Supplemental Material for
Unsupervised Classification During Time Series Model Selection

Authors: Gates*, Kathleen M., Lane, Stephanie T., Varangis, E., Giovanello, K., & Guiskewicz, K. M.
University of North Carolina, Chapel Hill, NC, USA
*E-mail: gateskm@email.unc.edu, Telephone: 919.962.4947

Contents

Estimation of uSEM within GIMME

In practice, the **A** (containing contemporaneous effects) and Φ (containing lagged effects) matrices above are contained within the **B** matrix and can be reduced to:

$$[\eta_{t-1}||\eta_t] = B[\eta_{t-1}||\eta_t] + \zeta_t \quad (1)$$

Here, $[\eta_{t-1}||\eta_t]$ indicates the horizontally concatenated variables at a previous time point (*t-1*) and at the next time point (t). Considering the **B** matrix in this way, the full SEM can be represented using standard SEM notation:

$$S = \Lambda(I - B)^{-1}\Psi(I - B)^{-1}\Lambda' + \Theta \quad (2)$$

where Ψ is the covariance matrix of the variables at a lag of one and the regression errors (with the diagonals being the variance), Λ the matrix of loadings for the observed variables on the latent variable eta, **S** the observed sample covariance matrix, and Θ the covariance matrix of measurement errors. An important set of constraints is followed to enable estimation. One, the search restricts the candidate paths in the beta matrix such that directed pathways cannot be backwards in time (i.e., η_t cannot predict η_{t-1}). Two, the relations among the η_{t-1} variables are constrained to zero in the **B** matrix and are captured as covariances in Ψ. Three, in the absence of a measurement model (i.e., latent constructs indicated by multiple variables), Λ is a $2p \times 2p$ identity matrix and the variance for measurement errors in Θ is set to zero. In the **B** submatrices Φ and **A**, there are only p variables (see Kim et al., 2007). The following

matrices depict the structures of the candidate paths, or paths that will be considered in the model search, in the **B** matrix and the estimated elements in the Ψ matrix:

$$B = \underbrace{\begin{bmatrix} 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ \Phi_{11} & \Phi_{12} & \Phi_{1p} & 0 & A_{12} & A_{1p} \\ \Phi_{21} & \ddots & A_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & A_{(p-1)p} \\ \Phi_{p1} & \Phi_{p(p-1)} & \Phi_{pp} & A_{p1} & A_{p(p-1)} & 0 \end{bmatrix}}_{\substack{\textit{Lagged} \quad \textit{Contemporaneous}}}$$

$$\Psi = \underbrace{\begin{bmatrix} \Psi_{11} & & & & & \\ \Psi_{21} & \ddots & & & & \\ \vdots & \ddots & & & & \\ \Psi_{p1} & \Psi_{p(p-1)} & & & & \\ 0 & \cdots & 0 & & & \\ \vdots & & \vdots & 0 & & \\ 0 & \cdots & 0 & 0 & \cdots & \Psi_{(2p)(2p)} \end{bmatrix}}_{\substack{\textit{Lagged} \quad \textit{Contemporaneous}}}$$

Since the rows are not independent, the estimation of these models is considered to be quasi-maximum likelihood estimation. This approach is optimal for model searching since it is computationally inexpensive yet yields parameter estimates which have about the same properties as those obtained with full maximum likelihood estimation (cf. Hamaker, Dolan, & Molenaar, 2002). It is well known that autoregressive parameters in a stationary ARMA(p,0) process may be estimated by LISREL analysis of the block-Toeplitz matrix (van Buuren, 1997) and given the equivalence in solutions between LISREL and lavaan, it follows that lavaan also can estimate AR models. Assuming normality, the parameters obtained are moment estimates that have been shown to approximate true ML estimates well (Hamaker, Dolan, & Molenaar, 2002), but may yield slightly greater bias (Chow et al., 2010). In the case of missing data, full information maximum likelihood is conducted as it is found to enable robust estimation (Enders & Bandalos, 2001) and is the default approach used by S-GIMME.

Empirical fMRI Data Example

fMRI Data Acquisition Procedures

Participants were all right-handed native English speakers without a history of psychiatric illness or neurological disorder. Before participating in the study, all persons gave written informed consent in

accordance with the requirements of the Institutional Review Board at the University of North Carolina (UNC) at Chapel Hill. The in-scanner working memory paradigm utilized in this study was an N-back task (Table 1; Kirchner, 1958). The stimuli for this task included both upper- and lower-case letters that appeared on the screen, one at a time, for 3 seconds each. Letters were mixed upper- and lower-case to avoid responses based on purely perceptual features of the stimuli, and to encourage a deeper level of working memory processing associated with task success. Stimuli were presented in six blocks, with 72 letters appearing in each block. Prior to the scanning sessions, participants were familiarized with the task, provided with verbal instructions for the task, and completed practice sessions for each of the three variants of the N-back task used in the scanner. Participants only began the scan session once they could successfully complete all three tasks outside of the scanner with 100% accuracy.

The task was divided into three conditions defined by working memory load: 0-back, 1-back, and 2-back. Each condition was presented over two consecutive blocks within the scanner, and condition order was randomized across participants. Prior to beginning each block, participants were provided with written instructions for the upcoming task to re-familiarize them with response criteria and task instructions. In each task condition, participants were told to respond yes or no with their dominant hand for each letter that appeared on the screen. The 1-back data were used for the present example. During the 1-back condition (moderate working memory load), participants were told to respond yes any time the letter on the screen was the same as the previous letter, and no for any letter that was not the same as the previous letter.

MRIs were acquired using a Siemens Allegra 3-T scanner. Participants' heads were held in place using cushions and a headrest. An initial localizing scan was followed by a high resolution T1-weighted structural scan for anatomical visualization (160 1-mm slices, TR=1750 msec, TE=4.38 msec). Next, functional scans were collected during memory retrieval. Whole brain, gradient-echo, echo planar images (50 interleaved 3-mm slices, TR=3sec, TE=30 msec, Flip angle=90 degrees, 3Œ3Œ3mm) were acquired at an angle parallel to the long axis of the hippocampus, identified during the T1 scan.

fMRI Data Preprocessing

Images were preprocessed and analyzed using Statistical Parametic Mapping (SPM)8 (Wellcome Department of Cognitive Neurology, London, UK) software implemented as a suite of commands in MATLAB (MathWorks, Natick, MA). Images were co-registered with each person's anatomical scan, slice-time corrected, realigned, normalized, and smoothed (convolved to an isotropic Gaussian kernel). Brain regions of interest (ROIs) within the fronto-parietal network (which is often associated with working memory) were identified using coordinates from previous research (Dosenbach, et al., 2011). Specifically, the bilateral ROIs selected were: dorsolateral prefrontal cortext (DLPFC), precuneous, intraparietal sulcus (IPS), inferior parietal lobule (IPL), and the frontal cortex. The midcingulate was also used, for a total of 11 ROIs. Time series were pulled from these ROIs using MarsBar MATLAB toolbox, with 4mm spheres used for the parietal and midcingulate ROIs and 10mm spheres for the frontal ROIs

Walktrap

As a random walk approach, Walktrap was originally created to enable greater computational efficiency for larger matrices without sacrificing the reliability for smaller matrices. Walktrap begins by estimating the transition probability of going from one individual ("node") in a given graph to another individual in a random walk of a given number of steps. For example, in a random walk of one the transition probability

between two individuals would be the probability of one of the individuals going to the other individual given all possible individuals in the graph (e.g., brain regions of interest or variables). This probability should be relatively high if individuals are in the same subgroup and low if individuals are in different subgroups. The sequence that the random walk takes across multiple steps is a Markov chain. Here, we use a walk of length 4, as this has been shown to provide a reasonable balance between being too long (and thus being identical to certain properties already available in the similarity matrix) and too short to provide information about the topology of the matrix under a series of simulated data (Pons & Latapy, 2006). Hence each element $r_{between,ij}$ (or s_{ij}) in a given similarity matrix $\mathbf{r_{between}}$ (or $\mathbf{s}$, depending on the feature selection approach) is recast into a transition matrix $\mathbf{P}^4$, with P^4_{ij} depicting the probability of going from individual i to j in 4 steps. The rationale behind using the probability as opposed to the raw correlation or count matrix (as would be done in traditional clustering) rests heavily on the observations that the two individuals i and j in the same cluster tend to have: (1) high transition probabilities and (2) similar probabilities to other individuals in the cluster (Pons & Latapy, 2006).

The distances between all individuals are then computed using the probability vectors $P^4_{i.}$ and $P^4_{j.}$ for each set of individuals at a walk of 4 is:

$$z_{ij} = \|D^{-1/2}P^4_{i.} - D^{-1/2}P^4_{j.}\| \tag{3}$$

where $\mathbf{D}$ is the average weight in the correlation matrix or count of edges for the count matrices (Rubinov & Sporns, 2010), and $\|.\|$ is the Euclidean norm of R^n. By extension, $P_{C_1 j}$ indicates the probability of going from cluster C_1 (equivalently, "community," "subgroup," or "class") to individual j, and $P_{C_1 C_2}$ the probability of going from cluster C_1 to C_2.

From the resulting distance matrix individuals are merged in an agglomerative manner based on Wards criterion (Ward, 1963) to merge two clusters (with each individual being in its own community to start):

$$\sigma_q = 1/2 \sum_{C \in P_q} \sum_{i \in C_q} z^2_{iC} \tag{4}$$

through a greedy search which minimizes σ_q for each step q, where P_q indicates the partition solution at that step. Further details of the algorithm and properties can be found in Pons & Latapy (2006).

These steps are conducted in an iterative fashion by merging communities at each step that decreases σ_q and selecting the P_q corresponding to the lowest σ_q as the final solution for that step. Modularity is then used to evaluate and choose the best partition (Pons & Latapy, 2006). When the average strength of path weights for individuals within a subgroup is larger than what is expected for individuals who are in subgroups, modularity rises (Newman, 2006; Rubinov & Sporns, 2010). Hence, the researcher does not have to define *a priori* the number of subgroups, nor does the researcher have to subjectively decide where in the dendrogram to make the cut. The present paper evaluates the performance of recovering the underlying subgroup structure using Walktrap on matrices obtained from the three approaches for feature selection described above (i.e., lag-0, lag-1/lag0, and S-GIMME).

Detailed Results from Omitted Variables Analysis

Supplemental Table 4 provides information on recall and precision given the omission of each variable separately. Of course, certain variables will likely have a greater impact on the precision of results if they relate to a large number of other variables. For these variables, an increase in false positive relations may occur given that an indirect effect will be induced for two given variables that are in turn related to a third

variable (but not themselves directly related). In particular, variables that are common causes for two other variables or mediate two variable may evidence a greater rate of false positives. The absence of the mediating or causal variable third variable may erroneously surface as direct effects between these two variables.

As seen in Figure 5 of the main text, variables 1, 3, and 5 serve as mediators between two other variables at the group level. Removing these from analysis consistently decreased precision in the recovery of effects (see Table 1) from the average 97% for this condition (see Supplemental Table 3) for path precision when all variables are included. However, notable differences emerged. The omission of variable 5 still had highly acceptable path precision (96%), while omitting variables 1 and 3 evidence more deleterious impacts (86% and 91% average recall precision, respectively). In terms of common causes, only variable 8 was a common cause of two other variables for all individuals simulated. Omission of this variable resulted in a greater number of false positives (as indicated by a path precision of 90%) than the omission of variable 3 (91%) but fewer than seen for omitting variable 1 (86%). The explanation for these findings seems to be that variables 1, 3, and 8 are the only variables that have 6 paths directly connecting them to other variables at the group and subgroup levels. They also had the lowest path precision. Variable 5 only has 4 paths relating it to other variables, which might explain why omitting this variable did not have deleterious effects even though it is a mediator. Hence it seems that the extent to which an omitted variable results in spurious connections relates more so to the number of variables to which it was connected than the specific nature of that relation (i.e., mediator or common cause).

References

Chow, S.-M., Ho, M. R., Hamaker, E. L., & Dolan, C. V. (2010). Equivalence and differences between structural equation modeling and state-space modeling techniques. *Structural Equation Modeling: A Multidisciplinary Journal, 17*(2), 303332. doi:10.1080/10705511003661553

Enders, C., & Bandalos, D. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 8*(3), 430457. doi:10.1207/S15328007SEM0803_5

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2002). On the nature of SEM estimates of ARMA parameters. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(3), 347368. doi:10.1207/S15328007SEM0903_3

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), 8593. doi:10.1002/hbm.20259

Newman, M. E. J. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics, 74*(3). doi:10.1103/PhysRevE.74.036104

Pons, P. & Latapy, M. (2006). Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*. doi:10.7155/jgaa.00124

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage, 52*(3), 10591069. doi:10.1016/j.neuroimage.2009.10.003

van Buuren, S. (1997). Fitting ARMA time series by structural equation models. *Psychometrika, 62*(2), 215236. doi:10.1007/BF02295276

Supplemental Tables

Supplemental Table 1. Hubert-Arabie Adjusted Rand Indices. ***($ARAI_{HA}$).*** *Note: N indicates sample size; K indicates number of subgroups; Lag0 indicates the lag-zero correlation matrices; Lag1 indicates the block-Toeplitz matrix with lag-0 and lag-1 cross-correlations; averages for each method are followed by standard deviations in parentheses.*

Condition	Equal Subgroup Sizes			Unequal Subgroup Sizes		
	Lag0	Lag1	S−GIMME	Lag0	Lag1	S−GIMME
N=25, K=2	.93 (12)	.92 (.11)	.98 (.05)	.91 (.17)	.89 (.24)	.98 (.07)
N=25, K=3	.45 (.17	.43 (.15)	.91 (.14)	.55 (.14)	.47 (.16)	.93 (.12
N= 25, K= 4	.36 (.16	.33 (.12)	.80 (.19)	.48 (.22)	.38 (.22)	.73 (.16)
N= 75, K=2	.94 (.07)	.91 (.09)	.99 (.03)	.94 (.09)	.93 (.09)	.99 (.05)
N=75, K=3	.58 (.18)	.55 (.15)	.98 (.04)	.63 (.21)	.59 (.17)	.98 (.04)
N=75, K=4	.50 (.16	.53 (.16)	.94 (.10)	.55 (.17)	.59 (.14)	.83 (.13)
N=150, K=2	.96 (.04)	.95 (.04)	.99 (.02)	.9 (.19)	.91 (.12)	1.00 (.02)
N=150, K=3	.57 (.21)	.54 (.12)	.98 (.03	.55 (.22)	.56 (.14)	.99 (.02)
N=150, K=4	.52 (.14)	.58 (.12)	.96 (.06)	.56 (.16)	.61 (.12)	.90 (.11)

Supplemental Table 2. Average Ramsey Indices for GIMME. *Note: N indicates sample size; K indicates number of subgroups; Eq. indicates equal subgroup sizes; Uneq. indicates unequal subgroup sizes; averages for each method are followed by standard deviations in parentheses.*

Condition	GIMME			
	Path Recall	Path Precision	Dir. Recall	Dir. Precision
N=25, K=2, Eq.	88.28 (1.23)	96.92 (1.55)	85.43 (1.5)	94.1 (3.83)
N=25, K=3, Eq.	89.45 (0.92)	98.11 (1.22)	86.87 (1.27)	95.99 (2.83)
N=25, K=4, Eq.	90.16 (0.86)	97.88 (1.28)	87.49 (0.91)	96.28 (2.12)
N=25, K=2, Uneq.	91.29 (1.62)	95.44 (1.5)	87.39 (2.44)	91.94 (4.39)
N=25, K=3, Uneq.	89.64 (1.07)	97.55 (1.85)	86.78 (1.63)	95.11 (3.77)
N=25, K=4, Uneq.	90.12 (1.08)	97.05 (1.67)	86.88 (1.66)	94.12 (3.6)
N=75, K=2, Eq.	87.94 (0.77)	97.1 (1.76)	85.21 (1.32)	94.7 (4.08)
N=75, K=3, Eq.	89.21 (0.67)	98.54 (0.57)	86.83 (0.88)	97.06 (1.4)
N=75, K=4, Eq.	90.17 (0.58)	98.53 (0.62)	87.66 (0.64)	97.4 (0.84)
N=75, K=2, Uneq.	90.82 (1.13)	95.87 (1.13)	86.6 (1.19)	92.18 (2.24)
N=75, K=3, Uneq.	89.65 (0.81)	98.41 (0.71)	86.66 (1.21)	95.98 (2.3)
N=75, K=4, Uneq.	90.47 (0.78)	97.87 (0.96)	87.27 (1.17)	95.62 (2.09)
N=150, K=2, Eq.	87.74 (0.68)	97.42 (1.24)	85.24 (0.94)	95.58 (2.38)
N=150, K=3, Eq.	89.34 (0.55)	98.68 (0.38)	86.98 (0.84)	97.22 (1.34)
N=150, K=4, Eq.	90.14 (0.48)	98.59 (0.33)	87.63 (0.52)	97.48 (0.43)
N=150, K=2, Uneq.	91.28 (0.62)	95.95 (0.92)	86.31 (1.05)	91.51 (2.64)
N=150, K-3, Uneq.	89.64 (0.8)	98.29 (0.65)	86.41 (1.02)	95.63 (2.06)
N=150, K=4, Uneq.	90.46 (0.72)	98.19 (0.79)	86.9 (0.99)	95.36 (2.15)

Supplemental Table 3. Average Ramsey Indices for S-GIMME. *Note: N indicates sample size; K indicates number of subgroups; Eq. indicates equal subgroup sizes; Uneq. indicates unequal subgroup sizes; averages for each method are followed by standard deviations in parentheses.*

Condition	S−GIMME			
	Path Recall	Path Precision	Dir. Recall	Dir. Precision
N=25, K=2, Eq.	88.28 (1.23)	96.92 (1.55)	85.43 (1.5)	94.1 (3.83)
N=25, K=3, Eq.	89.45 (0.92)	98.11 (1.22)	86.87 (1.27)	95.99 (2.83)
N=25, K=4, Eq.	90.16 (0.86)	97.88 (1.28)	87.49 (0.91)	96.28 (2.12)
N=25, K=2, Uneq.	91.29 (1.62)	95.44 (1.5)	87.39 (2.44)	91.94 (4.39)
N=25, K=3, Uneq.	89.64 (1.07)	97.55 (1.85)	86.78 (1.63)	95.11 (3.77)
N=25, K=4, Uneq.	90.12 (1.08)	97.05 (1.67)	86.88 (1.66)	94.12 (3.6)
N=75, K=2, Eq.	87.94 (0.77)	97.1 (1.76)	85.21 (1.32)	94.7 (4.08)
N=75, K=3, Eq.	89.21 (0.67)	98.54 (0.57)	86.83 (0.88)	97.06 (1.4)
N=75, K=4, Eq.	90.17 (0.58)	98.53 (0.62)	87.66 (0.64)	97.4 (0.84)
N=75, K=2, Uneq.	90.82 (1.13)	95.87 (1.13)	86.6 (1.19)	92.18 (2.24)
N=75, K=3, Uneq.	89.65 (0.81)	98.41 (0.71)	86.66 (1.21)	95.98 (2.3)
N=75, K=4, Uneq.	90.47 (0.78)	97.87 (0.96)	87.27 (1.17)	95.62 (2.09)
N=150, K=2, Eq.	87.74 (0.68)	97.42 (1.24)	85.24 (0.94)	95.58 (2.38)
N=150, K=3, Eq.	89.34 (0.55)	98.68 (0.38)	86.98 (0.84)	97.22 (1.34)
N=150, K=4, Eq.	90.14 (0.48)	98.59 (0.33)	87.63 (0.52)	97.48 (0.43)
N=150, K=2, Uneq.	91.28 (0.62)	95.95 (0.92)	86.31 (1.05)	91.51 (2.64)
N=150, K-3, Uneq.	89.64 (0.8)	98.29 (0.65)	86.41 (1.02)	95.63 (2.06)
N=150, K=4, Uneq.	90.46 (0.72)	98.19 (0.79)	86.9 (0.99)	95.36 (2.15)

Supplemental Table 4. Average Ramsey Indices for Omitted Variable Analysis. *Note: Dir. indicates direction. Averages for are followed by standard deviations in parentheses.*

Omitted Variable	Path Recall	Path Precision	Dir. Recall	Dir. Precision
1	96.46 (0.58)	86.12 (1.33)	93.88 (0.84)	79.9 (1.76)
2	96.17 (0.25)	99.64 (0.1)	93.56 (0.29)	99.35 (0.24)
3	96.06 (0.46)	91.35 (0.88)	93.47 (0.51)	87.23 (0.89)
4	96.23 (0.3)	92.71 (0.88)	93.53 (0.35)	88.98 (1.12)
5	95.95 (0.3)	95.51 (0.32)	93.23 (0.81)	89.51 (1.4)
6	96.17 (0.29)	97.68 (0.27)	93.49 (0.28)	97.33 (0.37)
7	96.41 (0.38)	95.01 (1.15)	94.28 (0.55)	88.69 (1.16)
8	95.91 (0.97)	89.76 (0.98)	92.04 (1.13)	81.58 (1.6)
9	96.17 (0.36)	96.82 (1.05)	93.72 (0.4)	94.81 (1.78)
10	96.18 (0.33)	99.41 (0.61)	93.55 (0.67)	94.45 (1.42)