR effect in lag-2 became less prominent once they are averaged across people within the same cluster (as shown in Figure 5). I only focused on the 3-cluster solution generated by LO strategy with GMM algorithm, which offered the most comprehensive summary of data. Table 6 presented the cluster-specific dynamics for the 3-cluster solution obtained using the LO strategy in conjunction with GMM. In addition, Table 5 and Figure 6 presented the cluster-wise average scores on external variables for the 3-cluster solution obtained by LO strategy with GMM. Table 6 contained demographic information and the trait affect, respectively. Finally, Figure 7 showed the network of correlations between trait affects, dynamic coefficients, and big-five personality traits for three clusters obtained by LO strategy with GMM. Detailed results of clustering solution obtained by the HO strategy with GMM algorithm were shown in Appendix B.

I conducted an ANOVA analysis to examine the differences in mean scores of PA and NA among the three clusters. The results revealed significant differences among the clusters for PA scores, $F$ (2,224) = 4.42, $p$ = 0.013. The post-hoc tests indicated that Cluster 3 had a significantly higher PA score than Cluster 2 ($p$ = 0.01). Additionally, significant differences were also observed for NA scores, $F$ (2, 224) = 7.48, $p$ < .001. The post-hoc tests showed that the NA score of Cluster 3 was significantly lower than both Cluster 1 ($p$ = .008) and Cluster 2 ($p$ < .001).

The relationship between mean PA and NA differed across the three clusters, as shown in Table 6. Specifically, a negative correlation was observed between the mean scores of PA and NA of individuals in Cluster 1 ($r$ = -0.57, $p$ < 0.01) and Cluster 2 ($r$ = -0.58, $p$ < 0.01), indicating that individuals with higher PA tended to have lower NA. However, the correlation between PA

and NA within Cluster 3 ($r$ = 0.43) was not significant. The details of interpretation and description of the characteristics of the three clusters were shown in following sections.

The majority of participants (64.3%) were assigned to cluster 1. Participants in this cluster showed intermediate levels of mood symptoms (e.g., anxiety), personality traits, and affect dynamic patterns. Specifically, they displayed relatively high AR coefficients for both PA and NA, with values of 0.41 and 0.38, respectively. These relatively strong AR coefficients indicated that the individuals in this cluster tended to experience emotions that persist and are stable over time. In contrast, the CR coefficients between the PA and NA were weak, ranging from -0.006 to -0.098, suggesting PA and NA are independent in this cluster.

Figure 7a displayed the correlations between the temporal dynamic effects, mean PA, and mean NA for individuals in three clusters. The edges in the figure are color-coded to represent the type of correlation, with red edges indicating positive correlations and blue edges indicating negative correlations. The thickness of each edge reflects the strength of the correlation. Only the correlations that are statistically significant are included in the figure. The results for Cluster 1, showed two significant positive correlations. Firstly, individuals with a higher AR effect for PA were also found to have a higher AR effect for NA ($r$ = 0.27, $p < 0.001$). This suggests that the persistence of positive and negative affect within individuals in Cluster 1 is positively related. Secondly, individuals with a higher AR effect for NA were observed to have a higher CR effect from PA to NA ($r$ = 0.2, $p$=0.015), Furthermore, the negative correlation between the AR effect for NA and the mean level of PA ($r$ = -0.25, $p$ = 0.002) suggested that individuals with less carry-over of NA from one-time point to the next tend to have a higher trait level for PA. Similarly, the negative correlation between the mean of NA and the CR effect for PA on NA ($r$=-

0.35, $p$<0.001) was found. It is noticed that the CR effect for PA on NA is negative in Cluster 1. It implied that the lower trait level for NA may result from the ability to shield NA from PA

I further explored the correlations between the AR and CR effects, the mean of PA, the mean of NA, and individuals' personality traits. No significant correlations were found between the AR and CR effects and personality traits. However, the mean of PA demonstrated significant positive correlations with extraversion ($r$=0.35, $p$<0.001) and conscientiousness ($r$=0.21, $p$=0.031) and a significant negative correlation with neuroticism ($r$=-0.48, $p$<0.001). On the other hand, the mean of NA exhibited opposite patterns, with negative correlations observed with extraversion ($r$=-0.37, $p$<0.001) and conscientiousness ($r$=-0.25, $p$=0.001) and a positive correlation with neuroticism ($r$=0.53, $p$<0.001). These findings are consistent with previous research in this area (e.g., Gutiérrez, Jiménez, Hernández & Pcn, 2005).

Cluster 2 comprised 29.5% of the participants and exhibited distinct characteristics compared to the other clusters. Individuals in Cluster 2 displayed relatively weaker AR coefficients on PA, with a value of 0.31, indicating less persistence in positive emotions. In contrast, their NA demonstrated remarkable stability, with an AR effect of 0.58, suggesting prolonged negative emotions. This pattern aligned with their higher scores on depression, anxiety, and stress, as well as their lower scores on happiness. The larger proportion of females in Cluster 2 may have contributed to some differences between this cluster and Clusters 1 and 3, as women generally tend to experience slightly more persistent negative emotions than men (e.g., Verduyn & Lavrijsen, 2015). In terms of personality traits, individuals in cluster 2 have the highest value on neuroticism, which corresponded to their higher AR coefficients on NA. Additionally, they exhibited relatively lower levels of extraversion, which aligned with their weaker AR effect on PA. Regarding the CR coefficients, Cluster 2 showed a negative temporal

link from the previous NA to the current PA, with a coefficient of -0.17. This indicated that when these participants experienced anxiety or fatigue (NA), they were less likely to feel relaxed or energetic (PA) six hours later.

Within Cluster 2, the positive correlation between the AR effect for PA and the AR effect for NA was also observed ($r$=0.26, $p$=0.031) (shown in Figure 7b). In contrast to Cluster 1, there was a negative correlation between the AR effect for PA and the CR effect from PA to NA ($r$=-0.43, $p$<.001), suggesting that individuals with lower levels of carry-over effect of PA tend to exhibit a stronger spill-over effect from PA to NA. The higher mean NA was associated with less carry-over in PA ($r$=-0.31, $p$=0.001) and a higher effect from PA to NA ($r$=0.48, $p$<0.0001). This pattern seems to suggest that higher trait levels of NA of Cluster 2 may emerge from the ability of the spill-over from PA into NA. Conversely, a lower mean of PA was associated with a stronger CR effect from PA on NA ($r$=-.44, $p$<0.001).

Regarding the association between personality traits and trait affects in Cluster 2, individuals with higher levels of neuroticism tended to have a lower PA ($r$=-0.61, $p$<.001) but a higher NA ($r$=-0.49, $p$<.001) at the trait level. Unlike the individuals within Cluster 1, the individual with a higher level of openness tends to have a lower AR effect on PA ($r$=-0.33, $p$=0.03).

Cluster 3 comprised the remaining 6.2% of participants. Individuals in this cluster exhibited the strongest AR coefficient for PA, indicating a persistent experience of positive emotions over time, with a value of 0.48. In contrast, their NA had the weakest AR coefficient with a value of 0.19, indicating that their negative emotions were less likely to persist over time. This emotional stability pattern aligned with their highest score on happiness and lowest scores on depression, stress, and anxiety. Furthermore, individuals in Cluster 3 had the highest scores

on agreeableness and conscientiousness, which have been associated with higher levels of well-being and lower levels of negative affect in previous research (Soto, 2015). In terms of temporal dynamics, Cluster 3 exhibited a stronger CR coefficient from the previous NA to the current PA. This suggests that when participants in this cluster experienced dullness or irritation (NA), they were more likely to feel enthusiastic or calm (PA) in the subsequent measurement event.

In contrast to cluster 1, 3 exhibited a stronger CR coefficient between PA and NA, suggesting that PA and NA are not distinct and independent factors within individuals. These findings are consistent with previous research (Rush & Hofer, 2014), which suggests that the relationship between PA and NA is not distinctly independent at the individual level. However, it is important to note that the distinct and independent relationship between PA and NA becomes evident at the between-person level, rather than within individuals. Because the CR coefficients fade once they are averaged across all the participants. The results highlighted the importance of identifying the heterogeneity of between-individual differences when summarizing the dynamics of a sample of people.

The relationships between AR and CR effects of affect within Cluster 3 (shown in Figure 7c) did not yield any significant findings. However, there were significant associations between personality traits and affect dynamics. Specifically, extraversion was positively correlated with the AR effect for PA ($r$=0.57, $p$=0.05) and the CR effect from NA to PA ($r$=0.72, $p$=0.001). This suggests that individuals with higher levels of extraversion tend to have higher carry-over effects on PA and spill-over effects from NA to PA, from the previous 6 hours to the next. These patterns may indicate that individuals within Cluster 3, who have the highest scores on extraversion, are able to maintain higher levels of PA by both sustaining positive emotions and transferring from NA into PA. In addition, a higher level of conscientiousness was related to a

stronger CR effect from NA to PA ($r$=0.6, $p$=0.04), suggesting that individuals with higher levels of conscientiousness also maintain higher trait levels of PA.

In summary, the analysis of the three clusters revealed distinct dynamic patterns of affect and their associations with personality traits and mood symptoms. The clusters provided meaningful insights into the heterogeneity of individuals' affective dynamics and highlighted important differences between individuals belonging to each cluster. The application of the LO method with model-based clustering to real data yielded a meaningful and informative clustering solution.

## Discussion and Conclusion

I compared two temporal order selection strategies, HO and LO, in identification of unknown clusters of participants who share the same dynamics, when the participants follow the dynamics process with different temporal orders. In addition, I examined whether there are differences in the clustering performance when applying HO or LO in combination with two widely used clustering algorithms: the non-probabilistic clustering algorithm (e.g. *k-means*) and the probabilistic clustering algorithm (e.g., GMM). Simulation data were generated by varying (1) the proportion of ILDs within each dataset generated by VAR(2) the model ; (2) the number of time-ordered observations; (3)sample size; (4) the number of clusters; (5) distance of dynamics between clusters and (6) the number of cluster-specific effective coefficients (or the number of cluster-specific dynamic coefficients that are differences across clusters).

The simulation study results demonstrated that LO, in conjunction with the GMM algorithm, consistently achieved the best clustering performance in terms of recovery of the number of clusters, cluster membership, and cluster-specific dynamics for each level of all design factors. On the other hand, as the proportion of ILDs generated by higher temporal order models increased, the improvement in clustering performance with LO was not as pronounced as with HO. Among the design factors, the number of time-ordered observations played a crucial role in enhancing the clustering performance of both LO and HO methods. Particularly, the clustering accuracy of the HO strategy showed marked improvement with an increasing number of time-ordered observations. However, even with this improvement, the clustering accuracy achieved with LO remained higher than that of the HO method. Furthermore, the GMM algorithm consistently outperformed the *k-means* algorithm in terms of clustering performance.

However, when HO was applied, GMM showed more vulnerability with smaller numbers of participants.

Model-based ILD clustering is a powerful method. It assumes all the participants' ILD are generated by the same type of time-series model (e.g., VAR model) with the same temporal order. The underlying time-series model is utilized to estimate dynamic coefficients, which capture the relationships and patterns in the data. These dynamic coefficients are then used to calculate the similarities between individuals, forming the basis for clustering. For accurate clustering results, it is crucial to specify the correct model that best represents the underlying dynamics of the ILDs. (e.g., Coke & Tsao, 2010; Gao et al., 2021; Ren &Barentt,2022; Wang & Tsay, 2019; Xiong & Yeung, 2002). In the current study, it was found that when the multivariate ILDs were generated by the dynamic processes with higher temporal order, the performance of model-based clustering using the true temporal order (HO strategy) was inferior to that using a lower temporal order (LO strategy). The inconsistent conclusion results from most existing studies on model-based clustering primarily focused on univariate time-series data. However, when multivariate ILDs follow the dynamic process with higher temporal order, applying model-based clustering methods presents two challenges: (1) the estimation of a large number of dynamic coefficients and (2) the utilization of a large number of features in the clustering algorithm, which may contribute to the aforementioned inconsistent conclusion.

Accurately estimating a large number of dynamic coefficients requires an adequate number of time-ordered observations to ensure precise estimation. When a model is overfitted, such as fitting a time-series model with higher temporal order than the true order, the issue lies in the large variance of parameter estimation rather than biased estimation (Babyak,2004; McNeish, 2015). Even if a true underlying data-generating model is correctly specified to fit a single

sample dataset, the problem of large variance can persist if there are insufficient observations relative to the number of parameters (Babyak,2004; Coolen et al., 2017). As the number of parameters that need to be estimated increases combined with the decrease of number of time-ordered observations or effect size of parameters, the problem of large variance becomes more pronounced (Coolen et al., 2017; Yarkoni & Westfall, 2017). Conversely, when a model is underfitted, such as fitting a time-series model with lower temporal order than the true order, the estimation of the parameter tends to be biased but with smaller variance. It is important to notice that bias refers to the tendency for a model to consistently produce an incorrect value in a particular direction rather than the difference between the estimated and population parameter value. On the other hand, variance refers to the extent to which a model' s estimated parameters tend to deviate from their central tendency across different datasets. (Yarkoni & Westfall, 2017).

Previous studies demonstrated that inaccurate estimation of dynamic coefficients and poor prediction of person-specific multivariate time series models could arise due to a combination of factors. These factors include the need to estimate a large number of dynamic effects resulting from a time-series model with a higher temporal order and limited availability of time-ordered observations (Coolen et al., 2017; Ivanov & Kilian, 2005; Kilian, 2000). For example, in a study by Lawford & Stamatogiannis (2009), a series of VAR (1) models were simulated with varying dimensions (i.e., number of ILD data) and different numbers of measurement occasions. The simulated data were then used to fit to the corresponding data-generation models, and the estimated dynamic coefficients were compared to the true dynamic coefficients of the underlying data-generating model. The findings revealed that even when true models were employed, the variance of the estimated dynamic coefficient increased with the number of dynamic coefficients in the underlying model. On the other hand, the accuracy of

coefficient estimation improved and approached the values of the underlying model as the number of time-ordered observations increased. In the study by Bulteel et al. (2018), it was demonstrated that model underfitting led to more accurate predictions and parameter estimations. They conducted a comparison between AR (1) and VAR (1) models using varying numbers of time-ordered observations. Interestingly, even when the true underlying data-generating model was VAR (1), AR (1) consistently outperformed in terms of parameter estimation and prediction accuracy, when the number of time-ordered observations is small. However, as the number of observations increased, the accuracy of the VAR (1) model eventually surpassed that of the AR (1) model.

Furthermore, Kilian (2000) and Ventzislav & Kilian (2005) studies shed light on the asymmetric effects of fitting a VAR model with a lag order lower than the optimal model order (i.e., underfitting) and a model with extra higher-order dynamics (i.e., overfitting). These studies investigated the influence of temporal order selection on the accuracy of impulse response estimators based on dynamic coefficients derived from the VAR($p$) model. The results indicated that incorporating extra higher-order lags consistently raises the error of impulsive response function parameters. The relationship between the number of dynamic coefficients and the error was nonlinear, indicating that the higher-order dynamics were included, the greater the impact on parameter accuracy. Conversely, the studies also found that a moderate level of underfitting, where the temporal order was lower than the optimal model order, could actually decrease the error when only a few time-ordered observations were available. In such scenarios, the bias caused by underfitting is considered to be less critical compared to the variance resulting from estimating a large number of parameters. However, as the number of time-ordered observations increases, the negative effects of both overfitting and underfitting become more symmetric. In

such cases, the cost of underfitting becomes more pronounced, especially as the number of time-ordered observations grows.

More important, during the clustering stage, the dynamic coefficients obtained from a multivariate time-series model are utilized as features to assess the similarities between individuals. The substantial difference between estimated dynamic coefficients and cluster-specific true coefficients can obscure the true similarities across individuals' dynamic coefficients, thereby reducing the accuracy of clustering, particularly for clustering algorithms that heavily rely on the distance among individuals, such as *k-means*. The results of the simulation study support this assumption. With a limited number of time-ordered observations (e.g., $t$=50), the k-means algorithm exhibited poorer recovery in terms of cluster membership (ARI=0.06 for $t$=50) compared to the GMM algorithm (ARI=0.18 for $t$=50) when HO strategy was applied. Furthermore, with an increasing number of time-ordered observations (e.g., t=1000), the clustering performance of *k-means* exhibits a substantial improvement (ARI=0.79 for $t$=1000).

Related to this point, increasing the number of time-ordered observations can help reduce the variance in the estimation of dynamic coefficients, leading to more accurate clustering results. The results of the current study support this finding, as the increment in the number of time-ordered observations showed a significant improvement in clustering performance for the HO strategy. However, it is important to note that increasing the number of time-ordered observations does not address the second issue of the large number of features used in the clustering algorithm. This issue can still pose challenges in accurately clustering individuals based on their dynamic coefficients.

The second issue with model-based clustering for multivariate ILDs is the utilization of a large number of features in the clustering analysis. This leads to data sparsity in high-dimensional space, which impairs the assessment of similarities among individuals (Steinley, 2006). With a large number of features, individuals tend to be equidistant from one another, rendering any similarity assessment for clustering purposes meaningless (Beyer et al., 1999). Furthermore, the inclusion of additional features introduces both new information and noise, thereby diminishing the accuracy of similarity assessments (Assent, 2012; Steinbach, Ertöz & Kumar, 2004). Consequently, clustering algorithms that rely on similarities, such as *k-means*, face restrictions in their performance.

In the case of a probabilistic clustering algorithm like GMM, the increasing number of features not only leads to data sparsity (Scott &Thompson, 1983) but also results in a higher number of parameters that need to be estimated in the clustering algorithm. This, in turn, necessitates a larger sample size to ensure an accurate estimation of the clustering algorithm's parameters (Bouveyron & Brunet-Saumard, 2014; Pavlenko, 2003; Pavlenko and Von Rosen, 2001). As a result, the findings of the simulation study indicated that when using the HO method, the GMM algorithm was more sensitive to the sample size compared to k-means.

In addition to the higher temporal order, the inclusion of a large number of intensive longitudinal variables in clustering analysis introduces a substantial number of dynamic coefficients. Even when using a parsimonious model such as VAR (1), the resulting dynamic coefficients can be numerous. To achieve satisfactory clustering performance, particularly for the GMM algorithm, one solution is collecting a large number of time-ordered observations to ensure accurate estimation of the dynamic coefficients, as well as a sizable sample size for precise parameter estimation of the clustering algorithm. Unfortunately, collecting a large

number of time-ordered observations or recruiting a sufficient number of participants can pose a big challenge in most social and behavioral science research.

An alternative approach to enhance clustering performance for ILDs is to reduce the number of dynamic coefficients and features used in the clustering analysis. This can be accomplished through the utilization of variable selection techniques in both the stage of dynamic coefficients estimation (Nicholson et al., 2020) and the clustering stage (Bouveyron & Brunet-Saumard, 2014). By carefully selecting relevant variables, it can improve the accuracy and efficiency of the model estimation and clustering process while addressing the difficulties associated with high-dimensional data. Furthermore, several techniques have been developed to address the challenges of estimating high-temporal order time-series models with a limited number of time-ordered observations (e.g., Basu & Michailidis, 2015; Brodinová, et al.,2019; Hsu, Hung & Chang, 2008; Raftery & Dean, 2006; Song & Bickel,2011) and clustering with a large number of features and limited sample size (e.g. Levina, Rothman & Zhu, 2008; Maugis, Celeux, & Martin-Magniette,2009). These techniques have been the subject of extensive research and offer potential solutions for improving clustering accuracy in multivariate ILDs.

A critical drawback of the two-step clustering method is its assumption that the importance of AR effect coefficients and CR effect coefficients are the same across all lags. However, in reality, AR coefficients may contain more informative data than CR coefficients, and the importance of coefficients tends to decrease as the lag order increases. To address this limitation, Kalpakis, Gada, and Puttagunta (2001) proposed using cepstral distance as a similarity measurement, which considers the decreasing importance of AR coefficients with higher order. Although cepstral distance has proven effective for univariate ILDs, there is currently no known similarity measurement for multivariate ILDs that incorporates this concept.

Another limitation of the current clustering method is its applicability only to completed ILDs. In practice, ILDs often involve missing data due to participant non-compliance or study design. Therefore, it is crucial to handle missing data appropriately before conducting a clustering analysis. In the empirical study, I employed multiple imputation methods to address missing data due to multiple imputations has been shown to be an accurate and efficient method for handling the missingness issue (Ji, et al., 2018). However, the previous study didn’ t examine the accuracy of multiple imputations under some special conditions that commonly occur in real data. For instance, there may be instances where all the values of intensive longitudinal variables are missing for a given person at a specific time point or when consecutive missing values occur over a specific period (e.g., a whole week) for a given person. Future research should explore the performance and robustness of missing methods handling methods, including multiple imputations, machine learning methods, and deep learning methods under different missing data patterns and conditions, and also investigate the impact of these methods on ILDs clustering.

The simulation study had several limitations to mention. Firstly, one issue is the number of variables used in the simulation study. Previous simulation studies investigating clustering performance on ILD have typically employed a range of four to six variables (e.g., Bulteel et al., 2016; Ernst et al., 2021; Ernst, Albers, Jeronimus & Timmerman, 2020; Takano et al., 2021). To achieve accurate estimation of a large number of dynamic coefficients necessitates a substantial number of time-ordered observations (e.g., Bulteel et al., 2018; Kilian, 2000; Ventzislav & Kilian, 2005). Furthermore, the high-dimensionality of dynamic coefficients poses a restriction on clustering algorithms' performance (Assent, 2012; Beyer et al., 1999; Bouveyron & Brunet-Saumard, 2014; Steinbach, Ertöz & Kumar, 2004; Steinley, 2006). Consequently, the present

study set the number of intensive longitudinal variables as four. However, even when using VAR (1) models to fit the ILDs, the chosen number of variables resulted in numerous dynamic coefficients that required estimation. This factor could potentially impact the clustering performance, particularly when comparing the superiority of the LO strategy over HO, even with a large number of time-ordered observations (e.g., $t$=1000) generated by a higher temporal order model. Therefore, it would be valuable to explore the clustering performance by varying the number of ILD variables.

Another limitation of the simulation study relates to the he positioning of between-cluster differences in the case of ILDs following a VAR (2) process. In the current simulations, it was assumed that the between-cluster differences occur evenly in both lags. However, the study did not explore the clustering performance of the LO approach when the between-cluster differences were exclusively present in higher temporal orders, such as lag-2. Although such conditions are uncommon in real data, further investigations about how the LO approach performs under such circumstances are still needed.

Thirdly, the simulations were limited to the VAR models, and it is important to note that the results may differ from other time-series models. The VAR model assumes stationarity of the ILDs, but in practice, this assumption may be violated. Preprocessing techniques such as differencing, detrending, and deseasoning can be applied to address non-stationarity in ILDs. However, the impact of preprocessing ILDs on clustering accuracy remains unexplored and warrants further investigation. Moreover, the VAR model assumes a fixed dynamic relationship over time. However, in reality, the dynamic relationship among variables may vary across time. Model-based clustering approaches that incorporate time-varying dynamic relationships (e.g.,

Bringmann et al., 2018; Commandeur & Koopman, 2007) could be explored to improve clustering accuracy.

**Recommendations and Suggestions**

To sum up, when dealing with a limited number of time-ordered observations, it is advisable to employ a parsimonious model, such as VAR(1), for all participants. This helps to mitigate the challenges posed by parameter estimation and high-dimensionality issues in the clustering stage. Furthermore, when working with a large number of intensive longitudinal variables and a limited sample size, it is recommended to use non-probabilistic clustering algorithms like *k-means*. Alternatively, incorporating variable selection techniques can help reduce the number of features and enhance clustering performance.

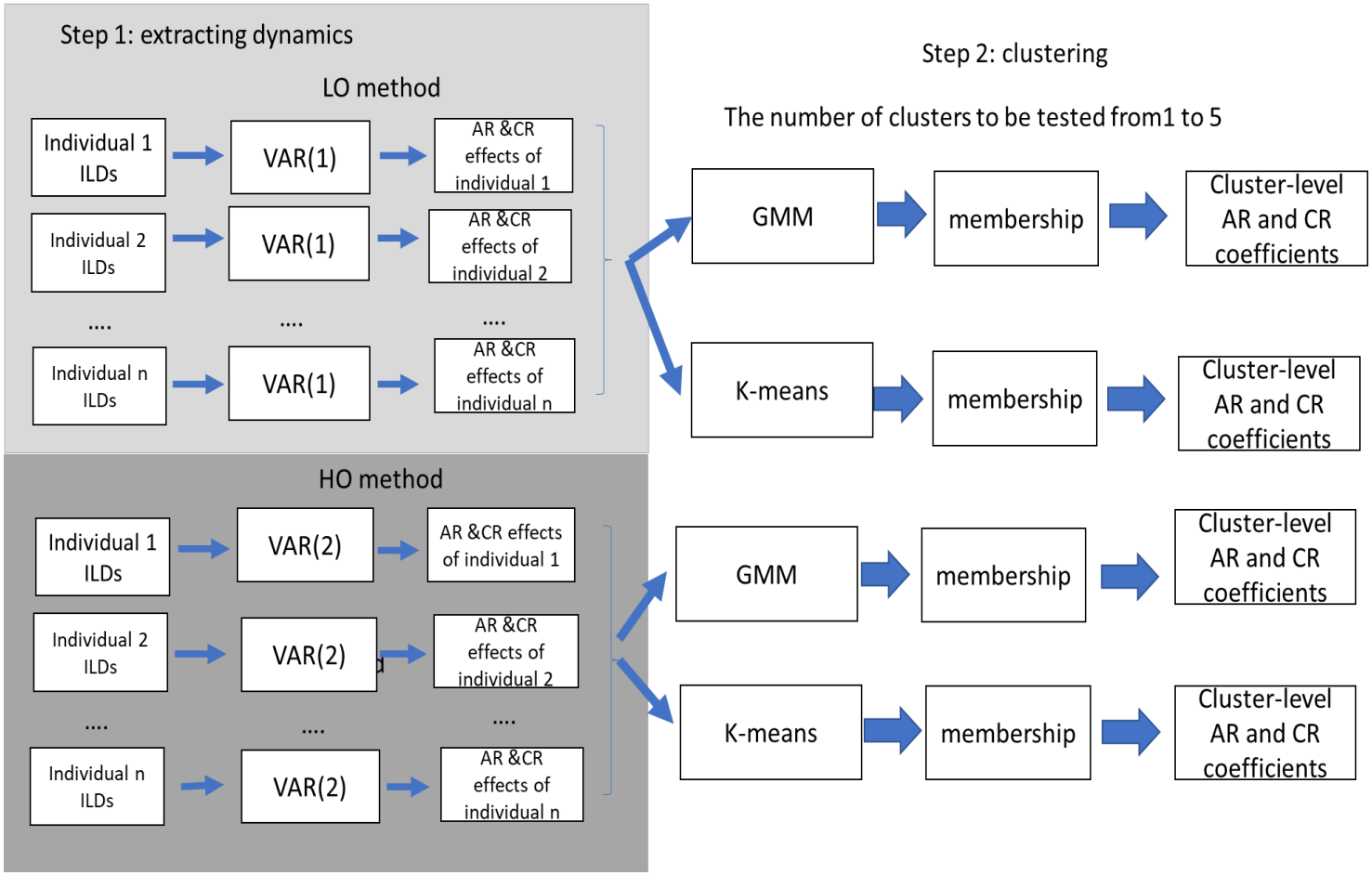



Figure 1. The procedures of two-steps model-based clustering method

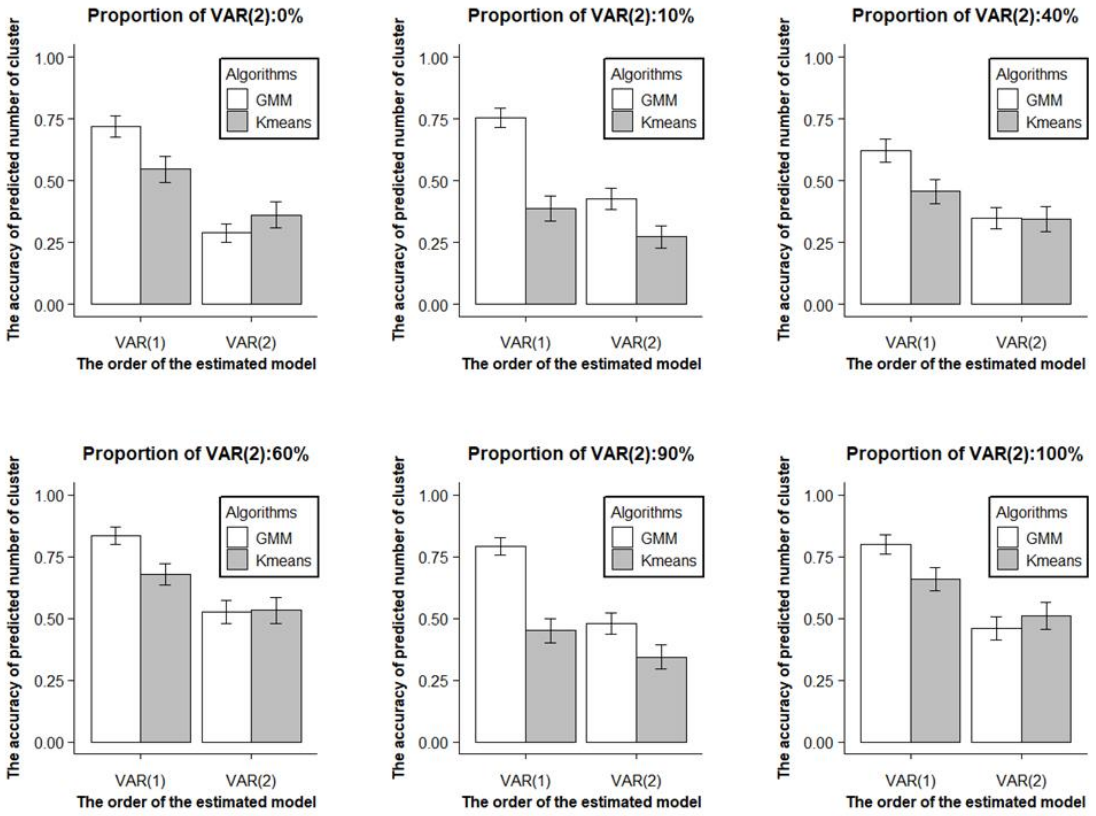



Figure 2. The accuracy of recovered number of clusters of under HO/LO and the GMM/k-means clustering algorithms.

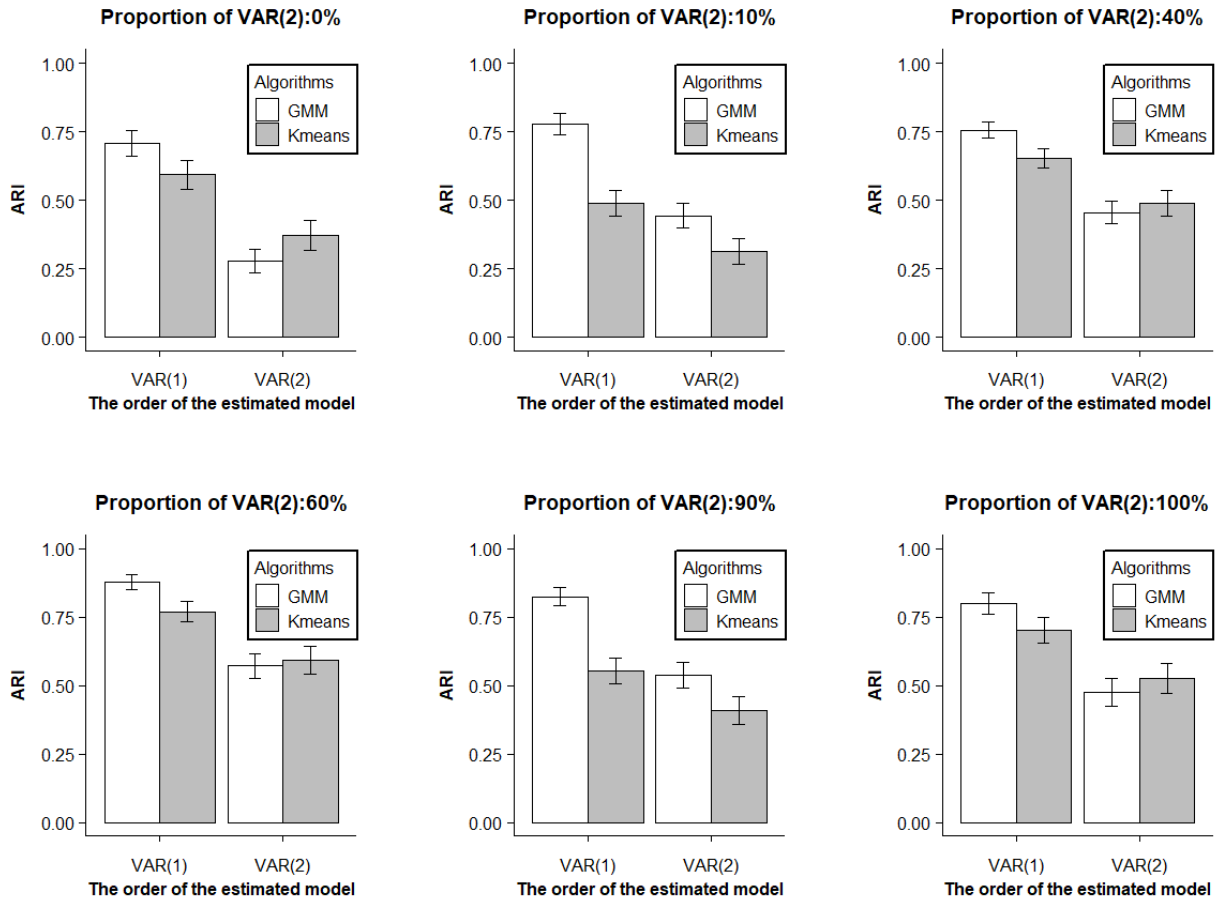



Figure 3. The accuracy of recovered cluster membership (ARI) under HO/LO and the GMM/k-means clustering algorithms.

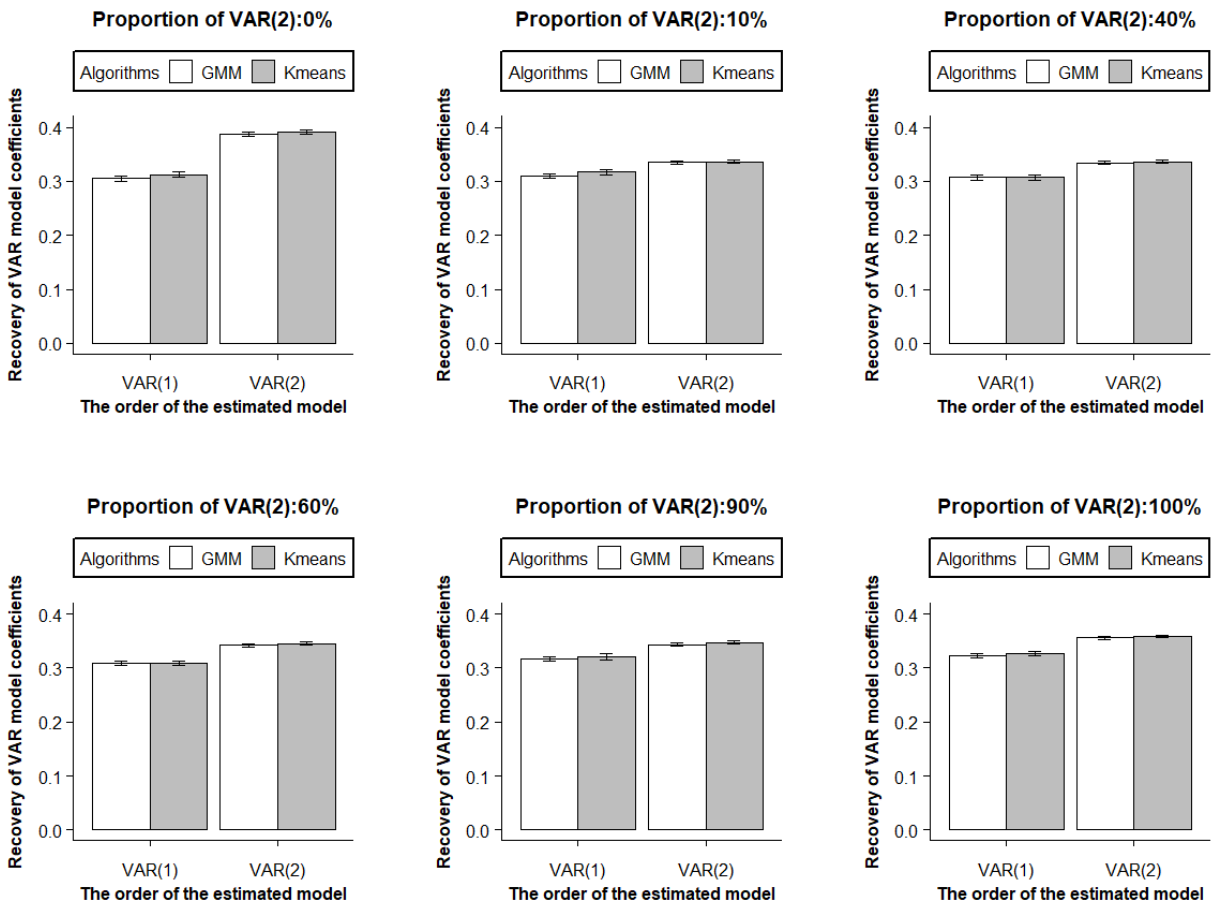



Figure 4. The Euclidian distance between the true and estimated VAR coefficients under HO/LO and the GMM/k-means clustering algorithms

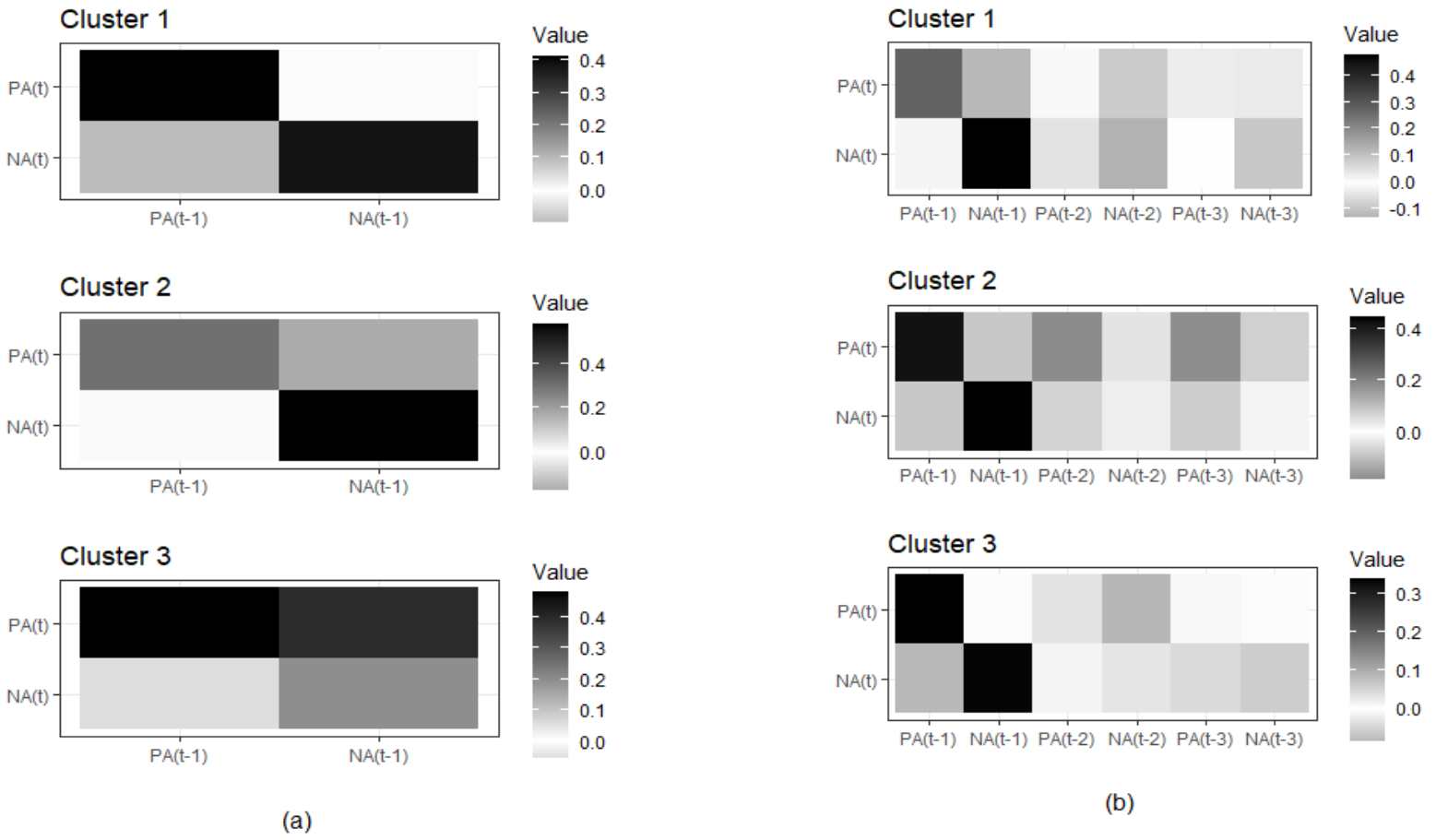



Figure 5. Cluster-specific dynamical coefficients for Positive Affect (PA) and Negative Affect (NA)

Note. (a) cluster-specific AR effects and CR effects through the GMM algorithm in conjunction with LO method. (b) cluster-specific AR effects and CR effects through the GMM algorithm in conjunction with HO method

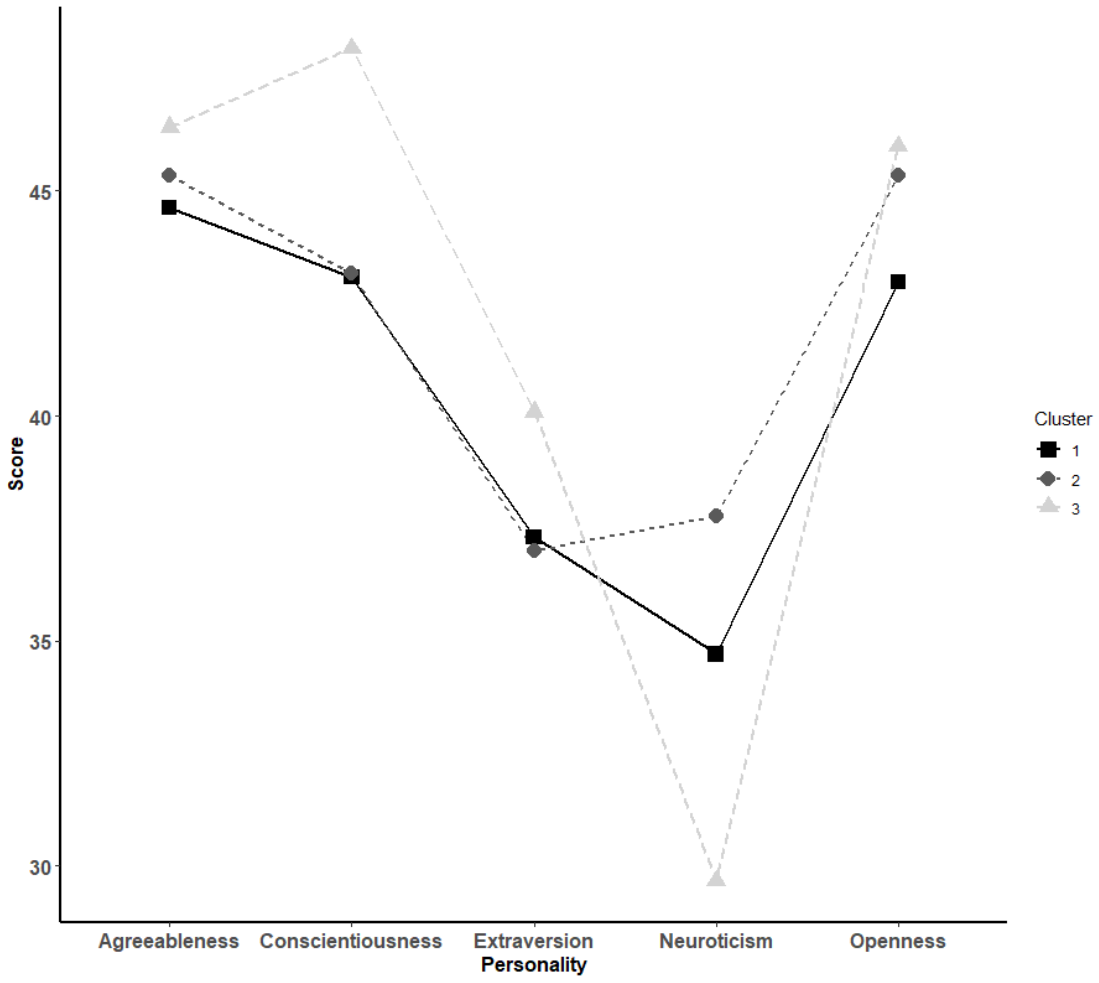



Figure 6. Score on personality measurements of three clusters obtained by LO strategy with GMM.

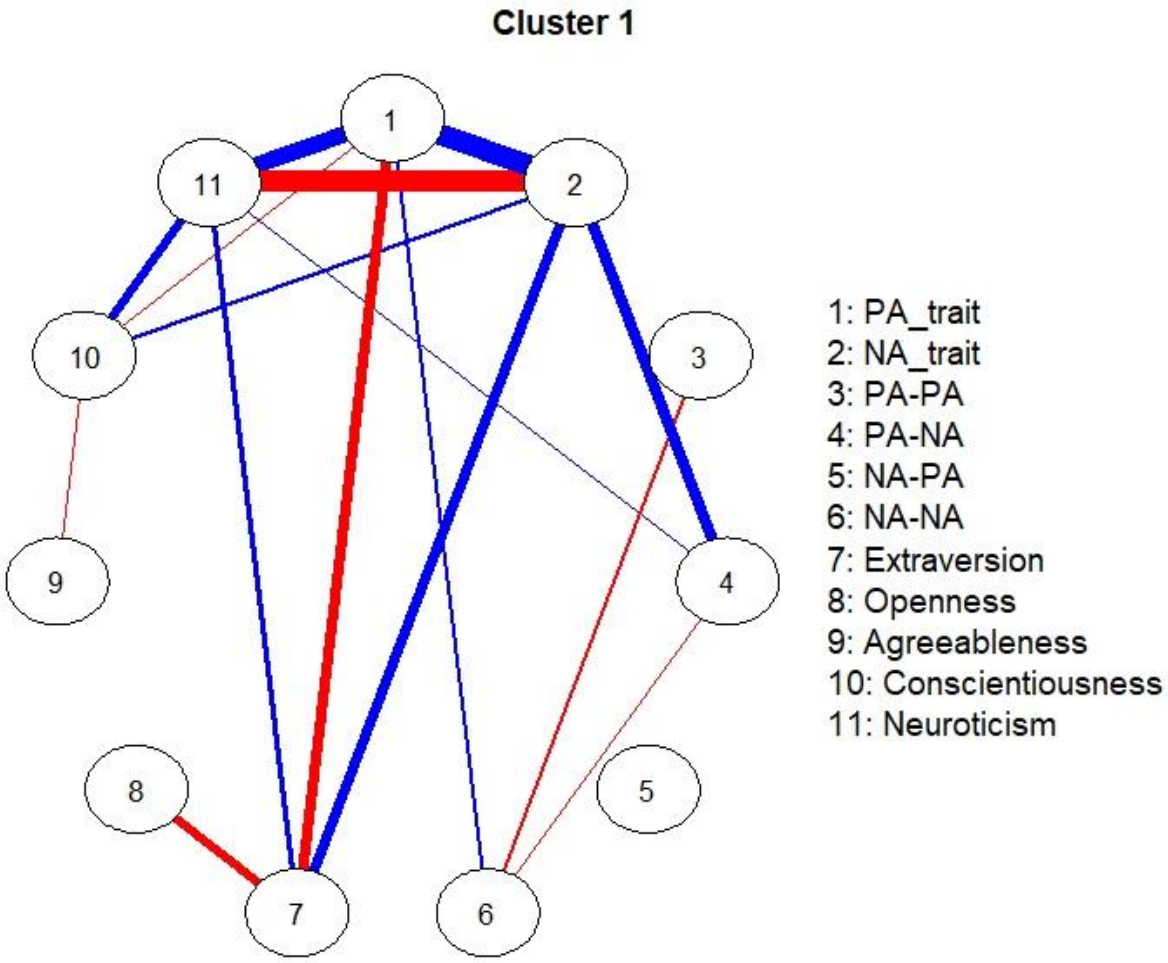



Figure 7a. Network of correlations between trait affects, AR and CR coefficients, and big-five personality traits for Cluster 1 obtained by LO strategy with GMM

Note. The node PA_trait and NA_trait represent the mean of PA and mean of NA, respectively. PA-PA and NA-NA refers to the AR effects for PA and AR effects for NA, respectively. PA-NA represents the CR effect from PA on NA and NA-PA represents the CR effect from NA on PA. The node from 7-9 represents the 5 domains of big-five personality traits.

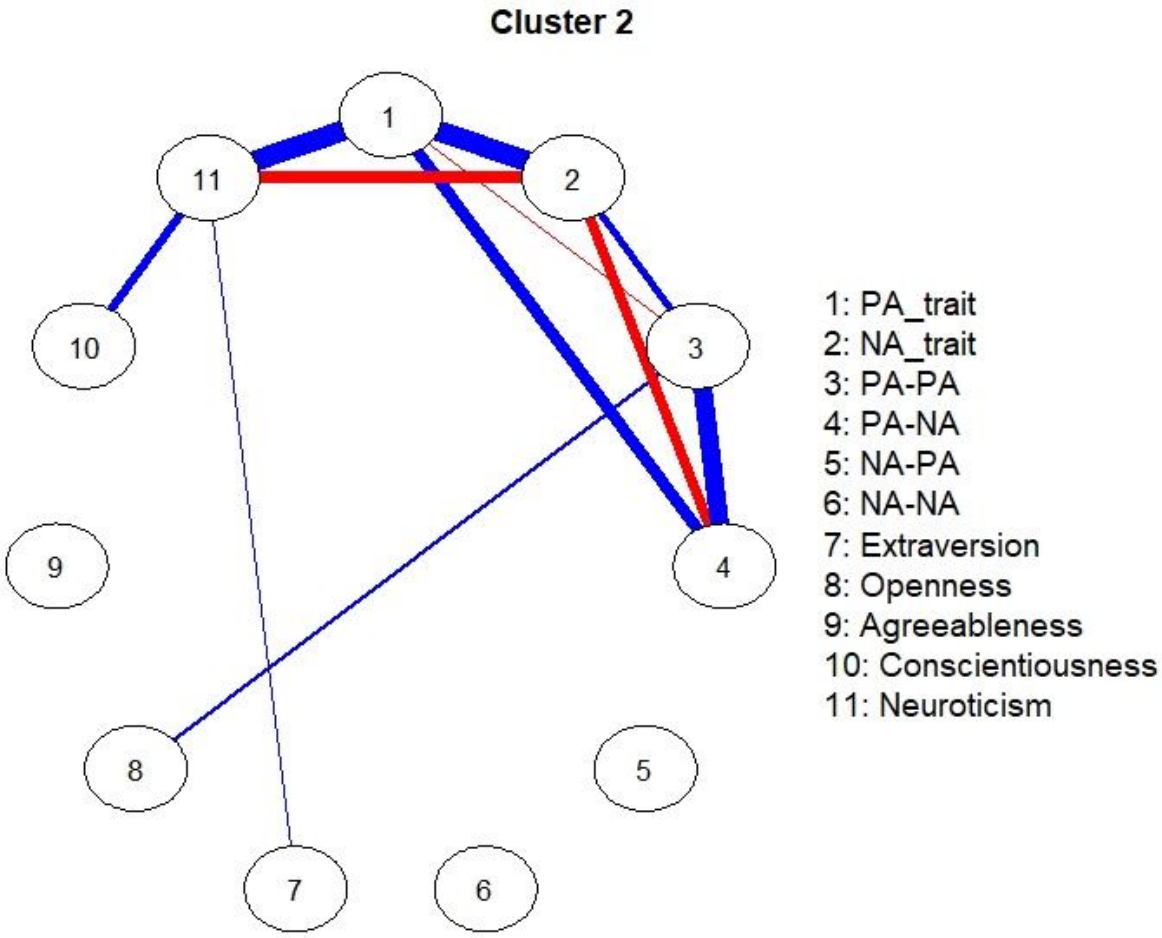



Figure 7b. Network of correlations between trait affects, AR and CR coefficients, and big-five personality traits for Cluster 2 obtained by LO strategy with GMM

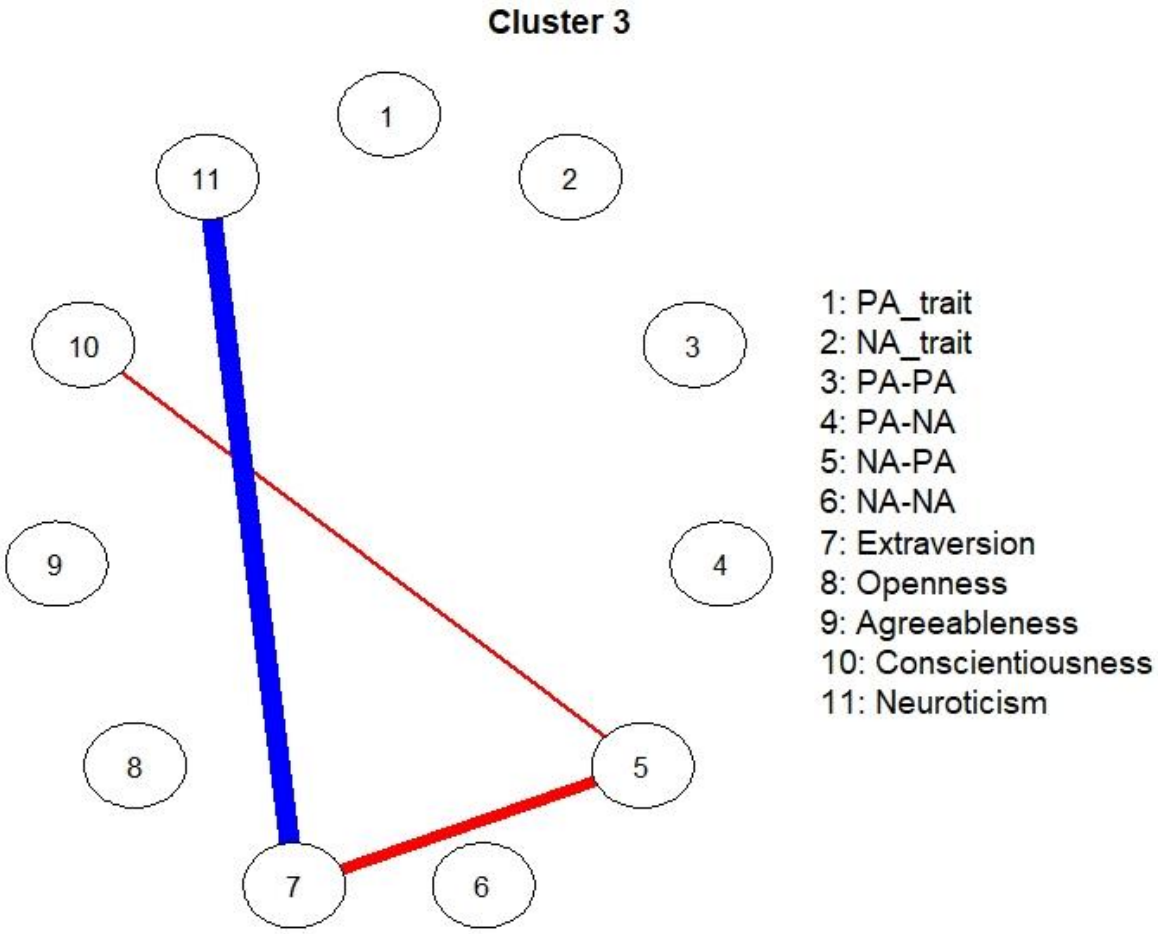



Figure 7c. Network of correlations between trait affects, AR and CR coefficients, and big-five personality traits for Cluster 3 obtained by LO strategy with GMM.

Table 1. Average probabilities of correctly predicted number of clusters.

| Factors | Levels | Clustering algorithm | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | GMM | | | | K-means | | | |
| | | Temporal order selection method | | | | | | | |
| | | LO | | HO | | LO | | HO | |
| | | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| **Proportion of ILDs generated by VAR(2)** | 0% | 0.72 | 0.37 | 0.29 | 0.32 | 0.55 | 0.43 | 0.36 | 0.45 |
| | 10% | 0.76 | 0.32 | 0.43 | 0.36 | 0.39 | 0.41 | 0.27 | 0.38 |
| | 40% | 0.62 | 0.39 | 0.35 | 0.37 | 0.46 | 0.40 | 0.34 | 0.42 |
| | 60% | 0.83 | 0.28 | 0.53 | 0.39 | 0.68 | 0.36 | 0.53 | 0.45 |
| | 90% | 0.79 | 0.30 | 0.48 | 0.37 | 0.45 | 0.41 | 0.34 | 0.42 |
| | 100% | 0.80 | 0.32 | 0.46 | 0.40 | 0.66 | 0.39 | 0.51 | 0.46 |
| **Number of clusters** | 2 | 0.88 | 0.20 | 0.53 | 0.36 | 0.46 | 0.39 | 0.37 | 0.42 |
| | 3 | 0.75 | 0.34 | 0.39 | 0.36 | 0.60 | 0.41 | 0.44 | 0.45 |
| | 4 | 0.64 | 0.40 | 0.34 | 0.38 | 0.53 | 0.44 | 0.38 | 0.45 |
| **Sample size** | 60 | 0.67 | 0.34 | 0.35 | 0.30 | 0.54 | 0.42 | 0.39 | 0.44 |
| | 120 | 0.76 | 0.35 | 0.34 | 0.38 | 0.53 | 0.42 | 0.40 | 0.44 |
| | 240 | 0.84 | 0.31 | 0.58 | 0.39 | 0.52 | 0.41 | 0.40 | 0.44 |
| **Number of time-ordered observations** | 50 | 0.53 | 0.36 | 0.22 | 0.25 | 0.21 | 0.28 | 0.05 | 0.14 |
| | 100 | 0.68 | 0.36 | 0.27 | 0.30 | 0.40 | 0.37 | 0.20 | 0.31 |
| | 500 | 0.88 | 0.25 | 0.53 | 0.38 | 0.71 | 0.38 | 0.62 | 0.43 |
| | 1000 | 0.92 | 0.21 | 0.65 | 0.38 | 0.80 | 0.34 | 0.71 | 0.40 |
| **Distance of dynamics between clusters** | 0.05 | 0.54 | 0.38 | 0.23 | 0.28 | 0.28 | 0.34 | 0.15 | 0.28 |
| | 0.1 | 0.71 | 0.36 | 0.35 | 0.35 | 0.44 | 0.42 | 0.33 | 0.42 |
| | 0.15 | 0.80 | 0.30 | 0.44 | 0.38 | 0.56 | 0.41 | 0.43 | 0.45 |
| | 0.2 | 0.85 | 0.27 | 0.52 | 0.38 | 0.65 | 0.39 | 0.50 | 0.45 |
| | 0.25 | 0.88 | 0.25 | 0.57 | 0.38 | 0.72 | 0.37 | 0.57 | 0.44 |
| **Number of effective effects** | 2 | 0.61 | 0.38 | 0.26 | 0.30 | 0.38 | 0.40 | 0.28 | 0.39 |
| | 4 | 0.71 | 0.36 | 0.35 | 0.34 | 0.47 | 0.41 | 0.34 | 0.43 |
| | 6 | 0.77 | 0.33 | 0.41 | 0.37 | 0.52 | 0.41 | 0.39 | 0.44 |
| | 8 | 0.79 | 0.32 | 0.46 | 0.39 | 0.57 | 0.41 | 0.42 | 0.45 |
| | 10 | 0.82 | 0.30 | 0.50 | 0.39 | 0.61 | 0.41 | 0.45 | 0.45 |
| | 12 | 0.83 | 0.30 | 0.53 | 0.39 | 0.63 | 0.40 | 0.48 | 0.45 |

Table 2. Average agreement between the true and the recovered cluster membership (ARIs).

| Factors | Levels | Clustering algorithm | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | GMM | | | | K-means | | | |
| | | Temporal order selection method | | | | | | | |
| | | LO | | HO | | LO | | HO | |
| | | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| **Proportion of ILDs generated by VAR(2)** | 0% | 0.71 | 0.39 | 0.28 | 0.35 | 0.59 | 0.44 | 0.37 | 0.46 |
| | 10% | 0.78 | 0.32 | 0.44 | 0.38 | 0.49 | 0.40 | 0.31 | 0.39 |
| | 40% | 0.76 | 0.25 | 0.45 | 0.35 | 0.65 | 0.30 | 0.49 | 0.39 |
| | 60% | 0.88 | 0.22 | 0.57 | 0.38 | 0.77 | 0.31 | 0.59 | 0.43 |
| | 90% | 0.82 | 0.27 | 0.54 | 0.39 | 0.55 | 0.39 | 0.41 | 0.42 |
| | 100% | 0.80 | 0.33 | 0.47 | 0.42 | 0.70 | 0.39 | 0.53 | 0.47 |
| **Number of clusters** | 2 | 0.83 | 0.28 | 0.48 | 0.40 | 0.55 | 0.38 | 0.40 | 0.43 |
| | 3 | 0.80 | 0.31 | 0.46 | 0.39 | 0.69 | 0.38 | 0.50 | 0.45 |
| | 4 | 0.74 | 0.33 | 0.44 | 0.39 | 0.64 | 0.38 | 0.45 | 0.44 |
| **Sample size** | 60 | 0.70 | 0.34 | 0.34 | 0.33 | 0.62 | 0.39 | 0.44 | 0.44 |
| | 120 | 0.81 | 0.30 | 0.35 | 0.38 | 0.63 | 0.38 | 0.45 | 0.44 |
| | 240 | 0.86 | 0.26 | 0.69 | 0.35 | 0.63 | 0.38 | 0.46 | 0.44 |
| **Number of time-ordered observations** | 50 | 0.55 | 0.35 | 0.18 | 0.26 | 0.30 | 0.33 | 0.06 | 0.18 |
| | 100 | 0.74 | 0.32 | 0.32 | 0.34 | 0.53 | 0.37 | 0.25 | 0.36 |
| | 500 | 0.92 | 0.18 | 0.61 | 0.36 | 0.81 | 0.29 | 0.70 | 0.37 |
| | 1000 | 0.95 | 0.14 | 0.73 | 0.33 | 0.87 | 0.23 | 0.79 | 0.32 |
| **Distance of dynamics between clusters** | 0.05 | 0.56 | 0.38 | 0.24 | 0.30 | 0.36 | 0.37 | 0.20 | 0.31 |
| | 0.1 | 0.73 | 0.33 | 0.37 | 0.37 | 0.53 | 0.40 | 0.37 | 0.43 |
| | 0.15 | 0.84 | 0.26 | 0.48 | 0.39 | 0.66 | 0.36 | 0.48 | 0.45 |
| | 0.2 | 0.89 | 0.21 | 0.57 | 0.38 | 0.76 | 0.32 | 0.56 | 0.44 |
| | 0.25 | 0.92 | 0.17 | 0.64 | 0.37 | 0.82 | 0.28 | 0.64 | 0.41 |
| **Number of effective effects** | 2 | 0.65 | 0.36 | 0.28 | 0.32 | 0.46 | 0.40 | 0.32 | 0.41 |
| | 4 | 0.75 | 0.33 | 0.38 | 0.37 | 0.56 | 0.39 | 0.39 | 0.43 |
| | 6 | 0.80 | 0.30 | 0.45 | 0.39 | 0.63 | 0.38 | 0.44 | 0.44 |
| | 8 | 0.83 | 0.28 | 0.51 | 0.39 | 0.67 | 0.37 | 0.48 | 0.44 |
| | 10 | 0.85 | 0.26 | 0.55 | 0.40 | 0.71 | 0.36 | 0.52 | 0.44 |
| | 12 | 0.86 | 0.25 | 0.58 | 0.39 | 0.73 | 0.35 | 0.55 | 0.44 |

Table 3. Average Euclidean distances between the true and estimated cluster-specific dynamics.

| Factors | Levels | Clustering algorithm | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | GMM | | | | K-means | | | |
| | | Temporal order selection method | | | | | | | |
| | | LO | | HO | | LO | | HO | |
| | | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| **Proportion of ILDs generated by VAR(2)** | 0% | 0.31 | 0.04 | 0.39 | 0.04 | 0.31 | 0.04 | 0.39 | 0.03 |
| | 10% | 0.31 | 0.03 | 0.33 | 0.02 | 0.32 | 0.05 | 0.34 | 0.03 |
| | 40% | 0.31 | 0.04 | 0.33 | 0.02 | 0.31 | 0.05 | 0.34 | 0.02 |
| | 60% | 0.31 | 0.03 | 0.34 | 0.02 | 0.31 | 0.04 | 0.34 | 0.02 |
| | 90% | 0.32 | 0.03 | 0.34 | 0.02 | 0.32 | 0.05 | 0.35 | 0.03 |
| | 100% | 0.32 | 0.03 | 0.36 | 0.02 | 0.33 | 0.04 | 0.36 | 0.02 |
| **Number of clusters** | 2 | 0.31 | 0.03 | 0.35 | 0.03 | 0.31 | 0.04 | 0.35 | 0.03 |
| | 3 | 0.31 | 0.04 | 0.35 | 0.03 | 0.32 | 0.04 | 0.35 | 0.03 |
| | 4 | 0.31 | 0.04 | 0.35 | 0.03 | 0.32 | 0.04 | 0.35 | 0.03 |
| **Sample size** | 60 | 0.31 | 0.04 | 0.34 | 0.03 | 0.32 | 0.04 | 0.35 | 0.03 |
| | 120 | 0.31 | 0.03 | 0.35 | 0.03 | 0.32 | 0.04 | 0.35 | 0.03 |
| | 240 | 0.31 | 0.03 | 0.35 | 0.03 | 0.31 | 0.04 | 0.35 | 0.03 |
| **Number of time-ordered observations** | 50 | 0.32 | 0.04 | 0.35 | 0.03 | 0.33 | 0.06 | 0.35 | 0.04 |
| | 100 | 0.31 | 0.04 | 0.35 | 0.03 | 0.32 | 0.04 | 0.35 | 0.03 |
| | 500 | 0.31 | 0.03 | 0.35 | 0.03 | 0.31 | 0.04 | 0.35 | 0.03 |
| | 1000 | 0.31 | 0.03 | 0.35 | 0.03 | 0.31 | 0.03 | 0.35 | 0.03 |
| **Distance of dynamics between clusters** | 0.05 | 0.29 | 0.04 | 0.34 | 0.03 | 0.30 | 0.05 | 0.34 | 0.03 |
| | 0.1 | 0.30 | 0.03 | 0.34 | 0.03 | 0.30 | 0.04 | 0.35 | 0.03 |
| | 0.15 | 0.31 | 0.03 | 0.35 | 0.03 | 0.31 | 0.04 | 0.35 | 0.03 |
| | 0.2 | 0.32 | 0.03 | 0.35 | 0.03 | 0.32 | 0.04 | 0.35 | 0.03 |
| | 0.25 | 0.34 | 0.03 | 0.36 | 0.03 | 0.34 | 0.04 | 0.36 | 0.03 |
| **Number of effective effects** | 2 | 0.30 | 0.04 | 0.34 | 0.03 | 0.30 | 0.04 | 0.34 | 0.03 |
| | 4 | 0.30 | 0.03 | 0.34 | 0.03 | 0.31 | 0.04 | 0.35 | 0.03 |
| | 6 | 0.31 | 0.04 | 0.35 | 0.03 | 0.31 | 0.04 | 0.35 | 0.03 |
| | 8 | 0.31 | 0.03 | 0.35 | 0.03 | 0.32 | 0.04 | 0.35 | 0.03 |
| | 10 | 0.32 | 0.03 | 0.35 | 0.03 | 0.32 | 0.04 | 0.36 | 0.03 |
| | 12 | 0.32 | 0.04 | 0.35 | 0.03 | 0.33 | 0.04 | 0.36 | 0.03 |

Table 4. Cluster-specific AR and CR effects for the PA and NA of the three clusters obtained by LO strategy with GMM.

| | **Cluster 1** | | **Cluster 2** | | **Cluster 3** | |
|---|---|---|---|---|---|---|
| | $PA_{t-1}$ | $NA_{t-1}$ | $PA_{t-1}$ | $NA_{t-1}$ | $PA_{t-1}$ | $NA_{t-1}$ |
| **$PA_t$** | 0.414 | -0.006 | 0.307 | -0.173 | 0.477 | 0.392 |
| **$NA_t$** | -0.098 | 0.389 | 0.010 | 0.580 | -0.057 | 0.195 |

Table 5. Means of the external variables for the three clusters of PA and NA obtained by LO strategy with GMM.

| Cluster | Depression (SD) (0-42) | Anxiety (SD) (0-42) | Stress (SD) (0-42) | Happiness (SD) (0-10) |
|---|---|---|---|---|
| **1** | 7.18(7.33) | 3.64(4.53) | 8.88(6.75) | 6.72(1.45) |
| **2** | 8.34(7.73) | 4.70(5.05) | 11.1(7.52) | 6.47(1.41) |
| **3** | 2.31(3.07) | 0.85(0.99) | 5.92(3.93) | 7.86(0.86) |

Table 6. Characteristics for the three clusters of PA and NA obtained by LO strategy with GMM.

| Cluster | Proportion | Age (SD) | Female% | PA(SD) | NA(SD) | Correlation of PA and NA |
|---|---|---|---|---|---|---|
| 1 | 64.3% (146) | 41.9(0.21) | 78.7% | 54.9(15.5) | 18.3(16.8) | -0.57 ($p$<0.001) |
| 2 | 29.5% (67) | 38(0.06) | 93.6% | 52.5(17.1) | 21.7(17.1) | -0.58 ($p$<0.001) |
| 3 | 6.2% (14) | 48.4(0.24) | 75.9% | 60.8(15.0) | 7.86(9.03) | -0.23 ($p$=0.43) |

## Footnote

1. ILD is more commonly referred to as a time-series data (Box & Jenkins, 1970) in other fields.
2. The BIC used for the optimal number of clusters and clusters structure in GMM algorithms was expressed as:

$$BIC_{ck} = 2L_{ck}(X, \hat{\theta}) - m_c \log(n)$$

where $L_{ck}(X, \hat{\theta})$ is the loglikelihood given the data x and model parameters $\hat{\theta}$ of GMM model. k in the subscript refers to different numbers of clusters. $m_c$ is the number of parameters of the GMM need to be estimated. n is sample size. Each $BIC_{ck}$ corresponds to the cth clusters structure with k clusters. The final number of clusters and cluster structure are chosen to maximize the BIC (Dasgupta & Raftery, 1998; Fraley & Raftery, 1998; Law, Figueiredo, & Jain, 2004; Martinez & Martinez, 2005, McLachlan & Peel, 2000, Raftery & Dean, 2006).

3. It is noticed that the total distance between a pair of clusters Φ matrix is equal to the value of Euclidean distance × the number of effective coefficients. When the temporal orders are identical for all clusters, the total distance between any pair of clusters remains equal. While, in cases where the temporal orders vary (e.g. some individuals' ILD follows the VAR (1) process, while the rest of individuals' ILD follows the VAR(2) process), the distance between any pair of clusters Φ matrices in lag 1 is maintained as equal.
4. Given the dataset X {a, b, c, d, e, f} contains two true underlying clusters: C1 {a, b, c} and C2 {e, d, f}. The clustering analysis resulted in two partitions: {a, e} and {b, c, d, f}. However, it is unclear which sets in these partitions should be labeled as C1, leading to

varying misclassification rates. The arbitrary assignment of cluster labels results in an arbitrary misclassification rate. The manipulation of cluster label assignments can yield either a more favorable or less favorable classification rate. Additionally, partitions can only be compared if they have the same number of clusters.

## Reference


Aghabozorgi, S., Shirkhorshidi, A. S., & Wah, T. Y. (2015). Time-series clustering–a decade review. *Information Systems*, *53*, 16–38.

Aghabozorgi, S., & Teh, Y. W. (2014). Stock market co-movement assessment using a three-phase clustering method. *Expert Systems with Applications*, *41*(4), 1301–1314.

Almeida, D. M., Wethington, E., & McDonald, D. A. (2001). Daily variation in paternal engagement and negative mood: Implications for emotionally supportive and conflictual interactions. *Journal of Marriage and Family*, *63*(2), 417–429.

Arabie, P., & Hubert, L. J. (1992). Combinatorial data analysis. *Annual Review of Psychology*, *43*(1), 169–203.

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 359–388.

Assent, I. (2012). Clustering high dimensional data. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *2*(4), 340–350.

Axelson, D. A., Bertocci, M. A., Lewin, D. S., Trubnick, L. S., Birmaher, B., Williamson, D. E., Ryan, N. D., & Dahl, R. E. (2003). Measuring mood and complex behavior in natural environments: Use of ecological momentary assessment in pediatric affective disorders. *Journal of Child and Adolescent Psychopharmacology*, *13*(3), 253–266.

Babbin, S. F., Velicer, W. F., Aloia, M. S., & Kushida, C. A. (2015). Identifying longitudinal patterns for individuals and subgroups: An example with adherence to treatment for obstructive sleep apnea. *Multivariate Behavioral Research*, *50*(1), 91–108.

Babyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine*, *66*(3), 411–421.

Banfield, J. D., & Raftery, A. E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics*, 803–821.

Barrett, L. F., & Russell, J. A. (1999). The structure of current affect: Controversies and emerging consensus. *Current Directions in Psychological Science*, *8*(1), 10–14.

Basu, S., & Michailidis, G. (2015). *Regularized estimation in sparse high-dimensional time series models*.

Beck, N., & Katz, J. N. (2011). Modeling dynamics in time-series–cross-section political economy data. *Annual Review of Political Science*, *14*, 331–352.

Bell, B. A., Morgan, G. B., Schoeneberger, J. A., Loudermilk, B. L., Kromrey, J. D., & Ferron, J. M. (2010). Dancing the sample size limbo with mixed models: How low can you go. *SAS Global Forum*, *4*, 11–14.

Beltrame, T., Amelard, R., Wong, A., & Hughson, R. L. (2017). Prediction of oxygen uptake dynamics by machine learning analysis of wearable sensors during activities of daily living. *Scientific Reports*, *7*(1), 1–8.

Berndt, D. J., & Clifford, J. (1994). Using dynamic time warping to find patterns in time series. *KDD Workshop*, *10*(16), 359–370.

Beyer, K., Goldstein, J., Ramakrishnan, R., & Shaft, U. (1999). When is "nearest neighbor" meaningful? *Database Theory—ICDT'99: 7th International Conference Jerusalem, Israel, January 10–12, 1999 Proceedings 7*, 217–235.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, *54*(1), 579–616.

Bolger, N., DeLongis, A., Kessler, R. C., & Wethington, E. (1989). The contagion of stress across multiple roles. *Journal of Marriage and the Family*, 175–183.

Bosley, H. G., Sandel, D. B., & Fisher, A. J. (2020). Idiographic dynamics of positive affect in GAD. *European Journal of Psychological Assessment*.

Bouveyron, C., & Brunet-Saumard, C. (2014). Model-based clustering of high-dimensional data: A review. *Computational Statistics & Data Analysis*, *71*, 52–78.

Braun, P. A., & Mittnik, S. (1993). Misspecifications in vector autoregressions and their effects on impulse responses and variance decompositions. *Journal of Econometrics*, *59*(3), 319–341.

Bringmann, L. F., Ferrer, E., Hamaker, E. L., Borsboom, D., & Tuerlinckx, F. (2018). Modeling nonstationary emotion dynamics in dyads using a time-varying vector-autoregressive model. *Multivariate Behavioral Research*, *53*(3), 293–314.

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., Borsboom, D., & Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PloS One*, *8*(4), e60188.

Brodinová, Š., Filzmoser, P., Ortner, T., Breiteneder, C., & Rohm, M. (2019). Robust and sparse k-means clustering for high-dimensional data. *Advances in Data Analysis and Classification*, *13*, 905–932.

Browne, R. P., & McNicholas, P. D. (2014). Estimating common principal components in high dimensions. *Advances in Data Analysis and Classification*, *8*, 217–226.

Bulteel, K., Mestdagh, M., Tuerlinckx, F., & Ceulemans, E. (2018). VAR (1) based models do not always outpredict AR (1) models in typical psychological applications. *Psychological Methods*, *23*(4), 740.

Bulteel, K., Tuerlinckx, F., Brose, A., & Ceulemans, E. (2016). Clustering vector autoregressive models: Capturing qualitative differences in within-person dynamics. *Frontiers in Psychology*, *7*, 1540.

Busk, P. L., & Marascuilo, L. A. (1988). Autocorrelation in single-subject research: A counterargument to the myth of no autocorrelation. *Behavioral Assessment*.

Carstensen, L. L., Pasupathi, M., Mayr, U., & Nesselroade, J. R. (2000). Emotional experience in everyday life across the adult life span. *Journal of Personality and Social Psychology*, *79*(4), 644.

Celeux, G., & Govaert, G. (1995). Gaussian parsimonious clustering models. *Pattern Recognition*, *28*(5), 781–793.

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2015). Determining the best number of clusters in a data set. *Recuperado de Https://Cran. Rproject. Org/Web/Packages/NbClust/NbClust. Pdf*.

Chow, S.-M., Ferrer, E., & Nesselroade, J. R. (2007). An unscented Kalman filter approach to the estimation of nonlinear dynamical systems models. *Multivariate Behavioral Research*, *42*(2), 283–321.

Chow, S.-M., Hamagani, F., & Nesselroade, J. R. (2007). Age differences in dynamical emotion-cognition linkages. *Psychology and Aging*, *22*(4), 765.

Chun, C. A. (2016). The expression of posttraumatic stress symptoms in daily life: A review of experience sampling methodology and daily diary studies. *Journal of Psychopathology and Behavioral Assessment*, *38*, 406–420.

Coke, G., & Tsao, M. (2010). Random effects mixture models for clustering electrical load series. *Journal of Time Series Analysis*, *31*(6), 451–464.

Commandeur, J. J., & Koopman, S. J. (2007). *An introduction to state space time series analysis*. Oxford university press.
Coolen, A., Barrett, J., Paga, P., & Perez-Vicente, C. (2017). Replica analysis of overfitting in regression models for time-to-event data. *Journal of Physics A: Mathematical and Theoretical*, *50*(37), 375001.
Cuturi, M., & Blondel, M. (2017). Soft-dtw: A differentiable loss function for time-series. *International Conference on Machine Learning*, 894–903.
daSilva, A. W., Huckins, J. F., Wang, W., Wang, R., Campbell, A. T., & Meyer, M. L. (2021). Daily perceived stress predicts less next day social interaction: Evidence from a naturalistic mobile sensing study. *Emotion*, *21*(8), 1760.
de Haan-Rietdijk, S., Voelkle, M. C., Keijsers, L., & Hamaker, E. L. (2017). Discrete-vs. Continuous-time modeling of unequally spaced experience sampling method data. *Frontiers in Psychology*, *8*, 1849.
Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1–22.
D'Urso, P., & Maharaj, E. A. (2009). Autocorrelation-based fuzzy clustering of time series. *Fuzzy Sets and Systems*, *160*(24), 3565–3589.
Elangasinghe, M., Singhal, N., Dirks, K., Salmond, J., & Samarasinghe, S. (2014a). Complex time series analysis of PM10 and PM2. 5 for a coastal site using artificial neural network modelling and k-means clustering. *Atmospheric Environment*, *94*, 106–116.
Elangasinghe, M., Singhal, N., Dirks, K., Salmond, J., & Samarasinghe, S. (2014b). Complex time series analysis of PM10 and PM2. 5 for a coastal site using artificial neural network modelling and k-means clustering. *Atmospheric Environment*, *94*, 106–116.
Ernst, A. F., Timmerman, M. E., Jeronimus, B. F., & Albers, C. J. (2021). Insight into individual differences in emotion dynamics with clustering. *Assessment*, *28*(4), 1186–1206.
Ernst, A. F., Albers, C. J., Jeronimus, B. F., & Timmerman, M. E. (2020). Inter-individual differences in multivariate time-series. *European Journal of Psychological Assessment*. 36, 482-491. https://doi.org/10.1027/1015-5759/a000578.
Fairbairn, C. E., & Sayette, M. A. (2013). The effect of alcohol on emotional inertia: A test of alcohol myopia. *Journal of Abnormal Psychology*, *122*(3), 770.
Ferreira, L. N., & Zhao, L. (2016). Time series clustering via community detection in networks. *Information Sciences*, *326*, 227–242.
Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., & Constable, R. T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, *18*(11), 1664–1671.
Fordyce, M. W. (1988). A review of research on the happiness measures: A sixty second index of happiness and mental health. *Social Indicators Research*, *20*, 355–381.
Fraley, C., & Raftery, A. E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal*, *41*(8), 578–588.
Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, *97*(458), 611–631.
Fulcher, B. D. (2017). Feature-based time-series analysis. *ArXiv Preprint ArXiv:1709.08055*.
Fuller, J. A., Stanton, J. M., Fisher, G. G., Spitzmüller, C., Russell, S. S., & Smith, P. C. (2003). A lengthy look at the daily grind: Time series analysis of events, mood, stress, and

satisfaction. *Journal of Applied Psychology*, *88*(6), 1019.
Galeano, P., & Peña, D. (2001). *Multivariate analysis in vector time series*.
Gates, K. M., Lane, S. T., Varangis, E., Giovanello, K., & Guiskewicz, K. (2017). Unsupervised classification during time-series model building. *Multivariate Behavioral Research*, *52*(2), 129–148.
Gates, K. M., Molenaar, P. C., Hillary, F. G., Ram, N., & Rovine, M. J. (2010). Automatic search for fMRI connectivity mapping: An alternative to Granger causality testing using formal equivalences among SEM path modeling, VAR, and unified SEM. *NeuroImage*, *50*(3), 1118–1125.
Gu, F., Preacher, K. J., & Ferrer, E. (2014). A state space modeling approach to mediation analysis. *Journal of Educational and Behavioral Statistics*, *39*(2), 117–143.
Gutiérrez, J. L. G., Jiménez, B. M., Hernández, E. G., Pcn, C., & others. (2005). Personality and subjective well-being: Big five correlates and demographic variables. *Personality and Individual Differences*, *38*(7), 1561–1569.
Hamaker, E. L., Asparouhov, T., Brose, A., Schmiedek, F., & Muthén, B. (2018). At the frontiers of modeling intensive longitudinal data: Dynamic structural equation models for the affective measurements from the COGITO study. *Multivariate Behavioral Research*, *53*(6), 820–841.
Hartmann, D. P., Gottman, J. M., Jones, R. R., Gardner, W., Kazdin, A. E., & Vaught, R. S. (1980). Interrupted time-series analysis and its application to behavioral data. *Journal of Applied Behavior Analysis*, *13*(4), 543–559.
Hamilton, J. D. (1994). Time Series Analysis. Princeton, NJ: Princeton University Press
Hay, E. L., & Diehl, M. (2011). Emotion complexity and emotion regulation across adulthood. *European Journal of Ageing*, *8*, 157–168.
Hays, W. L. (1981). Statistics, 2nd Edn. New York, NY: Holt, Rinehart, and Winston.
Hoekstra, H., & De Fruyt, F. (2014). *Dutch manual of the NEO-PI-3 en NEO-FFI-3 questionnaires*. Amsterdam, Netherlands: Hogrefe Uitgevers.
Honaker, J., & King, G. (2010). What to do about missing values in time-series cross-section data. *American Journal of Political Science*, *54*(2), 561–581.
Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, *45*, 1–47.
Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, *141*(4), 901.
Hsu, N.-J., Hung, H.-L., & Chang, Y.-M. (2008). Subset selection for vector autoregressive processes using lasso. *Computational Statistics & Data Analysis*, *52*(7), 3645–3657.
Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*, 193–218.
Ivanov, V., & Kilian, L. (2005). A practitioner's guide to lag order selection for VAR impulse response analysis. *Studies in Nonlinear Dynamics & Econometrics*, *9*(1).
Jacques, J., & Preda, C. (2014). Functional data clustering: A survey. *Advances in Data Analysis and Classification*, *8*, 231–255.
Jahng, S., Wood, P. K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods*, *13*(4), 354.
Ji, L., Chow, S.-M., Schermerhorn, A. C., Jacobson, N. C., & Cummings, E. M. (2018). Handling missing data in the modeling of intensive longitudinal data. *Structural Equation*

*Modeling: A Multidisciplinary Journal*, *25*(5), 715–736.
Kalpakis, K., Gada, D., & Puttagunta, V. (2001a). Distance measures for effective clustering of ARIMA time-series. *Proceedings 2001 IEEE International Conference on Data Mining*, 273–280.
Kalpakis, K., Gada, D., & Puttagunta, V. (2001b). Distance measures for effective clustering of ARIMA time-series. *Proceedings 2001 IEEE International Conference on Data Mining*, 273–280.
Karalunas, S. L., & Nigg, J. T. (2020). Heterogeneity and subtyping in attention-deficit/hyperactivity disorder—Considerations for emerging research using person-centered computational approaches. *Biological Psychiatry*, *88*(1), 103–110.
Kaya, H., & Gündüz-Öğüdücü, Ş. (2015). A distance based time series classification framework. *Information Systems*, *51*, 27–42.
Kilian, L. (2001). Impulse response analysis in vector autoregressions with unknown lag order. *Journal of Forecasting*, *20*(3), 161–179.
Kim, J., Zhu, W., Chang, L., Bentler, P. M., & Ernst, T. (2007). Unified structural equation modeling approach for the analysis of multisubject, multivariate functional MRI data. *Human Brain Mapping*, *28*(2), 85–93.
Knott, M., & Bartholomew, D. J. (1999). *Latent variable models and factor analysis* (Vol. 7). Edward Arnold.
Krieke, L. V. D., Jeronimus, B. F., Blaauw, F. J., Wanders, R. B., Emerencia, A. C., Schenk, H. M., Vos, S. D., Snippe, E., Wichers, M., Wigman, J. T., & others. (2016). HowNutsAreTheDutch (HoeGekIsNL): A crowdsourcing study of mental symptoms and strengths. *International Journal of Methods in Psychiatric Research*, *25*(2), 123–144.
Kuppens, P., Sheeber, L. B., Yap, M. B., Whittle, S., Simmons, J. G., & Allen, N. B. (2012). Emotional inertia prospectively predicts the onset of depressive disorder in adolescence. *Emotion*, *12*(2), 283.
Lane, S. T., Gates, K. M., Pike, H. K., Beltz, A. M., & Wright, A. G. (2019). Uncovering general, shared, and unique temporal patterns in ambulatory assessment data. *Psychological Methods*, *24*(1), 54.
Lawford, S., & Stamatogiannis, M. P. (2009). The finite-sample effects of VAR dimensions on OLS bias, OLS variance, and minimum MSE estimators. *Journal of Econometrics*, *148*(2), 124–130.
Levina, E., Rothman, A., & Zhu, J. (2008). *Sparse estimation of large covariance matrices via a nested lasso penalty*.
Liao, T. W. (2005). Clustering of time series data—A survey. *Pattern Recognition*, *38*(11), 1857–1874.
Liu, S., Rovine, M. J., Cousino Klein, L., & Almeida, D. M. (2013). Synchrony of diurnal cortisol pattern in couples. *Journal of Family Psychology*, *27*(4), 579.
Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, *33*(3), 335–343.
Lütkepohl,H, New introduction to multiple time series analysis, Springer, Berlin, Heidelberg, 2015.
Ma, Q., Li, S., Zhuang, W., Wang, J., & Zeng, D. (2020). Self-supervised time series clustering with model-based dynamics. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(9), 3942–3955.

Maas, C. J., & Hox, J. J. (2004). The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational Statistics & Data Analysis*, *46*(3), 427–440.
MacQueen, J. (1967). Classification and analysis of multivariate observations. *5th Berkeley Symp. Math. Statist. Probability*, 281–297.
Maharaj, E. A. (2000). Cluster of Time Series. *Journal of Classification*, *17*(2).
Masten, A. S., & Cicchetti, D. (2010). Developmental cascades. *Development and Psychopathology*, *22*(3), 491–495.
Maugis, C., Celeux, G., & Martin-Magniette, M.-L. (2009). Variable selection for clustering with Gaussian mixture models. *Biometrics*, *65*(3), 701–709.
McLachlan, G. J. (2011). Commentary on Steinley and Brusco (2011): Recommendations and cautions. *Psychological Methods*, *16*(1), 80–81. https://doi.org/10.1037/a0021141
McLachlan, G. J., & Basford, K. E. (1988). *Mixture models: Inference and applications to clustering* (Vol. 38). M. Dekker New York.
McLachlan, G. J., & Peel, D. (2000). Mixtures of factor analyzers. *Proceedings of the Seventeenth International Conference on Machine Learning*, 599–606.
McNeish, D., & Hamaker, E. L. (2020). A primer on two-level dynamic structural equation models for intensive longitudinal data in Mplus. *Psychological Methods*, *25*(5), 610.
McNeish, D. M. (2015). Using lasso for predictor selection and to assuage overfitting: A method long overlooked in behavioral sciences. *Multivariate Behavioral Research*, *50*(5), 471–484.
Milligan, G. W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, *45*, 325–342.
Milligan, G. W., & Cooper, M. C. (1986). A study of the comparability of external criteria for hierarchical cluster analysis. *Multivariate Behavioral Research*, *21*(4), 441–458.
Milligan, G. W., & Sokol, L. M. (1980). A two-stage clustering algorithm with robust recovery characteristics. *Educational and Psychological Measurement*, *40*(3), 755–759.
Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, *2*(4), 201–218.
Nakamura, T., Tanizawa, T., & Small, M. (2016). Constructing networks from a dynamical system perspective for multivariate nonlinear time series. *Physical Review E*, *93*(3), 032323.
Nicholson, W. B., Wilms, I., Bien, J., & Matteson, D. S. (2020). High dimensional forecasting via interpretable vector autoregression. *The Journal of Machine Learning Research*, *21*(1), 6690–6741.
Paparrizos, J., & Gravano, L. (2015). k-shape: Efficient and accurate clustering of time series. *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, 1855–1870.
Pavlenko, T. (2003). On feature selection, curse-of-dimensionality and error probability in discriminant analysis. *Journal of Statistical Planning and Inference*, *115*(2), 565–584.
Pavlenko, T., & Von Rosen, D. (2001). Effect of dimensionality on discrimination. *Statistics*, *35*(3), 191–213.
Pe, M. L., & Kuppens, P. (2012). The dynamic interplay between emotions in daily life: Augmentation, blunting, and the role of appraisal overlap. *Emotion*, *12*(6), 1320.
Piccolo, D. (1990). A distance measure for classifying ARIMA models. *Journal of Time Series Analysis*, *11*(2), 153–164.

Popivanov, I., & Miller, R. J. (2002). Similarity search over time-series data using wavelets. *Proceedings 18th International Conference on Data Engineering*, 212–221.
Price, R. B., Gates, K., Kraynak, T. E., Thase, M. E., & Siegle, G. J. (2017). Data-driven subgroups in depression derived from directed functional connectivity paths at rest. *Neuropsychopharmacology*, *42*(13), 2623–2632.
Priebe, S., Reininghaus, U., McCabe, R., Burns, T., Eklund, M., Hansson, L., Junghan, U., Kallert, T., van Nieuwenhuizen, C., Ruggeri, M., & others. (2010). Factors influencing subjective quality of life in patients with schizophrenia and other mental disorders: A pooled analysis. *Schizophrenia Research*, *121*(1–3), 251–258.
Raftery, A. E., & Dean, N. (2006). Variable selection for model-based clustering. *Journal of the American Statistical Association*, *101*(473), 168–178.
Ren, B., & Barnett, I. (2022). Autoregressive mixture models for clustering time series. *Journal of Time Series Analysis*, *43*(6), 918–937.
Rosmalen, J. G., Wenting, A. M., Roest, A. M., de Jonge, P., & Bos, E. H. (2012). Revealing causal heterogeneity using time series analysis of ambulatory assessments: Application to the association between depression and physical activity after myocardial infarction. *Psychosomatic Medicine*, *74*(4), 377–386.
Rush, J., & Hofer, S. M. (2014). Differences in within-and between-person factor structure of positive and negative affect: Analysis of two intensive measurement studies using multilevel structural equation modeling. *Psychological Assessment*, *26*(2), 462.
Schenk, H. M., Bos, E. H., Slaets, J. P., de Jonge, P., & Rosmalen, J. G. (2017). Differential association between affect and somatic symptoms at the between-and within-individual level. *British Journal of Health Psychology*, *22*(2), 270–280.
Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 461–464.
Schwarz, N., & Sudman, S. (2012). *Autobiographical memory and the validity of retrospective reports*. Springer Science & Business Media.
Scott, D. W., & Thompson, J. R. (1983). Probability density estimation in higher dimensions. *Computer Science and Statistics: Proceedings of the Fifteenth Symposium on the Interface*, *528*, 173–179.
Scott, S. B., Sliwinski, M. J., & Blanchard-Fields, F. (2013). Age differences in emotional responses to daily stress: The role of timing, severity, and global perceived stress. *Psychology and Aging*, *28*(4), 1076.
Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, *8*(1), 289.
Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol.*, *4*, 1–32.
Snippe, E., Nyklíček, I., Schroevers, M. J., & Bos, E. H. (2015). The temporal order of change in daily mindfulness and affect during mindfulness-based stress reduction. *Journal of Counseling Psychology*, *62*(2), 106.
Song, H., & Ferrer, E. (2012). Bayesian estimation of random coefficient dynamic factor models. *Multivariate Behavioral Research*, *47*(1), 26–60.
Song, H., & Zhang, Z. (2014). Analyzing multiple multivariate time series data using multilevel dynamic factor models. *Multivariate Behavioral Research*, *49*(1), 67–77.
Song, S., & Bickel, P. J. (2011). Large vector auto regressions. *ArXiv Preprint ArXiv:1106.3915*.
Soto, C. J. (2015). Is happiness good for your personality? Concurrent and prospective relations of the big five with subjective well-being. *Journal of Personality*, *83*(1), 45–55.

Stefanovic, M., Rosenkranz, T., Ehring, T., Watkins, E. R., & Takano, K. (2022). Is a high association between repetitive negative thinking and negative affect predictive of depressive symptoms? A clustering approach for experience-sampling data. *Clinical Psychological Science*, *10*(1), 74–89.

Steinbach, M., Ertöz, L., & Kumar, V. (2004a). The challenges of clustering high dimensional data. *New Directions in Statistical Physics: Econophysics, Bioinformatics, and Pattern Recognition*, 273–309.

Steinbach, M., Ertöz, L., & Kumar, V. (2004b). The challenges of clustering high dimensional data. *New Directions in Statistical Physics: Econophysics, Bioinformatics, and Pattern Recognition*, 273–309.

Steinley, D. (2004). Properties of the hubert-arable adjusted rand index. *Psychological Methods*, *9*(3), 386.

Steinley, D., & Brusco, M. J. (2011). Evaluating mixture modeling for clustering: Recommendations and cautions. *Psychological Methods*, *16*(1), 63.

Stetco, A., Zeng, X., & Keane, J. (2013). Fuzzy cluster analysis of financial time series and their volatility assessment. *2013 IEEE International Conference on Systems, Man, and Cybernetics*, 91–96.

Takano, K., Stefanovic, M., Rosenkranz, T., & Ehring, T. (2021). Clustering individuals on limited features of a vector autoregressive model. *Multivariate Behavioral Research*, *56*(5), 768–786.

Tran, D., & Wagner, M. (2002). Fuzzy c-means clustering-based speaker verification. *Advances in Soft Computing—AFSS 2002: 2002 AFSS International Conference on Fuzzy Systems Calcutta, India, February 3–6, 2002 Proceedings*, 318–324.

Verduyn, P., & Lavrijsen, S. (2015). Which emotions last longest and why: The role of event importance and rumination. *Motivation and Emotion*, *39*(1), 119–127.

Vermunt, J. K. (2011). K-means may perform as well as mixture model clustering but may also be much worse: Comment on Steinley and Brusco (2011). *Psychological Methods*, *16*(1), 82–88. https://doi.org/10.1037/a0020144

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

Volkmar, F. R., Lord, C., Bailey, A., Schultz, R. T., & Klin, A. (2004). Autism and pervasive developmental disorders. *Journal of Child Psychology and Psychiatry*, *45*(1), 135–170.

Von Leupoldt, A., Riedel, F., & Dahme, B. (2006). The impact of emotions on the perception of dyspnea in pediatric asthma. *Psychophysiology*, *43*(6), 641–644.

Wang, X., Mueen, A., Ding, H., Trajcevski, G., Scheuermann, P., & Keogh, E. (2013). Experimental comparison of representation methods and distance measures for time series data. *Data Mining and Knowledge Discovery*, *26*(2), 275–309. https://doi.org/10.1007/s10618-012-0250-5

Wang, X., Smith, K. A., Hyndman, R., & Alahakoon, D. (2004). A scalable method for time series clustering. *Unrefereed Research Papers*, *1*.

Wang, X., Smith, K., & Hyndman, R. (2006). Characteristic-based clustering for time series data. *Data Mining and Knowledge Discovery*, *13*, 335–364.

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*(301), 236–244.

Whalen, P. J., Johnstone, T., Somerville, L. H., Nitschke, J. B., Polis, S., Alexander, A. L., Davidson, R. J., & Kalin, N. H. (2008). A functional magnetic resonance imaging predictor of treatment response to venlafaxine in generalized anxiety disorder. *Biological Psychiatry*, *63*(9), 858–863.

Wood, P., & Brown, D. (1994). The study of intraindividual differences by means of dynamic factor models: Rationale, implementation, and interpretation. *Psychological Bulletin*, *116*(1), 166.

Wright, A. G. C., & Simms, L. J. (2016). Stability and Fluctuation of Personality Disorder Features in Daily Life. *Journal of Abnormal Psychology*, *125*(5), 641–656. https://doi.org/10.1037/abn0000169

Wright, A. G. C., & Zimmermann, J. (2019). Applied ambulatory assessment: Integrating idiographic and nomothetic principles of measurement. *Psychological Assessment*, *31*(12), 1467–1480. https://doi.org/10.1037/pas0000685

Wright, A. G., & Simms, L. J. (2016). Stability and fluctuation of personality disorder features in daily life. *Journal of Abnormal Psychology*, *125*(5), 641.

Xiong, Y., & Yeung, D.-Y. (2002). Mixtures of ARMA models for model-based time series clustering. *2002 IEEE International Conference on Data Mining, 2002. Proceedings.*, 717–720.

Xiong, Y., & Yeung, D.-Y. (2004). Time series clustering with ARMA mixtures. *Pattern Recognition*, *37*(8), 1675–1689.

Yahyaoui, H., & Al-Mutairi, A. (2016). A feature-based trust sequence classification algorithm. *Information Sciences*, *328*, 455–484. https://doi.org/10.1016/j.ins.2015.08.008

Yan, W., Zhao, M., Fu, Z., Pearlson, G. D., Sui, J., & Calhoun, V. D. (2022). Mapping relationships among schizophrenia, bipolar and schizoaffective disorders: A deep classification and clustering framework using fMRI time series. *Schizophrenia Research*, *245*, 141–150.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122.

Yik, M. S., Russell, J. A., & Barrett, L. F. (1999). Structure of self-reported current affect: Integration and beyond. *Journal of Personality and Social Psychology*, *77*(3), 600.

Zhang, Q., Wu, J., Zhang, P., Long, G., & Zhang, C. (2018). Salient subsequence learning for time series clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(9), 2193–2207.

Zhang, X., Liu, J., Du, Y., & Lv, T. (2011). A novel clustering method on time series data. *Expert Systems with Applications*, *38*(9), 11891–11900. https://doi.org/10.1016/j.eswa.2011.03.081

Zheng, Y., Wiebe, R. P., Cleveland, H. H., Molenaar, P. C., & Harris, K. S. (2013). An idiographic examination of day-to-day patterns of substance use craving, negative affect, and tobacco use among young adults in recovery. *Multivariate Behavioral Research*, *48*(2), 241–266.

Zimmermann, J., Woods, W. C., Ritter, S., Happel, M., Masuhr, O., Jaeger, U., Spitzer, C., & Wright, A. G. C. (2019). Integrating structure and dynamics in personality assessment: First steps toward the development and validation of a personality dynamics diary. *Psychological Assessment*, *31*, 516–531. https://doi.org/10.1037/pas0000625

## Appendix A

The least-square estimation for VAR (1) model.

Given

$$Y = (y_2, y_3, ...., y_t) \quad (1)$$

$$B = (a, \Phi) \quad (2)$$

$$z = (\begin{bmatrix} 1 \\ y_1 \end{bmatrix}, \begin{bmatrix} 1 \\ y_2 \end{bmatrix}, ....., \begin{bmatrix} 1 \\ y_{t-1} \end{bmatrix}) \quad (3)$$

The VAR(1) model can be rewritten as:

$$Y = BZ + \mathrm{E} \quad (4)$$

The dynamics of VAR(1) model can be obtained by solving the following closed-form expression:

$$\hat{B} = YZ^{'}(ZZ^{'})^{-1} \quad (5)$$

where Y represents time-ordered observations from time-point 2 to time-point t for a specific person i. Each $y_j$ corresponds to a V x 1 vector, representing the time-ordered observations at time-point j for V intensive longitudinal variables. V denotes the number of intensive longitudinal variables. B is the matrix of dynamic coefficients, which includes the intercept AR, CR effects and intercepts. E represents innovations. It is worth noting that the estimation step is equivale to separate ordinary least squares (LS) estimation for each equation (Lütkepohl, 2005).

## Appendix B

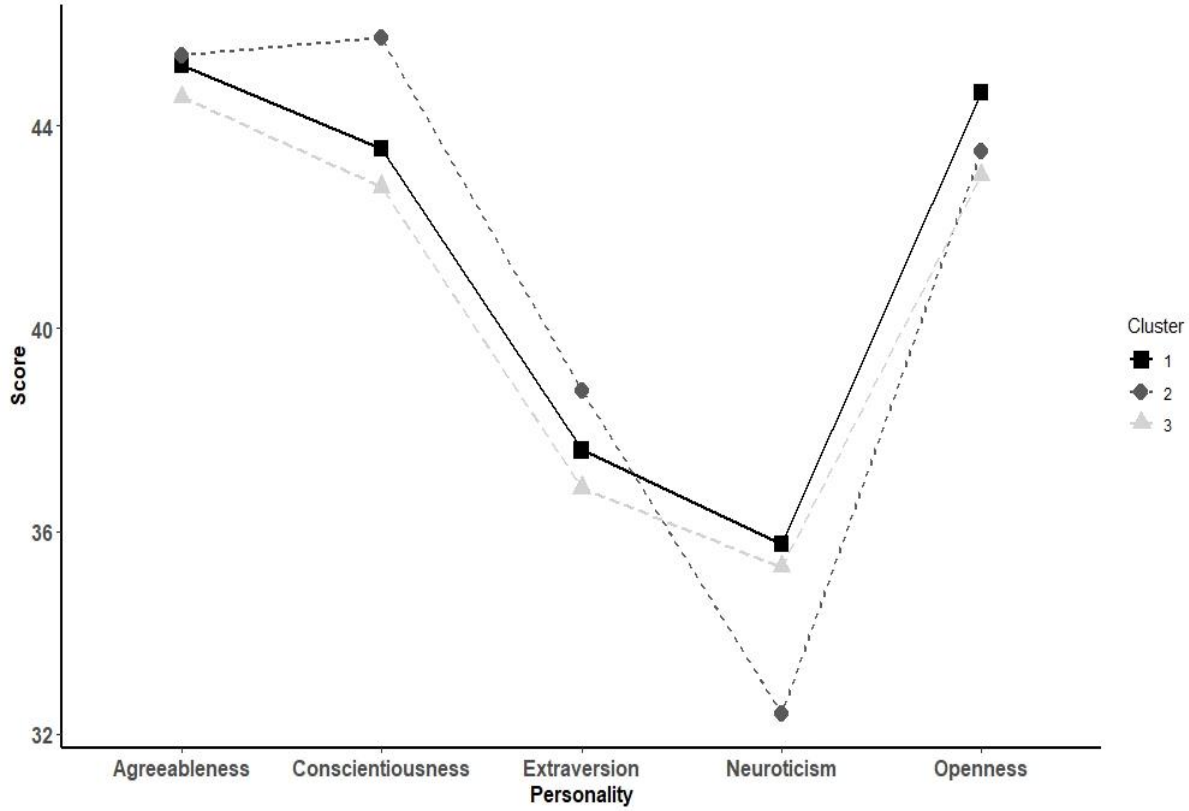



Figure B1. Score on personality measurements of three clusters obtained by HO strategy with GMM

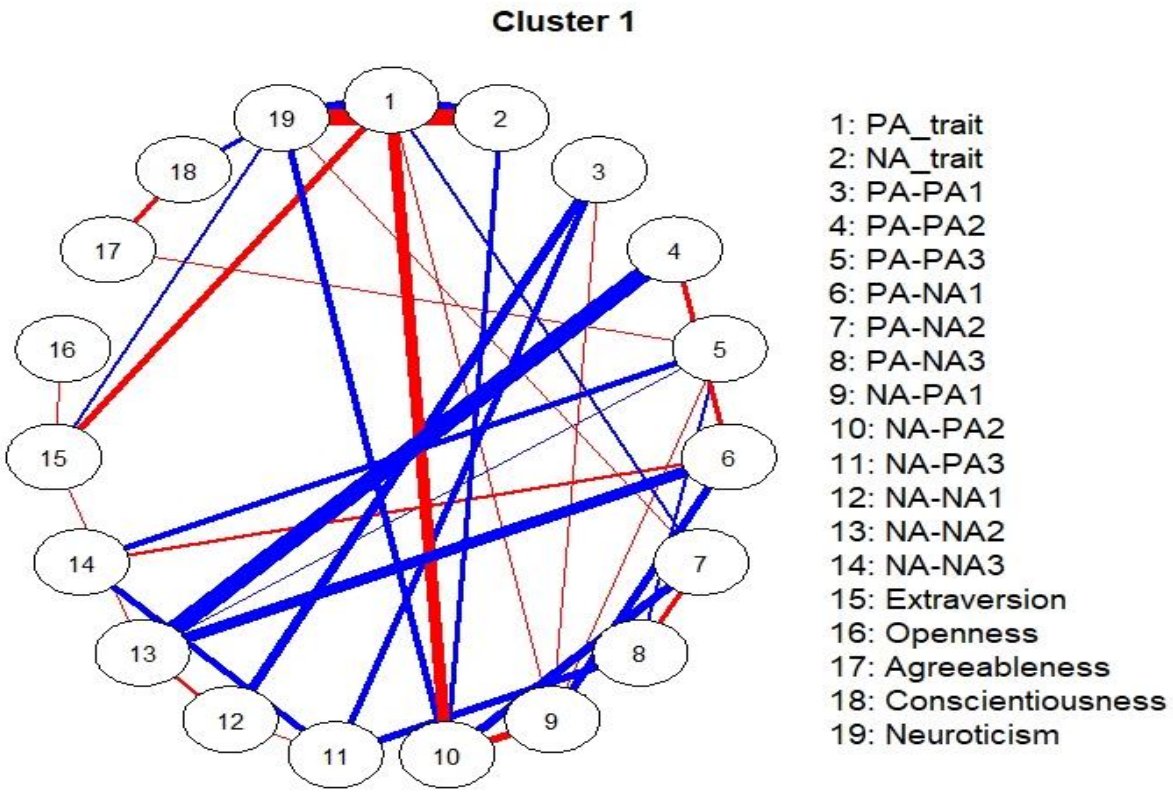



Figure B2a. Network of correlations between trait affects, AR and CR coefficients, and big-five personality traits for Cluster 1 obtained by HO strategy with GMM

Note. The node PA_trait and NA_trait represent the mean of PA and mean of NA, respectively. PA-PA1 and NA-NA1 refer to the AR effects for PA and AR effects for NA at lag-1, respectively. PA-PA2 and NA-NA2 refer to the AR effects for PA and AR effects for NA at lag-2. PA-PA3 and NA-NA3 refer to the AR effects for PA and AR effects for NA at lag-3. PA-NA1 represents the CR effect from PA on NA at lag-1 and NA-PA1 represents the CR effect from NA on PA at lag-1. PA-NA2 represents the CR effect from PA on NA at lag-2 and NA-PA2 represents the CR effect from NA on PA at lag-2. PA-NA3 represents the CR effect from PA on NA at lag-3 and NA-PA3 represents the CR effect from NA on PA at lag-3. The node from 15-19 represents the 5 domains of big-five personality traits.

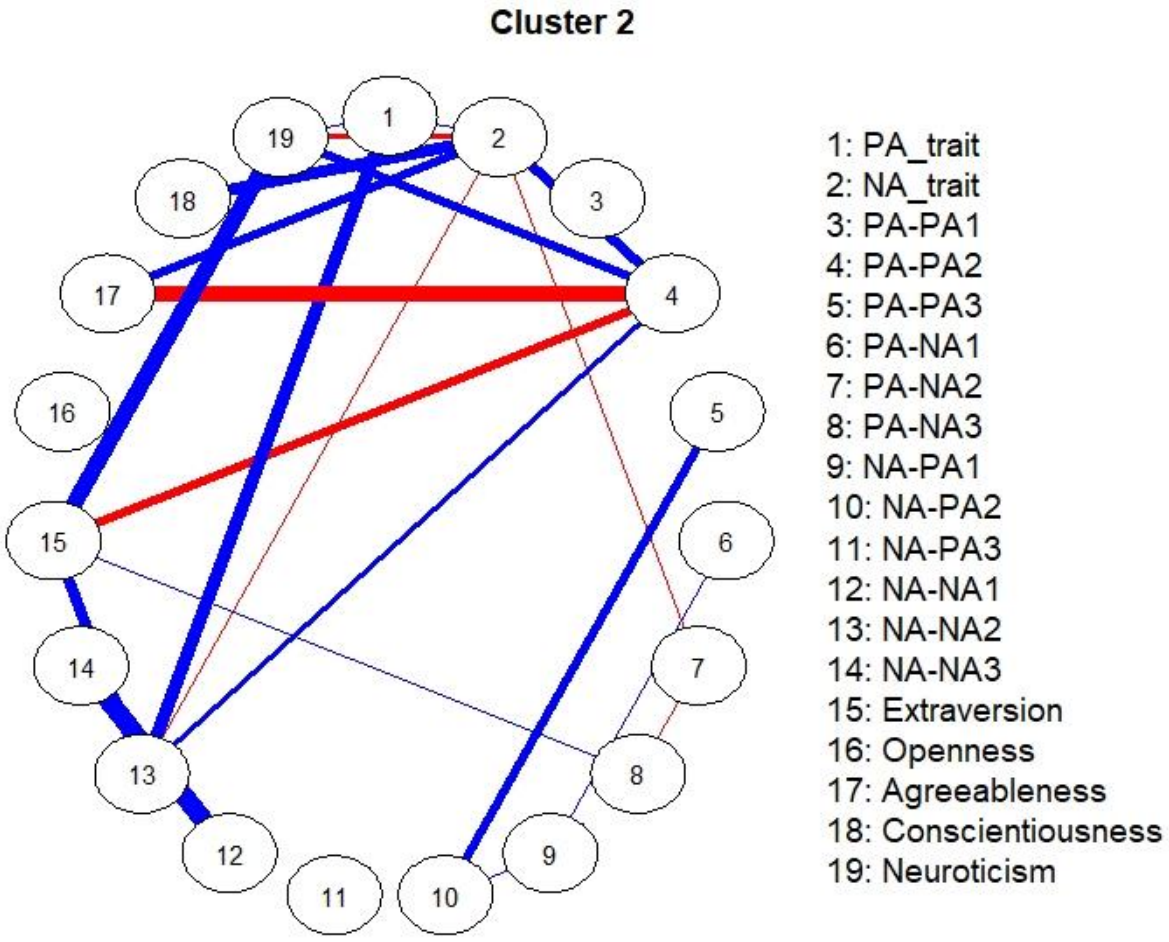



Figure B2b. Network of correlations between trait affects, AR and CR *coefficients,* and big-*five* personality traits for Cluster 2 obtained by HO strategy with GMM.

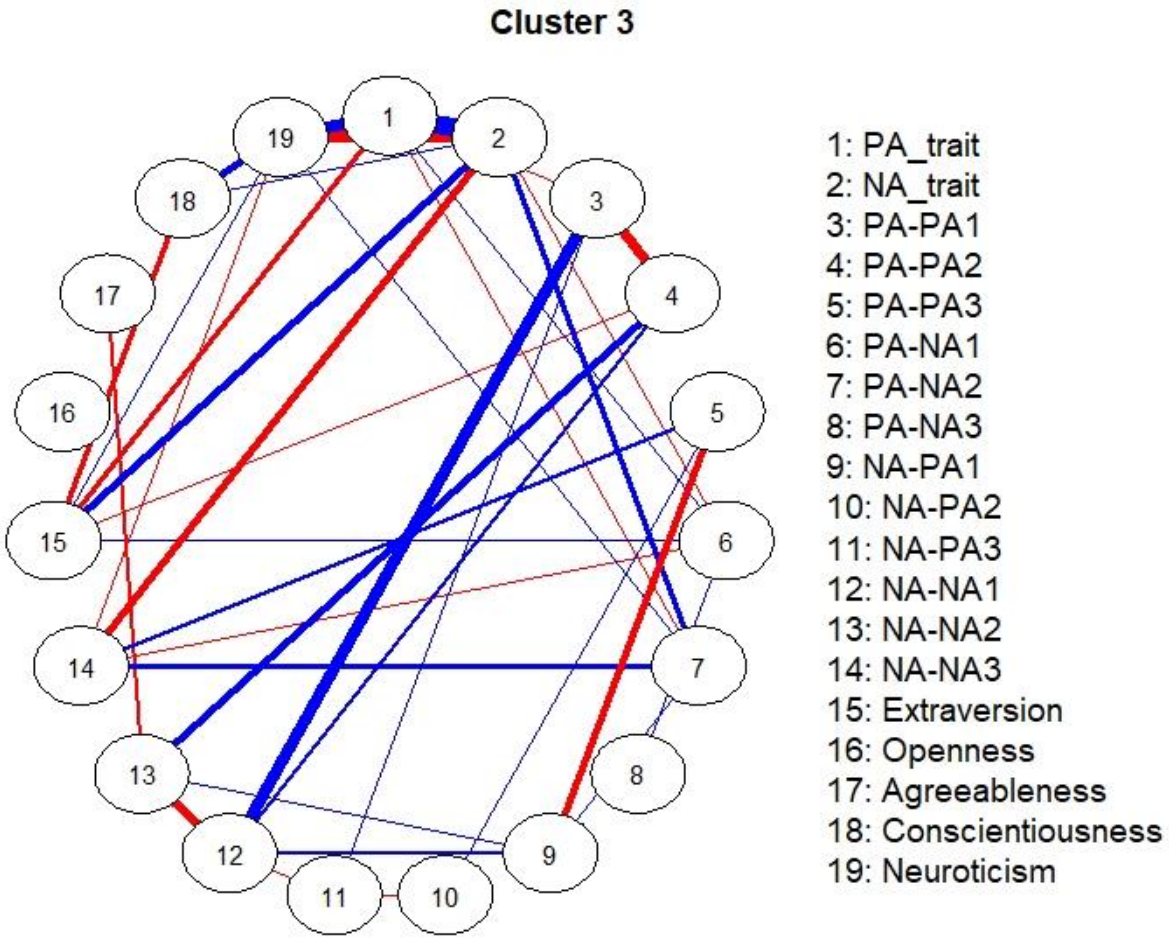



Figure B2c. Network of correlations between trait affects, AR and CR coefficients, and big-five personality traits for Cluster 3 obtained by HO strategy with GMM

Table B1a. Cluster-specific AR and CR effects of PA and NA for the Cluster 1 obtained by HO strategy with GMM

| | **Cluster 1** | | | | | |
|---|---|---|---|---|---|---|
| | $PA_{t-1}$ | $NA_{t-1}$ | $PA_{t-2}$ | $NA_{t-2}$ | $PA_{t-3}$ | $NA_{t-3}$ |
| **$PA_t$** | 0.358 | -0.059 | -0.032 | 0.050 | 0.001 | -0.066 |
| **$NA_t$** | -0.057 | 0.451 | -0.030 | -0.116 | 0.040 | 0.118 |

Table B1b. Cluster-specific AR and CR effects of PA and NA for the Cluster 2 obtained by HO strategy with GMM.

| | **Cluster 2** | | | | | |
|---|---|---|---|---|---|---|
| | $PA_{t-1}$ | $NA_{t-1}$ | $PA_{t-2}$ | $NA_{t-2}$ | $PA_{t-3}$ | $NA_{t-3}$ |
| **$PA_t$** | 0.504 | 0.273 | -0.169 | -0.264 | 0.051 | 0.008 |
| **$NA_t$** | -0.075 | 0.307 | 0.046 | 0.056 | 0.065 | 0.074 |

Table B1c. Cluster-specific AR and CR effects of PA and NA for the cluster 3 obtained by HO strategy with GMM.

| | **Cluster 3** | | | | | |
|---|---|---|---|---|---|---|
| | $PA_{t-1}$ | $NA_{t-1}$ | $PA_{t-2}$ | $NA_{t-2}$ | $PA_{t-3}$ | $NA_{t-3}$ |
| **$PA_t$** | 0.450 | -0.074 | -0.182 | -0.075 | 0.157 | 0.046 |
| **$NA_t$** | -0.091 | 0.457 | 0.065 | 0.013 | -0.045 | 0.019 |

Table B2. Means of the external variables for three clusters of PA and NA obtained by HO strategy with GMM.

| Cluster | Depression (SD) (0-42) | Anxiety (SD) (0-42) | Stress (SD) (0-42) | Happiness (SD) (0-10) |
|---|---|---|---|---|
| **1** | 6.71 | 7.33 | 4.23 | 9.88 |
| **2** | 7.00 | 4.64 | 2.95 | 8.55 |
| **3** | 6.65 | 7.74 | 3.35 | 8.76 |

Table B3. Characteristics for three clusters of PA and NA obtained by HO strategy with GMM.

| **Cluster** | Proportion | Age (SD) | **Fem**ale% | **PA(SD)** | **NA(SD)** | **Correlation of PA and NA** |
|---|---|---|---|---|---|---|
| 1 | 45.8% (104) | 40.96 | 85% | 54.3(16.1) | 18.5(15.9) | -0.52 ($p$<0.001) |
| 2 | 12.3% (28) | 38.89 | 89% | 57.5(14.7) | 12.9(13.1) | -0.29 ($p$=0.127) |
| 3 | 41.8% (95) | 40.87 | 80% | 54.0(15.6) | 20.5(18.3) | -0.69 ($p$<0.001) |

Note. The ANOVA analysis was conducted to investigate whether significant differences exist among three clusters produced by the GMM algorithm with the HO method. The results showed no significant differences between the three clusters on PA score, $F(2,224)=1.37$, $p=0.256$. Significant differences existed between the three clusters on NA score, $F(2,224)=3.98$, $p=0.02$. The post-hoc test results showed that the NA score of Cluster 3 is significantly higher than the score of Cluster 2($p$=.015).