

Evaluating Group-Level Recovery of Contemporaneous and Lagged Connectivity in GIMME: A Simulation Study

Yijia Xue and Linlin Liu

December 4, 2025

Group Iterative Multiple Model Estimation (GIMME) is a unified structural equation modeling framework that estimates directed connectivity networks at both the group and individual level. This project uses simulations to examine how the number of observations per subject (T) affects group-level recovery of contemporaneous edges (A) and lagged edges (Φ). We generated multivariate time series for $N = 100$ subjects with $p = 8$ nodes under a known uSEM model and applied GIMME for $T \in \{50, 150, 300\}$, repeating each condition 20 times. Performance was summarized with sensitivity (TPR) and false positive rate (FPR) for the estimated group-level A and Φ matrices. GIMME recovered contemporaneous edges well when time series were moderately long. Mean TPR for A increased from 0.00 at $T = 50$ to 0.67 at $T = 150$ and 0.94 at $T = 300$, with FPR close to zero across all T . Lagged edges were more difficult to detect, with TPR for Φ rising only from about 0.35 to 0.58 over the same range. Overall, the study suggests that reliable group-level estimation of contemporaneous connectivity with GIMME requires at least 150–300 time points per subject, while lagged connections remain harder to estimate accurately.

1 INTRODUCTION

The broad statistical problem motivating this project is the challenge of estimating directed connectivity networks from multivariate time-series data. In fields such as neuroscience,

psychology, and biomedicine, researchers seek to understand how activity in one process influences activity in another either within the same time point or across time. For example, in functional MRI studies, identifying directed interactions among brain regions is essential for understanding cognitive processes and neural organization. Traditional group-level approaches often aggregate data across subjects and assume a single "average" connectivity network, which can obscure important individual differences and lead to misleading conclusions when the underlying structure varies across participants.

Group Iterative Multiple Model Estimation (GIMME) is designed to address this limitation by combining group-level and individual-level modeling within a unified structural equation modeling (uSEM) framework. The algorithm first identifies connections that consistently improve model fit across a large proportion of subjects and treats these as group-level edges. It then adds person-specific connections to account for individual variation. This hybrid strategy enables researchers to draw population-level inferences while still capturing subject-specific dynamics, making GIMME particularly well suited for heterogeneous populations.

In this project, we focus on a fundamental and practically important question: How does the length of the time series T affect GIMME's ability to recover the true group-level connectivity structure? To investigate this, we evaluate GIMME's performance in recovering two types of directed edges in a known generative model: (i) contemporaneous edges in matrix A , representing directed relationships within the same time point; and (ii) lagged edges in matrix Φ , representing directed influences from time $t - 1$ to time t .

Understanding how time series length affects recovery is important for experimental design. Collecting long time series is costly and sometimes infeasible in real data applications, so determining the minimum number of observations needed for reliable network estimation has direct implications for planning empirical studies.

To address this question, we conduct a controlled simulation study in which multivariate time series are generated for $N = 100$ subjects with $p = 8$ variables under a known uSEM model with correlated noise. For each choice of $T \in \{50, 150, 300\}$, we generate 20 replications and apply GIMME to estimate the group-level networks. Performance is evaluated using sensitivity, specificity, false positive rate, and related metrics for both contemporaneous and lagged edges.

2 METHODOLOGY

2.1 MODEL AND NOTATION

We work within the unified structural equation modeling (uSEM) framework used by GIMME. For subject i with p variables, let $\eta_{i,t} \in \mathbb{R}^p$ denote the p -dimensional state vector at time t . The data-generating model is

$$\eta_{i,t} = A_i \eta_{i,t} + \Phi_i \eta_{i,t-1} + \zeta_{i,t}, \quad (2.1)$$

where

- A_i is a $p \times p$ matrix of contemporaneous effects within-time directed edges,
- Φ_i is a $p \times p$ matrix of lag-1 effects between-time directed edges, and
- $\zeta_{i,t} \sim N(0, \Psi_i)$ is the white noise with subject-specific covariance matrix Ψ_i .

Solving for $\eta_{i,t}$ gives the simulation form

$$\eta_{i,t} = (I - A_i)^{-1} (\Phi_i \eta_{i,t-1} + \zeta_{i,t}), \quad (2.2)$$

which is what we implement in the code.

2.2 GROUP-LEVEL STRUCTURE AND INDIVIDUAL HETEROGENEITY

For each simulation, we first specify group-level matrices A_{common} and Φ_{common} , which represent the edges shared across subjects. The contemporaneous matrix A_{common} follows a directed ring structure among eight nodes: node j predicts node $j+1$ with moderate positive weights sampled from $\text{Unif}(0.2, 0.35)$, and all diagonal entries are fixed to zero. The lagged matrix Φ_{common} is intentionally denser, with autoregressive self-loops on all nodes sampled from $\text{Uniform}(0.1, 0.2)$, supplemented by first- and second-off-diagonal bands and several randomly placed edges with smaller weights to create a richer temporal structure.

Subject-specific matrices A_i and Φ_i are then generated by adding a small number of individual edges to these group-level templates, introducing controlled heterogeneity across subjects. Individual-level variation is introduced using a helper function that (i) randomly selects a small number of zero entries in the common matrix, (ii) fills them with positive weights in a specified range, and (iii) enforces constraints. This leads to sparse subject-specific deviations from the group network.

2.3 NOISE COVARIANCE Ψ_i

For each subject, we generate a structured noise covariance matrix Ψ_i via the helper function `generate_Psi_i()`:

- Start from a diagonal matrix with common variance.
- Select a random subset of off-diagonal pairs where A_i has no edge and insert small positive covariances drawn from $\text{Uniform}(0.01, 0.05)$.
- If needed, add a small jitter to the diagonal to ensure Ψ_i is positive definite.

This creates correlated noise in directions that do not overlap with contemporaneous edges, mimicking realistic background covariance structure.

2.4 SIMULATION DESIGN

The main design factors are:

- Number of subjects: $N = 100$ per data set.
- Number of variables: $p = 8$.
- Observed time points per subject: $T_{\text{obs}} \in \{50, 150, 300\}$.
- Burn-in: 100 initial time points simulated and discarded to allow the process to approach stationarity.
- Replications: for each value of T_{obs} , we run $R = 20$ independent replications, for a total of 60 data sets.

For each subject, the function `simulate_subject_total()` generates the full time series of length $T_{\text{total}} = T_{\text{obs}} + 100$ using the uSEM recursion, and we keep the last T_{obs} rows. The time series of each subject is saved as a separate text file, which is the input format expected by the `gimme` R package.

2.5 GIMME ESTIMATION

For each simulated data set, we call

```
fit <- gimme(  
  data      = "sim_data",  
  out       = "sim_results",  
  ar        = TRUE,  
  subgroup  = FALSE,  
  groupcutoff = group_cutoff, # 0.75  
  subcutoff = 0.50,  
  standardize = TRUE,  
  plot      = FALSE  
)
```

Key choices:

- `ar = TRUE` lets GIMME automatically include first-order autoregressive terms, consistent with the uSEM framework.
- `groupcutoff = 0.75` requires that a candidate path significantly improves model fit for at least 75% of subjects before it is added to the group model. This reflects the idea that group-level edges should be fairly common.
- `subcutoff = 0.50` controls subject-specific paths; my focus here is on group-level recovery.

GIMME writes a summary path count matrix, which tallies how many subjects show each contemporaneous and lagged path in their final models. We post-process this file to recover estimated group-level adjacency matrices:

- Columns containing the string "lag" correspond to lagged (Φ) edges.
- The remaining columns correspond to contemporaneous (A) edges.
- A group-level edge is declared present if its count is at least $\text{groupcutoff} \times N_{\text{subj}}$.

The resulting binary matrices A_{est} and Φ_{est} are compared with the true A_{common} and Φ_{common} .

2.6 PERFORMANCE METRICS

For each simulation and each matrix type (A or Φ), we compute:

- True positives (TP): edges present in both true and estimated matrices.
- False positives (FP): edges absent in truth but present in the estimate.
- False negatives (FN): true edges that were missed.
- True negatives (TN): zero edges correctly estimated as zero.

From these counts, we derive:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{Spec} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \quad \text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}, \quad \text{FNR} = \frac{\text{FN}}{\text{TP} + \text{FN}}.$$

We then average these metrics across the 20 replications at each T_{obs} .

3 IMPLEMENTATION

All analyses were implemented in R using the packages `gimme`, `MASS`, `dplyr`, `ggplot2`, and `qgraph`. The code is organized into three logical blocks:

1. Helper functions.

- `generate_Psi_i()` constructs subject-specific noise covariance matrices.
- `generate_random_individual_edges()` adds a random number of subject-specific edges to the common A or Φ while respecting constraints.
- `simulate_subject_total()` simulates the full multivariate time series using the uSEM recursion.
- `compare_mats()` computes TP, FP, FN, TN, and derived metrics for any pair of true vs. estimated matrices.

2. **Single-simulation driver.** The function `run_one_sim()` takes T_{obs} and other parameters, generates the full synthetic data set for N subjects, runs GIMME, extracts A_{est} and Φ_{est} , and returns the performance metrics. An optional flag `return_nets = TRUE` additionally returns the true and estimated networks for plotting.
3. **Monte Carlo loop and summaries.** An outer loop runs `run_one_sim()` for each combination of T_{obs} and replication index. Results are combined into a single data frame and summarized using `dplyr`.

Several coding challenges arose during implementation. First, it was necessary to manage directory creation and cleanup carefully. The folders for simulation data and results must be cleared at the beginning of each simulation loop to ensure that no files from previous runs contaminate subsequent analyses. Second, the summary path count matrix produced by `gimme()` required careful parsing. Because contemporaneous and lagged paths are intermixed in this matrix, we relied on consistent column-name patterns to separate counts corresponding to the A matrix from those corresponding to the Φ matrix, and then reconstructed the estimated adjacency matrices accordingly. Finally, runtime considerations played an important role. Each GIMME call is computationally intensive, especially with $N = 100$ subjects and moderately long time series, so we selected a replication size of $R = 20$ to strike a balance between Monte Carlo precision and manageable computation time. No external baseline method was implemented. Performance was evaluated relative to the known truth encoded in A_{common} and Φ_{common} .

4 DATA ANALYSIS

4.1 OVERALL TPR PATTERNS FOR A AND Φ

Figure 4.1 and the corresponding summary tables show a clear improvement in recovery as the length of the time series increases.

For lagged edges (Φ), the mean TPR across 20 replications increases gradually with T :

T_{obs}	mean TPR (Φ)	sd TPR (Φ)
50	0.35	0.01
150	0.38	0.04
300	0.58	0.03

Even with only 50 time points, GIMME identifies about one-third of the true lagged edges at the group level. Increasing to 150 observations yields only a modest gain, but by 300 observations the average lagged TPR rises to about 0.58. The standard deviations are fairly small, suggesting stable performance across replications.

For contemporaneous edges (A), the dependence on T is much stronger:

T_{obs}	mean TPR	mean FPR	mean Spec	mean FN	mean FNR
50	0.00	0.00	1.000	7.00	1.00
150	0.67	0.009	0.999	2.30	0.33
300	0.94	0.026	0.997	0.45	0.06

At $T = 50$, GIMME fails to include any group-level contemporaneous edges: TPR and FPR are essentially 0, and all true A edges are treated as false negatives. At $T = 150$, performance improves substantially, and by $T = 300$ recovery is excellent: mean TPR is approximately 0.94 and mean FNR is only 0.06, while FPR remains below 0.03.

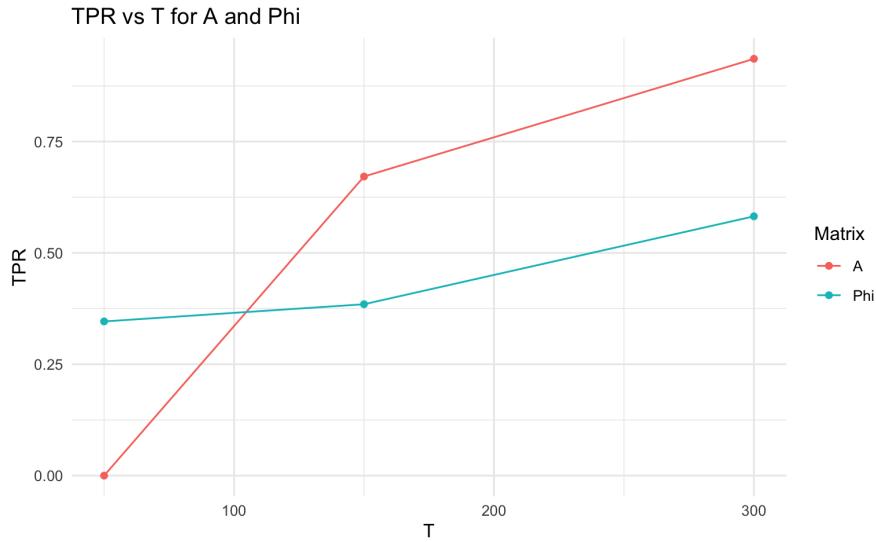


Figure 4.1: True positive rate (TPR) versus T for contemporaneous (A) and lagged (Φ) group-level edges.

Figure 4.2 emphasizes that the mean TPR for A rises steeply from 0 at 50 time points to near 1 at 300, with decreasing variability.

Interpretation. These results suggest that, under the current simulation design and group cutoff, GIMME needs moderately long time series (150–300 points) to reliably detect group-level contemporaneous connections, whereas lagged edges are consistently harder to recover. The combination of fewer observations per parameter and the more complex lagged structure in Φ likely contributes to the lower TPR for lagged edges.

4.2 NETWORK PLOTS: TRUE VS. ESTIMATED A

Figure 4.3 compares the true and estimated group-level contemporaneous networks for a representative simulation at each value of T_{obs} . The true A matrix is identical across all

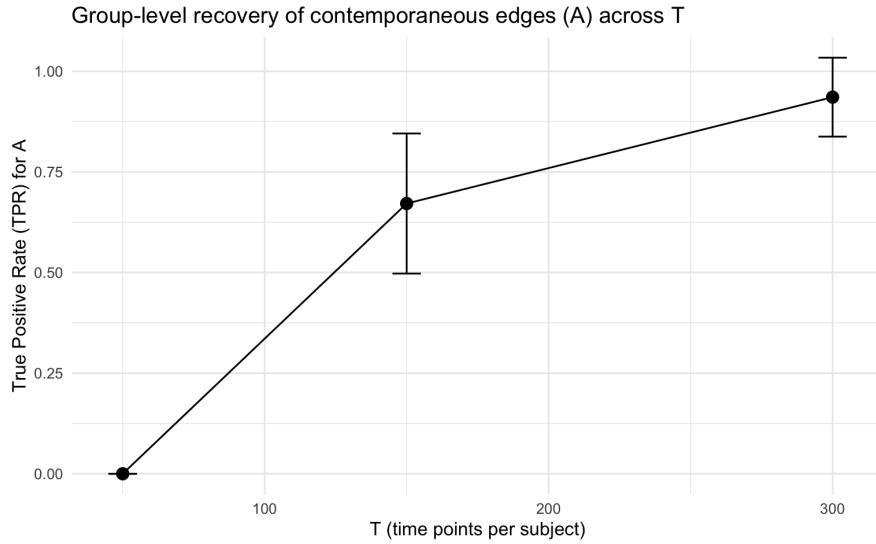


Figure 4.2: Mean TPR for group-level contemporaneous edges (A) across different time series lengths T , with one standard deviation error bars.

conditions and follows a directed ring structure in which node 1 influences node 2, node 2 influences node 3, and so on through node 8, with moderately strong positive weights. At $T = 50$, the estimated group network is almost empty: with so few observations per subject, very few individual models include the ring edges reliably enough to exceed the 0.75 group cutoff. When T increases to 150, the estimated network begins to resemble the true ring more clearly, although several edges remain undetected, consistent with the mid-range TPR of approximately 0.67. At $T = 300$, the ring structure is nearly perfectly recovered, with all or almost all true edges present and very few spurious connections, matching the high TPR and low false-negative rate observed in the numerical summaries.

4.3 NETWORK PLOTS: TRUE VS. ESTIMATED Φ

Figure 4.4 presents the corresponding comparisons for the lagged network Φ . The true Φ matrix is substantially denser than A , containing autoregressive self-loops on all nodes as well as numerous off-diagonal lagged connections that create a more complex temporal structure. At both $T = 50$ and $T = 150$, however, the estimated group-level lagged networks are extremely sparse, typically containing at most a few weak edges or none at all, with the vast majority of true lagged paths being missed. Even at $T = 300$, where contemporaneous recovery is excellent, the estimated Φ network remains much sparser than the truth: although a handful of key edges begin to appear, large portions of the true lagged structure remain unrecovered. These plots visually explain the more modest TPR values for Φ . The combination of a denser true lagged network, the group cutoff, and competition between contemporaneous and lagged paths during model selection appears to make lagged edges particularly difficult to recover.

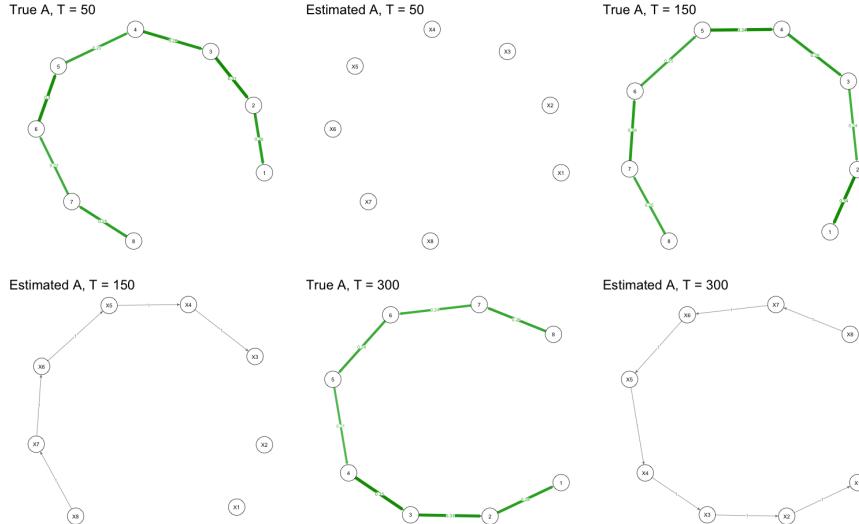


Figure 4.3: True and estimated group-level contemporaneous networks A for $T = 50$, $T = 150$, and $T = 300$.

5 CONCLUSIONS

This simulation study used synthetic multivariate time series to evaluate how the number of observations per subject affects GIMME's ability to recover group-level connectivity in a uSEM setting.

KEY FINDINGS

Overall, the simulation results show a clear and consistent relationship between time series length and GIMME's ability to recover group-level connectivity. Recovery of contemporaneous edges was highly sensitive to T : with only 50 observations per subject, GIMME did not identify any group-level contemporaneous connections under the 0.75 group cutoff, whereas at $T = 150$ it successfully recovered roughly two-thirds of the true A edges, and by $T = 300$ recovery was excellent with a mean TPR of approximately 0.94 and minimal false positives. In contrast, lagged edges in the denser Φ matrix were substantially more difficult to estimate. Although TPR increased with T , it rose only from about 0.35 at $T = 50$ to about 0.58 at $T = 300$, indicating that many true lagged paths fail to meet the group-level threshold even with relatively long time series. These patterns highlight that GIMME is inherently conservative when T is small: noisy individual fits combined with a strict group cutoff produce sparse group networks that avoid false positives but exhibit high false negative rates. The network visualizations reinforce these quantitative findings, showing that the contemporaneous ring structure is absent at $T = 50$, partially visible at $T = 150$, and nearly perfectly reconstructed at $T = 300$, while the lagged network remains noticeably under-estimated across all conditions.

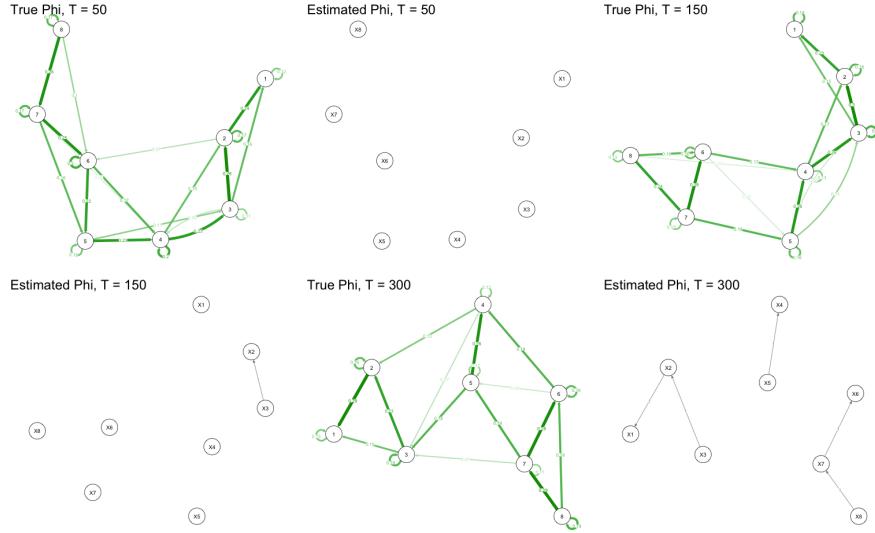


Figure 4.4: True and estimated group-level lagged networks Φ for $T = 50$, $T = 150$, and $T = 300$.

PRACTICAL IMPLICATIONS

The findings of this study carry several practical implications for researchers considering the use of GIMME in applications involving multivariate time-series data. First, the results suggest that at least 150 time points per subject are necessary before GIMME begins to consistently recover meaningful group-level contemporaneous structure. With fewer observations, the algorithm tends to produce extremely sparse group networks due to insufficient evidence for paths meeting the group-level threshold. In contrast, time series of approximately 300 observations appear to provide strong and reliable recovery of contemporaneous edges, with high sensitivity and minimal false positives. However, even with 300 observations, lagged connections remain considerably more difficult to identify, as reflected in the noticeably lower TPR values for the Φ matrix. Researchers whose primary interest lies in modeling lagged dynamics may therefore require substantially longer time series, more permissive group-level cutoffs, or alternative modeling strategies that emphasize temporal structure.

FUTURE DIRECTIONS

This work suggests several avenues for further investigation. A natural extension would be to examine a wider range of group cutoff values to characterize how sensitivity–specificity trade-offs change under different criteria for declaring group-level paths. It would also be useful to vary the density and strength of true lagged edges in the data-generating process to identify conditions under which the Φ matrix becomes easier to recover. Incorporating richer forms of heterogeneity would better approximate real datasets and clarify how heterogeneity interacts with time-series length. Finally, comparing GIMME with alternative connectivity estimation approaches under matched simulations would provide a more comprehensive assessment of its strengths and limitations.

TEAM MEMBER CONTRIBUTION

Both team members contributed substantially and fairly to the completion of this project. While Yiji Xue took primary responsibility for developing and refining the R code used for data generation, simulation, and model estimation. Linlin Liu led the preparation of the written report. We jointly created the presentation slides for our own parts. Throughout the project, we worked together, regularly reviewing each other's work and discussing design decisions. Although our specific tasks differed, the overall workload was shared equally, and the final project reflects the combined and collaborative effort of both team members.

REFERENCES

- Fair, D. A., Posner, J., Nagel, B. J., Bathula, D., Dias, T. G. C., Mills, K. L., Blythe, M. S., Giwa, A., Schmitt, C. F., & Nigg, J. T. (2010). Atypical default network connectivity in youth with attention-deficit/hyperactivity disorder. *Biological Psychiatry*, 68(12), 1084–1091.
- Hillary, F. G., Medaglia, J. D., Gates, K. M., Molenaar, P. C. M., Slocomb, J., Peechatka, A., & Good, D. C. (2011). Examining working memory task acquisition in a disrupted neural network. *Brain: A Journal of Neurology*, 134, 1555-1570.
- Gates, K. M., Molenaar, P. C. M., Hillary, F. G., Ram, N., & Rovine, M. J. (2010). Automatic search in fMRI connectivity mapping: an alternative to Granger causality using formal equivalences between SEM path modeling, VAR, and unified SEM. *NeuroImage*, 53, 1118-1125.
- Gates, K. M., Molenaar, P. C. M., Hillary, F. G., & Slobounov, S. (2011). Extended unified SEM approach for modeling event-related fMRI data. *NeuroImage*, 54, 1151-1158.
- Gates, K.M., & Molenaar, P.C.M. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage*, 63, 310–319.
- Kim, J., Zhu, W., Chang, L., Bentler, P.M., & Ernst, T. (2007). Unified structural equation modeling for the analysis of multisubject, multivariate functional fMRI data. *Human Brain Mapping*, 28, 85–93.
- Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: bringing the person back into scientific psychology, this time forever. *Measurement*, 2(4), 201–218.
- Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., Ramsey, J. D., & Woolrich, M. W. (2011). Network modeling methods for FMRI. *NeuroImage*, 54, 875–891.