

The Gaussian Graphical Model in Cross-sectional and Time-series Data

Supplementary Materials

Contents

1	Notation	1
2	Two-step multi-level VAR	2
2.1	Multi-level modeling	2
2.2	Extending multi-level VAR: two-step multi-level VAR	3
3	Simulation Studies	5
4	Stationary distribution	8
5	References	8

1 Notation

Throughout the paper we employ the following notation. Roman letters indicate observed variables, and Greek letters indicate parameters or latent variables. Nonboldface letters indicate a single value. An uppercase nonboldface letter indicates a random variable, and a lowercase nonboldface letter indicates a realization. We use t to denote measurement occasion and T to denote a random measurement occasion,¹ i ($i \in \{1, 2, \dots, m\}$) to denote item administered, and p ($p \in \{1, 2, \dots, n\}$) to denote a subject and P to denote a random subject. We will use lowercase boldface letters to denote column vectors and uppercase boldface vectors to denote matrices. Subscripts will denote if these are random or fixed. For example, \mathbf{B}_p will denote a fixed matrix for subject p , and \mathbf{B}_P will denote the matrix of random subject P (which has a distribution).

Because we are interested in finding dynamics between items, we use vector \mathbf{y} to denote the set of all items.² For the observed variables, we will use consistent subscripts (measurement, subject) to denote which items are contained in the vector. For example, $\mathbf{y}_{[t,p]}$ denotes all responses of subject p at time point t , and $\mathbf{y}_{[T,p]}$ denotes all responses of subject p at a random time point T . A set in this subscript indicates multiple responses.

¹Mostly we assume measurements are nested in subjects, and two subjects might have a different number of measurement occasions. As such, $t = 1$ for subject $p = 1$ might not correspond to $t = 1$ for subject $p = 2$.

²If researchers are interested in dyadic interactions (Ferrer, 2016), for example, then a dyadic pair can be seen as a “subject,” and items can be the item responses from both subjects.

For example, we will use $\mathbf{y}_{[\{t-1,t\},p]}^\top = [\mathbf{y}_{[t-1,p]}^\top \quad \mathbf{y}_{[t,p]}^\top]$ to denote a set of lagged and current responses from subject p around time point t . If only one observation or subject is measured, we will drop the square brackets (e.g., $\mathbf{y}_P = \mathbf{y}_{[1,P]}$ indicates the cross-sectional response pattern of a random subject). When it is unclear if the set of items corresponds to a random person or a random measurement occasion, we refer to C as a random case, with c as a particular case, and subset the data either as \mathbf{y}_C to describe a random response pattern or \mathbf{y}_c to describe a realization—in cross-sectional data $\mathbf{y}_C = \mathbf{y}_P$ and in $N = 1$ time-series data $\mathbf{y}_C = \mathbf{y}_T$. C could also indicate a set of multiple responses. Other subscripts denote subsets of a vector or matrix, with notation $-(\dots)$ indicating the subset of everything except $\{\dots\}$.

2 Two-step multi-level VAR

In this appendix, we will outline two-step multi-level VAR, which we propose as a methodology to estimate the GVAR model using multi-level estimation. This method builds on the work of Bringmann et al. (2013), and extends their proposed algorithm by including between-subject effects (Hamaker & Grasman, 2014) and estimating the contemporaneous network by performing a second multi-level estimation on the residuals of the temporal model (the second “step”). To reiterate the paper, the model to estimate is:

$$\mathbf{y}_{[T,p]} \mid \mathbf{y}_{[T-1,p]} = \mathbf{y}_{[t-1,p]} \sim N\left(\boldsymbol{\mu}_p + \mathbf{B}_p\left(\mathbf{y}_{[t-1,p]} - \boldsymbol{\mu}_p\right), \boldsymbol{\Theta}_p\right).$$

In particular, we are interested in estimating between-subjects network $\mathbf{K}^{(\Omega)} = \boldsymbol{\Omega}^{-1} = \text{Var}(\boldsymbol{\mu}_P)^{-1}$ and the (distributions of) temporal networks \mathbf{B}_p and contemporaneous networks $\mathbf{K}_p^{(\Theta)} = \boldsymbol{\Theta}_p^{-1}$.

2.1 Multi-level modeling

The fixed effects and random effect variances and covariances can be estimated by estimating a VAR model for every subject, pooling the parameter estimates, and computing the mean (fixed effects) and variance–covariance matrix (random effects distribution). This estimation, however, is separate for every subject. To combine all observations in a single model, we can assign distributions over the parameters; in which case, we make use of multilevel modeling. Assigning distributions has two main benefits. First, instead of having a single parameter per subject, we now only need to estimate the parameters of the distribution. For example, when we model observations from 100 subjects, instead of estimating each parameter 100 times, we now only need to estimate its mean and variance. Second, the multilevel structure acts as a prior distribution in Bayesian estimation procedures—in case we wish to obtain person-specific parameter estimates post hoc. In particular, multilevel modeling leads to *shrinkage*; parameter values that are very different from the fixed effects are likely to be estimated closer to the fixed effect in multilevel modeling than when using a separate model for every subject. For example, if we estimate a certain temporal regression in five people and find the values 1.1, 0.9, 0.7, 1.3, and 10, it is likely that the fifth statistic, 10, is an outlier. Ideally, we would estimate this value to be closer to the other values.

Bringmann et al. (2013) proposed a sequential univariate method for estimating temporal VAR models. Because the joint conditional distribution of $\mathbf{y}_{[T,p]} \mid \mathbf{y}_{[T-1,p]} = \mathbf{y}_{[t-1,p]}$ is

normal, it follows that the marginal distribution of every variable is univariate normal and can be obtained by dropping all other parameters from the distribution:

$$y_{[T,p,i]} \mid \mathbf{y}_{[T-1,p]} = \mathbf{y}_{[t-1,p]} \sim N\left(\mu_{[p,i]} + \boldsymbol{\beta}_{[p,i]} \left(\mathbf{y}_{[t-1,p]} - \boldsymbol{\mu}_p\right), \theta_{[p,i]}\right),$$

in which $y_{[T,p,i]}$ denotes the i th element of $\mathbf{y}_{[T,p]}$, $\boldsymbol{\beta}_{[p,i]}$ indicates the row vector of the i th row of \mathbf{B}_p , and $\theta_{[p,i]}$ denotes the i th diagonal element of $\boldsymbol{\Theta}_p$. When drawn as a temporal network, the edges point to node i . Many software packages do not allow the estimation of $\boldsymbol{\mu}_p$ as described above. In this case, the sample means of every subject, $\bar{\mathbf{y}}_p$, can be taken as a substitute for $\boldsymbol{\mu}_p$ (Hamaker & Grasman, 2014). The model then becomes a univariate multilevel regression model with within-subject centered predictors, estimable by functions such as the `lmer` in `lme4` (Bates, Mächler, Bolker, & Walker, 2015). The Level 1 model becomes

$$\begin{aligned} y_{[t,p,i]} &= \mu_{[p,i]} + \boldsymbol{\beta}_{[p,i]} \left(\mathbf{y}_{[t-1,p]} - \bar{\mathbf{y}}_p\right) + \varepsilon_{[t,p,i]} \\ \varepsilon_{[T,p,i]} &\sim N(0, \theta_{[p,i]}), \end{aligned} \quad (1)$$

and the Level 2 model becomes

$$\begin{bmatrix} \mu_{[P,i]} \\ \boldsymbol{\beta}_{[P,i]} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ \boldsymbol{\beta}_{*i} \end{bmatrix}, \begin{bmatrix} \omega_{\mu_i} & \boldsymbol{\omega}^{(\boldsymbol{\beta}_i \boldsymbol{\mu}_i)^\top} \\ \boldsymbol{\omega}^{(\boldsymbol{\beta}_i \boldsymbol{\mu}_i)} & \boldsymbol{\Omega}(\boldsymbol{\beta}_i) \end{bmatrix}\right).$$

Estimation of such univariate models requires integrating over a simpler integral than estimation of multivariate models. As a result, sequential estimation using univariate models have been used in estimating multilevel VAR models (Bringmann et al., 2013). A downside, however, is that not all parameters are included in the model. In particular, correlations between means (between-subject effects) and between contemporaneous covariances are not retained, as well as the correlations between temporal edges pointing to different nodes. A second downside is that estimating correlated random effects does not work well for models with many predictors. In particular, `lmer` becomes very slow with approximately more than eight predictors. As such, networks with more than eight nodes are hard to estimate. To estimate larger networks (e.g., 20 nodes), we can choose to estimate uncorrelated random effects, which we term *orthogonal estimation*.

2.2 Extending multi-level VAR: two-step multi-level VAR

The methodology of Bringmann et al. (2013) does not estimate contemporaneous or between-subjects networks. Therefore, we propose extensions to the algorithm to estimate these networks. We propose a two-step method. Step 1 follows the procedure of (Bringmann et al., 2013) with the addition that between-subject effects are included (Hamaker & Grasman, 2014). This leads to estimates of the temporal and between-subjects networks. The second step involves taking the residuals of step 1 in order to obtain contemporaneous networks.

Step 1: Temporal and between-subjects networks. To obtain estimates of between-subject effects, the sample means of every subject, $\bar{\mathbf{y}}_p$ in Equation (1), can be included as predictors at the subject level (except for the mean of the dependent variable;

Hamaker & Grasman, 2014; Hoffman & Stawski, 2009; Curran & Bauer, 2011). With this extension, the Level 2 model for the person-specific mean of the i th variable now becomes

$$\mu_{[p,i]} = \boldsymbol{\beta}_i^{(\mu)} \bar{\mathbf{y}}_{[p,-(i)]} + \varepsilon_{[p,i]}^{(\mu)}, \quad (2)$$

in which we use $\boldsymbol{\beta}_i^{(\mu)}$ to denote the i th row (without the diagonal element i) of an $m \times m$ matrix $\mathbf{B}^{(\mu)}$, and $\bar{\mathbf{y}}_{[p,-(i)]}$ denotes the vector $\bar{\mathbf{y}}_p$ without the i -th element. Because $\bar{y}_{[p,i]}$ is itself an estimate of $\mu_{[p,i]}$, Equation (2) takes the form of a multiple regression model. As such, these estimates can be used to estimate a GGM between the means (Lauritzen, 1996; Meinshausen & Bühlmann, 2006)—the between-subjects network:

$$\mathbf{K}^{(\mu)} \approx \mathbf{D}^{(\mu)} \left(\mathbf{I} - \mathbf{B}^{(\mu)} \right),$$

with $d_{ii}^{(\mu)} = 1/\text{Var}(\varepsilon_{[p,i]}^{(\mu)})$. Due to the estimation in a multilevel framework, the resulting matrix will not be perfectly symmetric and must be made symmetric by averaging lower and upper triangular elements. Thus, each edge (i.e., partial correlation) in the between-subjects network is estimated by standardizing and averaging two regression parameters: the parameter denoting how well mean A predicts mean B and the regression parameter denoting how well mean B predicts mean A . Obtaining the between-subjects effects using regression coefficients rather than correlating the estimated means leads to standard errors that can be used to select significant edges.

Step 2: Contemporaneous networks. An estimate for contemporaneous networks can be obtained in a second step by investigating the residuals of the multilevel model that estimate the temporal and between-subject effects. These residuals can be used to run multilevel models that predict the residuals of one variable from the residuals of other variables at the same time point. Let $\hat{\varepsilon}_{[t,p,i]}$ denote the estimated residual of variable i at time point t of person p , and let $\hat{\boldsymbol{\varepsilon}}_{[t,p,-(i)]}$ denote the vector of residuals of all other variables at this time point. The Level 1 model then becomes

$$\hat{\varepsilon}_{[t,p,i]} = \boldsymbol{\beta}_{[p,i]}^{(\Theta)} \hat{\boldsymbol{\varepsilon}}_{[t,p,-(i)]} + \varepsilon_{[t,p,i]}^{(\Theta)}, \quad (3)$$

in which $\boldsymbol{\beta}_{[p,i]}^{(\Theta)}$ represents the i -th row (without the diagonal element i) of an $m \times m$ matrix, $\mathbf{B}_p^{(\Theta)}$, and $\varepsilon_{[t,p,i]}^{(\Theta)}$ represents a residual. In the Level 2 model, we again assign a multivariate normal distribution to parameters in $\boldsymbol{\beta}_i^{(\Theta)}$. It can be seen that Equation (3) also takes the form of a multiple regression model. Thus, this model can again be seen as the node-wise GGM estimation procedure:

$$\mathbf{K}_p^{(\Theta)} \approx \mathbf{D}_p^{(\Theta)} \left(\mathbf{I} - \mathbf{B}_p^{(\Theta)} \right),$$

with $d_{[p,i]}^{(\Theta)} = 1/\text{Var}(\varepsilon_{[T,p,i]}^{(\Theta)})$. Again the matrices need to be made symmetric by averaging upper and lower triangle elements. By using univariate multi-level regressions, rather than simply correlating the residuals, we can impose multilevel structure on the partial correlations in order to estimate fixed and random effects.³ Fixed effects can be obtained by using

³Estimating correlated random effects for regression coefficients is straightforward while estimating correlated random effects on covariances or correlations is not due to a requirement that these must add to a positive definite variance–covariance matrix; fixed effects covariances plus multivariate normally distributed random effects may lead to intra-individual variance–covariances that are not positive-definite.

the fixed effects matrices instead in the expression above. As with the temporal network, orthogonal estimation can be used when the number of variables is large (i.e., larger than approximately eight).

Thresholding. After estimating network structures, researchers may be interested in removing edges that may be spurious and due to sampling error. By setting edge weights to zero, effectively removing edges from a network, a sparse network is obtained that is more easily interpretable. One method of doing so is by removing all edges that are not significantly different from zero. For fixed effects, multilevel software returns standard errors and p -values, allowing for this thresholding. For the temporal networks, each edge is represented by one parameter and thus by one p -value. The contemporaneous and between-subjects networks, however, are a function of two parameters that are standardized and averaged: a regression parameter for the multiple regression model of the first node and a regression parameter for the multiple regression model of the second node. As such, for every edge, two p -values are obtained. We can choose to retain edges of which at least one of the two p -values is significant, termed the “or” rule, or we can choose to retain edges in which both p -values are significant, termed the “and” rule (Barber, Drton, & Others, 2015).

Summary. In sum, the above described two-step estimation method proposes to estimate a multilevel model per variable, using within-person centered lagged variables as within-subject predictors and the sample means as between-subject predictors. These models can be used to obtain estimates for the temporal network and between-subjects network. In a second step, the contemporaneous networks can be estimated by estimating a second multilevel on the residuals of the first multilevel model. The *mlVAR* R package implements these methods (Epskamp, Deserno, & Bringmann, 2017). In this package, temporal coefficients can be estimated as being “unique” per subject (unique VAR models per subject), “correlated” (estimating correlations between temporal effects), “orthogonal” (assuming temporal effects are not correlated), or “fixed” (no multilevel structure on temporal effects). The contemporaneous effects can also be estimated as being “fixed” (all residuals are used to obtain one GGM), “correlated” (second step multilevel model with correlated random effects), “orthogonal” (second step multilevel model with uncorrelated random effects), or “unique” (residuals are used to obtain a GGM per subject). The *mlVAR* package can also be used to plot the estimated networks, in which significance thresholding is used by default with a significance level of $\alpha = 0.05$.

3 Simulation Studies

In this section, we present simulations to assess the performance of *mlVAR* and *graphicalVAR* in performing the above-described methods for estimating network structures on ESM data of multiple subjects. Simulation studies on the described methods for cross-sectional and $n = 1$ studies are available elsewhere (Abegaz & Wit, 2013; Epskamp, 2016; Foygel & Drton, 2010). We varied the following conditions:

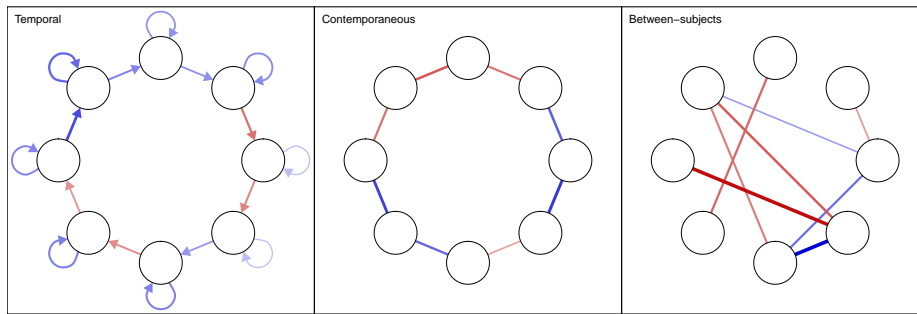
- **Number of nodes.** The number of nodes was set to be either 8 or 16, to be representative of plausible variable numbers in psychological ESM studies.
- **Sample size.** The sample size was varied between 50, 100 and 200. These values were chosen to represent plausible values in a psychological study.

- **Number of observations.** The number of observations per person was varied between 50, 100 and 200. These values were chosen to represent plausible values in a psychological study.
- **Fixed structure.** The contemporaneous network structure was simulated to be a chain graph (e.g., 1 – 2 – 3) and the between-subjects network was simulated to be a random network with the same number of edges. The temporal network was simulated to be either a chain graph (condition 1) or a chain graph in which every node is connected to the second to next node (e.g., 1 → 3 → 5; condition 2). These conditions were chosen such that in condition 1 there is a temporal edge whenever there is a contemporaneous edge, and in condition 2 there is not a temporal edge whenever there is a contemporaneous edge.
- **Individual structure.** Edges in the individual networks (temporal and contemporaneous) were either rewired with 50% probability per person or kept stable. The rewiring condition ensured that people may have vastly different network structures, thus reducing the strength of borrowing information from other subjects to estimate individual networks.

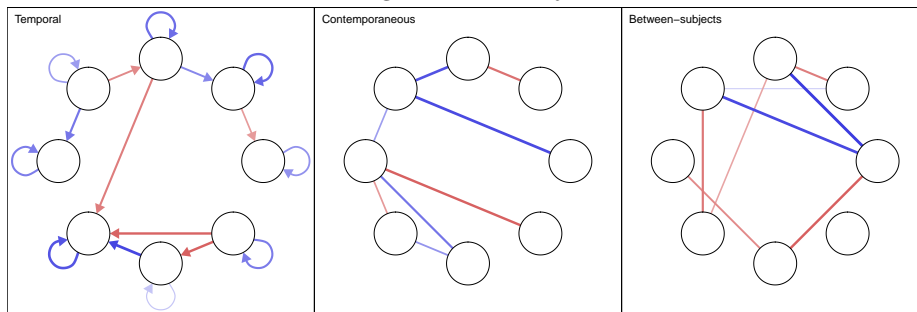
After generating network structures, and before rewiring, 50% of all edges were made negative. Next, individual edge parameters (for all networks) were drawn from normal distributions with mean of 0.35 and standard deviation of 0.1. Thus, when edges were not rewired, for different subjects in temporal and contemporaneous networks were equal in structure (which edge was present and the sign of the edge) but not in weight. In the condition in which edges were rewired, individual networks differed in both weight and structure. Each condition was replicated 100 times, leading to 72,000 total simulated datasets.

We used both *mlVAR* (two-step multi-level VAR) and *graphicalVAR* (pooled and individual LASSO estimation) to estimate fixed and subject-specific network structures. In *mlVAR*, correlated random effects were estimated in the 8-node condition and orthogonal random effects were estimated in the 16-node condition. In addition, an “and”-rule was used to threshold significant edges. In *graphicalVAR*, we varied 10 (temporal) by 10 (contemporaneous) LASSO tuning parameters in selecting the optimal GVAR model and 100 LASSO tuning parameters in selecting the optimal between-subjects GGM. The optimal tuning parameters were selected by minimizing the EBIC with $\gamma = 0.25$. To save computing time, we only estimated one individual subject network per replication in the *graphicalVAR* condition (fixed effects were based on all subjects), and thus base the results of individual network estimation performance in both methods on one network per replication. The true fixed effects were set to the mean of all individual networks created.

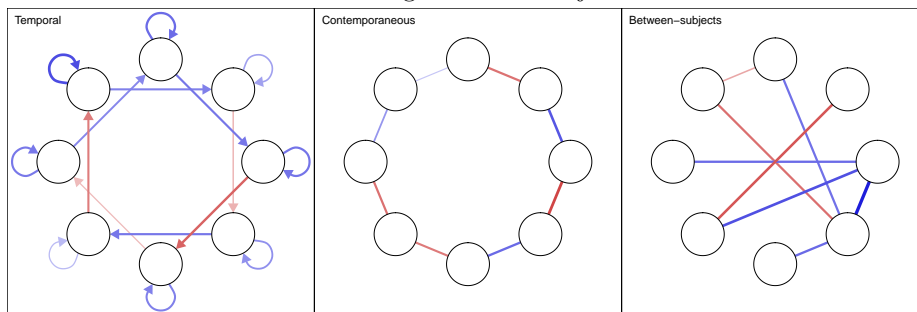
In order to assess how well the estimated networks resemble the true networks, we computed for each dataset the correlations between true and estimated fixed temporal, contemporaneous, and between-subjects networks and the correlations between true and estimated subject-specific temporal and contemporaneous networks—because the between-subjects network does not have random effects. In line with other studies on assessing how well a method retrieves the structure of a network (e.g., Epskamp, Rhemtulla, & Borsboom, 2017; van Borkulo et al., 2014), we computed the *sensitivity* (true positive rate) and the *specificity* (true negative rate). In addition, we computed the mean squared error and the average bias (mean absolute deviation of true to estimated parameters) per dataset.



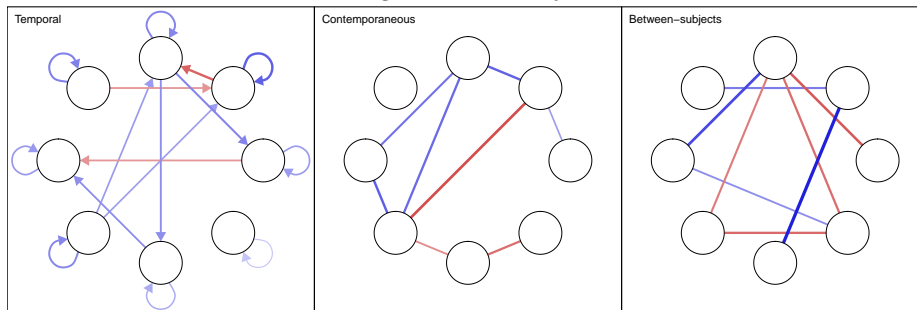
(a) Temporal condition 1 with no re-wiring of within-subject network structures.



(b) Temporal condition 1 with 50% re-wiring of within-subject network structures.



(c) Temporal condition 2 with no re-wiring of within-subject network structures.



(d) Temporal condition 2 with 50% re-wiring of within-subject network structures.

Figure 1. Examples of generated networks for one subject (temporal and contemporaneous) and overall between-subjects effects. Contemporaneous networks were always simulated to be a chain graph, and temporal networks were simulated to be a chain graph (condition 1) or a chain connected to each second node (e.g., $1 \rightarrow 3 \rightarrow 5$; condition 2). Edges in the within subject network were either rewired (ensuring every person had a different network structure) or kept the same.

Appendix A shows the results of the simulation study in the condition where edges were not rewired. It can be seen that performance was generally good in both methods. Fixed effects of the temporal and contemporaneous networks were well estimated (high correlations), most edges in the true network were detected (high sensitivity), and few edges were detected to be nonzero that were, in truth, zero (high specificity). In addition, the bias and mean squared error were generally low. The between-subjects network was better estimated with more people. Using *graphicalVAR* for estimating individual networks showed that at low sample-sizes, the method lacked power to detect true edges (low sensitivity) but did not estimate false edges (high specificity). No model selection is performed in *mlVAR* on subject-specific networks, leading to the specificity of 0 (all edges were always included in the network). The between-subjects estimation using *graphicalVAR* featured a moderate specificity, indicating some false edges were detected. It should be noted that the simulations used EBIC tuning parameter $\gamma = 0.25$, which errs more on the side of discovery than the often used $\gamma = 0.5$ value (Foygel & Drton, 2010). Of note is that the two-step multi-level procedure performed comparable, if not better, in estimating between-subject network structures than LASSO estimation based on the aggregated scores per person. Appendix B shows the results in the condition where edges were rewired, and shows here too a good performance for both methods. *mlVAR* estimation performed poorer than when the structure was the same over all subjects, and *graphicalVAR* performed identically. This was expected given that the *graphicalVAR* method does not take information of other subjects into account when estimating a network for one subject, while the *graphicalVAR* method does. A table with further detailed results from the simulation studies can be found in the other supplementary materials.

4 Stationary distribution

The GVAR model implies the following expression for the variance-covariance matrix of \mathbf{y}_t (dropping matrix indexing subscripts for notational clarity):

$$\begin{aligned}\text{Var}(\mathbf{y}_T) &= \text{Var}(\mathbf{B}\mathbf{y}_{T-1} + \boldsymbol{\varepsilon}_T) \\ \boldsymbol{\Sigma} &= \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^\top + \boldsymbol{\Theta},\end{aligned}$$

in which we make use of the assumption of stationarity and the assumption that residuals $\boldsymbol{\varepsilon}_T$ are uncorrelated with \mathbf{y}_{T-1} . Now, we can make use of the vectorization operator Vec and the Kronecker product \otimes to obtain (Kim, Nelson, et al., 1999):

$$(\mathbf{I} - \mathbf{B} \otimes \mathbf{B})^{-1} \text{Vec}(\boldsymbol{\Theta}) = \text{Vec}(\boldsymbol{\Sigma}) \quad ,$$

which gives an expression for the elements of $\boldsymbol{\Sigma}$ in terms of \mathbf{B} and $\boldsymbol{\Theta}$.

5 References

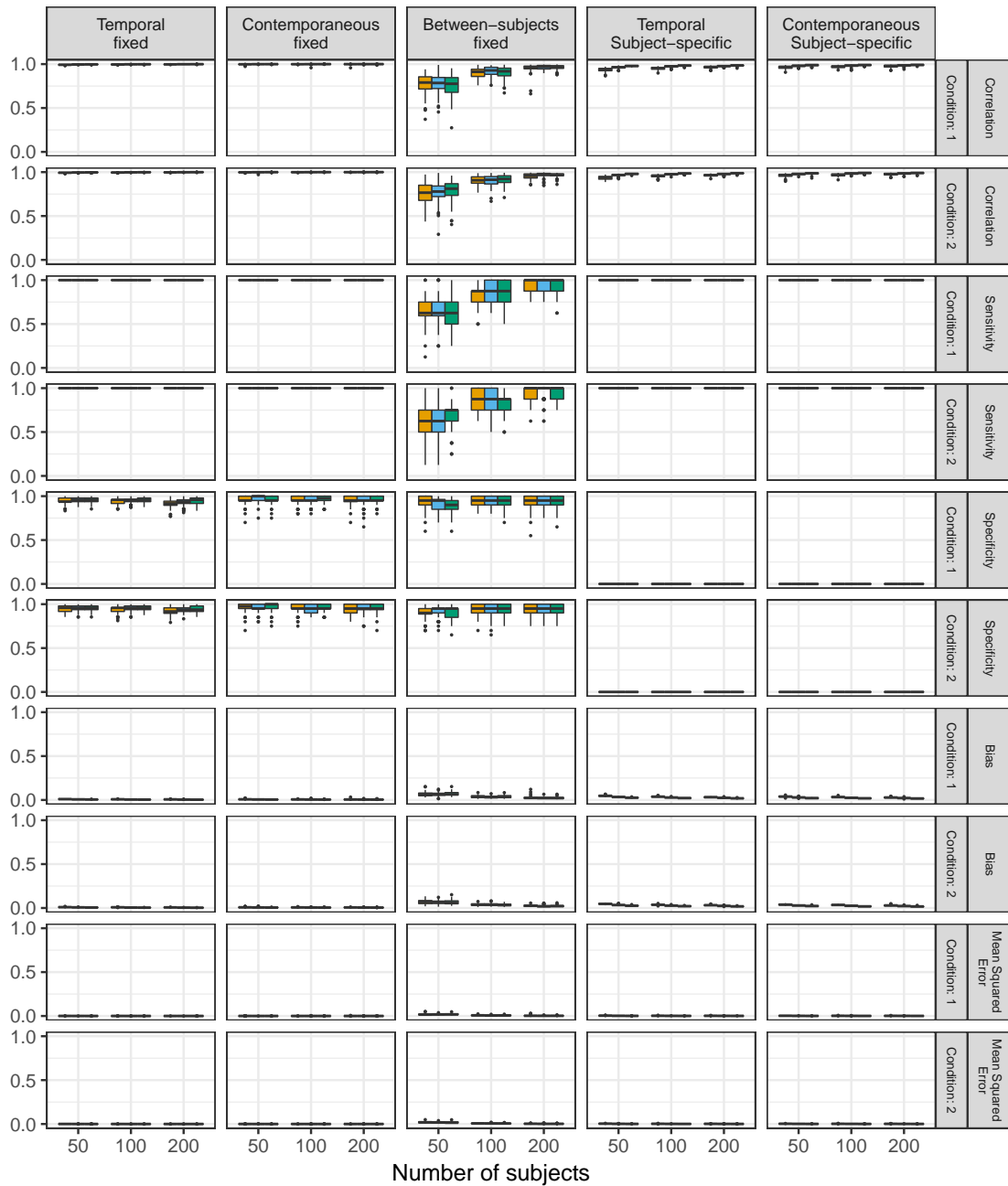
- Abegaz, F., & Wit, E. (2013). Sparse time series chain graphical models for reconstructing genetic networks. *Biostatistics*, kxt005. doi: 10.1093/biostatistics/kxt005
- Barber, R. F., Drton, M., & Others. (2015). High-dimensional Ising model selection with Bayesian information criteria. *Electronic Journal of Statistics*, 9(1), 567–607.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi: 10.18637/jss.v067.i01

- Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., ... Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLOS ONE*, *8*(4), e60188.
- Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual review of psychology*, *62*, 583–619. doi: 10.1146/annurev.psych.093008.100356
- Epskamp, S. (2016). Regularized Gaussian Psychological Networks: Brief Report on the Performance of Extended BIC Model Selection. *arXiv preprint*, arXiv:1606.05771.
- Epskamp, S., Deserno, M. K., & Bringmann, L. F. (2017). mlVAR: Multi-level vector autoregression [Computer software manual]. (R package version 0.4)
- Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika*, *82*(4), 904–927.
- Ferrer, E. (2016). Exploratory approaches for studying social interactions, dynamics, and multivariate processes in psychological science. *Multivariate behavioral research*, *51*(2-3), 240–256.
- Foygel, R., & Drton, M. (2010). Extended Bayesian information criteria for Gaussian graphical models. *Advances in Neural Information Processing Systems*, *23*, 2020–2028. Retrieved from <https://arxiv.org/abs/1011.6640>
- Hamaker, E. L., & Grasman, R. P. (2014). To center or not to center? investigating inertia with a multilevel autoregressive model. *Frontiers in Psychology*, *5*, 1492. doi: 10.3389/fpsyg.2014.01492
- Hoffman, L., & Stawski, R. S. (2009). Persons as contexts: Evaluating between-person and within-person effects in longitudinal analysis. *Research in Human Development*, *6*(2-3), 97–120. doi: 10.1080/15427600902911189
- Kim, C.-J., Nelson, C. R., et al. (1999). *State-space models with regime switching: classical and Gibbs-sampling approaches with applications*. Cambridge, MA, USA: The MIT press.
- Lauritzen, S. L. (1996). *Graphical models*. Oxford, UK: Clarendon Press.
- Meinshausen, N., & Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, *34*(3), 1436–1462. doi: 10.1214/009053606000000281
- van Borkulo, C. D., Borsboom, D., Epskamp, S., Blanken, T. F., Boschloo, L., Schoevers, R. A., & Waldorp, L. J. (2014). A new method for constructing networks from binary data. *Scientific Reports*, *4*(5918), 1–10. doi: 10.1038/srep05918

Appendix A
Simulation results (no rewiring)

Two-step multi-level estimation (mIVAR)

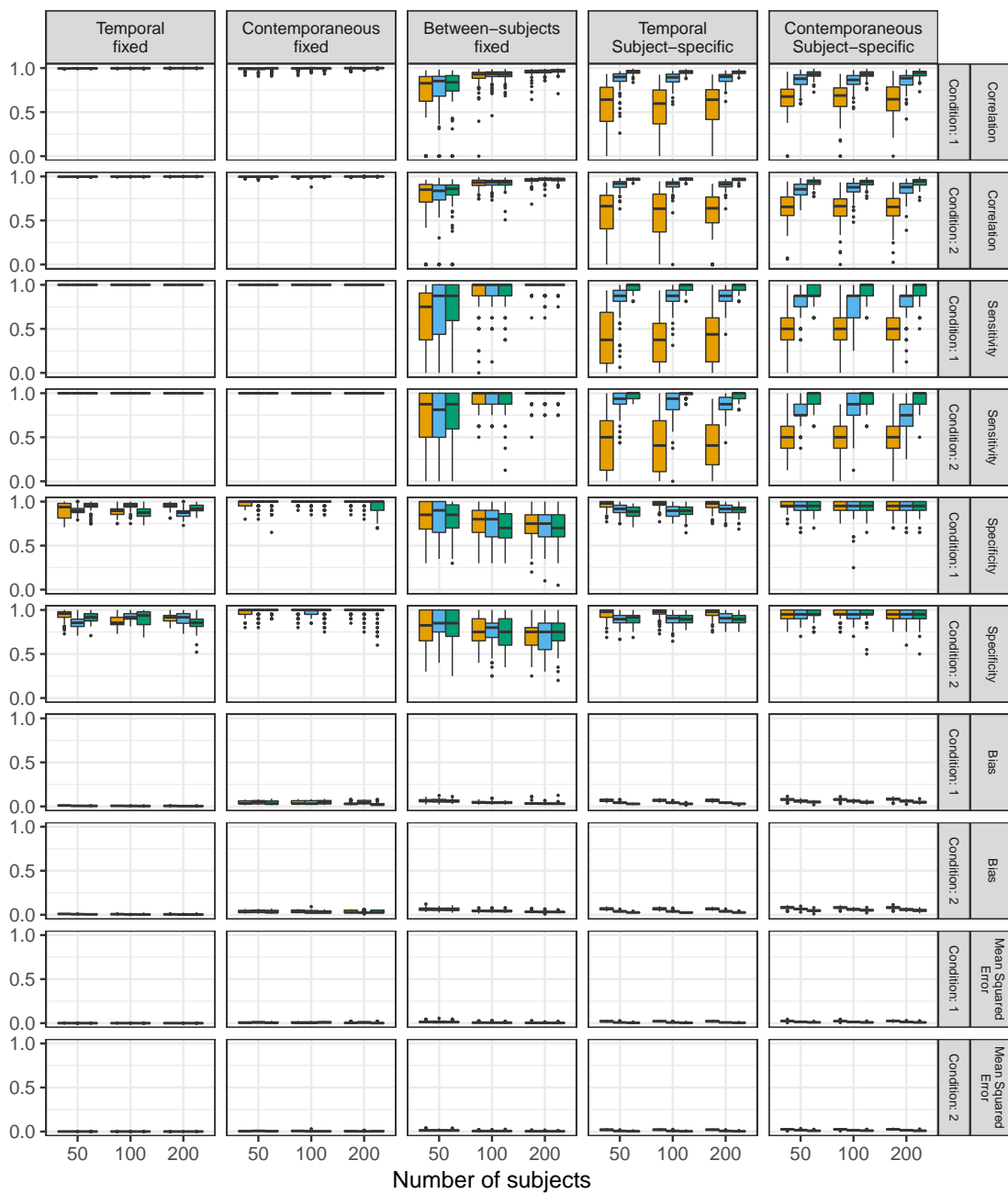
8-node networks, no rewiring (all subjects have same network structure)



Time ■ 50 ■ 100 ■ 200

Pooled and individual LASSO estimation (graphicalVAR)

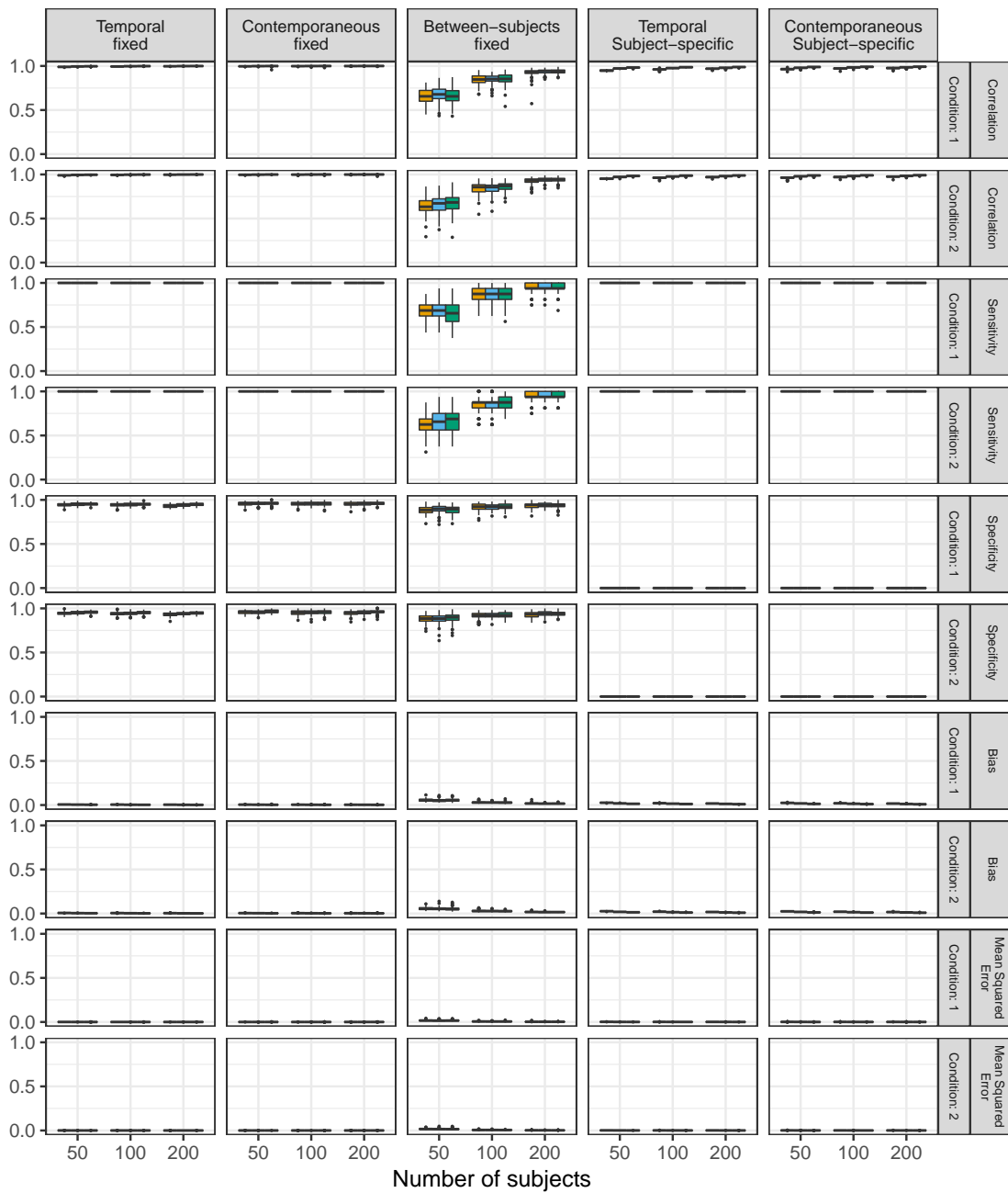
8-node networks, no rewiring (all subjects have same network structure)



Time 50 100 200

Two-step multi-level estimation (mIVAR)

16-node networks, no rewiring (all subjects have same network structure)



Time ■ 50 ■ 100 ■ 200

Pooled and individual LASSO estimation (graphicalVAR)

16-node networks, no rewiring (all subjects have same network structure)



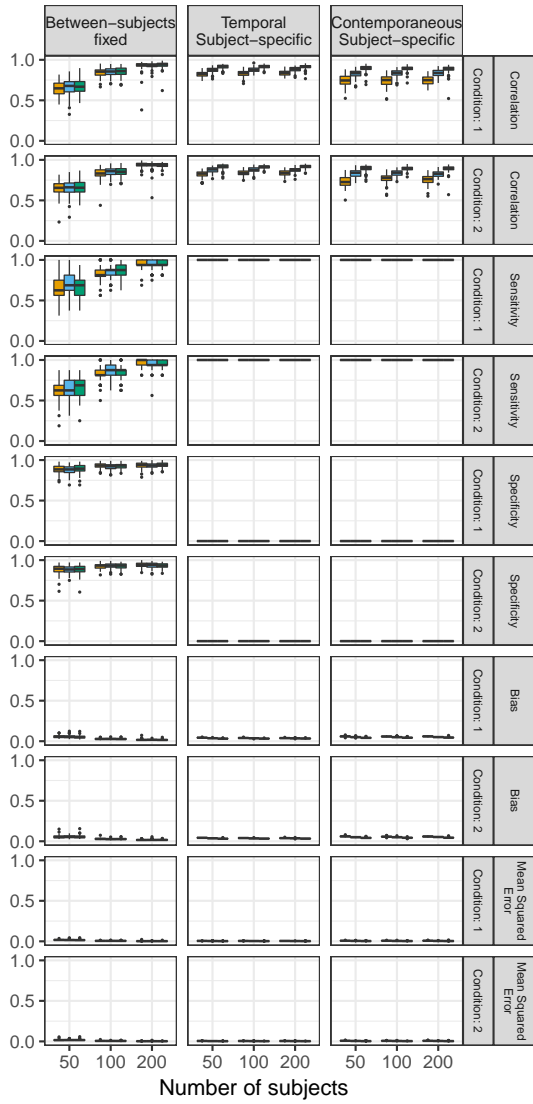
Time ■ 50 ■ 100 ■ 200

Appendix B

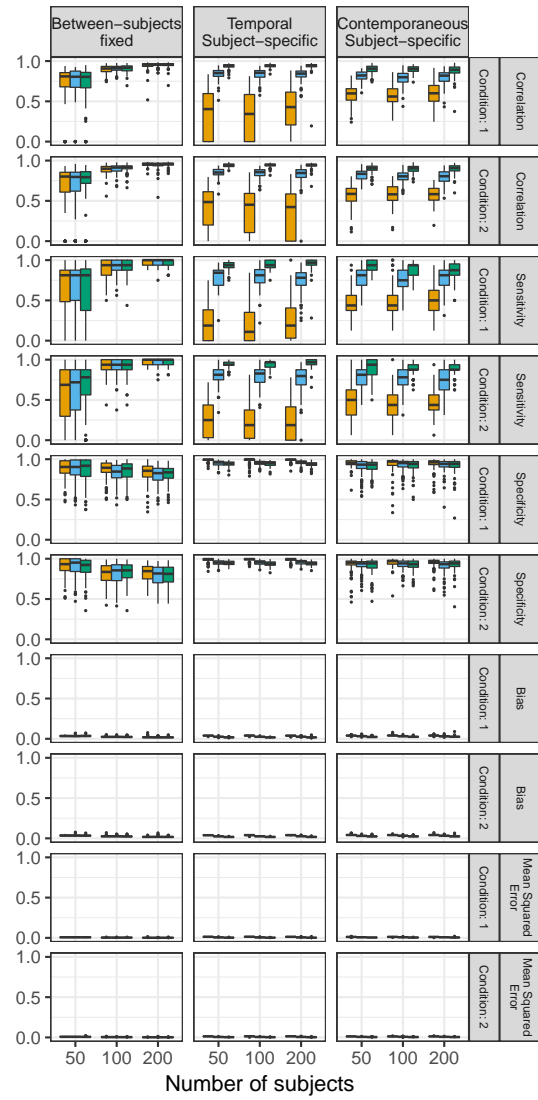
Simulation results (50% rewiring)



mIVAR
16-node networks, 50% rewiring



graphicalVAR
16-node networks, 50% rewiring



Time ■ 50 ■ 100 ■ 200