

# Discovering Psychological Dynamics: The Gaussian Graphical Model in Cross-sectional and Time-series Data

Sacha Epskamp,<sup>1</sup> Lourens J. Waldorp,<sup>1</sup> René Möttus,<sup>2</sup> Denny Borsboom<sup>1</sup>

1. University of Amsterdam: Department of Psychological Methods

2. University of Edinburgh: Department of Psychology

<b>Abstract</b> . . . . .	<b>1</b>
<b>Introduction</b> . . . . .	<b>1</b>
<b>The Gaussian Graphical Model</b> . . . . .	<b>2</b>
The Gaussian Graphical Model and Causal Generating Structures . . . . .	4
The Gaussian Graphical Model and Multiple Regression . . . . .	6
Estimation . . . . .	6
<b>Cross-Sectional Data Analysis</b> . . . . .	<b>7</b>
<b>Temporally Ordered Data</b> . . . . .	<b>8</b>
Structural VAR and graphical VAR . . . . .	10
Estimation . . . . .	11
<b>Temporally Ordered Data of Multiple Subjects</b> . . . . .	<b>12</b>
Estimation . . . . .	13
<b>Empirical Example 1: Reanalysis of Möttus et al. (2017)</b> . . . . .	<b>16</b>
<b>Empirical Example 2: Reanalysis of Bringmann et al. (2013)</b> . . . . .	<b>16</b>
<b>Within- and between-subjects effects</b> . . . . .	<b>20</b>
<b>Discussion</b> . . . . .	<b>22</b>
Limitations and Challenges . . . . .	22
Conclusion . . . . .	23
<b>Acknowledgements</b> . . . . .	<b>23</b>
<b>Supplementary 1: Notation</b> . . . . .	<b>24</b>
<b>Supplementary 2: Two-step multi-level VAR</b> . . . . .	<b>24</b>
Supplementary 2.1: Multi-level modeling . . . . .	24
Supplementary 2.2: Extending multi-level VAR: two-step multi-level VAR . . . . .	25
<b>Supplementary 3: Simulation Studies</b> . . . . .	<b>26</b>
Supplementary 3.1: Simulation Study 1: Cross-sectional Analysis and Repeated Measures . . . . .	26
Supplementary 3.2: Simulation Study 2: Two-step Multi-level VAR and Pooled and Aggregated LASSO Estimation . . . . .	27
<b>Supplementary 4: Structural VAR and Graphical VAR</b> . . . . .	<b>30</b>
<b>Supplementary 5: Stationary distribution</b> . . . . .	<b>30</b>
<b>Appendix A. Glossary of terms</b> . . . . .	<b>31</b>
<b>References</b> . . . . .	<b>32</b>

## Abstract

We discuss the Gaussian graphical model (GGM; an undirected network of partial correlation coefficients) and detail its utility as an exploratory data analysis tool. The GGM can highlight potential causal relationships between observed variables—psychological dynamics—in addition to showing which variables predict one-another. We describe the utility in 3 kinds of psychological datasets: datasets in which consecutive cases are assumed independent (e.g., cross-sectional data), temporally ordered datasets (e.g.,  $n = 1$  time series), and a mixture of the 2 (e.g.,  $n > 1$  time series). In time-series analysis, the GGM can be used to model the residual structure of a vector-autoregression analysis (VAR), also termed *graphical VAR*. Two network models can then be obtained: a temporal network and a contemporaneous network. When analyzing data from multiple subjects, a GGM can also be formed on the covariance structure of stationary means—the between-subjects network. We discuss the interpretation of these models and propose estimation methods to obtain these networks, which we implement in the R packages *graphical-VAR* and *mlVAR*. The methods are showcased in two empirical examples, and simulation studies on these methods are included in the supplementary materials.

## Introduction

There has been a surge of network models being applied to psychological datasets in recent years. This is consistent with a general call to conceptualize observed psychological processes not merely as indicative of latent common causes but rather as emergent behavior of complex, dynamical systems in which psychological, biological, and sociological components directly interact with each other (Borsboom, Cramer, Schmittmann, Epskamp, & Waldorp, 2011; Cramer et al., 2012; Cramer, Waldorp, van der Maas, & Borsboom, 2010; Schmittmann et al., 2013; van der Maas et al., 2006). Such relationships are typically not known, and probabilistic network models (Koller & Friedman, 2009) are used to explore potential causal relationships between observables (Epskamp, Maris, Waldorp, & Borsboom, in press; van Borkulo et al., 2014)—the dynamics of psychology. In this paper we aim to provide a methodological introduction to a powerful probabilistic network model applicable in exploratory data analysis, the Gaussian graphical model (GGM), and to propose how it can be used and interpreted in the analysis of time-series data.

**Two lines of network research in psychology.** We can currently distinguish two distinct and mostly separate lines of research in which networks are utilized on psychological datasets: the modeling of cross-sectional data and the modeling of intensive repeated measures in relatively short time frames (e.g., several times per day during several weeks). In cross-sectional modeling, a model is applied to a dataset

in which multiple subjects are measured only once. The most popularly used methods estimate undirected network models, indicating pairwise interactions—so-called pairwise Markov random fields (Epskamp et al., in press; Murphy, 2012). When the data are continuous and assumed normally distributed, the GGM can be estimated. The GGM estimates a network of *partial correlation coefficients*—the correlation between two variables after conditioning on all other variables in the dataset (Epskamp, Borsboom, & Fried, 2017). This model is applied extensively to psychological data (e.g., Cramer et al., 2012; Fried, Epskamp, Nesse, Tuerlinckx, & Borsboom, 2016; Isvoranu et al., 2017; Kossakowski et al., 2015; McNally et al., 2015; van Borkulo et al., 2015).

Researchers can obtain time-series data by using the experience sampling method (ESM; Myin-Germeys et al., 2009), in which subjects are asked several times per day to fill out a short questionnaire using a device or smartphone app. Often in ESM data, repeated measures of one or multiple participants are modeled through the use of (multilevel) vector autoregressive (VAR) models, which estimate how well each variable predicts the measured variables at the next time point (Borsboom & Cramer, 2013). These models are increasingly popular in assessing intraindividual dynamical structures (e.g., Bringmann et al., 2013; Bringmann, Lemmens, Huibers, Borsboom, & Tuerlinckx, 2015; Wigman et al., 2015). Estimating the GGM is not limited to cross-sectional data; the model merely does not take temporal information into account. As such, the lines of research on network modeling of cross-sectional data and time-series data can naturally be combined. First, GGM models can readily be estimated on repeated measures, if these can be assumed to be temporally independent. Second, as the VAR model can be seen as a generalization of the GGM that takes violations of independence between consecutive cases into account; the GGM can be used to model the contemporaneous time level of a time-series analysis. Finally, the between-subjects effects of  $n > 1$  studies can also be modeled through the use of the GGM.

**Outline.** We show that in time-series modeling the GGM allows researchers to extend the modeling framework to incorporate contemporaneous and between-subjects effects. We do this by building up the model complexity in three steps: (1) when cases can be assumed to be independent (e.g., cross-sectional data or repeated measures in which no auto-regression is assumed), (2) temporally ordered data (e.g.,  $n = 1$  time-series data or  $n > 1$  time-series data where no individual differences are assumed), and (3) temporally ordered data from multiple subjects (e.g.,  $n > 1$  time series). The final level of model complexity leads to a novel contribution of this paper: separation of variance into *contemporaneous*, *temporal*, and *between-subjects* network structures. We propose novel estimation procedures to estimate these models, which we have implemented in two free software pack-

ages: *mlVAR*,<sup>1</sup> and *graphicalVAR*.<sup>2</sup> We furthermore expand on existing literature by providing a comprehensive methodological discussion of the GGM, by comparing the GGM to approaches of causal modeling in time-series analysis, by providing overviews of estimation methods and software packages useable in each kind of dataset and by discussing the interpretation of networks estimated at the contemporaneous and between-subjects levels. We showcase network models estimated from  $n > 1$  time-series data in two empirical examples by reanalyzing existing datasets (Bringmann et al., 2013; Geschwind, Peeters, Drukker, van Os, & Wichers, 2011; Möttus, Epskamp, & Francis, 2017). In the supplementary materials, we provide codes to perform the analyses and we assess the performance of these methods in large-scale simulation studies. To aid the reader in the various different terms used in this paper, we have included a glossary of terms in Appendix A.

### The Gaussian Graphical Model

Let  $\mathbf{y}_C^\top = [Y_{C1} \ Y_{C2} \ \dots \ Y_{Cm}]$  denote a random vector with  $\mathbf{y}_C$  as its realization.<sup>3</sup> We assume  $\mathbf{y}_C$  is centered and normally distributed with some variance–covariance matrix  $\Sigma$ :

$$\mathbf{y}_C \sim N(\mathbf{0}, \Sigma). \quad (1)$$

The subscript  $C$  denotes a case (a row in the spreadsheet). We currently do not define the nature of the observed variables. Thus,  $\mathbf{y}_C$  can consist of variables relating to one or more subjects, could contain repeated measures on one or more variables, could contain variables of a single subject that do not vary within-subject, and so forth. Consider three examples: (1)  $Y_1$  could represent the level of anxiety of subject  $p$  on day 1 and  $Y_2$  the level of anxiety of subject  $p$  on day 2, (2)  $Y_1$  could represent the length of subject  $p$  and  $Y_2$  the number of times subject  $p$  bumps his or her head, and (3)  $Y_1$  could represent the number of cigarettes subject  $p$  smokes per day and  $Y_2$  the number of cigarettes another subject  $p + 1$  smokes per day (case  $C$  then represents a dyadic pair).

**Partial correlation networks.** Assuming multivariate normality,  $\Sigma$  encodes all the information necessary to determine how the observed measures relate to one another. However, we will not focus on  $\Sigma$  in this paper but rather on its inverse—the *precision matrix*  $\mathbf{K}$ ,

$$\mathbf{K} = \Sigma^{-1}.$$

Of particular importance is that the precision matrix can be standardized to encode partial correlation coefficients of two variables, given all other variables (dropping subscript  $C$  for notational clarity; Lauritzen, 1996):<sup>4</sup>

$$\text{Cor}(Y_i, Y_j | \mathbf{y}_{-(i,j)}) = -\frac{\kappa_{ij}}{\sqrt{\kappa_{ii}} \sqrt{\kappa_{jj}}}, \quad (2)$$

in which  $\kappa_{ij}$  denotes an element of  $\mathbf{K}$ , and  $\mathbf{y}_{-(i,j)}$  denotes the set of variables without  $i$  and  $j$ . These partial correlations

can be graphically displayed in a weighted network, in which each variable  $Y_i$  is represented as a node, and connections (edges) between these nodes represent the partial correlation between two variables. When the partial correlation (thus the corresponding element in  $\mathbf{K}$ ) equals zero, no edge is drawn. Thus, modeling the inverse variance–covariance matrix, such that every nonzero element is treated as a freely estimated parameter, allows for a sparse model for  $\Sigma$  (i.e., every element in  $\Sigma$  may be nonzero while some elements in  $\mathbf{K}$  are zero; Epskamp, Rhemtulla, & Borsboom, 2017). Such a model is termed a GGM (Lauritzen, 1996).

When drawing a GGM as a network (often termed a *partial correlation network*, positive partial correlations are typically visualized with blue or green edges and negative partial correlations with red edges,<sup>5</sup> and the absolute strength of a partial correlation is represented by the width and saturation of an edge (Epskamp et al., 2012). When a partial correlation is zero, we draw no edge between two nodes. As such, the GGM can be seen as a network model of conditional associations; no edge indicates that two variables are independent after conditioning on all other variables in the dataset. This allows us to model conditional associations, which we might expect to be zero, rather than marginal associations, which we rarely expect to be zero (Meehl, 1990).

To exemplify the above, suppose for three variables the true variance–covariance matrix is:

$$\Sigma = \begin{bmatrix} 1 & -0.26 & 0.31 \\ -0.26 & 1 & -0.08 \\ 0.31 & -0.08 & 1 \end{bmatrix}.$$

To estimate this matrix, we need six parameters (three covariances and three variances). The corresponding true precision

<sup>1</sup>CRAN link: <http://cran.r-project.org/package=mlVAR>  
Github link (developmental): <http://www.github.com/SachaEpskamp/mlVAR>.

<sup>2</sup>CRAN link: <http://cran.r-project.org/package=graphicalVAR>  
Github link (developmental): <http://www.github.com/SachaEpskamp/graphicalVAR>.

<sup>3</sup>We use capitalized subscripts to denote random variables and lower case subscripts to denote fixed variables. A variable can potentially be fixed with respect to one subscript but random with respect to another. Supplementary materials section 1 contains a complete overview of the notation used in this paper.

<sup>4</sup>This relationship can be traced back much further. For example, Heiser (2017) traced this relationship back to the work of Guttman et al. (1938).

<sup>5</sup>Many publications make use of the default color setup used in *qgraph* (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012): green for positive edges and red for negative edges. A later version of *qgraph* includes the option `theme = "colorblind"` using a more colorblind-friendly coloring scheme and setting the positive edge color to blue. This option has been used for all graphs in this paper. Note that some publications (e.g., Schuurman, 2016) also use blue and red edges but use red to denote positive and blue to denote negative effects akin to a heat map.

matrix becomes:

$$\mathbf{K} = \mathbf{\Sigma}^{-1} = \begin{bmatrix} 1.18 & 0.28 & -0.34 \\ 0.28 & 1.07 & 0 \\ -0.34 & 0 & 1.11 \end{bmatrix}.$$

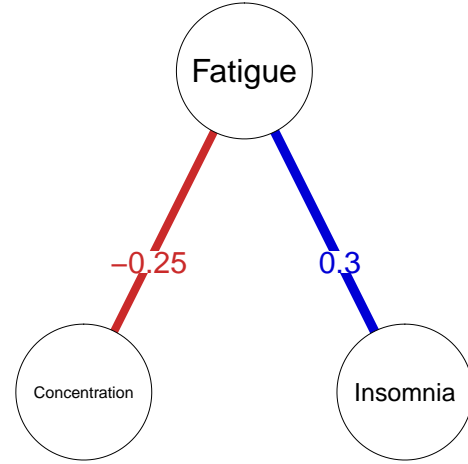
Similar to structural equation modeling (SEM; Kaplan, 2000; Wright, 1921), a model can be devised that perfectly explains this pattern using only five parameters, because one of the elements in  $\mathbf{K}$  can be constrained to be zero (Epskamp, Rhemtulla, & Borsboom, 2017). We can now standardize this matrix and make the off-diagonal elements negative (Equation (2)) to obtain the partial correlation matrix, which we will denote  $\mathbf{R}$ :

$$\mathbf{R} = \begin{bmatrix} 1 & -0.25 & 0.3 \\ -0.25 & 1 & 0 \\ 0.3 & 0 & 1 \end{bmatrix}.$$

This matrix can be used to draw a network as is shown in Figure 1. This figure shows that someone who is tired is also more likely to suffer from concentration problems and insomnia. Furthermore, this network shows that concentration problems and insomnia are conditionally independent given the level of fatigue.

**Interpreting GGMs.** In SEM, a causal model can be tested through specifying a sparse directed network, typically leading to nonzero expected elements in  $\mathbf{\Sigma}$  with positive degrees of freedom. Because elements in the GGM can be zero while allowing for all elements of  $\mathbf{\Sigma}$  to be nonzero, the GGM offers a similar way of modeling  $\mathbf{\Sigma}$  in a sparse manner, but, in contrast to SEM, now by using undirected networks that do not rely on a strong causal theory on the direction of the effects (Epskamp, Rhemtulla, & Borsboom, 2017). This paper concerns the exploratory estimation of such models from data, without prior knowledge on the model structure. The use of undirected network models, rather than directed network models, in such an exploratory setting helps us to discover psychological dynamics in several ways:

1. The GGM is well identified and does not feature equivalent models. Many different directed causal models and an infinite number of latent variable models can lead to one exact GGM. Therefore, exploratory search algorithms perform well in identifying a GGM model without prior knowledge (e.g., latent variables) or certain assumptions (e.g., acyclicity).
2. Causal effects between two variables lead to an edge in the GGM. Furthermore, assuming all interacting variables are observed without measurement error (both unobserved variables as well as measurement error can induce spurious edges), the absence of a causal effect between two variables leads to no edge in the GGM, except when both variables jointly cause a third variable in the model. Therefore, an edge in the GGM can be seen as indicative of potential causal pathways.



*Figure 1.* A hypothetical example of a GGM on psychological variables. Nodes represent someone's ability to concentrate, someone's level of fatigue, and someone's level of insomnia. Connections between the nodes, termed *edges*, represent partial correlation coefficients between two variables after conditioning on the third. Blue edges indicate positive partial correlations, red edges indicate negative partial correlations, and the width and saturation of an edge corresponds to the absolute value of the partial correlation.

3. When an edge is spurious, for example, due to the presence of an unobserved cause, the undirected network does not feature a troublesome directional interpretation (Baumert et al., in press). Instead, the presence of common-cause structures in a dataset is represented by clusters (a set of nodes that are all connected to one another) in the GGM.
4. Unlike directed causal models, which strongly rely on the causal interpretation, the GGM can also be interpreted without any causal interpretation and used merely as a tool to show which variables predict one another. Interpreting the parameters associated with the model  $A \rightarrow B \rightarrow C$  requires a causal interpretation, while the predictive quality between these nodes can directly be obtained from the equivalent GGM  $A - B - C$ . As such, the GGM can always be interpreted to show predictive effects and offers a powerful exploratory tool to gain insight into potential mediation paths and to map out multicollinearity.
5. Undirected network models also allow for a causal interpretation, one of genuine symmetric effects, and

have been used thusly in diverse scientific fields such as statistical physics (Murphy, 2012; Epskamp et al., in press).

We outline these points in more detail below. First, we discuss Points 1 to 3 by exploring the relationship between the GGM and causal generating mechanisms. Next, we discuss Point 4 by showing how partial correlation coefficients correspond to multiple regression coefficients. Point 5 follows from observing that the GGM is directly related to similar undirected models such as the Ising model (Ising, 1925). A discussion on the causal interpretation of such models is beyond the scope of this paper, and we refer the reader for this topic to Epskamp et al. (in press) and van Borkulo et al. (2014).

### The Gaussian Graphical Model and Causal Generating Structures

Let  $\eta_C$  represent a set of unobserved variables, which we assume to be jointly normally distributed with  $y_C$ . Then, we can form an encompassing framework for several possible causal generating models:

$$\begin{aligned} y_C &= B y_C + \Lambda \eta_C + \varepsilon_C \\ \varepsilon_C &\sim N(\mathbf{0}, \Theta), \end{aligned} \quad (3)$$

in which  $\Theta$  is a diagonal matrix, indicating that after conditioning on all causes the variables are independent,  $B$  is a square matrix with zeros on the diagonal of causal effects between observed variables, and  $\Lambda$  is a factor-loading matrix. The variance–covariance matrix of  $\eta_C$  may in turn be modeled in various ways to achieve complicated model setups. The expression above is well-known in SEM, which allows for confirmatory testing of causal models.

**Problems with estimating causal models.** Suppose there are no unobserved causes to any of the variables in  $y_C$ , and the variables in  $y_C$  are only caused by other variables in  $y_C$ . The corresponding model for  $\Sigma$  becomes

$$\Sigma = (I - B)^{-1} \Theta (I - B)^{-1\top}. \quad (4)$$

In this expression,  $B$  can now be seen to encode the causal model (Pearl, 2000). Table 1 summarizes the comparison between such causal models and GGMs. Although useful for generating data, we can immediately see several problems in exploratory estimation of  $B$  without any prior knowledge. First, if  $m$  variables are included,  $\Sigma$  contains  $m(m+1)/2$  elements, while  $\Theta$  contains  $m$  parameters and  $B$  contains  $m(m-1)$  parameters. As a result, the model above is under-identified without stringent restrictions on  $B$ . One assumption is that  $y_C$  can be ordered such that  $B$  is lower triangular, indicating that if this matrix is used to draw a directed graph—a graph in which  $A \rightarrow B$  indicates that  $A$  causes  $B$ —that graph does not contain any cycles. Such a graph is

called a directed acyclic graph (DAG; Kalisch & Bühlmann, 2007; Pearl, 2000), meaning that directed edges cannot be traced from any node back to itself (e.g.,  $A \rightarrow B \rightarrow A$ ). Although cycles can be identified when exogenous variables are present (such as the weather, time, or, depending on the modeling framework, lagged variables; Rigdon, 1995), interpreting such cycles is still not without problems (Hayduk, 2009). Several software packages exist that aim to find such a DAG (e.g., *pcalg*, Kalisch, Mächler, Colombo, Maathuis, & Bühlmann, 2012; *bnlearn*, Scutari, 2010). However, the assumption of acyclicity is debatable in the context of psychological variables (Schmittmann et al., 2013) because many effects can be plausibly assumed cyclic (e.g., fatigue  $\rightarrow$  concentration problems  $\rightarrow$  stress  $\rightarrow$  fatigue).

Second, the same structure for  $\Sigma$  can be obtained under various different specifications of  $B$ . Thus, many equivalent models can lead to exactly the same fit. This can be seen because several matrix decompositions of  $\Sigma$ , such as a Cholesky decomposition or an eigendecomposition, can be used to produce equivalently fitting  $B$ . The problem of equivalent models is also well-known in the literature on directed networks and SEM (MacCallum, Wegener, Uchino, & Fabrigar, 1993; Pearl, 2000). For example, the following three causal models are not statistically distinguishable:

1. Concentration  $\rightarrow$  Fatigue  $\rightarrow$  Insomnia
2. Concentration  $\leftarrow$  Fatigue  $\rightarrow$  Insomnia
3. Concentration  $\leftarrow$  Fatigue  $\leftarrow$  Insomnia

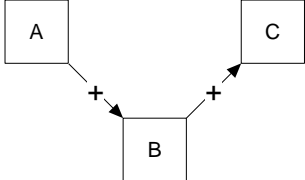
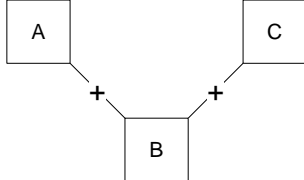
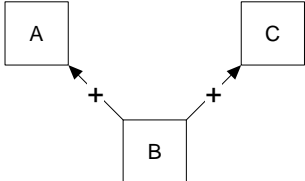
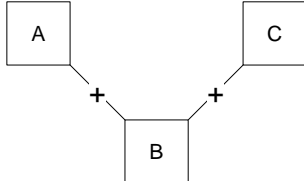
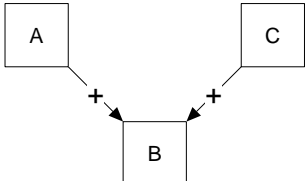
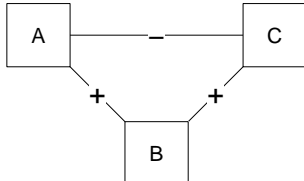
All three models only imply that concentration and insomnia are conditionally independent given fatigue. With more variables, the number of potential equivalent models increases drastically, making it evident that model search is likely to fail. At best, exploratory estimation can result in a set of equally plausible DAGs (an *equivalence class*; Drton & Maathuis, 2017), each differently parameterized and each leading to different strong causal hypotheses.

Third, exploratory search of causal networks critically relies on the assumption that no latent variables caused some of the covariation in the dataset. For example, suppose a single common cause was responsible for all covariation between observed variables. Then, even when conditioning on all other observed variables two variables in the dataset would still feature a correlation; thus, the optimal fitting equivalence class would include every possibly fully connected DAG (which is a saturated model). Every one of these fully connected DAGs comes with strong causal hypotheses; all of them are wrong.

**Causal modeling and the GGM.** We consider equivalent models, poorly identified directions of effect (potentially leading to the wrong direction to be estimated), and the reliance on no latent variables to be considerable downsides to

Table 1

Overview of Causal Models (Directed Networks) and Gaussian Graphical Models (Undirected Networks)

	Causal model	Gaussian graphical model
$\Sigma^{-1} =$	$(\mathbf{I} - \mathbf{B})^{\top} \Theta^{-1} (\mathbf{I} - \mathbf{B}) =$	$\mathbf{K}$
$A \perp\!\!\!\perp C \mid B$		
$A \perp\!\!\!\perp C \mid B$		
$A \not\perp\!\!\!\perp C \mid B$		
R packages (confirmatory)	Any SEM package	<i>lvnet</i> (fit measures); <i>qgraph</i> (fit measures); <i>ggm</i> (estimation only); <i>glasso</i> (estimation only)
R packages (exploratory)	<i>pcalg</i> ; <i>bnlearn</i>	<i>qgraph</i> (EBICglasso function); <i>glasso</i> (no automatic tuning parameter selection); <i>huge</i> ; <i>parcor</i> ; <i>BDgraph</i> ; <i>lvnet</i> (for GGM at latent or residual level of SEM)
Pros	Causal interpretation; allows for confirmatory testing of causal hypotheses	No equivalent models; fast structure and parameter estimation using LASSO; edges parametrizable as partial correlation coefficients; edges interpretable as predictive effects; latent variables result in clusters; edges can be indicative of potential causal effects
Cons	Exploratory estimation requires assumption of acyclicity; many equivalent models; direction of effect poorly or not identified; strongly depends on assumption of no latent variables	No direction of effect; common effect structure can induce spurious edge; LASSO estimation assumes true model is sparse

using causal networks or DAGs in the discovery of potential psychological dynamics, especially given that the added direction of effect comes with a strong causal hypothesis on what would happen under intervention. To this end, the undirected GGM offers an attractive alternative to exploratory DAG estimation: the GGM is saturated rather than overidentified if all edges are present, does not feature equivalent models, does not suffer from a questionable direction of causal effect, does not require the assumption of acyclicity, and is easily parameterized using partial correlation coefficients (Epskamp, Rhemtulla, & Borsboom, 2017). In the presence of latent variables, the GGM also shows a fully connected cluster of nodes. However, edges in this cluster do not impose causal direction. Thus, the GGM gives insight into potential latent variables through its clustering, naturally complementing exploratory factor modeling (Golino & Epskamp, 2017).

To investigate the structure of a GGM under the causal model of Equation (4), in which observed variables can only be caused by other observed variables, we can invert that expression to obtain

$$\mathbf{K} = (\mathbf{I} - \mathbf{B})^\top \mathbf{\Theta}^{-1} (\mathbf{I} - \mathbf{B}), \quad (5)$$

in which  $\mathbf{\Theta}^{-1}$  is still a diagonal matrix. It becomes evident that there is no longer a matrix inversion needed and that the sparsity in  $\mathbf{B}$  directly corresponds to the sparsity in  $\mathbf{K}$ . Unlike marginal covariances, the GGM thus acts on the same level as causal modeling. We can derive that  $\kappa_{ij}$  equals zero if there is no directed edge between node  $i$  and  $j$  (e.g.,  $Y_i \rightarrow Y_j$  or  $Y_i \leftarrow Y_j$ ) and if there is no common effect of node  $i$  and node  $j$  (e.g.,  $Y_i \rightarrow Y_k \leftarrow Y_j$ ; Koller & Friedman, 2009). Thus, assuming a causal model as in Equation (4) generated the data, an edge in a GGM emerges as a result of a causal effect or due to a common effect. Links in the GGM can therefore be indicative of potential causal effects. In the case of a spurious edge due to a common effect, the sign of the edge can be informative: two positive causal effects from two variables on a third lead to a negative partial correlation. Important to note is that the GGM does not estimate a *skeleton graph*, a causal network with arrowheads removed, as the common effect structure leads to induced edges in the GGM. Such networks can be estimated (e.g., Kalisch, Maechler, & Colombo, 2017), but are not parameterized and rely on many separate conditional independence tests, potentially leading to power issues.

## The Gaussian Graphical Model and Multiple Regression

An edge in a GGM indicates that one node predicts a connected node after controlling for all other nodes in the network. This can also be shown in the relationship between coefficients obtained from least-squares prediction and the inverse variance–covariance matrix. Let  $\mathbf{\Gamma}$  represent an  $k \times k$  matrix with zeros on the diagonal. Each row of  $\mathbf{\Gamma}$ , without the

diagonal element  $\mathbf{\Gamma}_{i,-(i)}$ , contains the regression coefficients obtained in a multiple regression model:<sup>6</sup>

$$y_{ci} = \tau + \mathbf{\Gamma}_{i,-(i)} \mathbf{y}_{c,-(i)} + \varepsilon_{ci}.$$

As such,  $\gamma_{ij}$  encodes how well the  $j$ th variable predicts the  $i$ th variable. This predictive effect is naturally symmetric; if knowing someone’s level of insomnia predicts his or her level of fatigue, then conversely knowing someone’s level of fatigue allows us to predict his or her level of insomnia. As a result,  $\gamma_{ij}$  is proportional to  $\gamma_{ji}$ . There is a direct relationship between these regression coefficients and the inverse variance–covariance matrix (Meinshausen & Bühlmann, 2006). Let  $\mathbf{D}$  denote a diagonal matrix on which the  $i$ th diagonal element is the inverse of the  $i$ th residual variance:  $d_{ii} = 1/\text{Var}(\varepsilon_{ci})$ . As a result, it can be shown (Pourahmadi, 2011) that<sup>7</sup>

$$\mathbf{K} = \mathbf{D}(\mathbf{I} - \mathbf{\Gamma}). \quad (6)$$

Thus,  $\kappa_{ij}$  is proportional to  $\gamma_{ij}$ ; a zero in the inverse variance–covariance matrix indicates that one variable does not predict another. Consequently, the network tells us something about the extent to which variables predict each other. In addition to aiding the interpretation of GGM models, this relationship between multiple regression and undirected network edges plays a crucial role in many network estimation procedures (Meinshausen & Bühlmann, 2006; van Borkulo et al., 2014; Haslbeck & Waldorp, 2016b), including the methods discussed below in this paper.

## Estimation

A GGM can be estimated in datasets where cases can be assumed to be independent. Two common examples of such data are cross-sectional data, in which every subject is only measured once on a set of response items, or  $n = 1$  time-series data that feature large intervals between measurement occasions. In time-series data featuring shorter intervals, a GGM can be estimated as well; in this case, the network could be termed a contemporaneous network. However, as we argue in the next section on temporally ordered data, better methods exist that take temporal information into account in addition to modeling the contemporaneous effects in a GGM. In cross-sectional data analysis, only one observation per subject is available; thus, we cannot expect to estimate

<sup>6</sup>This expression should not be confused with Equation 3, which we used as generating model. Here, we do not assume error terms are independent and obtain  $\mathbf{\Gamma}_{i,-(i)}$  by univariate multiple regressions.

<sup>7</sup>This expression may differ by a scalar, depending on the estimation method. For example, by default R computes the variance–covariance matrix by using  $n - 1$  in the denominator, but computes  $\text{Var}(\varepsilon_{ci})$  by using  $n - m$  in the denominator. This denominator is cancelled out in Equation (2) when standardizing to partial correlation coefficients.

subject-specific means or GGM networks. It is typically assumed that the subjects all share the same distribution. That is,

$$\mathbf{y}_P \sim N(\mathbf{0}, \mathbf{\Sigma}),$$

in which  $\mathbf{y}_P$  denotes the random response of subject  $P$  on all items. Similarly, in  $n = 1$  time-series data we can make a similar assumption:

$$\mathbf{y}_T \sim N(\mathbf{0}, \mathbf{\Sigma}),$$

in which  $\mathbf{y}_T$  denotes the random response of a subject on all items at time point  $T$ . In both cases, the full likelihood can be readily obtained, and the variance–covariance matrix  $\mathbf{\Sigma}$  can reliably be estimated using maximum likelihood estimation (MLE), least-squares estimation, or Bayesian estimation.

**Regularization.** The MLE solution of  $\mathbf{K}$ —the precision matrix encoding a GGM—can be obtained by standardizing the inverse sample variance–covariance as per Equation (2). To obtain a sparse network (i.e.,  $\mathbf{K}$  contains zeros), model search can be performed by iteratively adding and removing edges and fitting confirmatory GGMs. In recent literature, it has become increasingly popular to use regularization techniques, such as penalized MLE, to jointly estimate model structure and parameter values (Costantini et al., 2015; van Borkulo et al., 2014). The *least absolute shrinkage and selection operator* (LASSO; Tibshirani, 1996) has been shown to perform well in quickly estimating model structure and parameter estimates of a sparse GGM (Friedman, Hastie, & Tibshirani, 2008; Meinshausen & Bühlmann, 2006; Yuan & Lin, 2007). A particularly popular variant of LASSO is the *graphical LASSO* (glasso; Friedman et al., 2008), which directly penalizes elements of the inverse variance–covariance matrix (Witten, Friedman, & Simon, 2011; Yuan & Lin, 2007). The glasso algorithm is useful as it is typically faster than other GGM estimation algorithms (which conduct multiple separate regressions and then combine the results using Equation 6), and requires only an estimate of the variance–covariance matrix rather than raw data (Epskamp & Fried, 2017). LASSO utilizes a tuning parameter which can be chosen in a way that optimizes cross-validated prediction accuracy or that minimizes information criteria such as the extended Bayesian information criterion (EBIC; Chen & Chen, 2008). Estimating a GGM with the glasso algorithm in combination with EBIC model selection has been shown to work well in retrieving the true network structure (Epskamp, 2016; Foygel & Drton, 2010). For an introduction to this methodology aimed at empirical researchers, we refer the reader to Epskamp and Fried (2017).

**Software.** Several software packages allow for GGM estimation as described above. MLE can be performed in any programming language and in many statistical programs by inverting and subsequently standardizing the sample variance–covariance matrix. In the open-source statistical programming language R (R Core Team, 2017), au-

tomated procedures have been implemented in the *corpcor* package (Schafer et al., 2017) and the *qgraph* (Epskamp et al., 2012) package. The *qgraph* package also supports thresholding via significance testing or false discovery rates. The glasso algorithm is implemented in the *glasso* (Friedman, Hastie, & Tibshirani, 2014) and *huge* (Zhao et al., 2015) packages. EBIC-based tuning parameter selection using the glasso package has been implemented in the *qgraph* package. The *huge* package also allows for selection of the tuning parameter using cross validation or EBIC. The *parcor* package (Krämer, Schäfer, & Boulesteix, 2009) implements other LASSO variants to estimate the GGM. The *BDgraph* package (Mohammadi & Wit, 2015) implements a Bayesian method to estimate the undirected structure. Finally, fitting an estimated GGM to data can be done in the R packages *ggm* (Marchetti, Drton, & Sadeghi, 2015) and *lvnet* (Epskamp, Rhemtulla, & Borsboom, 2017).

## Cross-Sectional Data Analysis

**Within- and between-subjects variation.** A type of data to which the GGM is now often applied is data belonging to multiple subjects that are all measured only once (e.g., Isvoranu et al., 2017; van Borkulo et al., 2015). Such a dataset is often termed cross-sectional data, and such an analysis is often termed a between-subjects analysis. However, the term between-subjects analysis might not be warranted, as it is difficult to distinguish between within-subject variation around an individual’s stable mean and between-subject variation of such stable within-subject means using only cross-sectional data (Hamaker, 2012). It is well known that subjects might respond differently when measured multiple times (Lord, Novick, & Birnbaum, 1968). As such, the single observation per subject leads to the time point and the subject being random:  $\mathbf{y}_{[T,P]}$ . We might make the argument that two distinct sources of variation cause the outcome (Bolger & Laurenceau, 2013). Repeated measures of a subject (here  $p$ ) are distributed according to a unique within-subject model:

$$\mathbf{y}_{[T,p]} \sim N(\boldsymbol{\mu}_p, \boldsymbol{\Theta}_p),$$

That is, of a particular response, the subject’s score is a composite of the average stationary score  $\boldsymbol{\mu}_p$  and random deviation.<sup>8</sup> These average stationary scores also differ in the population. Thus, we need to model the average stationary scores of a random subject  $P$  with a separate distribution:

$$\boldsymbol{\mu}_P \sim N(\mathbf{0}, \mathbf{\Omega}),$$

<sup>8</sup>Section 5 of the supplementary materials show that when consecutive cases ( $t$  and  $t+1$ ) are assumed dependent, such a zero-order network may result from a mixture of temporal and contemporaneous effects. The discussion here does not yet concern estimation of model parameters and hence does not require an assumption of independence of cases.



in which we can assume, without loss of generality, an overall mean of  $\mathbf{0}$ . We can invert the variance–covariance matrix  $\mathbf{\Omega}$  to obtain a GGM:

$$\mathbf{K}^{(\mathbf{\Omega})} = \mathbf{\Omega}^{-1}.$$

We call this a between-subjects network. The matrix  $\mathbf{\Theta}_p$  can also be inverted and standardized to a GGM to obtain a within-subject network:

$$\mathbf{K}_p^{(\mathbf{\Theta})} = \mathbf{\Theta}_p^{-1}.$$

We will term this network a *within-subjects network*.

**The value of a cross-sectional analysis.** It is immediately clear that with only one response per subject we cannot hope to estimate subject-specific variance-covariance matrices  $\mathbf{\Theta}_p$  (and as a result individual GGMs). Moreover, even if we assume that within-subject effects are equal across subjects and drop the subscript  $p$ , this still leaves us without an estimable model because  $\boldsymbol{\mu}$  is also assumed to be normally distributed. The co-variation between responses thus becomes an unidentified blend between  $\mathbf{\Theta}$  and  $\mathbf{\Omega}$ :  $A$  and  $B$  may correlate in cross-sectional data because people who score on average high on  $A$  also score on average high on  $B$  (trait-level variation in  $\mathbf{\Omega}$ ), or because when people deviate from their average on  $A$  they also tend to deviate from their average on  $B$  (state-level variation in  $\mathbf{\Theta}$ ). Even when within- and between-subjects effects are assumed not to correlate, the GGM estimated on such data becomes

$$\mathbf{K} = (\mathbf{\Theta} + \mathbf{\Omega})^{-1},$$

which is not a simple function of the between-subjects GGM and the within-subjects GGM. Only when no short-term within-subject variation,  $\mathbf{\Theta} = \mathbf{O}$ , or no between-subjects variation,  $\mathbf{\Omega} = \mathbf{O}$ , is assumed does the cross-sectional GGM correspond exactly to one of the two networks.

Cross-sectional data analysis thus cannot disentangle between-subjects relationships from short term within-subjects relationships (Hamaker, 2012). For example, cross-sectional analysis cannot distinguish whether or not fatigue and concentration correlate because whenever people feel fatigued they also concentrate poorly (a within-subjects effect) or because people who are on average fatigued also tend to concentrate poorly on average (a between-subjects effect). We conducted a simulation study to assess the performance of a cross-sectional GGM estimation, where we generated data from completely different within- and between-subjects networks. The results are reported in Section 3.1 of the supplementary materials and show that the resulting cross-sectional GGM generally does not contain edges that are not present in either the within- or between-subjects network. Depending on the ratio of within-to-between person variance, the cross-sectional analysis will pick up the within-subject network, the between-subject network, or a mixture

of the two. We furthermore conducted a simulation study with two uncorrelated repeated measures per subject (rather than just one). Now, estimating a GGM on within-subjects centered pooled data adequately captures the within-subjects network (assuming it is the same for every subject). Furthermore, the GGM estimated on the subject-specific means can estimate the between-subjects networks. These simulation studies do rely on the assumption of no auto-regression between the consecutive measurements, which might be tenable or not based on the research design. These simulation studies are the first of its kind investigating cross-sectional and repeated measurements GGM estimation, and should be expanded to include more conditions. We encourage further simulation studies investigating the performance of cross-sectional and repeated measurements analysis.

#### *Cross-sectional analysis as between-subjects analysis.*

An important consideration is that a typical cross-sectional questionnaire or interview is vastly different than a typical ESM questionnaire, and many cross-sectional studies aim to measure variables that are more stable over time and for which a time-series analysis might not make sense. Good examples of this are recent network analyses in the area of schizophrenia (Isvoranu, Borsboom, van Os, & Guloksuz, 2016; Isvoranu et al., 2017), in which the impact of environmental factors (e.g., childhood trauma, urbanization) on psychotic symptoms and general psychopathology was studied. Such variables do not vary much over time; therefore, a cross-sectional analysis seems more suitable here. Other examples are questionnaires asking participants to rate symptoms over a period of several weeks or to describe themselves as “I am a person who...” In such cases, the cross-sectional network can be interpreted as a between-subjects network, of which we discuss the interpretation below.

### **Temporally Ordered Data**

In line with a call for more intraindividual and person-based research (Molenaar, 2004), an increasingly popular form of data pertains to  $n = 1$  time series, in which a single individual is measured repeatedly over a period of time. One such situation is in clinical practice (Kroeze et al., 2017; Epskamp, van Borkulo, et al., 2017), where a patient can be measured several times per day over a period of a few weeks. We will limit our discussion to data obtained in a relatively short time-frame so that we can reasonably assume the model will remain stable over time. Then, we can apply the methodology above to obtain a GGM for the  $n = 1$  data. However, such an analysis does not take temporal ordering of data into account (i.e., relationships between measurement occasions) and only investigates contemporaneous relationships between variables (e.g., within the same measurement occasion). This is important for several reasons. First, valuable information, especially in the context of discovering psychological dynamics, might be contained at the temporal level

rather than at the contemporaneous level. Second, not taking temporal ordering into account might bias the estimated contemporaneous relationships (see Section 5 of the supplementary materials). For example, if one variable causes itself and another variable at the next time, then not taking temporal ordering into account turns that variable into a latent cause, which would produce an edge in the GGM. Third, temporal information is needed when constructing the joint likelihood over time (e.g., to obtain the information retained in a system over time; Epskamp, 2017a; Quax, Kandhai, & Sloot, 2013). Finally, temporal information can aide in distinguishing reciprocal and cyclic effects by regarding these as acyclic effects unfolding over time.

**Vector Auto-regression.** The simplest way to deal with temporal ordering of cases is to incorporate the effect between consecutive measurements (Shumway & Stoffer, 2010; Hamilton, 1994; Chatfield, 2016). This is called a Lag-1 model because it includes both measurements at the current time point  $t$  as well as measurements from the previous time point  $t - 1$ . We will focus our discussion on Lag-1 models, noting that everything below also generalizes to more complicated models. In intraindividual analysis, VAR (Brandt & Williams, 2007; Rosmalen, Wenting, Roest, de Jonge, & Bos, 2012) has gained substantive footing in visualizing temporal information through networks.

We can combine current and lagged measurements in a single vector  $\mathbf{y}_{\{t-1,t\}}^\top = [\mathbf{y}_{t-1}^\top \quad \mathbf{y}_t^\top]$  and form a general modeling framework:

$$\mathbf{y}_{\{t-1,t\}} = \mathbf{B}\mathbf{y}_{\{t-1,t\}} + \boldsymbol{\varepsilon}_{\{t-1,t\}},$$

in which  $\mathbf{y}_{t-1}$  are treated as exogenous. Thus,  $\mathbf{B}$  does not contain any effects towards any of the lagged variables,

$$\mathbf{B} = \begin{bmatrix} \mathbf{O} & \mathbf{O} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix},$$

and  $\boldsymbol{\varepsilon}$  is distributed normally,

$$\boldsymbol{\varepsilon}_{\{T-1,T\}} \sim N(\mathbf{0}, \boldsymbol{\Theta}).$$

We do not set  $\boldsymbol{\Theta}$  to be diagonal but instead model its inverse as a GGM, which we can arbitrarily separate into four blocks:

$$\boldsymbol{\Theta}^{-1} = \mathbf{K}^{(\boldsymbol{\Theta})} = \begin{bmatrix} \mathbf{K}_{11}^{(\boldsymbol{\Theta})} & \mathbf{O} \\ \mathbf{O} & \mathbf{K}_{22}^{(\boldsymbol{\Theta})} \end{bmatrix}.$$

As the inverse of a block-diagonal matrix is similarly block-diagonal, inverting  $\mathbf{K}_{11}^{(\boldsymbol{\Theta})}$  will lead to the variance-covariance matrix of  $\mathbf{y}_{t-1}^\top$  and inverting  $\mathbf{K}_{22}^{(\boldsymbol{\Theta})}$  will lead to the contemporaneous variance-covariance matrix: the residual structure after taking  $\mathbf{B}$  into account (also termed *innovation*). Every generating model from Equation (3) can be represented in this framework, where covariation between exogenous or latent variables can be represented in the variance-covariance

structure of the residuals. Using this framework, we can construct several modeling frameworks that allow us to jointly model temporal and contemporaneous effects in a sparse manner. Such effects can then be summarized in two network models: the temporal network, discussed below, and the contemporaneous network, which is either a directed or undirected network of effects in the same window of measurement.

**Temporal networks.** Temporal networks, encoded by  $\mathbf{B}_{21}$  in both below mentioned modeling frameworks, have grown popular in recent psychological literature (e.g., Bringmann et al., 2013, 2015; Wigman et al., 2015; Bos et al., 2017; Snippe et al., 2017; Klippel et al., 2017). A temporal network is formed by combining a lagged variable  $y_{t-1}$  and current variable  $y_t$  into a single node, connected with directed edges which are weighted according to the regression parameters contained in  $\mathbf{B}_{21}$ .<sup>9</sup> Thus, an edge in the temporal network indicates that a node predicts another node (or itself in the common case of self-loops) at the next measurement occasion, after controlling for all other variables at the previous measurement occasion. Temporal prediction is termed *Granger causality* in the economic literature (Eichler, 2007; Granger, 1969), and it satisfies at least the temporal requirement for causation (i.e., the cause must precede the effect). Temporal networks may thus highlight potential causal pathways. An additional useful property of temporal networks is that they need not be acyclic (the network can be seen as a summary of an acyclic network over time).

**Contemporaneous networks.** In addition to temporal effects, VAR analyses also include contemporaneous effects, which can be modeled in different ways. First, using *structural VAR* (SVAR; Chen et al., 2011), also called *unified SEM* (Gates, Molenaar, Hillary, Ram, & Rovine, 2010), we specify

$$\mathbf{B} = \begin{bmatrix} \mathbf{O} & \mathbf{O} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix},$$

and

$$\boldsymbol{\Theta}^{-1} = \mathbf{K}^{(\boldsymbol{\Theta})} = \begin{bmatrix} \mathbf{K}_{11}^{(\boldsymbol{\Theta})} & \mathbf{O} \\ \mathbf{O} & \mathbf{K}_{22}^{(\boldsymbol{\Theta})} \end{bmatrix},$$

in which we keep  $\mathbf{K}_{11}^{(\boldsymbol{\Theta})}$  saturated and force  $\mathbf{K}_{22}^{(\boldsymbol{\Theta})}$  to be diagonal. The contemporaneous effects may be identified by including all diagonal elements of  $\mathbf{B}_{21}$  (autoregressions). Model selection can be used to structure temporal effects in  $\mathbf{B}_{21}$  and contemporaneous effects in  $\mathbf{B}_{22}$ .

Second, the conditional distribution of  $\mathbf{y}_T$  given  $\mathbf{y}_{T-1} = \mathbf{y}_{t-1}$  can be modeled as a GGM. Doing so is equivalent to modeling the residual structure of a regular VAR model as

<sup>9</sup>Note, in graph theory it is common to encode a network using a *weights matrix* in which the row indicates the node of origin and the column indicates the row of destination. As such, to obtain the directed weights matrix to draw a temporal network  $\mathbf{B}_{21}$  needs to be transposed.

a GGM. We will term this modeling framework (a VAR model with contemporaneous effects explicitly modeled and portrayed as a GGM) *graphical VAR* (GVAR; Wild et al., 2010).<sup>10</sup> We obtain the GVAR model by setting

$$B = \begin{bmatrix} O & O \\ B_{21} & O \end{bmatrix}$$

and

$$\Theta^{-1} = K^{(\Theta)} = \begin{bmatrix} K_{11}^{(\Theta)} & O \\ O & K_{22}^{(\Theta)} \end{bmatrix},$$

in which we keep  $K_{11}^{(\Theta)}$  saturated and perform model selection on temporal effects in  $B_{21}$  and contemporaneous effects in  $K_{22}^{(\Theta)}$ . It can be seen that when  $B_{21} = O$ , the GVAR model is exactly the same as the GGM model described above for independent cases. Thus, the GVAR model can be seen as a generalization of the GGM model to temporally ordered data. GVAR only differs from regular VAR in that the contemporaneous structure is modeled and represented as a GGM, instead of being saturated. This leads to a strikingly different interpretation of the VAR model; the VAR model can be seen as an inclusion of temporal effects on a GGM.

**Temporal and contemporaneous information.** Figure 2 shows a hypothetical example of the two network structures obtained in a GVAR analysis and shows how they might plausibly differ. The left panel shows the temporal network. The self-loop shows that whenever the subject in question felt energetic (or tired) this person also felt more (or less) energetic in the next measurement. The temporal network also shows us that after exercising, this person felt less energetic. The contemporaneous network in the right panel shows a plausible reverse relationship: Whenever this person exercised, he or she felt more energetic in the same measurement occasion. In psychology, there will likely be many causal relationships that occur much faster than the lag interval of a typical ESM study; in which case, these pathways will be captured in the contemporaneous network. For example, if someone is experiencing bodily discomfort, that will immediately negatively affect that person's ability to enjoy him or herself (Epskamp, van Borkulo, et al., 2017). Especially when the measurement is on blocks of time (e.g., "since the last measurement did you feel ..."), such effects are likely to be caught in the contemporaneous network, regardless of if they are modeled using a directed or an undirected network.

### Structural VAR and graphical VAR

Section 4 of the supplementary materials shows that the SVAR and GVAR models can be equated, such that estimating one model can lead to an equivalent conceptualization in the other framework (Lütkepohl, 2005; Beltz & Molenaar, 2016; Molenaar & Lo, 2016). For every SVAR model, one unique GVAR conceptualization exists. Conversely however, for every GVAR model multiple possible SVAR models

would explain the data equally well. Table 2 summarizes this information. At the contemporaneous level, the same equivalence Equation (5) shows between GGMs and DAGs holds between GVAR and SVAR: the sparsity of the contemporaneous model in SVAR directly relates to the sparsity of the contemporaneous model in GVAR. If data are generated via a SVAR model, the corresponding contemporaneous network in the GVAR model will feature an edge whenever there is an edge between two variables in the contemporaneous network of the SVAR model or whenever there is a common effect between two variables in the contemporaneous network of the SVAR model. The temporal structure between GVAR and SVAR differs in that in GVAR mediators at the contemporaneous time level are not taken into account in the estimation of the temporal model. As Table 2 shows, if  $A$  predicts itself over time, and  $A$  predicts  $B$  in the same measurement occasion, then  $A$  predicts  $B$  over time as well. When correctly modeled, such a mediator is controlled for in the temporal network obtained from SVAR but not in the temporal network obtained from GVAR. In sum, when a SVAR model generated the data, spurious connections can appear in the corresponding GVAR model due to (1) a common effect structure at the contemporaneous time level or (2) an unmodelled mediator between a temporal and a contemporaneous effect.

**GVAR as alternative to SVAR.** The GVAR thus contains more edges than SVAR in addition to losing information on the direction of effect at the contemporaneous level. Still, the GVAR model offers an attractive alternative to SVAR modeling. Most importantly, the GVAR model is always well identified and has only one unique optimal solution, whereas the SVAR model is potentially unidentified. As can be derived from the equivalence in Table 2, when a saturated SVAR model is estimated, any matrix decomposition (typically a Cholesky decomposition is used) of the inverse contemporaneous variance-covariance matrix leads to an equivalent SVAR model, whereas the structural model in the GVAR model is directly obtained from the same inverse contemporaneous variance-covariance matrix. Constraining elements in the contemporaneous SVAR network based on prior theory or model search (i.e., removing edges by using unified SEM; Gates et al., 2010) overcomes this problem. When temporal effects are present, lagged variables act as exogenous predictors, making cycles in the contemporaneous model identified (Rigdon, 1995). As a result, the contemporaneous model  $A \rightarrow B$  can be distinguishable from  $B \rightarrow A$ . While autoregressions are to be expected, they are not guaranteed. As such, even when modeled, multiple optimal SVAR models

<sup>10</sup>Wild et al. (2010) do not use the term graphical VAR in the exact same way we do, and use it more to refer to graphical modeling in a VAR framework, including structural VAR. We use the term here as described because having an explicit term helps in contrasting GVAR from, e.g., SVAR.

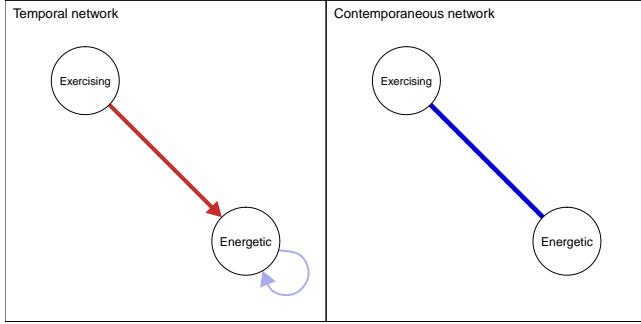


Figure 2. A hypothetical example of two network structures obtained from a GVAR analysis. The network on the left indicates the temporal network, demonstrating that a variable predicts another variable at the next time point. The network on the right indicates the contemporaneous network, demonstrating that two variables predict each other at the same time point.

could exist. In the presence of weak auto-regressions, such competing models might be hard to distinguish statistically. For a principled way of dealing with such multiple solutions, we refer the reader to Beltz and Molenaar (2016).

A second reason is that, like the GGM compared to a DAG, GVAR handles covariation due to outside sources better than SVAR. In both frameworks, an unmodeled outside variable causing covariation at the contemporaneous time level will lead to an edge or cluster of edges. Such a cluster of edges would show up as an interpretable cluster of undirected edges in the GVAR model, and as a cluster of directed effects generating strong incorrect causal hypotheses in the SVAR model.<sup>11</sup>

Finally, although less relevant for the discussion in this paper (as our context is explanation rather than prediction), the temporal model in GVAR is optimal for forecasting future responses. While the structural model at contemporaneous time helps in predicting what would happen under interventions on a time-series, it is not needed in merely forecasting new responses without interventions. As contemporaneous effects are modeled as external unpredictable shocks to the system, the temporal network of the GVAR model is sufficient for forecasting data. As GVAR only differs from regular VAR in that the contemporaneous effects are modeled as partial correlations rather than marginal correlations, the same is true for regular VAR models. It is for this reason that VAR is extensively utilized in prediction.

## Estimation

Estimating saturated (fully connected temporal and contemporaneous networks) SVAR or GVAR models is straightforward. First, one needs to estimate temporal effects of a regular VAR model by performing multivariate multiple regression of all variables on the previous measurement occa-

sion,

$$\mathbf{y}_t = \mathbf{B}_{21}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t,$$

or by estimating univariate models for every variable,

$$y_{ti} = \beta_i \mathbf{y}_{t-1} + \varepsilon_{ti},$$

in which  $\beta_i$  denotes the  $i$ th row of  $\mathbf{B}_{21}$ . Next one can invert the variance-covariance matrix of the residuals to obtain a GVAR model or one can transform this variance-covariance matrix (e.g., by using a Cholesky decomposition on its inverse; Lütkepohl, 2005) to obtain an SVAR model (the temporal effects need be transformed as well after obtaining a contemporaneous model. This technique of obtaining an SVAR model leads to multiple solutions (Beltz & Molenaar, 2016). Step-wise model selection can also be used to estimate sparse SVAR or GVAR models. SVAR models can be estimated by using step-wise model selection in (unified) SEM (Gates et al., 2010) and GVAR models can be estimated by using step-wise model selection in latent network models (a generalization of SEM including GGMs; Epskamp, Rhemtulla, & Borsboom, 2017). Missing data can be handled in default ways of SEM or regression models (e.g., list-wise deletion and full-information maximum likelihood), or by using more sophisticated techniques such as Bayesian estimation (Schuurman, Grasman, & Hamaker, 2016) or the Kalman filter (Harvey, 1990; Kim, Nelson, et al., 1999).

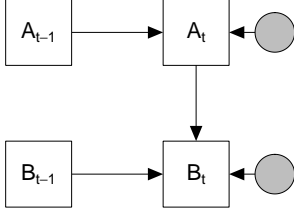
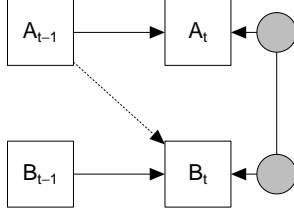
**Novel estimation methods.** A promising recent method for estimating VAR models is the Bayesian dynamical SEM implementation in version 8 of Mplus (Muthén & Muthén, 2017; Asparouhov, Hamaker, & Muthén, 2016), which includes handling of missing data, measurement invariance and latent variables. Mplus can be used to estimate saturated GVAR and SVAR models, and to perform model selection in SVAR models or the temporal network of a GVAR model. Model selection in the contemporaneous network of a GVAR model is not yet implemented, but credibility intervals around contemporaneous effects can be obtained by manually inverting each sampled residual variance-covariance matrix (these can be stored using the BPARAMETERS option).

When estimating GVAR models regularization methods can be used similar to the estimation of GGMs on non-temporally ordered data. Abegaz and Wit (2013) proposed to apply LASSO estimation to jointly estimate the temporal and contemporaneous network structures using the multivariate regression with the covariance estimation (MRCE) algorithm described by Rothman, Levina, and Zhu (2010). MRCE involves iteratively optimizing  $\mathbf{B}_{21}$ , using cyclical-coordinate descent, and  $\mathbf{K}_{22}$ , using the glasso algorithm

<sup>11</sup>It should be noted that if the data is generated by auto-correlated unmodelled common causes, both SVAR and GVAR will feature spurious temporal edges in addition to spurious contemporaneous edges.

Table 2

Overview of Two Different Methods for Modeling Temporal and Contemporaneous Effects in Time-Series Data

	Structural VAR	Graphical VAR
Example		
Model	$\mathbf{K}^{(\Theta)}$ set to be diagonal	$\mathbf{B}_{21} = \mathbf{O}$ and $\mathbf{K}_{22}^{(\Theta)}$ as GGM
Temporal network	$\mathbf{B}_{21}$	$\mathbf{B}_{21}$
Contemporaneous network	$\mathbf{B}_{22}$ (directed)	proportional to $\mathbf{K}_{22}^{(\Theta)}$
Contemporaneous equivalence	$(\mathbf{I} - \mathbf{B}_{22})^\top \mathbf{K}_{22}^{(\Theta)} (\mathbf{I} - \mathbf{B}_{22})$	$\mathbf{K}_{22}^{(\Theta)}$
Temporal equivalence	$(\mathbf{I} - \mathbf{B}_{22})^{-1} \mathbf{B}_{21}$	$\mathbf{B}_{21}$
Pros	Can incorporate lag-0 causal effects; temporal network estimation takes contemporaneous effects into account; cyclic effects are identified with auto-regressions.	Well identified; well defined saturated model; allows sparse modeling of contemporaneous effects; latent variables result in clusters of undirected edges.
Cons	Multiple solutions possible; strong reliance on the assumption of no latent variables; strong interpretation of the direction of effect.	Potentially spurious edges in both the temporal and contemporaneous networks; No direction of effect in contemporaneous network. Limited software for confirmatory estimation.

Note. The dashed line indicates an edge that is present in the GVAR model but not in the corresponding SVAR model.

(Friedman et al., 2008, 2014). EBIC model selection can be used to obtain the best performing model. This methodology has been implemented in two open source R packages: *sparseTSCGM* (Abegaz & Wit, 2015), which aims to estimate the model on repeated multivariate genetic data, and *graphicalVAR* (Epskamp, 2017b), which was designed to estimate the model on the psychological data of a single subject. The *graphicalVAR* package also allows for unregularized multivariate estimation.

### Temporally Ordered Data of Multiple Subjects

A type of data that is increasingly common due to the emergence of ESM studies is time series of multiple subjects (e.g., Bringmann et al., 2013, 2015; Möttus et al., 2017; Schmiedek, Lövdén, & Lindenberger, 2010; Wigman et al., 2015). Such datasets pose a promising gateway to study both intraindividual dynamics and between-subjects overlap as well as their differences. Here, we assume that the number

of time points might differ per person and that measurement occasions are nested in people. We can model the temporal data of every person with an individual GVAR model (dropping subscripts for block matrices used above for notational clarity):

$$\begin{aligned}
 \mathbf{y}_{[t,p]} &= \boldsymbol{\mu}_p + \mathbf{B}_p (\mathbf{y}_{[t-1,p]} - \boldsymbol{\mu}_p) + \boldsymbol{\varepsilon}_{[t,p]} \\
 \boldsymbol{\varepsilon}_{[T,p]} &\sim N(\mathbf{0}, \boldsymbol{\Theta}_p) \\
 \boldsymbol{\Theta}_p^{-1} &= \mathbf{K}_p^{(\Theta)},
 \end{aligned}$$

in which  $\boldsymbol{\mu}_p$  indicates the stationary mean vector of subject  $p$  (which enters the model because we can no longer assume within-subject means are zero without loss of generality),  $\mathbf{B}_p$  encodes the person-specific temporal network, and  $\mathbf{K}_p^{(\Theta)}$  encodes the person-specific contemporaneous GGM.

**Multilevel modeling.** To gain insight in the general network structure over subjects we can investigate the individual networks at a second level. Doing so is termed *multilevel modeling*, explained in more detail in section 2.1 of the

supplementary materials. Let  $\mathbf{B}_*$  and  $\mathbf{K}_*^{(\Theta)}$  encode the expected temporal and contemporaneous network when selecting a person at random. Furthermore, we can assume without loss of generality that data are grand-mean centered. We then obtain

$$\begin{aligned}\mathcal{E}(\boldsymbol{\mu}_p) &= \mathbf{0} \\ \mathcal{E}(\mathbf{B}_p) &= \mathbf{B}_* \\ \mathcal{E}(\mathbf{K}_p^{(\Theta)}) &= \mathbf{K}_*^{(\Theta)}.\end{aligned}$$

Here,  $\mathbf{B}_*$  and  $\mathbf{K}_*^{(\Theta)}$  now encode the average parameters in the population: the *fixed effects*. Deviations from these fixed effects, such as  $\mathbf{B}_p - \mathbf{B}_*$ , are often called *random effects*. Besides the individual network structures, researchers often aim to estimate the structure and parameters of these fixed effects because these tell us something about the average intraindividual effect. Researchers also aim to estimate the variance–covariance structure of the random effects because it tells us something about individual differences (Bringmann et al., 2013).

The random effects can be modeled by assuming a second level normal distribution on all the parameters. This can be complicated, however, especially when modeling partial correlation coefficients in such a way (e.g., any hierarchical model for  $\mathbf{K}^{(\Theta)}$  needs to take into account that this matrix must remain positive definite). The interpretation of, for example, correlations between different temporal or contemporaneous edges is also difficult. Therefore, we only focus here on a subset of the parameters where we can easily interpret the second-level model: the mean structure. As a result, if a multivariate normal is assumed for all parameters, then it is also assumed for the marginal distribution of the means—regardless of other parameters:

$$\boldsymbol{\mu}_p \sim N(\mathbf{0}, \boldsymbol{\Omega}).$$

Again, we can invert the variance–covariance matrix to obtain a GGM,

$$\mathbf{K}^{(\Omega)} = \boldsymbol{\Omega}^{-1},$$

which corresponds to the between-subjects network, a network between stationary means of different subjects, described earlier (not to be confused with a cross-sectional network, see Appendix A). As such, estimating the GVAR model on  $n > 1$  time-series analysis allows for the separation of variance into three distinct network structures: temporal networks, contemporaneous networks, and the between-subjects network.

## Estimation

**Pooled and individual LASSO estimation.** Table 3 describes several estimation procedures for the GVAR model in  $n > 1$  datasets. First, we can estimate a GVAR model for every subject to obtain subject-specific estimates for the

temporal and contemporaneous networks. Similarly, we can estimate fixed-effects networks by estimating a GVAR model on the entire within-subjects centered dataset, using the sample means of every subject on every variable as a plug-in for the within-subject means. Consequently, we can estimate the between-subjects network by estimating a GGM on the sample means of each subject on all variables. We can readily apply the LASSO regularization methods described earlier for this purpose: the methodology outlined by Abegaz and Wit (2013) to estimate temporal and contemporaneous networks and the methodology outlined by Foygel and Drton (2010) to estimate the between-subjects GGM. We term this framework *pooled and individual LASSO estimation* and have implemented it in the R package graphicalVAR (Epskamp, 2017b). The performance of pooled and individual LASSO estimation is assessed in simulations reported in section 3.2 of the supplementary materials.

**Multilevel estimation.** The second and third procedures described in Table 3 make use of multilevel modeling (Hamaker, 2012). Two main benefits of this approach are (1) instead of estimating the VAR model in each subject, only the fixed effects and variance–covariance of the random effects need to be estimated, and (2) afterwards, estimates of subject-specific parameters can be obtained, which are somewhat pulled together (termed *shrinkage*). Shrinkage allows the estimation of the model for one subject to borrow information from other subjects. Multilevel estimation can be performed by integrating over the distribution of the random effects or by specifying the multivariate model using hierarchical Bayesian Monte-Carlo sampling methods (Gelman & Hill, 2006; Schuurman, Grasman, & Hamaker, 2016).

**Multivariate Bayesian multilevel.** Bayesian multivariate estimation has proven to be powerful in estimating such models, especially given its flexibility in adding measurement error, latent variables and in handling missing data (Schuurman, Houtveen, & Hamaker, 2015). Recently, the dynamic SEM methodology implemented in Mplus version 8 (Muthén & Muthén, 2017; Asparouhov et al., 2016) has made estimation of multivariate multilevel VAR models much faster and more user-friendly than other Bayesian software routines. Specifying a temporal VAR model with correlated random effects is straightforward and relatively fast to compute with a moderate number of variables (e.g., 6). At the time of writing, Mplus does not return partial correlations by default, but these can be obtained by using the BPARAMETERS option and manually inverting the sampled variance–covariance matrices. Mplus allows for specifying random effects on the contemporaneous covariances and thus, by extension, allows for estimating random contemporaneous networks in addition to random temporal networks. Specifying such a model can be done by specifying dummy latent variables for the residual covariance between each pair of variables (a prior guess on the sign of the co-

variance is needed). Doing so, however, can significantly increase computation length especially when all random effects are allowed to correlate. To facilitate estimation, we have implemented a function generating Mplus code for a multilevel GVAR model and subsequently running the model using the *MplusAutomation* package (Hallquist & Wiley, 2017) in version 0.4 of the *mlVAR* package, which can be called using `estimator = "Mplus"` and requires the Mplus program to be installed.

**Two-step multilevel VAR.** A downside of multivariate estimation is that the number of random effect covariances to be estimated increases quadratically with the number of variables. Forcing random effects to be uncorrelated helps, but places strict assumptions on the model. Bringmann et al. (2013) proposed to estimate multilevel VAR models using univariate models instead, using a frequentist estimation procedure. In this work, the multilevel VAR model is estimated by sequentially estimating univariate multilevel regression models of one variable given all lagged variables. Doing so ignores several correlations of random effects because many parameters are not estimated in the same model, simplifying the analysis: only correlations between incoming edges to the same node and the intercept of that node are included in the univariate models. This method scales up well to approximately six to eight variables when estimating correlated random effects and around 20 variables when estimating orthogonal random effects (or by using a moving window approach; Bringmann et al., 2015). Of note, when specifying orthogonal random effects not all random effects are assumed to be uncorrelated, merely the ones used in the same univariate model.

The methodology of Bringmann et al. (2013) does not estimate contemporaneous or between-subjects networks. To this end, we extended the algorithm in a framework we term *two-step multilevel VAR*. The details of this estimation procedure are explained in Section 2 of the supplementary materials. In short, we extend the methodology of Bringmann et al. (2013) by within-subject centering and by adding subject sample means as between-subjects predictors (as discussed by e.g., Hoffman & Stawski, 2009; Curran & Bauer, 2011; Hamaker & Grasman, 2014). This allows us to estimate between-subjects networks by collecting regression coefficients as in Equation (6) and to symmetrize the resulting matrix.<sup>12</sup> 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. In a second step, we take the residuals of the first analysis and again perform sequential univariate multilevel regression models to predict each residual from all other residuals in the same measurement occasion. Again, these can be collected, as in Equation (6), and symmetrized to obtain contemporaneous networks. By using univariate regressions, rather than simply correlating the residuals, we

can impose multilevel structure on the partial correlations in order to estimate fixed and random effects.<sup>13</sup> Networks can be thresholded by removing all effects that are not significant. For the between-subjects and contemporaneous networks, this results in two  $p$  values for every edge—either both can be required to be significant (“and” rule) or only one (“or” rule). Using the “and” rule means erring more on the side of caution (sparser network), whereas using the “or” rule means erring more on the side of discovery. We have implemented two-step multilevel VAR in the *mlVAR* package, which can be called using `estimator = "lmer"` (the default).

**Choosing the estimation method.** The choice of which estimator to use is not trivial and depends on the interests of the researcher. In Table 3 we list some pros and cons of each of the methodologies. In particular, multilevel estimation can be very complicated and is harder in high-dimensional settings. Assuming normally distributed parameters can also be problematic because doing so imposes that subjects cannot differ on the structure of the networks, merely on their parameterization. When a parameter (e.g., a temporal edge) is zero in some subjects but nonzero in others, then this parameter cannot be normally distributed (the distribution would peak at 0). Therefore, it is currently hard to estimate differently structured individual networks (different edges set to be exactly 0 between subjects) in multilevel estimation. Nonetheless, multilevel estimation particularly shines in that when estimating an individual network, researchers can borrow information from other subjects. We have performed simulation studies to assess the performance of the two proposed methods in this paper: pooled and individual LASSO estimation and two-step multilevel VAR. We report the results of these studies in Section 3.2 of the supplementary materials, which shows that both methods adequately detect the true network structures by increasing sample size. Having more time points per subject helps to estimate the contemporaneous and temporal networks, and having more subjects helps to estimate the between-subject networks. Two step multilevel VAR performs well in estimating intraindividual networks when the number of observations are low, but does not perform subject-specific model selection: all estimated intra-individual networks are saturated and contain all edges.

**GIMME.** Finally, when analyzing  $n > 1$  data, another option is to estimate SVAR models instead. A promising

<sup>12</sup>Standardizing regression parameters from nodewise multilevel models to partial correlation coefficients does not lead to perfectly identical estimates.

<sup>13</sup>Estimating 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.

Table 3  
*Three Methods of Estimating GVAR Models With  $n > 1$  Subjects*

	Pooled and individual LASSO estimation	Bayesian multilevel	Two-step frequentist multilevel
Software	<i>graphicalVAR</i> (Epskamp, 2017b); <i>sparseTSCGM</i> (Abegaz & Wit, 2015).	<i>MPlus</i> 8 (Muthén & Muthén, 2017; Asparouhov et al., 2016); <i>mlVAR</i> (wrapper around <i>Mplus</i> ).	<i>mlVAR</i> (Epskamp, Deserno, & Bringmann, 2017).
Estimation	(1) Joint multivariate LASSO estimation with EBIC model selection (Abegaz & Wit, 2013) of within-subjects centered data to obtain fixed effects temporal and contemporaneous networks. (2) glasso algorithm with EBIC model selection (Foygel & Drton, 2010) on sample means of subjects to obtain between-subject network. (3) Step (1) repeated for each individual dataset to obtain subject-specific networks.	MCMC sampling from multivariate hierarchical model (e.g., Schuurman, Grasman, & Hamaker, 2016).	(1) Sequential univariate multilevel regression models on previous measurement (similar to Bringmann et al., 2013), with within-subject centered lagged variables as within-subjects level predictors and sample-means of all other variables as between-subjects predictor. (2) Sequential multilevel regression models using the residuals of (1): residuals of one variable are predicted by residuals of all other variables in the same measurement occasion.
Pros	Fast estimation of fixed effects; scales up well to large numbers of nodes; model selection in individual networks; temporal and contemporaneous networks obtained in the same analysis.	Borrowing information in individual network estimation from other subjects; all model parameters and random-effect (co)variances can be estimated; credibility intervals can be obtained for edges and descriptive statistics (e.g., centrality; density); advanced extensions such as measurement error and latent variable modeling possible; powerful handling of missing values.	Borrowing information in individual network estimation from other subjects; scales up well to 8 nodes (correlated random effects) or 20 nodes (orthogonal random effects); many random effect variances correlations can be estimated; fast estimation of individual networks.
Cons	Fixed effects estimated on pooled data; Subject specific networks estimated without borrowing information from other subjects (no multilevel structure); between-subjects network estimated in a different model; very slow to estimate subject-specific networks; poor handling of missing values.	Relatively slow estimation, especially in higher dimensional models; no model selection (thresholding possible via credibility intervals); complicated to estimate contemporaneous random effects.	Slow estimation in larger datasets; no model selection (fixed effects can be thresholded using significance); combination of many different models; does not scale up well past 20 nodes; poor handling of missing values.

*Note.* The software listed only concerns user-friendly automated software because all these models could readily be implemented in most programming languages or Bayesian sampler packages.



estimation procedure to estimate such models over many individuals, while dealing with potential heterogeneity, is “group iterative multiple model estimation” (GIMME; Gates & Molenaar, 2012), which is implemented in R using the *gimme* package (Lane, Gates, Molenaar, Hallquist, & Pike, 2016). In GIMME, no multilevel structure is imposed and subject-specific networks are allowed to differ in structure. Information from other subjects is borrowed, however, in that the structure of individual networks can be based on other subjects (e.g., an edge can be included because it is present in many other subjects). No shrinkage is induced on the parameter estimates that are nonzero (as would be the case in multilevel or hierarchical Bayesian modeling). A variant of GIMME that estimates the GVAR or a combination of structural and GVAR models has not yet been developed and could pose a promising estimation technique for future research.

### Empirical Example 1: Reanalysis of Möttus et al. (2017)

We reanalyzed the data of Möttus et al. (2017) to provide an empirical example of the multilevel VAR methods described above. This data consists of two independent ESM samples, in which items tapping three of the five Five-Factor Model (neuroticism, extraversion, and conscientiousness; McCrae & John, 1992) domains were administered, as was an additional question that asked participants how much they had exercised since the preceding measurement occasion. Sample 1 consisted of 26 people providing 1,323 observations in total, and Sample 2 consisted of 62 people providing a total of 2,193 observations. Participants in Sample 1 answered questions three times per day, whereas participants in Sample 2 answered questions five times per day. In both samples, the minimum time between measurements was two hours. For more information about the samples and the specific questions asked, we refer readers to Möttus et al. (2017).

To obtain an easier and more interpretable example, we first only analyzed questions aimed to measure the extraversion trait and the question measuring exercise. This led to five variables of interest: questions pertaining to feeling outgoing, energetic, adventurous, or happy and the question measuring participants’ exercise habits. We analyzed the data using the two-step multilevel procedure as described in detail in the Section 2 of the supplementary materials. We used the *mlVAR* package, version 0.4, for the estimation of this model. Because the number of variables was small, we estimated the model using correlated temporal and contemporaneous random effects. We ran the model separately for both samples and computed the fixed effects for the temporal, contemporaneous, and between-subjects networks. Correlations of the edge weights indicated that all three networks showed high correspondence between the two samples (temporal network: 0.82, contemporaneous network: 0.94, between-subjects network: 0.70). Owing to the degree of

replicability, we combined the two samples and estimated the model on the combined data.

**Results.** Figure 3 shows the estimated fixed effects of the temporal, contemporaneous, and between-subjects network. In these figures, only significant edges ( $\alpha = 0.05$ ) are shown. In the contemporaneous and between-subjects networks, an edge was retained if one of the two regressions on which the partial correlation is based was significant (the so-called “or” rule; van Borkulo et al., 2014). These results are in line with the hypothetical example shown in Figure 2: People who exercised were more energetic while exercising and less energetic after exercising. In the between-subjects network, no relationship between exercising and energy was found. The between-subjects network, however, showed a strong relationship between feeling adventurous and exercising: People who, on average, exercised more also felt, on average, more adventurous. This relationship was not present in the temporal network and much weaker in the contemporaneous network. Also noteworthy is that people were less outgoing after exercising. Figure 4 shows the standard deviation of the random effects in the temporal and contemporaneous networks. Although not many differences can be detected in the temporal network, the contemporaneous network shows strong differences: People mostly differed in their relationship between exercising and feeling energetic.

In addition to using only the extraversion and exercise items, we also ran the model on all 17 administered items in the dataset. In this analysis, we used orthogonal random effects to estimate the model because correlated random effects cannot be estimated with such a large number of variables. Figure 5 shows the estimated fixed effects of the three network structures; it can be seen that indicators of the three traits tend to cluster together in all three networks. Regarding the node exercise, we found the same relationships between exercise, energetic, and adventurous (also found in the previous example) in the larger networks. Furthermore, we noted that exercising was connected to feeling angry in the between-subjects network but not in the other networks. Finally, there was a between-subjects connection between exercising and feeling self-disciplined: People who, on average, exercised more also felt, on average, more self-disciplined.

### Empirical Example 2: Reanalysis of Bringmann et al. (2013)

To showcase additional information that can be obtained using the GGM model, we reanalyzed the dataset used and made publicly available by Bringmann et al. (2013), which has been collected by Geschwind et al. (2011). This dataset contains ESM measures of 129 participants, which was collected in two periods over 6 days each: a baseline period and a posttreatment period (mindfulness treatment and a control group). Participants answered 60 measurements per period. Similar to Figure 1 of Bringmann et al. (2013), we

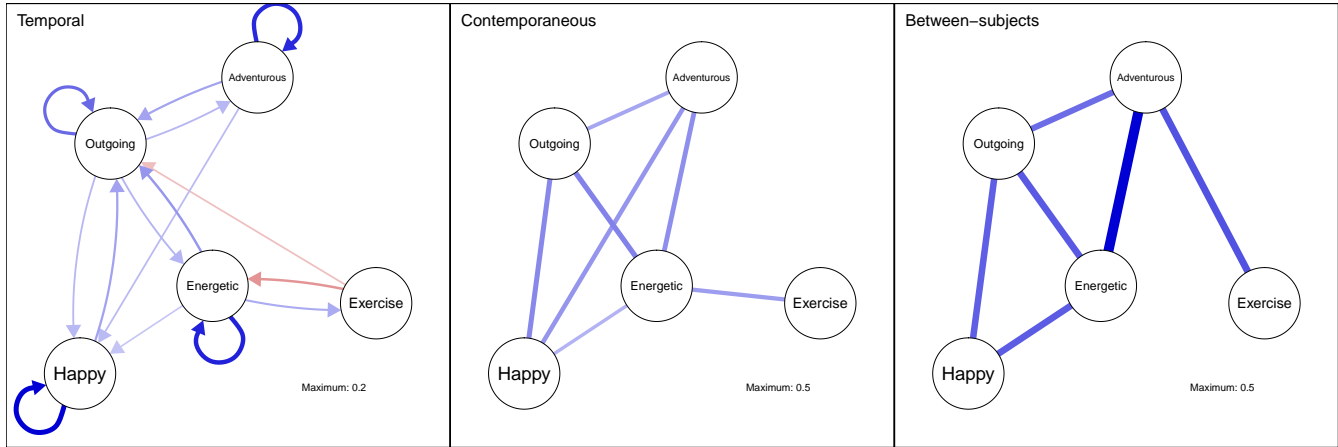


Figure 3. The estimated fixed effects of the three network structures obtainable in multilevel VAR. The model is based on ESM data of 88 people providing a total of 3,516 observations. Due to differences in the scale of the networks, the temporal network was drawn with a different maximum value (i.e., the value indicating the strongest edge in the network) than the contemporaneous and between-subjects networks. Edges that were not significantly different from zero were removed from the networks.

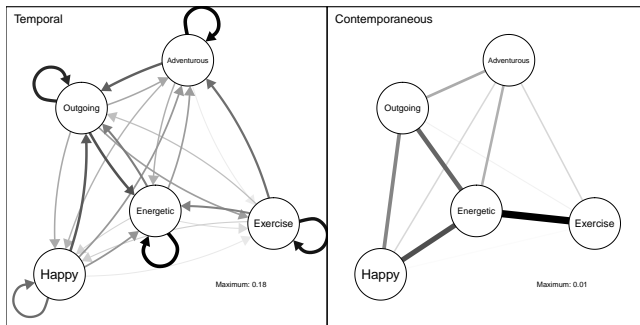


Figure 4. The networks showing the standard deviation of random effects in the temporal and contemporaneous networks. Due to scale differences, networks were plotted using different maximum values.

analyzed only the baseline dataset on the six items selected by Bringmann et al. (2013). We estimated the networks using three modeling frameworks discussed in Table 3. First, we analyze data using multilevel Bayesian estimation using Mplus version 8 (model generated using the *mlVAR* package). We estimated correlated random effects for the temporal effects but only fixed effects for the contemporaneous effects (making these random led to slow convergence). The model was estimated using three chains that ran until convergence. Nights were handled by adding a row of missing values between consecutive days. Second, we analyzed the data using two-step multilevel estimation as implemented in the *mlVAR* package, using an “and” rule and estimating correlated random temporal and contemporaneous effects. Finally, we estimated the data using pooled and individual LASSO estimation using the *graphicalVAR* package, using  $\gamma = 0.25$ .

In the final two analyses, we did not regress the first measurement of the day on the last measurement of the previous day, and removed all pairs of lagged and current variables that contained missing responses. The final sample size was 5,927 observations. Edges were retained if they were significant at the  $\alpha = 0.05$  level, or if 0 was not included in the 95% credibility interval.

**Results.** Figure 6 shows the resulting network structures, and shows that all three methods are mostly aligned. Unsurprisingly, the temporal networks are very similar to those reported by Bringmann et al. (2013).<sup>14</sup> Both the temporal and contemporaneous network are in line with what would be expected under a unidimensional auto-correlated latent variable model (many edges selected, low-rank structure, edges of expected sign) with the exception of the positive temporal edge from “fearful” to “pleasant” in the two-step multilevel network (which was not selected by the other methods). Of note is that Bayesian multilevel estimation resulted in a sparser temporal network. Remarkable is the positive edge between “sad” and “relaxed” in the two-step multilevel between-subjects network, which is based on two significant positive Level 2 regression coefficients ( $\beta = 0.202, p = 0.046$  and  $\beta = 0.151, p = 0.036$ ) where the estimated between-subjects correlation is strongly negative ( $-0.53$ ). This edge is especially remarkable since both nodes are otherwise strongly connected. The Bayesian multilevel between-subjects network showed a similar positive edge between “cheerful” and “worry”. This is noteworthy because under a unidimensional factor model, we would not expect

<sup>14</sup>The networks differ because the estimation of temporal effects differs in that measures are within-subjects centered and subject means are included as Level 2 predictors.

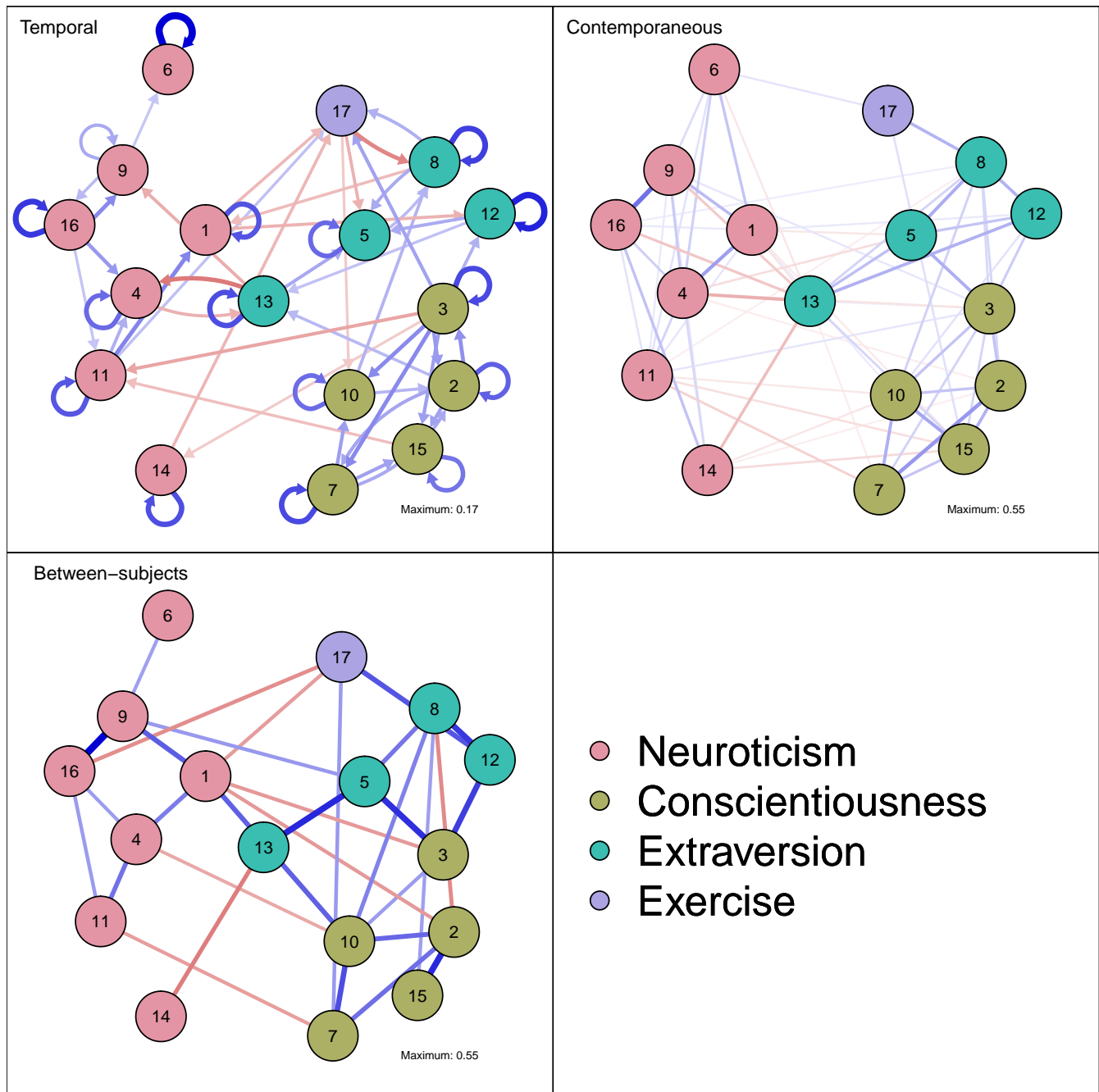
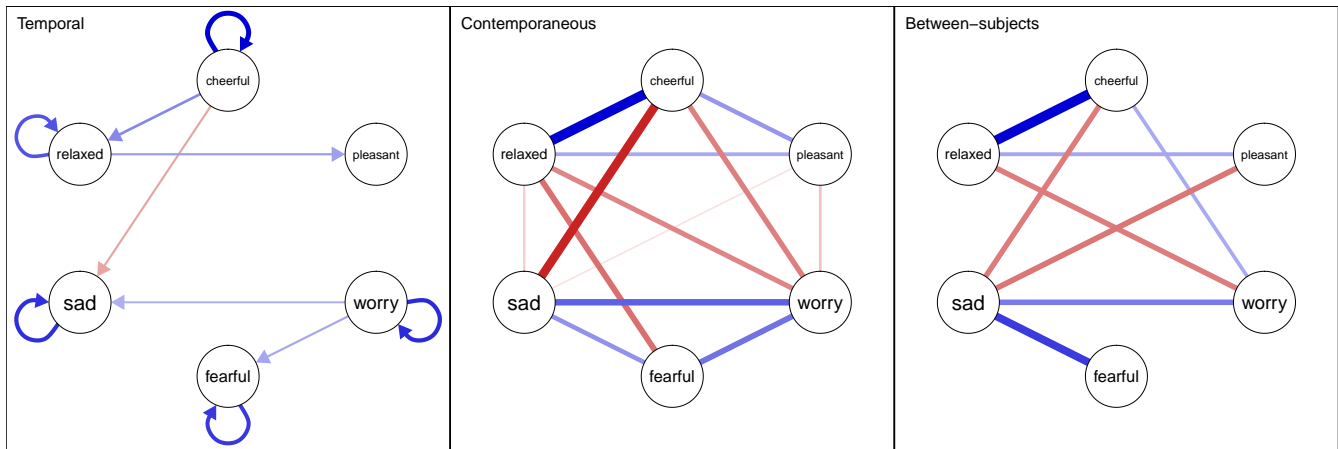
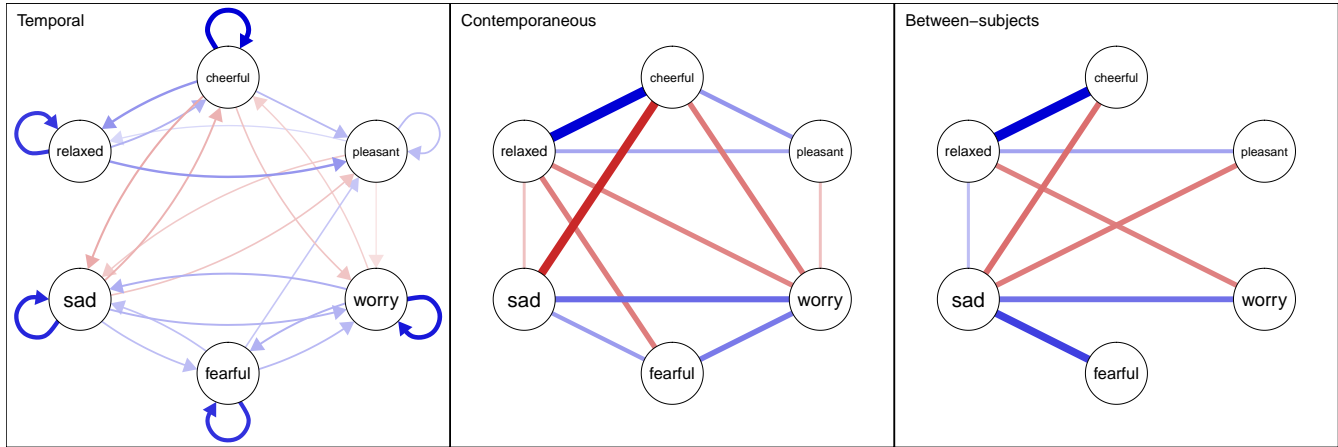


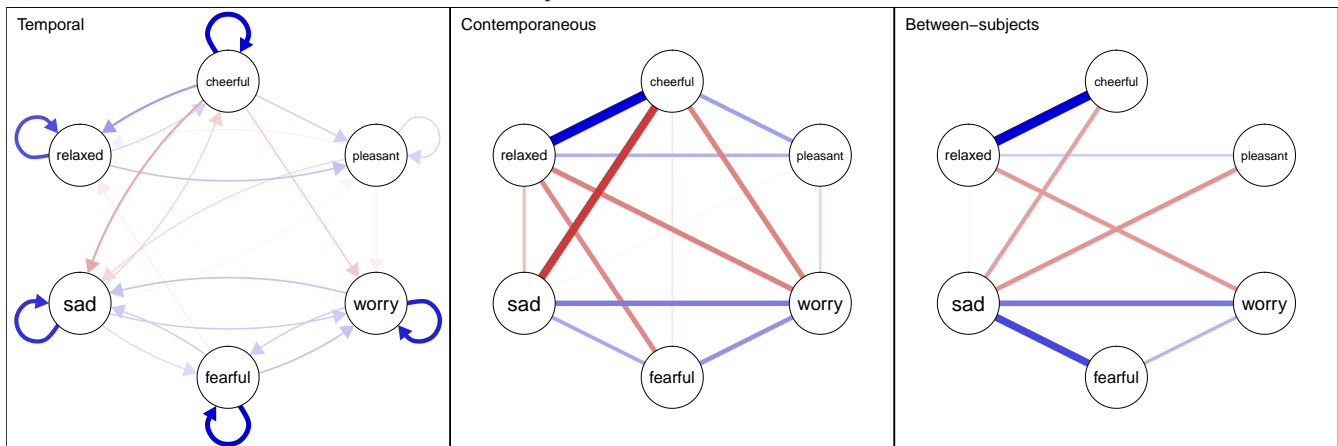
Figure 5. The estimated fixed effects of the three network structures based on all 17 variables administered. Only significant edges are shown. Legend: 1 = “Worried”; 2 = “Organized”; 3 = “Ambitious”; 4 = “Depressed”; 5 = “Outgoing”; 6 = “Self-Conscious”; 7 = “Self-Disciplined”; 8 = “Energetic”; 9 = “Frustrated”; 10 = “Focused”; 11 = “Guilty”; 12 = “Adventurous”; 13 = “Happy”; 14 = “Control”; 15 = “Achieved”; 16 = “Angry”; 17 = “Exercise.”



(a) Fixed effect network structures estimated via multilevel Bayesian estimation.



(b) Fixed effect network structures estimated via two-step multilevel estimation.



(c) Fixed effect network structures estimated via pooled and individual LASSO estimation

Figure 6. Reanalysis of the Geschwind et al. (2011) dataset used by Bringmann et al. (2013).

partial correlation coefficients to switch sign from marginal correlation coefficients (Holland & Rosenbaum, 1986; van Bork, Grasman, & Waldorp, 2016). A possible way the partial correlation coefficient switches sign is if it has been conditioned on one or more common effects between the two variables of interest (in this case, potentially “worry,” “pleasant,” or “fearful”). Of course, these effects must be interpreted with great care, especially given the high  $p$  values; we did not control for multiple comparisons, and the same edges are not retained in the other methods. Still, it is noteworthy that if this edge is weak or nonexistent, the between-subjects structure is still not in line with a unidimensional factor model. In such a factor model, “sad” and “relaxed” (which feature the most connections) would be expected to have a strong negative edge between them (a depression factor would lead to “sad” having a strong positive factor loading and “relaxed” having a strong negative factor loading).

### Within- and between-subjects effects

This paper characterized between-subjects effects (relationships between stationary means of subjects) in two occasions: first in our discussion of cross-sectional data and second in our discussion of  $n > 1$  temporally ordered data. In stark contrast to prior work, criticizing cross-sectional analysis for not reflecting within-subjects effects (Bos & Wanders, 2016; Bos et al., 2017; Molenaar, 2004), we interpreted cross-sectional networks to potentially reflect a between-subjects structure, depending on the type of questions asked. Furthermore, in contrast to prior work on multilevel VAR modeling (e.g., Bringmann et al., 2013, 2015; Wigman et al., 2015; Pe et al., 2015), we conceptualized between-subjects effects in addition to the within-subjects effects in a separate GGM. This raises the question on how such models could be interpreted. In particular, if edges in the GGM are interpreted as generating hypotheses to potential causal pathways, the question is raised how such causal effects can occur at the between-subjects level. This section discusses the topic of causation at the between-subjects level. Here, we interpret the stationary means as being locally stationary: the average of a subject in a relatively short time span of measurement (e.g., a few weeks). As such, we do not interpret the mean vector  $\mu_P$  as a lifetime average. Instead, we assume it could change, potentially due to experimental intervention. As a result, we argue that the between-subjects network can also be indicative of potential causal pathways—regardless of whether it is estimated from a cross-sectional interview concerning variables that are not expected to vary much over time or obtained from estimating the means from time-series data. To simplify the argumentation, we do not discuss separate temporal and contemporaneous networks but combined them into within-subjects networks (a GGM of within-subject data without taking temporal ordering into

account).

**Simpsons paradox.** Hamaker (2012) described an example of how within- and between-subject effects can strongly differ from each other. Suppose we let people write several texts, and we measure the number of spelling errors they make and the number of words per minute they type (typing speed). We would expect to see the seemingly paradoxical network structures shown in Figure 7, Panel (a). We would expect a positive relationship in the within-subjects network (e.g., typing faster than your average leads to making more errors). Conversely, we would expect a negative relationship in the between-subject network (e.g., people who type fast, on average, generally make fewer spelling errors). This is because people who type fast, on average, are likely to be more skilled in writing (e.g., a courtroom stenographer) and are less prone to make a lot of spelling errors, compared to someone who types infrequently. Panel (b) of Figure 7 shows another example in which the structures might differ (Hoffman, 2015; provided by Hamaker, 2017). These network structures show that when people exert more physical activity than their average they likely experience an elevated heart rate, while people who on average are often physically active likely have a lower average heart rate. Such a different effect depending on the level of analysis is well known in the statistical literature as Simpson’s paradox (Simpson, 1951).

**Interventionist accounts of causation.** The different ways of thinking about the effects of manipulations in time-series models can be organized in terms recently developed from interventionist accounts of causation (Woodward, 2005). According to Woodward, causation is fleshed out in terms of interventions:  $X$  is a cause of  $Y$  if an intervention (natural or experimental) on  $X$  leads to a change in  $Y$ . Statistically, the interventionist account is compatible with, for example, Pearl’s (2000) semantics in terms of a “do-operator.” Here, an intervention on  $X$  is represented as  $\text{Do}(X = x)$ , and the causal effect on  $Y$  is formally expressed as  $\mathbb{E}(Y \mid \text{Do}(X = x))$ . Pearl distinguished this from the classical statistical association, in which no intervention is present, and we get the ordinary regression  $\mathbb{E}(Y \mid \text{See}(X = x))$ . This notation is useful here, because it can be used to show how different kinds of causal manipulations, each at the intraindividual level, can produce a signal in either the between-subjects or the within-subjects network.

Cashing out causal effects in terms of interventions is useful for understanding the intervention  $\text{Do}(X = x)$ . We can think of this in terms of a random shock to the system, which sets  $X$  to value  $x$  at a particular time point and evaluates the effect on another variable  $Y$  shortly afterwards. If we want to gauge this type of causal relationship, we might look at the within-subjects VAR model. Consider Hamaker’s (2012) example regarding typing errors: If a researcher forced a person to type very fast, that researcher would need to evaluate the within-subject data, which would show a positive association

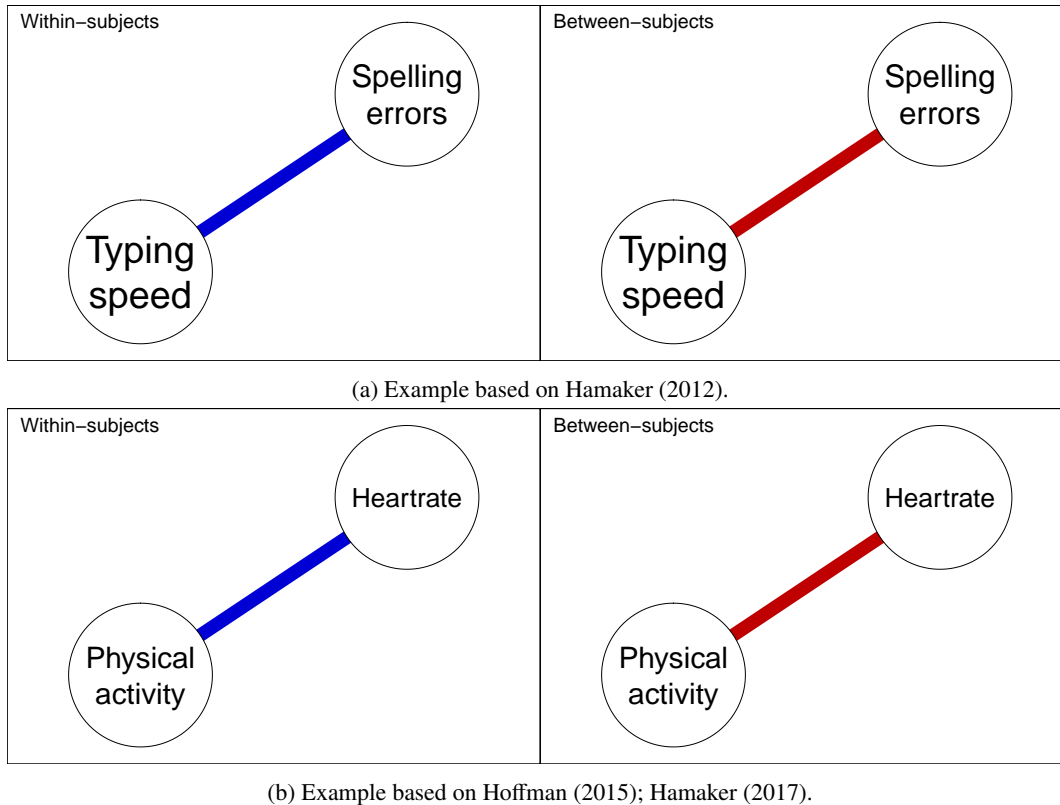


Figure 7. Two hypothetical examples of differing within- and between-subject networks. The networks on the left indicates the within-subject network, showing that personal deviations from the means predict each other at the same time point, and the networks on the right indicates the between-subjects network, showing how the *means* of different subjects relate to one another.

between typing speed and the number of errors. In this example, between-subjects data would be misleading because individual differences would probably yield a negative correlation between speed and accuracy—faster typists are more likely to make less errors.

**Interventions at the mean level.** However, we can also think of a manipulation that sets  $X$  to value  $x$  in a different way, for instance, by inducing a long-term change in the system that leads it to converge on  $X = x$  in expectation. To evaluate the effect of this type of intervention, it is important to consider the behavior of the system as it relates to the changes of the intercept of  $X$ . When analyzing time-series data gathered in a relatively short time-span, the within-subjects VAR network as discussed here cannot represent the relevant effects, because it assumes stationarity. However, such effects will be visible in the between-subjects network, which may thus contain important clues to the behavior of the system under potential changes in the intercept of one variable. In terms of Panel (b) of Figure 7, if we are interested in the effect of changing someone structurally—reducing the heart rate of a person on average—our preferred source of hypothesis generation would likely stem from the

between-subjects model, as the corresponding within-subject model using the methods described in this paper only models deviations from the stationary mean. Such hypotheses could then be further investigated by using experimental design or lengthier longitudinal data analysis.

Many such examples can be envisioned, especially in the field of psychopathology. For instance, short-term deviations from the mean in abusing a substance might not immediately develop tolerance or lead to one suffering from work or life inferences, but a subject who abuses a substance on average over a long time period might develop these problems (example based on variables used by Rhemtulla et al., 2016). A between-subjects network could similarly show that loneliness mediates the effect of losing a spouse on depressive symptoms (Fried et al., 2015) or highlight the possible effects of childhood trauma and urbanization on psychotic symptoms (Isvoranu et al., 2016, 2017)—both cases in which within-subjects networks based on short-term deviations from the average seem less applicable. This analysis is important because it shows that, even though relevant causal interventions in psychology will typically operate at the intra-individual level, evidence for the effect of such in-

terventions may arise at either the within- or the between-subjects level depending on the nature of the intervention.

### Discussion

We discussed the Gaussian graphical model (GGM; Lauritzen, 1996), an undirected network model of partial correlation coefficients, and discussed its utility in the analysis of psychological datasets. The GGM presents a promising exploratory data analysis tool because of several reasons: (1) The GGM is well identified, has no equivalent models, and exploratory search algorithms perform well in capturing the true GGM structure. On the other hand, directed network estimation is often poorly identified and leads to edges estimated in the wrong direction. (2) causal effects between variables result in an edge, whereas the lack of a causal effect results in no edge, except in the presence of latent variables or a common effect. The GGM can, therefore, be seen as hypothesis generating structures that highlight potential causal pathways. (3) Latent variables as common causes emerge in the GGM as low-rank clusters and can also be detected. (4) Edges of the GGM can also be interpreted without reliance on a causal interpretation and merely used to show which variables predict each other. (5) Undirected models can be used and interpreted as causal data-generating process and have been used as such in several fields of research.

The GGM can readily be estimated on any dataset that contains multiple observations of the same variables (e.g., multiple people in cross-sectional data or multiple responses in time-series data). LASSO regularization methods perform especially well in estimating such a GGM structure. In temporally ordered data (e.g.,  $n = 1$  time series), the graphical VAR (GVAR; Wild et al., 2010) model generalizes the GGM to incorporate temporal effects. We described that two network structures can be obtained: a temporal network, which is a directed network of regression coefficients between lagged and current variables, and a contemporaneous network, which is a GGM describing the relationships that remain after controlling for temporal effects. In temporally ordered data of multiple subjects (e.g.,  $n > 1$  time series), the natural combination of cross sectional and time-series data came by adding a third network structure: the between-subjects network, which is a GGM that describes relationships between the stationary means of subjects. We proposed two methods to estimate the three network structures: (1) two-step multilevel estimation, which we implemented in the open source R package *mlVAR*, and (2) pooled and individual VAR model estimations using LASSO regularization, which we implemented in the open source R package *graphicalVAR*.

### Limitations and Challenges

**Multilevel estimation.** The presented methods are not without problems and have several limitations. In regards

to multilevel estimation, first, multivariate estimation of the multilevel VAR model is not yet feasible for larger datasets. As such, the proposed two-step multilevel VAR combines univariate models. Doing so, however, means that not all parameters are in the same model. In addition, univariate models do not readily provide estimates of the contemporaneous networks, which must be estimated in a second step. Second, even when multivariate estimation is possible, it is still challenging to estimate a multilevel model on contemporaneous networks due to the requirement of positive definite matrices. Third, when more than approximately eight variables are measured, estimating the multilevel models with correlated random effects is no longer feasible in open source software. In this case, orthogonal random effects can be used, which enforce parsimony on the model that may not be plausible. Finally, even when orthogonal estimation is used, multilevel analysis runs very slowly in models with more than 20 variables. As such, multilevel VAR analysis of high-dimensional datasets is not yet feasible. To this end, we discussed pooling within-subject centered data and estimating fixed-effects models using LASSO regularization (Abegaz & Wit, 2013). This performed on par with multilevel estimation in higher sample sizes and allows researchers to scale up the analysis. However, individual network estimation using separate VAR models does not borrow information from other subjects and performs poorly in low sample sizes. Promising developments are new LASSO methods in which shrinkage from subject-specific parameters to their mean is attained through penalization rather than hierarchical modeling (Hastie, Tibshirani, & Wainwright, 2015). Future research should investigate the utility of such models in estimating individual network structures that might differ in structure but borrow information from other subjects in its estimation.

**VAR modeling assumptions.** These limitations on the estimation methods come with more limitations in the statistical models themselves. VAR modeling, especially, is not without problems and faces severe challenges (Hamaker, Ceulemans, Grasman, & Tuerlinckx, 2015; Hamaker & Wichers, 2017). We made several assumptions that can be problematic. For instance, in characterizing the likelihood of time-series data, we need to assume that the conditional distribution of variables at time  $t$  given time  $t - 1$  are the same for all  $t$ . That raises two distinct assumptions: (1) The difference in time between measurements are roughly equal, and (2) the parameters do not change over time. Equidistance in time is especially important for the interpretation of temporal networks. Promising work is being done in this area where VAR networks can be estimated on nonequidistant datasets (Driver, Oud, & Voelkle, 2017; Oravecz, Tuerlinckx, & Vandekerckhove, 2009; Oud & Jansen, 2000). The assumption of stationarity is needed to estimate structures when data are limited but might not be tenable especially in longer time series (Rovine & Walls, 2006). Promising time-varying es-

timization procedures are being developed (Bringmann et al., 2016; Haslbeck & Waldorp, 2016a), but are not yet extended to the GVAR framework. Furthermore, the interpretation of temporal coefficients when drawn as a network is not without discussion, and several different methods for standardization exist (Bulteel, Tuerlinckx, Brose, & Ceulemans, 2016; Schuurman, Ferrer, de Boer-Sonnenschein, & Hamaker, 2016).<sup>15</sup>

**Normality.** Another particularly important assumption made in this paper is that of multivariate normality. Indeed, Equation (1) makes this assumption and all other equations follow from this. The assumption of normality is not without problems (Terluin, de Boer, & de Vet, 2016). However, it is not one that is easily solved. This is because there can be many reasons why data are not normally distributed. When data are not normally distributed, then they cannot be explained properly using only the means vector and variance–covariance matrix. As a result, the GGM does not properly characterize the joint likelihood function. Three conceivable ways can lead to data not being distributed normally: (1) the data are measured on a different scale (Stevens, 1946), (2) the data are continuous but do not follow a normal density, and (3) there are nonlinear relationships between variables. In the first case, a different graphical model can be used, such as the Ising model for binary data (Epskamp et al., in press; van Borkulo et al., 2014) or a mixed graphical model for categorical and Poisson-distributed variables as well as binary and Gaussian variables (Haslbeck & Waldorp, 2016b). Such models have yet to be extended to time-series analysis, especially in separating temporal and contemporaneous effects as the GVAR model does. When data are continuous but not normal, multiple reasons can (again) attribute to this. When the underlying process is normal but the measured variables are on a transformed scale, transforming data back to normal should offer a solution (Liu, Lafferty, & Wasserman, 2009), but when the process itself is nonnormal, such as skewed residuals, the entire modeling framework does not correctly capture the likelihood. Finally, multivariate normality assumes all relationships between variables are linear. When this is not the case, the GGM and VAR model (which fit linear effects) will not properly describe the data. We encourage future researchers to focus on the problem of normality and to develop new methods of overcoming these challenges.

**Interpretation.** Finally, it should be noted that all network structures only generate hypotheses and are in no way confirmatory. The analyses showcased in this paper are exploratory and allow researchers to obtain insights into the predictive relationships present in the data—regardless of theory with respect to the data-generating model. Under the assumptions of multivariate normality, stationarity, and the Lag-1 factorization, the networks show how variables predict each other over time (temporal network), within time (contemporaneous network), and on average (between-subjects network). Furthermore, during the thresholding of

edges in the multilevel analyses, we did not apply a correction for multiple testing by default. We deliberately chose this because our aim was to present exploratory hypothesis-generating structures, and not correcting for multiple testing yields greater sensitivity.

## Conclusion

This paper provides a methodological overview of how the GGM can be used to gain exploratory insight into the potential dynamics present in psychological data. The GGM can be used to map out unique variance in cross-sectional data or at the contemporaneous and between-subjects levels of time-series analysis. We contrasted this method to exploratory estimation of causal models: DAGs and SVAR. While losing information on the direction of effect, estimating GGMs offers an attractive alternative in that these models are uniquely identified, well parameterized, closely related to causal models and also offer exploratory insight on predictive effects between observed variables. When the aim is to discover psychological dynamics, the GGM can be used as a hypothesis generating technique inspiring future research or therapy directions (Epskamp, van Borkulo, et al., 2017; Kroeze et al., 2017). For example, an effect found in a cross-sectional analysis could inspire a time-series study, a contemporaneous effect could inspire a shorter time-lag time-series study and a between-subjects effect could inspire lengthy longitudinal studies. All network structures may inspire experimental design, or to gather a mixture of observational and experimental data (Magliacane et al., 2017). The GGM thus provides a powerful addition to the exploratory toolbox in behavioral research.

## Acknowledgements

We would like to thank Laura Bringmann, Noémi Schuurman, Oisín Ryan and Ellen Hamaker for helpful tips and invigorating discussions, and Katharina Jorgensen for valuable comments on earlier versions of this paper.

<sup>15</sup>We standardized every dataset before analyzing and used the standardization of Wild et al. (2010) for temporal networks in  $n = 1$  and pooled temporal networks. GGMs are readily standardized by using partial correlation coefficients (Equation (2)), which have been used in all GGMs shown in this paper.



### Supplementary 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,<sup>16</sup>  $i$  ( $i \in \{1, 2, \dots, m\}$ ) 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.<sup>17</sup> 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. 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  $\{\dots\}$ .

### Supplementary 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:

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

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  $\mathbf{K}_p^{(\Theta)} = \boldsymbol{\Theta}_p^{-1}$ .

### Supplementary 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:

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

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

<sup>16</sup>Mostly 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  $p = 2$ .

<sup>17</sup>If 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.

taken as a substitute for  $\mu_p$  (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

$$\begin{aligned} y_{[t,p,i]} &= \mu_{[p,i]} + \beta_{[p,i]} (\mathbf{y}_{[t-1,p]} - \bar{\mathbf{y}}_p) + \varepsilon_{[t,p,i]} \\ \varepsilon_{[T,p,i]} &\sim N(0, \theta_{[p,i]}), \end{aligned} \quad (7)$$

and the Level 2 model becomes

$$\begin{bmatrix} \mu_{[p,i]} \\ \beta_{[p,i]} \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ \beta_{*i} \end{bmatrix}, \begin{bmatrix} \omega_{\mu_i} & \omega^{(\beta_i \mu_i)^T} \\ \omega^{(\beta_i \mu_i)} & \Omega^{(\beta_i)} \end{bmatrix} \right).$$

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*.

### Supplementary 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 (7), 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

$$\mu_{[p,i]} = \beta_i^{(\mu)} \bar{\mathbf{y}}_{[p,-(i)]} + \varepsilon_{[p,i]}^{(\mu)}, \quad (8)$$

in which we use  $\beta_i^{(\mu)}$  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  $\bar{\mathbf{y}}_{[p,i]}$  is itself an estimate of  $\mu_{[p,i]}$ , Equation (8) 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:

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

with  $d_{ii}^{(\mu)} = 1/\text{Var}(\varepsilon_{[p,i]}^{(\mu)})$ . Due to the estimation in a multi-level 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  $A$ .

**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

$$\hat{\varepsilon}_{[t,p,i]} = \beta_{[p,i]}^{(\boldsymbol{\varepsilon})} \hat{\boldsymbol{\varepsilon}}_{[t,p,-(i)]} + \varepsilon_{[t,p,i]}^{(\boldsymbol{\varepsilon})}, \quad (9)$$

in which  $\beta_{[p,i]}^{(\boldsymbol{\varepsilon})}$  represents the  $i$ -th row (without the diagonal element  $i$ ) of an  $m \times m$  matrix,  $\mathbf{B}_p^{(\boldsymbol{\varepsilon})}$ , and  $\varepsilon_{[t,p,i]}^{(\boldsymbol{\varepsilon})}$  represents a residual. In the Level 2 model, we again assign a multivariate normal distribution to parameters in  $\beta_i^{(\boldsymbol{\varepsilon})}$ . It can be seen that Equation (9) also takes the form of a multiple regression model. Thus, this model can again be seen as the node-wise GGM estimation procedure:

$$\mathbf{K}_p^{(\boldsymbol{\varepsilon})} \approx \mathbf{D}_p^{(\boldsymbol{\varepsilon})} (\mathbf{I} - \mathbf{B}_p^{(\boldsymbol{\varepsilon})}),$$

with  $d_{[p,i]}^{(\boldsymbol{\varepsilon})} = 1/\text{Var}(\varepsilon_{[T,p,i]}^{(\boldsymbol{\varepsilon})})$ . Again the matrices need to be made symmetric by averaging upper and lower triangle elements. Fixed effects can be obtained by using 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 *mIVAR* 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 *mIVAR* package can also be used to plot the estimated networks, in which significance thresholding is used by default with a significance level of  $\alpha = 0.05$ .

### Supplementary 3: Simulation Studies

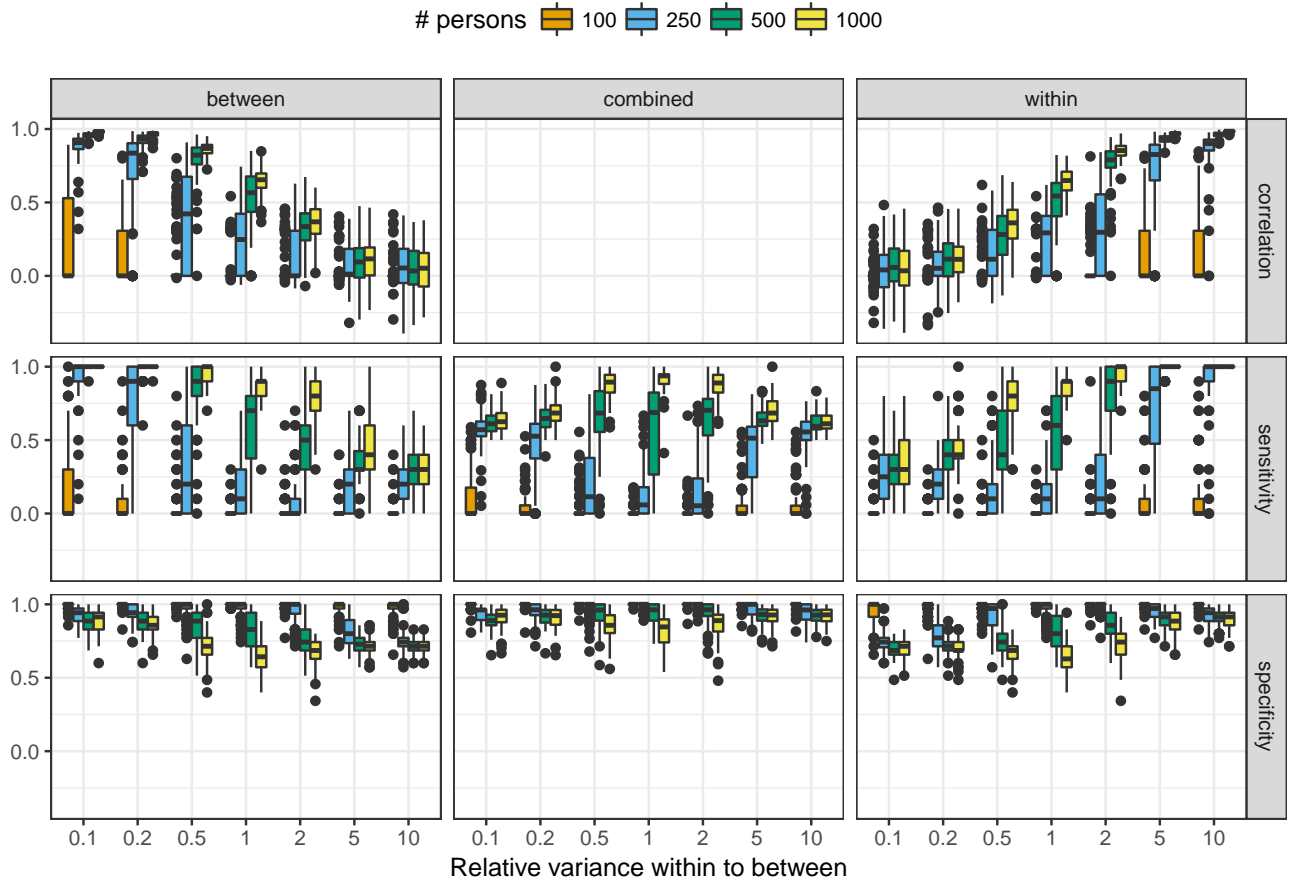
#### Supplementary 3.1: Simulation Study 1: Cross-sectional Analysis and Repeated Measures

We performed a simulation study to investigate the interpretability of GGMs obtained from cross-sectional data. We constructed two network structures: a within-subjects network and a between-subjects network. For the within-person network, we constructed a chain graph (as shown in Figure S3, panel (b)) in which each edge-weight was set to 0.25 and made negative with 50% probability. For the between-person network, we constructed the same structure but randomly rewired all the edges, such that a random graph was

obtained (as shown in Figure S3, panel (c)). Next, the within-person network was scaled such that the within-person variance of all nodes was 0.1, 0.2, 0.5, 1, 2, 5 or 10. The between-subjects variance was set to 1. We estimated the network structure using glasso in combination with EBIC model selection (Foygel & Drton, 2010), as implemented in the *qgraph* package (Epskamp et al., 2012; Epskamp & Fried, 2017). We set the EBIC hyperparameter  $\gamma$  to 0.25.

Figure S1 shows the results, in which the estimated network was compared to the true within and between subjects networks as well as an unweighted network that contained an edge whenever there was an edge in either true network. We did not compute the correlation of edge weights with the unweighted graph as it contained no weights. The figure shows that in large differences between within-person and between-person variance the cross-sectional analysis converges to one of the two networks. When the between-subjects variance was relatively high the within-subjects network was not retrievable and vice versa. When the two variances were approximately equal, edges were detected that were in either of the networks (high sensitivity in the combined graph) and not many edges were detected that were not present in either network (high specificity). This indicates that detected edges in a cross-sectional network can be interpreted to likely represent an edge in the within-subjects network or in the between-subjects network. It should be noted that specificity did go down with increased sample size, as the true network does not necessarily contain zeros anymore. These false edges were usually estimated to be very weak. Finally, these results are based on a simulation study in which the within- and between-subject networks completely differed from each other. Preliminary results, such as the empirical samples in this paper, seem to suggest this is not the case in empirical data. We expect cross-sectional data analysis to perform better when the two network structures align.

In order to start disentangling within- and between-subjects variance, one needs an estimate of the person-specific mean. The simplest way to obtain this is by averaging two repeated measures. To this end, we repeated the simulation study above using the exact same setup, but now generated two independent responses per subject. Next, we centered the responses per subject to compute a within-subject network and used the sample means per subject to compute a between-subjects network. Figure S2 shows that the within-subject network could now be well retrieved, while the between-subject network could only be retrieved when the between-subject variance was large. This seems to suggest that now the estimated network structure can be more reliably interpreted to not be confounded by between-subjects variance. Care should still be taken in interpreting such results, as, for instance, the analysis assumes both no autoregressions (these would greatly bias the estimated means per subject) and no individual differences in network structure



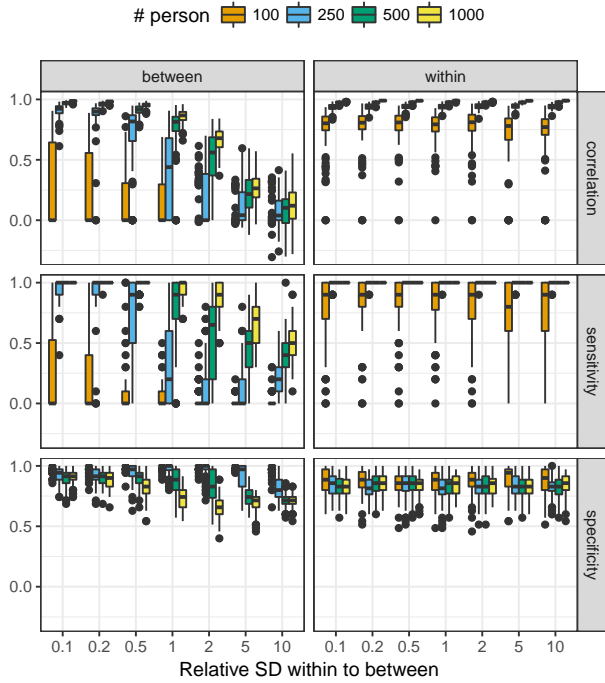
**Figure S1.** Simulations study showing the performance of cross-sectional analysis. Data were generated by summing scores generated from two distinct distributions: a chain graph GGM (within-person) and a random graph GGM (between-person). Left panels show performance compared to the between-subjects network and right panels show performance compared to the within-subjects network. The center panels show performance compared to an unweighted graph that contained an edge if there was an edge in the within-subjects network or the between-subjects network.

(which one can not expect to obtain with only two repeated measures). We encourage future researchers to study cross-sectional and repeated measures analysis in more detail by performing more extensive simulation studies.

### Supplementary 3.2: Simulation Study 2: Two-step Multi-level VAR and Pooled and Aggregated LASSO Estimation

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 simulated ESM data on 8 variables. First, we constructed a temporal, contemporaneous and between-subject network structure as shown in Figure S3, with 50% of non-diagonal parameters made negative.

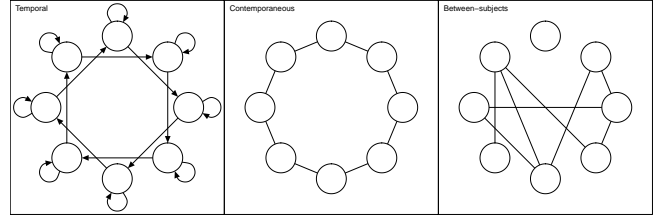
Next, we parameterized the networks (temporal and contemporaneous networks were parameterized separately per subject). Temporal coefficients and partial contemporaneous or between-subject correlations were drawn from a uniform distribution between 0.2 and 0.5 (or  $-0.5$  and  $-0.2$  for negative edges). The contemporaneous networks and the between-subjects networks were subsequently rescaled such that each node had a residual variance of 0.25 and a between-subjects variance of 1. Thus, for different subjects temporal and contemporaneous networks were equal in structure (which edge was present and the sign of the edge) but not in weight. To simulate subjects differing in network structure, we rewired intra-individual networks with probability 0 (no differences) or 1 (fully different within-person networks). Sample size and number of observations per subject were varied between 50, 100 and 200, leading to a total of 2 (rewiring)  $\times$  3 (sample size)  $\times$  3 (number of observations per subject) conditions. Each condition was replicated 100 times, leading to 18,000



**Figure S2.** Simulations study showing the performance of cross-sectional analysis. Data were generated by summing scores generated from two distinct distributions: a chain graph GGM (within-person) and a random graph GGM (between-person). Left panels show performance compared to the between-subjects network and right panels show performance compared to the within-subjects network. The center panels show performance compared to an unweighted graph that contained an edge if there was an edge in the within-subjects network or in the between-subjects network.

total simulated datasets. We used both *mlVAR* (two-step multi-level VAR) and *graphicalVAR* (pooled and aggregated LASSO estimation) to estimate fixed and subject-specific network structures. In *mlVAR*, orthogonal random effects in combination with an “and”-rule were used to threshold significant edges. In *graphicalVAR*, we tested  $10 \times 10$  tuning parameters (temporal by between) and selected the set of tuning parameters 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.

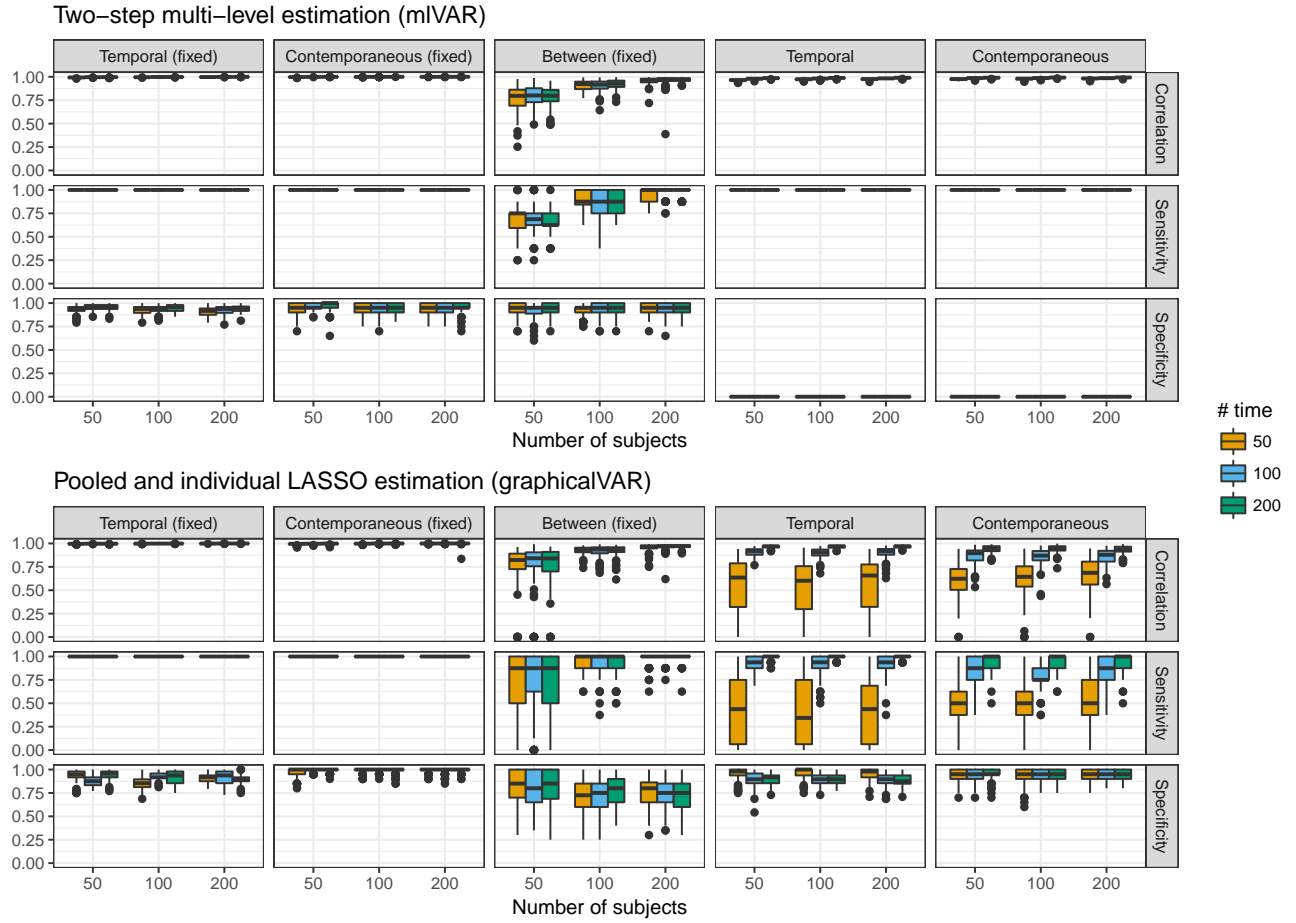
In order to assess how well the estimated networks resemble the true networks, we computed for each dataset the correlations between true and estimated fixed temporal, contemporaneous, and between-subjects networks and the correlations between true and estimated subject specific tempo-



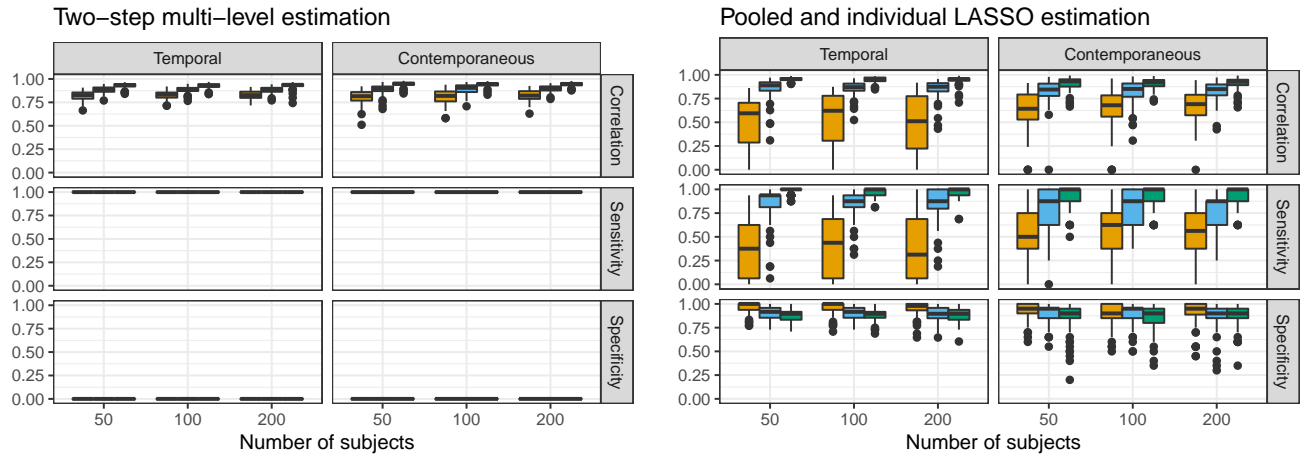
**Figure S3.** Example of simulated network models in the simulation study. Placing each node in a circle, the temporal network contained positive auto-regressions on each node and one cross-lagged regression on the node two positions to the right, the contemporaneous network was simulated as a chain graph, and the between-subjects network was simulated as a random network with the same number of edges as the contemporaneous network (a different random network was generated for each dataset). Non-diagonal elements were made negative with 50% probability.

ral 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 again computed the *sensitivity* (true positive rate) and the *specificity* (true negative rate).

Figure S4 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). 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). Figure S5 shows the results in the condition where edges were rewired, and shows here too a good performance for both methods. *mlVAR* estimation performed slightly poorer than when the structure was the same over all subjects, and *graphicalVAR* performed identically (as can be expected since no information of other subjects is used in the estimation).



*Figure S4.* Results of the simulation study using 8 nodes with the same network structure (but different parameters) for each subject. True network structures were generated as shown in Figure S3. Boxplots indicate the distribution of the measures over all the 100 simulated datasets per condition. The left panels relate to the fixed effect structures and the right panels relate to subject specific network structures.



*Figure S5.* Results of the simulation study using 8 nodes with the differing network structures for each subject. True network structures were generated as shown in Figure S3, but temporal and contemporaneous edges were rewired with 50% probability to other nodes at random. Only subject-specific networks are shown (as the fixed effect structure is ill-defined).

### Supplementary 4: Structural VAR and Graphical VAR

In this appendix we will show the equivalence between structural VAR (SVAR) and graphical VAR (GVAR). This equivalence is well known, as the only difference between GVAR and regular VAR is that the contemporaneous variance–covariance matrix is interpreted and modeled as a GGM, the equivalence between regular VAR and SVAR holds (e.g., Lütkepohl, 2005; Beltz & Molenaar, 2016; Molenaar & Lo, 2016). We supply the equivalence here because inverting this contemporaneous variance–covariance matrix simplifies the expression, and to keep the paper standalone readable. First, we can derive an expression for precision matrix  $\mathbf{K}$ :

$$\begin{aligned}\mathbf{K} &= (\mathbf{I} - \mathbf{\Gamma})^\top \mathbf{K}^{(\Theta)} (\mathbf{I} - \mathbf{\Gamma})^\top \\ &= \begin{bmatrix} \mathbf{K}_{11} - \mathbf{\Gamma}_{21}^\top \mathbf{K}_{22} - \mathbf{K}_{12} \mathbf{\Gamma}_{21} + \mathbf{\Gamma}_{21}^\top \mathbf{K}_{22} \mathbf{\Gamma}_{21} & \mathbf{K}_{12} (\mathbf{I} - \mathbf{\Gamma}_{22}) - \mathbf{\Gamma}_{21}^\top \mathbf{K}_{22} (\mathbf{I} - \mathbf{\Gamma}_{22}) \\ (\mathbf{I} - \mathbf{\Gamma}_{22})^\top \mathbf{K}_{21} - (\mathbf{I} - \mathbf{\Gamma}_{22})^\top \mathbf{K}_{22} \mathbf{\Gamma}_{21} & (\mathbf{I} - \mathbf{\Gamma}_{22})^\top \mathbf{K}_{22} (\mathbf{I} - \mathbf{\Gamma}_{22}) \end{bmatrix}\end{aligned}$$

We add superscript ( $S$ ) to denote a matrix following the SVAR setup and ( $G$ ) to denote a matrix following the GVAR setup. Then, we obtain:

$$\begin{bmatrix} \mathbf{K}_{11}^{(S)} + \mathbf{\Gamma}_{21}^{(S)\top} \mathbf{K}_{22}^{(S)} \mathbf{\Gamma}_{21}^{(S)} & -\mathbf{\Gamma}_{21}^{(S)\top} \mathbf{K}_{22}^{(S)} (\mathbf{I} - \mathbf{\Gamma}_{22}^{(S)}) \\ -(\mathbf{I} - \mathbf{\Gamma}_{22}^{(S)})^\top \mathbf{K}_{22}^{(S)} \mathbf{\Gamma}_{21}^{(S)} & (\mathbf{I} - \mathbf{\Gamma}_{22}^{(S)})^\top \mathbf{K}_{22}^{(S)} (\mathbf{I} - \mathbf{\Gamma}_{22}^{(S)}) \end{bmatrix} = \begin{bmatrix} \mathbf{K}_{11}^{(G)} + \mathbf{\Gamma}_{21}^{(G)\top} \mathbf{K}_{22}^{(G)} \mathbf{\Gamma}_{21}^{(G)} & -\mathbf{\Gamma}_{21}^{(G)\top} \mathbf{K}_{22}^{(G)} \\ -\mathbf{K}_{22}^{(G)} \mathbf{\Gamma}_{21}^{(G)} & \mathbf{K}_{22}^{(G)} \end{bmatrix}.$$

We can readily find:

$$(\mathbf{I} - \mathbf{\Gamma}_{22}^{(S)})^\top \mathbf{K}_{22}^{(S)} (\mathbf{I} - \mathbf{\Gamma}_{22}^{(S)}) = \mathbf{K}_{22}^{(G)},$$

in which  $\mathbf{K}_{22}^{(S)}$  is diagonal. We can recognize this form from equating a causal model to a GGM in the main article. Thus, the contemporaneous GGM in GVAR is the GGM form of the contemporaneous directed network used in SVAR: it has an undirected edge whenever there is a directed edge or a common effect in the SVAR model. Next, we can solve for  $\mathbf{\Gamma}_{21}^{(G)\top}$ :

$$\begin{aligned}-(\mathbf{I} - \mathbf{\Gamma}_{22}^{(S)})^\top \mathbf{K}_{22}^{(S)} \mathbf{\Gamma}_{21}^{(S)} &= -\mathbf{K}_{22}^{(G)} \mathbf{\Gamma}_{21}^{(G)} \\ (\mathbf{I} - \mathbf{\Gamma}_{22}^{(S)})^\top \mathbf{K}_{22}^{(S)} \mathbf{\Gamma}_{21}^{(S)} &= (\mathbf{I} - \mathbf{\Gamma}_{22}^{(S)})^\top \mathbf{K}_{22}^{(S)} (\mathbf{I} - \mathbf{\Gamma}_{22}^{(S)}) \mathbf{\Gamma}_{21}^{(G)} \\ \mathbf{\Gamma}_{21}^{(S)} &= (\mathbf{I} - \mathbf{\Gamma}_{22}^{(S)}) \mathbf{\Gamma}_{21}^{(G)} \\ (\mathbf{I} - \mathbf{\Gamma}_{22}^{(S)})^{-1} \mathbf{\Gamma}_{21}^{(S)} &= \mathbf{\Gamma}_{21}^{(G)}.\end{aligned}$$

### Supplementary 5: 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  $\text{Vec}$  and the Kronecker product  $\otimes$  to obtain (Kim et al., 1999):

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

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

**Appendix A. Glossary of terms**

Term	Explanation
Undirected network	A network model in which nodes are connected by edges (also termed links) without arrowheads.
Directed network	A network model in which nodes are connected by edges with arrowheads, assumed to display causal effects or temporal prediction.
Gaussian graphical model	An undirected network model in which observed variables are represented with nodes. Nodes are connected with an edge if two variables are not independent after conditioning on all other observed variables. Edges are parametrized by using partial correlation coefficients.
Causal model	A causal model of observed and unobserved variables that is assumed to generate the data.
Directed acyclic graph	A directed network in which one node does not eventually point to itself.
Within-subjects network	A network model explaining within-subject (co)variation from the stationary mean.
Between-subjects network	A network model explaining (co)variation between stationary means of different persons.
Cross-sectional network	A network model estimated on cross-sectional data. Can be shown to be a blend of the within-subjects and between-subjects networks. Can be interpreted as representative of within-subjects or between-subjects network based on the way in which data is gathered.
Vector auto-regression (VAR)	Multivariate regression of a set of variables on previous realizations of that set of variables.
Temporal network	A within-subject network model of effects between different measurement occasions, showing temporal prediction or potential causal pathways.
Contemporaneous network	A within-subject network model of effects between variables in the same measurement occasion, after taking temporal effects into account. Can be directed (SVAR) or undirected (GVAR), and is equal to a within-subjects network if there are no temporal effects.



## References

- Abegaz, F., & Wit, E. (2013). Sparse time series chain graphical models for reconstructing genetic networks. *Biostatistics*, kxt005. doi: 10.1093/biostatistics/kxt005
- Abegaz, F., & Wit, E. (2015). SparseTSCGM: Sparse time series chain graphical models (r package version 2.2) [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=SparseTSCGM>
- Asparouhov, T., Hamaker, E., & Muthén, B. (2016). *Dynamic structural equation models* (Tech. Rep.). Technical Report. In preparation. Retrieved from <https://www.statmodel.com/download/DSEM.pdf>
- 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
- Baumert, A., Schmitt, M., Perugini, M., Johnson, W., Blum, G., Borkenau, P., & Wrzus, C. (in press). Integrating personality structure, personality process, and personality development. *European Journal of Personality*.
- Beltz, A. M., & Molenaar, P. C. (2016). Dealing with multiple solutions in structural vector autoregressive models. *Multivariate behavioral research*, 51(2-3), 357–373. doi: 10.1080/00273171.2016.1151333
- Bolger, N., & Laurenceau, J. (2013). *Intensive longitudinal methods*. New York, NY, USA: Guilford.
- Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, 9, 91–121. doi: 10.1146/annurev-clinpsy-050212-185608
- Borsboom, D., Cramer, A. O. J., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. *PloS one*, 6(11), e27407. doi: 10.1371/journal.pone.0027407
- Bos, E. H., & Wanders, R. B. (2016). Group-level symptom networks in depression. *JAMA psychiatry*, 73(4), 411–411. doi: 10.1001/jamapsychiatry.2015.3103
- Bos, F. M., Snippe, E., de Vos, S., Hartmann, J. A., Simons, C. J. P., van der Krieke, L., ... Wichers, M. (2017). Can we jump from cross-sectional to dynamic interpretations of networks? implications for the network perspective in psychiatry. *Psychotherapy and Psychosomatics*, 86, 185–177. doi: 10.1159/000453583
- Brandt, P. T., & Williams, J. T. (2007). *Multiple time series models* (No. 148). Thousand Oaks, CA, USA: Sage Publications, Inc.
- Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2016). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological methods*. doi: 10.1037/met0000085
- Bringmann, L. F., Lemmens, L. H., Huibers, M. J., Borsboom, D., & Tuerlinckx, F. (2015). Revealing the dynamic network structure of the Beck Depression Inventory-II. *Psychological Medicine*, 45(04), 747–757. doi: 10.1017/S0033291714001809
- 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.
- Bulteel, K., Tuerlinckx, F., Brose, A., & Ceulemans, E. (2016). Using raw var regression coefficients to build networks can be misleading. *Multivariate behavioral research*, 51(2-3), 330–344. doi: 10.1080/00273171.2016.1150151
- Chatfield, C. (2016). *The analysis of time series: an introduction*. Boca Raton, FL, USA: CRC press.
- Chen, G., Glen, D. R., Saad, Z. S., Hamilton, J. P., Thomason, M. E., Gotlib, I. H., & Cox, R. W. (2011). Vector autoregression, structural equation modeling, and their synthesis in neuroimaging data analysis. *Computers in biology and medicine*, 41(12), 1142–1155. doi: 10.1016/j.combiomed.2011.09.004
- Chen, J., & Chen, Z. (2008). Extended bayesian information criteria for model selection with large model spaces. *Biometrika*, 95(3), 759–771. doi: 10.1093/biomet/asn034
- Costantini, G., Epskamp, S., Borsboom, D., Perugini, M., Möttus, R., Waldorp, L. J., & Cramer, A. O. J. (2015). State of the aRt personality research: A tutorial on network analysis of personality data in R. *Journal of Research in Personality*, 54, 13–29. doi: 10.1016/j.jrp.2014.07.003
- Cramer, A. O. J., Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., ... Borsboom, D. (2012). Dimensions of normal personality as networks in search of equilibrium: You can't like parties if you don't like people. *European Journal of Personality*, 26(4), 414–431. doi: 10.1002/per.1866
- Cramer, A. O. J., Waldorp, L., van der Maas, H., & Borsboom, D. (2010). Comorbidity: A Network Perspective. *Behavioral and Brain Sciences*, 33(2-3), 137–150. doi: 10.1017/S0140525X09991567
- 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
- Driver, C. C., Oud, J. H. L., & Voelkle, M. C. (2017). Continuous time structural equation modelling with r package

- ctsem. *Journal of Statistical Software*, 77, 1 – 35. doi: 10.18637/jss.v077.i05
- Drton, M., & Maathuis, M. H. (2017). Structure learning in graphical modeling. *Annual Review of Statistics and Its Application*, 4, 365–393. doi: 10.1146/annurev-statistics-060116-053803
- Eichler, M. (2007). Granger causality and path diagrams for multivariate time series. *Journal of Econometrics*, 137(2), 334–353. doi: 10.1016/j.jeconom.2005.06.032
- Epskamp, S. (2016). Regularized Gaussian Psychological Networks: Brief Report on the Performance of Extended BIC Model Selection. *arXiv preprint*, arXiv:1606.05771.
- Epskamp, S. (2017a). Chapter 12: Discussion: The road ahead. In *Network psychometrics*. Retrieved from <http://sachaepskamp.com/dissertation/Discussion.pdf>
- Epskamp, S. (2017b). graphicalVAR: Graphical VAR for experience sampling data (R package version 0.1.6) [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=graphicalVAR>
- Epskamp, S., Borsboom, D., & Fried, E. I. (2017). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*. Retrieved from arXiv:1604.08462 doi: 10.3758/s13428-017-0862-1
- Epskamp, S., Cramer, A., Waldorp, L., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, 48(1), 1–18. doi: 10.18637/jss.v048.i04
- Epskamp, S., Deserno, M. K., & Bringmann, L. F. (2017). mlVAR: Multi-level vector autoregression [Computer software manual]. (R package version 0.4)
- Epskamp, S., & Fried, E. I. (2017). A tutorial on estimating regularized psychological networks. *arXiv preprint*, arXiv:1607.01367.
- Epskamp, S., Maris, G., Waldorp, L., & Borsboom, D. (in press). Network psychometrics. In P. Irving, D. Hughes, & T. Booth (Eds.), *Handbook of psychometrics*. New York, NY, USA: Wiley. Retrieved from [arxiv.org/abs/1609.02818](http://arxiv.org/abs/1609.02818)
- Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika*. doi: 10.1007/s11336-017-9557-x
- Epskamp, S., van Borkulo, C. D., Isvoranu, M. N., Adela M. Servaas, van der Veen, D. C., Riese, H., & Cramer, A. O. J. (2017). Personalized network modeling in psychopathology: The importance of contemporaneous and temporal connections. *PsyArXiv pre-print*. doi: 10.17605/OSF.IO/JNPRZ
- 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>
- Fried, E. I., Bockting, C., Arjadi, R., Borsboom, D., Amshoff, M., Cramer, O. J., ... Stroebe, M. (2015). From loss to loneliness: The relationship between bereavement and depressive symptoms. *Journal of abnormal psychology*, 124(2), 256–265. doi: 10.1037/abn0000028
- Fried, E. I., Epskamp, S., Nesse, R. M., Tuerlinckx, F., & Borsboom, D. (2016). What are ‘good’ depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. *Journal of Affective Disorders*, 189, 314–320. Retrieved from 10.1016/j.jad.2015.09.005
- Friedman, J. H., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3), 432–441. doi: 10.1093/biostatistics/kxm045
- Friedman, J. H., Hastie, T., & Tibshirani, R. (2014). glasso: Graphical lasso- estimation of gaussian graphical models (R package version 1.8) [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=glasso>
- Gates, K. M., & Molenaar, P. C. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage*, 63(1), 310–319. doi: 10.1016/j.neuroimage.2012.06.026
- 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. doi: 10.1016/j.neuroimage.2009.12.117
- Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. New York, NY, USA: Cambridge University Press.
- Geschwind, N., Peeters, F., Drukker, M., van Os, J., & Wichers, M. (2011). Mindfulness training increases momentary positive emotions and reward experience in adults vulnerable to depression: a randomized controlled trial. *Journal of consulting and clinical psychology*, 79(5), 618–628. doi: 10.1037/a0024595
- Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: a new approach for estimating the number of dimensions in psychological research. *PlosOne*, 2(6), e0174035. doi: 10.1371/journal.pone.0174035

- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 424–438.
- Guttman, L., et al. (1938). A note on the derivation of formulae for multiple and partial correlation. *The Annals of Mathematical Statistics*, 9(4), 305–308. Retrieved from <http://www.jstor.org/stable/2957470>
- Hallquist, M., & Wiley, J. (2017). MplusAutomation: Automating mplus model estimation and interpretation [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=MplusAutomation> (R package version 0.7)
- Hamaker, E., Ceulemans, E., Grasman, R., & Tuerlinckx, F. (2015). Modeling affect dynamics: State of the art and future challenges. *Emotion Review*, 7(4), 316–322. doi: 10.1177/1754073915590619
- Hamaker, E. L. (2012). Why researchers should think “within-person”: A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). New York, NY: Guilford Press. Retrieved from <https://www.researchgate.net/publication/266896375>
- Hamaker, E. L. (2017). A brief history of dynamic modeling in psychology. In *82th annual meeting of the Psychometric Society (IMPS)*.
- 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
- Hamaker, E. L., & Wichers, M. (2017). No time like the present. *Current Directions in Psychological Science*, 26(1), 10–15. doi: 10.1177/0963721416666518
- Hamilton, J. D. (1994). *Time series analysis* (Vol. 2). Princeton, NJ, USA: Princeton university press.
- Harvey, A. C. (1990). *Forecasting, structural time series models and the kalman filter*. Cambridge, UK: Cambridge university press.
- Haslbeck, J. M. B., & Waldorp, L. J. (2016a). mgm: Structure estimation for time-varying mixed graphical models in high-dimensional data. *arXiv preprint*, arXiv:1510.06871.
- Haslbeck, J. M. B., & Waldorp, L. J. (2016b). Structure estimation for mixed graphical models in high dimensional data. *arXiv preprint*, arXiv:1510.05677.
- Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with sparsity: the lasso and generalizations*. Boca Raton, FL, USA: CRC Press.
- Hayduk, L. A. (2009). Finite feedback cycling in structural equation models. *Structural Equation Modeling*, 16(4), 658–675.
- Heiser, W. J. (2017). Early psychometric contributions to gaussian graphical modeling: A tribute to louis guttman (1916–1987). In *82th annual meeting of the Psychometric Society (IMPS)*.
- Hoffman, L. (2015). *Longitudinal analysis: Modeling within-person fluctuation and change*. New York, NY, USA: Routledge.
- 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
- Holland, P. W., & Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. *The Annals of Statistics*, 1523–1543. doi: 10.1214/aos/1176350174
- Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik A Hadrons and Nuclei*, 31(1), 253–258. doi: 10.1007/BF02980577
- Isvoranu, A. M., Borsboom, D., van Os, J., & Guloksuz, S. (2016). A Network Approach to Environmental Impact in Psychotic Disorders: Brief Theoretical Framework. *Schizophrenia Bulletin*, 42(4), 870–873. doi: 10.1093/schbul/sbw049
- Isvoranu, A. M., van Borkulo, C. D., Boyette, L., Wigman, J. T. W., Vinkers, C. H., Borsboom, D., & GROUP Investigators. (2017). A Network Approach to Psychosis: Pathways between Childhood Trauma and Psychotic Symptoms. *Schizophrenia Bulletin*, 43(1), 187–196. doi: 10.1093/schbul/sbw055
- Kalisch, M., & Bühlmann, P. (2007). Estimating high-dimensional directed acyclic graphs with the pc-algorithm. *Journal of Machine Learning Research*, 8(Mar), 613–636. Retrieved from <http://www.jmlr.org/papers/v8/kalisch07a.html>
- Kalisch, M., Mächler, M., Colombo, D., Maathuis, M. H., & Bühlmann, P. (2012). Causal inference using graphical models with the R package pcalg. *Journal of Statistical Software*, 47(11), 1–26. doi: 10.18637/jss.v047.i11
- Kalisch, M., Maechler, M., & Colombo, D. (2017). pcalg: Estimation of CPDAG/PAG and causal inference using the IDA algorithm [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=pcalg> (R package version 2.5-0)
- Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Thousand Oaks, CA, USA: Sage.
- 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.
- Klippel, A., Viechtbauer, W., Reininghaus, U., Wigman, J., van Borkulo, C., MERGE, ... Wichers, M. (2017). The cascade of stress: A network approach to explore differential dynamics in populations varying in risk

- for psychosis. *Schizophrenia Bulletin*, sbx037. doi: 10.1093/schbul/sbx037
- Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: Principles and techniques*. Cambridge, MA, USA: MIT press.
- Kossakowski, J. J., Epskamp, S., Kieffer, J. M., van Borkulo, C. D., Rhemtulla, M., & Borsboom, D. (2015). The application of a network approach to health-related quality of life (HRQoL): Introducing a new method for assessing hrqol in healthy adults and cancer patient. *Quality of Life Research*, 25, 781–92. doi: 10.1007/s11136-015-1127-z
- Krämer, N., Schäfer, J., & Boulesteix, A.-L. (2009). Regularized estimation of large-scale gene association networks using graphical gaussian models. *BMC Bioinformatics*, 10(1), 1–24. doi: 10.1186/1471-2105-10-384
- Kroeze, R., Van Veen, D., Servaas, M., Bastiaansen, J., Oude Voshaar, R., Borsboom, D., & Riese, H. (2017). Personalized feedback on symptom dynamics of psychopathology: a proof-of-principle study. *Journal for Person-Oriented Research*, 3(1), 1–10. doi: 10.17505/jpor.2017.01
- Lane, S., Gates, K., Molenaar, P., Hallquist, M., & Pike, H. (2016). *gimme: Group iterative multiple model estimation* [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=gimme> (R package version 0.1-7)
- Lauritzen, S. L. (1996). *Graphical models*. Oxford, UK: Clarendon Press.
- Liu, H., Lafferty, J. D., & Wasserman, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *The Journal of Machine Learning Research*, 10, 2295–2328. Retrieved from <https://arxiv.org/abs/0903.0649>
- Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). *Statistical theories of mental test scores*. Oxford, UK: Addison-Wesley.
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Berlin, Germany: Springer-Verlag.
- MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, 114(1), 185–199. doi: 10.1037/0033-2909.114.1.185
- Magliacane, S., van Ommen, T., Claassen, T., Bongers, S., Versteeg, P., & Mooij, J. M. (2017). Causal transfer learning. *arXiv preprint arXiv:1707.06422*.
- Marchetti, G. M., Drton, M., & Sadeghi, K. (2015). *ggm: Functions for graphical markov models* (R package version 2.3) [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=ggm>
- McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, 60(2), 175–215. doi: 10.1111/j.1467-6494.1992.tb00970.x
- McNally, R. J., Robinaugh, D. J., Wu, G. W., Wang, L., Deserno, M. K., & Borsboom, D. (2015). Mental disorders as causal systems a network approach to posttraumatic stress disorder. *Clinical Psychological Science*, 3(6), 836–849. doi: 10.1177/2167702614553230
- Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, 66(1), 195–244. doi: 10.2466/pr0.1990.66.1.195
- 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
- Mohammadi, A., & Wit, E. C. (2015). BDgraph: An R package for bayesian structure learning in graphical models. *arXiv preprint*, arXiv:1501.05108.
- 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. doi: 10.1207/s15366359mea0204\_1
- Molenaar, P. C., & Lo, L. L. (2016). Alternative forms of granger causality, heterogeneity and non-stationarity. In W. Wiedermann & A. von Eye (Eds.), *Statistics and causality: Methods for applied empirical research*. Wiley Hoboken, NJ. doi: 10.1002/9781118947074.ch9
- Möttus, R., Epskamp, S., & Francis, A. (2017). Within- and between individual variability of personality characteristics and physical exercise. *Journal of Research in Personality*, 69, 139–148. doi: 10.1016/j.jrp.2016.06.017
- Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. Cambridge, MA, USA: MIT press.
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide: Statistical analysis with latent variables. Version*. Retrieved from [https://www.statmodel.com/download/usersguide/MplusUserGuideVer\\_8.pdf](https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf)
- Myin-Germeyns, I., Oorschot, M., Collip, D., Lataster, J., Delespaul, P., & van Os, J. (2009). Experience sampling research in psychopathology: Opening the black box of daily life. *Psychological medicine*, 39(09), 1533–1547. doi: 10.1017/S0033291708004947
- Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2009). A hierarchical ornstein–uhlenbeck model for continuous repeated measurement data. *Psychometrika*, 74(3), 395–418. doi: 10.1007/s11336-008-9106-8
- Oud, J. H., & Jansen, R. A. (2000). Continuous time state space modeling of panel data by means of SEM. *Psychometrika*, 65(2), 199–215. Retrieved from 10.1007/

- BF02294374
- Pe, M. L., Kircanski, K., Thompson, R. J., Bringmann, L. F., Tuerlinckx, F., Mestdagh, M., ... Others (2015). Emotion-network density in major depressive disorder. *Clinical Psychological Science*, 3(2), 292–300. doi: 10.1177/2167702614540645
- Pearl, J. (2000). *Causality: Models, reasoning, and inference*. New York, NY: Cambridge University Press.
- Pourahmadi, M. (2011). Covariance estimation: The glm and regularization perspectives. *Statistical Science*, 26(3), 369–387. doi: 10.1214/11-STS358
- Quax, R., Kandhai, D., & Sloot, P. M. (2013). Information dissipation as an early-warning signal for the lehman brothers collapse in financial time series. *Scientific reports*, 3(1898). doi: 10.1038/srep01898
- R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Rhemtulla, M., Fried, E. I., Aggen, S. H., Tuerlinckx, F., Kendler, K. S., & Borsboom, D. (2016). Network analysis of substance abuse and dependence symptoms. *Drug and alcohol dependence*, 161, 230–237. doi: 10.1016/j.drugalcdep.2016.02.005
- Rigdon, E. E. (1995). A necessary and sufficient identification rule for structural models estimated in practice. *Multivariate Behavioral Research*, 30(3), 359–383. doi: 10.1207/s15327906mbr3003\_4
- 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. doi: 10.1097/PSY.0b013e3182545d47
- Rothman, A. J., Levina, E., & Zhu, J. (2010). Sparse multivariate regression with covariance estimation. *Journal of Computational and Graphical Statistics*, 19(4), 947–962. doi: 10.1198/jcgs.2010.09188
- Rovine, M. J., & Walls, T. A. (2006). Multilevel autoregressive modeling of interindividual differences in the stability of a process. *Models for intensive longitudinal data*, 124–147. doi: 10.3389/fpsyg.2017.00262
- Schafer, J., Opgen-Rhein, R., Zuber, V., Ahdesmaki, M., Silva, A. P. D., & Strimmer, K. (2017). corpcor: Efficient estimation of covariance and (partial) correlation [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=corpcor> (R package version 1.6.9)
- Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: Findings from the cogito study. *Frontiers in aging neuroscience*, 2, 27. doi: 10.3389/fnagi.2010.00027
- Schmittmann, V. D., Cramer, A. O. J., Waldorp, L. J., Episkamp, S., Kievit, R. A., & Borsboom, D. (2013). Deconstructing the construct: A network perspective on psychological phenomena. *New Ideas in Psychology*, 31(1), 43–53. doi: 10.1016/j.newideapsych.2011.02.007
- Schuurman, N. K. (2016). Chapter 5: Measurement error and person-specific reliabilities in multilevel autoregressive modeling. In *Multilevel autoregressive modeling in psychology: Snags and solutions*. Retrieved from [http://www.nkschuurman.com/NKSchuurman\\_dissertation.pdf](http://www.nkschuurman.com/NKSchuurman_dissertation.pdf)
- Schuurman, N. K., Ferrer, E., de Boer-Sonnenschein, M., & Hamaker, E. L. (2016). How to compare cross-lagged associations in a multilevel autoregressive model. *Psychological methods*, 21(2), 206. doi: 10.1037/met0000062
- Schuurman, N. K., Grasman, R. P. P. P., & Hamaker, E. L. (2016). A comparison of inverse-wishart prior specifications for covariance matrices in multilevel autoregressive models. *Multivariate Behavioral Research*, 51(2-3), 185–206. doi: 10.1080/00273171.2015.1065398
- Schuurman, N. K., Houtveen, J. H., & Hamaker, E. L. (2015). Incorporating measurement error in n = 1 psychological autoregressive modeling. *Frontiers in psychology*, 6. doi: 10.3389/fpsyg.2015.01038
- Scutari, M. (2010). Learning Bayesian Networks with the bnlearn R Package. *Journal of Statistical Software*, 35(3), 1–22. doi: 10.18637/jss.v035.i03
- Shumway, R. H., & Stoffer, D. S. (2010). *Time series analysis and its applications: with r examples*. New York, NY, USA: Springer Science & Business Media. doi: 10.1007/978-3-319-52452-8
- Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 238–241. Retrieved from <http://www.jstor.org/stable/2984065>
- Snippe, E., Viechtbauer, W., Geschwind, N., Klippel, A., De Jonge, P., & Wichers, M. (2017). The impact of treatments for depression on the dynamic network structure of mental states: Two randomized controlled trials. *Scientific Reports*, 7, 46523.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science, New Series*, 103(2684), 677–680. doi: 10.1126/science.103.2684.677
- Terluin, B., de Boer, M. R., & de Vet, H. C. (2016). Differences in connection strength between mental symptoms might be explained by differences in variance: Reanalysis of network data did not confirm staging. *PloS one*, 11(11), e0155205. doi: 10.1371/journal.pone.0155205

- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58, 267–288. Retrieved from <http://www.jstor.org/stable/2346178>
- van Bork, R., Grasman, R. P. P. P., & Waldorp, L. J. (2016). Unidimensional factor models imply weaker partial correlations than zero-order correlations. *arXiv preprint*, arXiv:1610.03375.
- 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
- van Borkulo, C. D., Boschloo, L., Borsboom, D., Penninx, B. W. J. H., Waldorp, L. J., & Schoevers, R. A. (2015). Association of symptom network structure with the course of depression. *JAMA Psychiatry*, 72(12), 1219–1226. doi: 10.1001/jamapsychiatry.2015.2079
- van der Maas, H. L., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, 113(4), 842–861. doi: 10.1037/0033-295X.113.4.842
- Wigman, J., van Os, J., Borsboom, D., Wardenaar, K., Epskamp, S., Klippel, A., ... Wichers, M. (2015). Exploring the underlying structure of mental disorders: Cross-diagnostic differences and similarities from a network perspective using both a top-down and a bottom-up approach. *Psychological Medicine*, 45(11), 2375–2387. doi: 10.1017/S0033291715000331
- Wild, B., Eichler, M., Friederich, H.-C., Hartmann, M., Zipfel, S., & Herzog, W. (2010). A graphical vector autoregressive modeling approach to the analysis of electronic diary data. *BMC Medical Research Methodology*, 10(1), 28. doi: 10.1186/1471-2288-10-28
- Witten, D. M., Friedman, J. H., & Simon, N. (2011). New insights and faster computations for the graphical lasso. *Journal of Computational and Graphical Statistics*, 20(4), 892–900. doi: 10.1198/jcgs.2011.11051a
- Woodward, J. (2005). *Making things happen: A theory of causal explanation*. Oxford, UK: Oxford University Press. Retrieved from <http://www.jstor.org/stable/3840611>
- Wright, S. (1921). Correlation and causation. *Journal of agricultural research*, 20(7), 557–585.
- Yuan, M., & Lin, Y. (2007). Model selection and estimation in the gaussian graphical model. *Biometrika*, 94(1), 19–35. doi: 10.1093/biomet/asm018
- Zhao, T., Li, X., Liu, H., Roeder, K., Lafferty, J., & Wasserman, L. (2015). huge: High-dimensional undirected graph estimation (R package version 1.2.7) [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=huge>