**Clustering Individuals on Limited Features of a Vector Autoregressive Model**

**Supplementary Materials**

**Study 1**

**Parameter recovery for the individualistic VAR estimation.** Our GMM approach used regression coefficients that were estimated for each participant as input features for clustering. Because the precision of the individualistic estimation affects subsequent GMM clustering, we tested the extent to which the simulated regression coefficients were recovered by this individualistic VAR estimation. We first computed correlations between the simulated and estimated regression coefficients in each simulation run. The results indicated that the mean of the correlations was *r* = 0.55 (*SD* = 0.07) across regression coefficients and simulation runs, which suggest that the estimation precision is typically at the moderate level. We also computed mean RMSEs between the simulated vs. estimated regression coefficients per simulation run. The mean across all simulation runs was 0.12 (*SD* = 0.002), which is not a trivial size of error given that a regression coefficient was sampled from a normal distribution with *SD* of 0.10.

**Dimension reduction for the Φ matrix.** Another factor that influenced GMM clustering could be the noisiness of our feature space, which had a number of features that did not contribute to the clustering. For example, in the simulation condition of 10 variables, there were only 1, 5, and 9 EFs out of 100 regression coefficients in the Φ matrix. All of these coefficients were put in GMM clustering simultaneously, so that over 90 % of the features were random noise for the clustering. We tested whether a further dimension reduction, i.e., factor analysis on the Φ matrix, improves performance of GMM clustering. For the model with 5 variables, this dimension reduction improved the accuracy identifying the correct number of groups (Table S1), which was, however, far from an acceptable level. An interesting finding here is that the accuracy predicting participants’ group membership is influenced by the number of factors that are suggested by the factor analysis (optimal coordinates) on the feature space. If the factor analysis indicated one or two factors (these are the “correct” number of factors given that EFs are moderately correlated with each other), GMM achieved the accuracy of > 70 % when there were prominent group differences (Table S2). Although factor analysis or principal component analysis is, in general, a common technique to reduce the high dimensionality of a feature space, this procedure may be suboptimal (or even inappropriate) when there are features that do not contribute to the clustering. An alternative approach is to select features that are locally optimal with group information (e.g., Scrucca & Raftery, 2014), which we were not able to implement in the current simulations because it requires too high computational resource.

**Reference**

Scrucca, L., & Raftery, A. E. (2014). clustvarsel : A Package Implementing Variable Selection for Model-based Clustering in R. https://arxiv.org/abs/1411.0606

Table S1

The Number of Simulation Runs Where GMM Identified the Correct Number of Groups (5 variables in a Model)

|  |  |  |  |
| --- | --- | --- | --- |
| ES conditions | EF =1 | EF = 5 | EF = 9 |
| Moderate ES | 2 | 0 | 0 |
| Large ES | 0 | 1 | 1 |
| Very large ES | 0 | 1 | 26 |

Note. ES = Effect Size; EF = Effective Features

Table S2

Accuracy Predicting Participants’ Group Membership as a Function of Number of Factors (with 5 Variables)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | N Factors | | | |
|  | 1 | 2 | 3 | > 4 |
| ES = very large, EF =1 | 0.500 | 0.500 | 0.500 | 0.500 |
| ES = very large, EF =5 | 0.500 | 0.519 | 0.500 | 0.500 |
| ES = very large, EF =9 | 0.934 | 0.715 | 0.598 | 0.500 |

Note. ES = Effect Size; EF = Effective Features.

**Study 2b**

Table S3

Specificity (and Sensitivity) Predicting Participants’ Group Membership

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Study / ES conditions | With 5 variables | | | With 10 variables | | |
|  | EF =1 | EF = 5 | EF = 9 | EF =1 | EF = 5 | EF = 9 |
| Moderate ES | .46 (.43) | .47 (.46) | .49 (.49) | .49 (.44) | .49 (.44) | .50 (.44) |
| Large ES | .49 (.44) | .49 (.51) | .60 (.67) | .51 (.46) | .48 (.45) | .52 (.52) |
| Very large ES | .50 (.50) | .80 (.88) | .97 (.93) | .49 (.46) | .59 (.66) | .77 (.88) |

Note. EF = Effective Features; ES = Effect size.