Network psychometrics

Epskamp, S.

Chapter 6

Discovering Psychological Dynamics

Abstract

This chapter outlines statistical network models in cross-sectional and time-series data, that attempt to highlight potential causal relationships between observed variables. The chapter describes three kinds of datasets. In cross-sectional data (1), one can estimate a Gaussian graphical model (GGM; a network of partial correlation coefficients). In single-subject time-series analysis (2), networks are typically constructed through the use of (multilevel) vector autoregression (VAR). VAR estimates a directed network that encodes temporal predictive effects—the temporal network. We show that GGM and VAR models are closely related: VAR generalizes the GGM by taking violations of independence between consecutive cases into account. VAR analyses can also return a GGM that encodes relationships within the same window of measurement—the contemporaneous network. When multiple subjects are measured (3), multilevel VAR estimates fixed and random temporal networks. We show that between-subject effects can also be obtained in a GGM network—the between-subjects network. We propose a novel two-step multilevel estimation procedure to obtain fixed and random effects for contemporaneous network structures. This procedure is implemented in the R package mlVAR. The chapter presents a simulation study to show the performance of *mlVAR* and showcases the method in an empirical example on personality inventory items and physical exercise.

6.1 Introduction

Network modeling of psychological data has increased in recent years. This is consistent with a general call to conceptualize observed psychological processes that are not merely indicative of latent common causes but rather reflect the emergent

This chapter has been adapted from: Epskamp, S., Waldorp, L.J., Mõttus, R., and Borsboom, D. (2016). Discovering Psychological Dynamics: The Gaussian Graphical Model in Cross-sectional and Time-series Data. *arXiv preprint*, arXiv:1609.04156.

behavior of complex, dynamical systems in which psychological, biological, and sociological components directly interact with each other (Borsboom et al., 2011; Cramer, Sluis, et al., 2012; Cramer et al., 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 et al., in press; van Borkulo et al., 2014)—the dynamics of psychology. This chapter provides a methodological overview of statistical network models in cross-sectional and time-series data. Furthermore, this chapter shows that the common network models for cross-sectional and time-series data are closely related. In time-series modeling, this relationship allows researchers to extend the modeling framework to incorporate contemporaneous and between-subject effects. We propose a novel estimation procedure to do so, which we implemented in the free software package *mlVAR*.[1]

We can distinguish two lines of research in which networks are utilized on psychological datasets: modeling of cross-sectional data and 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 persons are measured only once. The most popular method is to estimate undirected network models, indicating pairwise interactions—so-called pairwise Markov random fields (Lauritzen, 1996; Murphy, 2012). When the data are continuous, the Gaussian graphical model (GGM; Lauritzen, 1996) 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. This model is applied extensively to psychological data (e.g., Cramer, Sluis, et al., 2012; Fried, Epskamp, et al., 2016; Isvoranu, van Borkulo, et al., 2016; Kossakowski et al., 2016; McNally et al., 2015; van Borkulo et al., 2015).

Time-series data can be obtained by using the experience sampling method (ESM; Myin-Germeys et al., 2009), in which subjects are asked several times per day to fill in a short questionnaire through 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 growing increasingly popular in assessing intraindividual dynamical structures (e.g., Bringmann et al., 2013, 2015; Wigman et al., 2015). As will be shown below, the VAR model can be seen as a generalization of the GGM that takes violations of independence between consecutive cases into account. Thus, the lines of research on cross-sectional and time-series data can naturally be combined. The GGM is, however, not yet commonly used in time-series analysis.

In this chapter we present an overview of out-of-the-box methods, applicable to normally distributed data, that aim to map out the dynamics present in psychological data. We will do so in two distinct settings: cross-sectional data, in which observations are plausibly independent, and intensive repeated measures in a relatively short time span obtained through ESM data. In multivariate normal data,

[1] CRAN link: `http://cran.r-project.org/package=mlVAR`
Github link (developmental): `http://www.github.com/SachaEpskamp/mlVAR`

all relationships between variables are contained within the variance–covariance matrix. Thus, characterizing the covariances in estimable ways provides a method to characterize all relationships that are present. This information can then be represented in networks. In cross-sectional data, one network can be obtained—an undirected network of partial correlation coefficients. In time-series data, up to three networks can be obtained—the *temporal network*, a directed network indicating within-person relationships across time; the *contemporaneous network*, an undirected partial correlation network within the same measurement; and the *between-subjects network*, an undirected partial correlation network between the means of the subject's scores within the time span of measurement. We describe how all three networks can highlight potential causal pathways and thereby act as hypothesis-generating structures.

The chapter is structured in the following manner. We first characterize the joint likelihood of full ESM data in three steps: (a) when cases are deemed to be plausibly independent, (b) when the time-series data of a single subject contain plausibly nonindependent observations, and (c) when we have time-series data of several subjects that combine both independent and nonindependent observations. In each of these situations, we outline estimation procedures including a description of open-source software packages. We also implement the novel methods of this chapter, the methods for analyzing ESM data of multiple subjects, in the software package mlVAR. Furthermore, we show the functionality of this package in an empirical example by reanalyzing personality inventory items measured in an ESM design (Mõttus, Epskamp, & Francis, 2016). Finally, we assess the performance of these methods in a large-scale simulation study.

6.2 Characterizing Multivariate Gaussian Data

In this chapter we will model measurements of N subjects ($p \in 1, 2, \ldots, N$), in which subject p is measured T_p times ($t \in 1, 2, \ldots, T_p$) on I variables ($i \in 1, 2, \ldots, I$). Let $\boldsymbol{Y}$ represent the set of random variables measured in each subject:

$$\boldsymbol{Y} = \left\{ \boldsymbol{Y}^{(1)}, \boldsymbol{Y}^{(2)}, \ldots, \boldsymbol{Y}^{(N)} \right\}.$$

Element $\boldsymbol{Y}^{(p)}$ contains the random responses of a subject on all T_p measurements:

$$\boldsymbol{Y}^{(p)} = \begin{bmatrix} \boldsymbol{Y}_1^{(p)} \\ \boldsymbol{Y}_2^{(p)} \\ \vdots \\ \boldsymbol{Y}_{T_p}^{(p)} \end{bmatrix},$$

in which $\boldsymbol{Y}_t^{(p)}$ contains the row vector with random responses of subject p on time-point t, which we assume to be a *multivariate Gaussian* distribution with some mean vector $\boldsymbol{\mu}^{(p)}$ and some variance–covariance matrix $\boldsymbol{\Sigma}^{(p)}$:

$$\boldsymbol{Y}_t^{(p)} = \begin{bmatrix} Y_{t1}^{(p)} & Y_{t2}^{(p)} & \ldots & Y_{tI}^{(p)} \end{bmatrix} \sim N\left(\boldsymbol{\mu}^{(p)}, \boldsymbol{\Sigma}^{(p)} \right).$$

We will denote realizations of the above described random variables by lowercase letters (e.g., $\boldsymbol{y}$ and $y_{ti}^{(p)}$). It is important to note that measurements are nested in persons and not the other way around (two subjects can have a different number of measurements) and that the measurement t of one subject does not correspond to measurement t of another subject.[2] In typical data representation used in statistical software, $\boldsymbol{Y}_t^{(p)}$ will correspond to a row of responses from a given subject on all items. We will term these the *cases*: A case is the set of responses of a subject on a time point on all items.

Given a set of parameters $\boldsymbol{\xi}$, let $\mathcal{L}$ denote the likelihood function

$$\mathcal{L}(\boldsymbol{\xi}; \boldsymbol{y}) = f_{\boldsymbol{Y}}(\boldsymbol{y} \mid \boldsymbol{\xi}),$$

in which $f_{\boldsymbol{Y}}(\boldsymbol{y} \mid \boldsymbol{\xi})$ denotes a probability density function, which we will shorten to $f(\boldsymbol{y} \mid \boldsymbol{\xi})$ for the remainder of this chapter. The likelihood function is crucial in estimating the set of parameters $\boldsymbol{\xi}$, either by maximum likelihood estimation (MLE) or by playing a crucial role in the formation of the posterior distribution of $\boldsymbol{\xi}$ in Bayes' rule. An important assumption in computing such a likelihood function is the *assumption of independence*. Given a set of estimated parameters $\boldsymbol{\xi}$, and assuming the model is correct, we can reasonably assume that scores of subjects are independent, allowing us to write the joint likelihood as a product of marginal likelihoods:

$$f(\boldsymbol{y} \mid \boldsymbol{\xi}) = \prod_{p=1}^{N} f\left(\boldsymbol{y}^{(p)} \mid \boldsymbol{\xi}\right).$$

Suppose we measured subjects only once and on one variable—say, their IQ level. This assumption of independence indicates, for example, that knowing Peter has an IQ of 90 does not help us predict Sarah's IQ level, given that IQ has a mean of 100 and a standard deviation of 15.

In order to fully characterize the likelihood of all observations, we further need to characterize $f(\boldsymbol{y}^{(p)} \mid \boldsymbol{\xi})$, the joint likelihood of all cases of a single subject. This is easily done in the cross-sectional example described above because every subject has only one observation. When multiple cases of a subject can be assumed to be independent, this likelihood similarly can be expressed as a product of all the likelihoods of the cases. However, as we will detail below, often the assumption of independent cases is not valid. The remainder of this section will first describe graphical models based on cross-sectional data, in which cases can be assumed to be independent, followed by a description of dependent cases for a single subject ($N = 1$) as well as for multiple subjects ($N > 1$).

6.3 The Gaussian Graphical Model

In cross-sectional data, every subject is only measured once on a set of response items. In this case, as described above, we can reasonably assume that cases are independent and thus characterize the likelihood as factorized over subjects.

[2] Data cannot be represented as a box (Cattell, 1988), as would be the case if subjects were all measured at fixed measurement occasions (e.g., at baseline, one week after treatment, etc.).

Because only one observation per subject is available, however, we cannot expect to estimate subject-specific mean and variance–covariance structures. It is typically assumed that the cases all share the same distribution. That is,

$$\boldsymbol{Y}^{(p)} \sim N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right) \forall_p,$$

in which $\forall_p$ should be read as "for all subjects." Now, the full likelihood can be readily obtained and the mean vector $\boldsymbol{\mu}$ and variance–covariance matrix $\boldsymbol{\Sigma}$ can reliably be estimated using MLE, least squares estimation, or Bayesian estimation.

Our focus point is on $\boldsymbol{\Sigma}$. Because we assume multivariate normality, $\boldsymbol{\Sigma}$ encodes all the information necessary to determine how the observed measures relate to one another. It is to this end that great effort has been made to further model the structure of $\boldsymbol{\Sigma}$. Elements of this variance–covariance matrix can be standardized to *correlation coefficients*, allowing researchers to investigate marginal pairwise associations. This matrix, however, encodes more than just marginal associations. The *Schur complement* (Ouellette, 1981) shows that all conditional distributions of a set of variables, given another set of variables, can be obtained from blocks of $\boldsymbol{\Sigma}$. Therefore, in order to discover dynamics in psychological data, investigating the structure of $\boldsymbol{\Sigma}$ is of great importance.

However, we will not focus on $\boldsymbol{\Sigma}$ in this chapter but rather on its inverse—the *precision matrix* $\boldsymbol{K}$,

$$\boldsymbol{K} = \boldsymbol{\Sigma}^{-1},$$

also know as the GGM (Lauritzen, 1996). Of particular importance is that the standardized elements of the precision matrix encode partial correlation coefficients of two variables given all other variables:

$$\text{Cor}\left(Y_i, Y_j \mid \boldsymbol{Y}_{-(i,j)}\right) = -\frac{\kappa_{ij}}{\sqrt{\kappa_{ii}}\sqrt{\kappa_{jj}}},$$

in which κ_{ij} denotes an element of $\boldsymbol{K}$, and $\boldsymbol{Y}_{-(i,j)}$ denotes the set of variables without i and j (we dropped the person superscript for notational clarity). These partial correlations can be used as *edge weights* in a weighted network. Each variable Y_i is represented as a node, and connections (edges) between these nodes represent the partial correlation between two variables. When drawing such a network, positive partial correlations are typically visualized with green edges and negative partial correlations with red edges, and the absolute strength of a partial correlation is represented by the width and saturation of an edge (see Chapter 9). 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 other variables in the dataset.

Figure 6.1 shows a hypothetical example of such a GGM in psychology. Three nodes represent if someone is able to concentrate well, if someone is fatigued, or if someone is suffering from insomnia. 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. The GGM shown in Figure 6.1

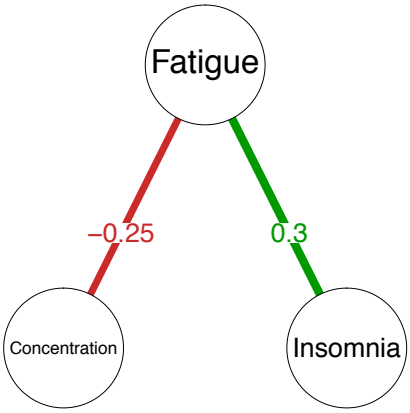

Figure 6.1: A hypothetical example of a GGM on psychological variables. Nodes represent someone's ability to concentrate, someone's level of fatigue, and someone's level of insomnia. Connections between the nodes, termed *edges*, represent partial correlation coefficients between two variables after conditioning on the third. Green 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.

can be interpreted in three different ways: (a) potential causal relationships, (b) predictive effects and predictive mediation, and (c) genuine mutual interactions.

First, a GGM can be taken to show potential causal relationships because the structure can be equivalent to three causal structures (Pearl, 2000):

1. Concentration → Fatigue → Insomnia
2. Concentration ← Fatigue → Insomnia
3. Concentration ← Fatigue ← Insomnia

In these structures, → denotes that what is on the left side of the arrow causes what is on the right side of the arrow. In observational data without temporal information, distinguishing between these models beyond only identifying the conditional independency is not possible. Thus, we may not know exactly why conditioning on fatigue leads to insomnia and concentration being independent—this finding may represent the smoke of a figurative causal fire. With more variables, the number of potential equivalent causal models can increase drastically (MacCallum et al., 1993).

The GGM is useful for generating hypotheses that can, at least in principle, later be experimentally tested. Specifically, the causal structures above hypothe-

size what happens when you intervene on the nodes (Pearl, 2000). For example, if we observe that someone is less able to concentrate after being fatigued in an experiment, the relationship fatigue → concentration appears more plausible. If we also observe that the person does not become more fatigued after we have experimentally impaired his or her ability to concentrate, the reverse causal relationship becomes less plausible.

Second, an edge in a GGM indicates that one node predicts a connected node after controlling for all other nodes in the network. This can be shown in the relationship between coefficients obtained from least-squares prediction and the inverse variance–covariance matrix. Let $\boldsymbol{\Gamma}$ represent an $I \times I$ matrix with zeros on the diagonal. Each row of $\boldsymbol{\Gamma}$, without the diagonal element $\boldsymbol{\gamma}_{i.-(i)}$, contains the regression coefficients obtained in

$$y_i = \tau + \boldsymbol{\gamma}_{i.-(i)}\boldsymbol{y}_{-(i)} + \varepsilon_i. \tag{6.1}$$

As such, γ_{ij} encodes how well the jth variable predicts the ith 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, γ_{ij} is proportional to γ_{ji}. There is a direct relationship between these regression coefficients and the inverse variance–covariance matrix (Meinshausen & Bühlmann, 2006). Let $\boldsymbol{D}$ denote a diagonal matrix on which the ith diagonal element is the inverse of the ith residual variance: $d_{ii} = 1/\text{Var}(\varepsilon_i)$. Then, it can be shown (Pourahmadi, 2011) that

$$\boldsymbol{K} = \boldsymbol{D}\left(\boldsymbol{I} - \boldsymbol{\Gamma}\right). \tag{6.2}$$

Thus, γ_{ij} is proportional to κ_{ij}. A zero in the inverse variance–covariance matrix indicates that one variable does not predict another variable. Consequently, the network tells us something about the extent to which variables predict each other. In the case of Figure 6.1, the network demonstrates that both insomnia and fatigue as well as fatigue and concentration predict each other. This does not mean that knowing someone's level of fatigue does not say anything about that person's concentration problems—because these nodes are connected via an indirect path, they may correlate with each other—but merely that fatigue mediates this predictive effect. When someone's level of fatigue is known, also knowing that person's level of insomnia does not add any predictive value to that person's ability to concentrate.

Third, the network structure found in a GGM can be interpreted as showing genuine mutual causation between two nodes of the network—manipulating one node can affect the other and vise versa. Mathematically, the GGM can be shown to have the same form as the Ising model (Ising, 1925) from statistical physics (see Chapter 8), except that the Ising model only models binary data and therefore has a different normalizing constant. This is because both models are part of a class of models, called pairwise Markov random fields (Lauritzen, 1996; Murphy, 2012), which have been extensively used to model complex behavior in physical systems. For example, the Ising model represents particles with nodes and the distance between particles with edges. Particles, in essence very small magnets, are then modeled to have their north pole face up or down. Particles tend to

oriented themselves randomly at normal temperatures but align at low temperatures. Because particles tend to align at low temperatures, one particle being aligned somehow causes an adjacent particle to align in that same way and vice versa. These relationships are naturally symmetric (see Chapter 8). Applying the analogy of the Ising model for particles to the GGM shown in Figure 6.1, we could say that these symptoms tend to be in the same state (of alignment) if there is a positive connection between them and if they tend to be in different states if there is a negative connection. In this system, a person suffering from fatigue could also suffer from insomnia as well as concentration problems (van Borkulo et al., 2014).

Estimation

The maximum likelihood solution of $\boldsymbol{K}$ can readily be obtained by standardizing the inverse sample variance–covariance matrix and by multiplying all off-diagonal elements by -1. An interesting problem pertains to elements of $\boldsymbol{K}$ which are close to, but not exactly, zero. In the interest of parsimony, researchers may want to remove these edges and obtain conditional independence with fewer parameters in the model. One way to obtain this is to use the sampling distribution of the partial correlation coefficients represented in $\boldsymbol{K}$ to test if edges are significantly different from zero. The network can then be thresholded by removing the nonsignificant edges (by fixing them at zero). Alternatively, lengthy model search algorithms can be applied to iteratively add and remove edges. In recent literature, it has become increasingly popular to use regularization techniques, such as penalized MLE, to jointly estimate model structure and parameter values (van Borkulo et al., 2014; see also Chapter 2). In particular, 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 et al., 2008; Meinshausen & Bühlmann, 2006; Yuan & Lin, 2007). A particularly fast 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). In addition, glasso utilizes a tuning parameter that controls the sparsity of the network: A sparse network is one with few edges (i.e., $\boldsymbol{K}$ contains mostly zeros). The tuning parameter 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 (Foygel & Drton, 2010; see also Chapter 2) and is currently the dominant method for estimating the GGM in psychological data (see also Chapter 2 for an introduction to this methodology aimed at empirical researchers).

Figure 6.2 shows an example of a GGM estimated using glasso in combination with EBIC model selection. This network was estimated on the `bfi` dataset from the *psych* R package (Revelle, 2010). This dataset contains the responses of 2,800 people on 25 items designed to measure the Big Five personality traits (McCrae & Costa, 1997). The network shows many meaningful connections, such as "make friends easily" being linked to "make people feel at ease," "don't talk

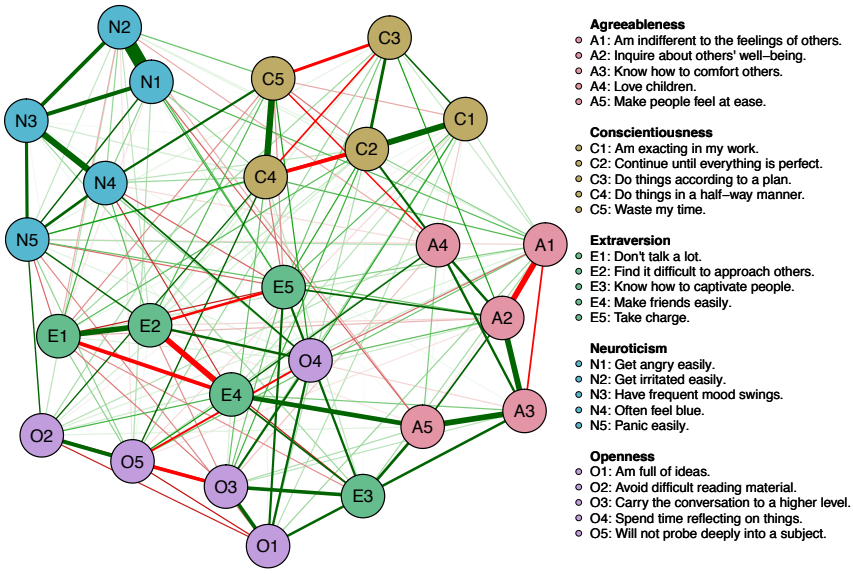

Figure 6.2: An example of a network model estimated on the BFI personality dataset from the psych package in R (cross-sectional data, N = 2,800). Nodes represent variables (in this case personality inventory items) and edges between the nodes represent partial correlation coefficients. The network was estimated using the glasso in combination with EBIC model selection, using the `EBICglasso` function in the *qgraph* package.

a lot" being linked to "find it difficult to approach others," and "carry the conversation to a higher level" being linked to "know how to captivate people." For a detailed discussion on the interpretation of such models in personality research (see Chapter 10).

The GGM can be estimated by inverting and standardizing the sample variance–covariance matrix, which can be done in the open-source statistical programming language R (R Core Team, 2016) by using the *corpcor* (Schäfer et al., 2015) or *qgraph* (Epskamp et al., 2012) R package. The *qgraph* package also supports thresholding via significance testing or false discovery rates. The glasso algorithm is implemented in the packages glasso (Friedman et al., 2014) and *huge* (Zhao et al., 2015). The huge package also allows for selection of the tuning parameter using cross-validation or EBIC. The EBIC-based tuning parameter selection with the glasso package, using only a variance–covariance matrix as input, has been implemented in the *qgraph* package. The *parcor* package (Krämer et al., 2009) implements other LASSO variants for estimating the GGM. Finally, fitting an estimated GGM to data can be done in the R packages *ggm* (Marchetti, Drton, & Sadeghi, 2015) and *lvnet* (see Chapter 7).

6.4 When Cases Are Not Independent: $n = 1$

Different kinds of data in psychology occur when only one subject is measured several times. The likelihood is fully characterized by the likelihood of the single subject:

$$f(\boldsymbol{y} \mid \boldsymbol{\xi}) = f(\boldsymbol{y}^{(1)} \mid \boldsymbol{\xi}).$$

We need to characterize $f(\boldsymbol{y}^{(p)} \mid \boldsymbol{\xi})$. In repeated measures of psychological constructs—assuming a reasonably short interval between consecutive measurements is typical in ESM studies—we cannot reasonably assume that cases are independent. For example, suppose we measured Peter multiple times on his level of fatigue, measured on a scale from 1 (*not at all fatigued*) to 10 (*extremely fatigued*). Suppose we know Peter has an average fatigue level of 5 with a standard deviation of 1. Knowing that Peter scored a 2 at some time point, we can make a better prediction regarding the level of Peter's fatigue at the next time point (it is probably still low a few hours later) than if we only knew his mean and standard deviation, which would predict this level most likely to be somewhere between 3 and 7. This is because someone's fatigue, like most psychological and physiological states, is likely to show some stability over a time interval of several hours.

It is important to note that $f(\boldsymbol{y} \mid \boldsymbol{\xi})$ cannot be computed by multiplying the marginal likelihoods of every response. Instead, we now need to express the full joint likelihood. When we drop superscript $^{(1)}$ denoting the single subject, this becomes

$$f(\boldsymbol{y} \mid \boldsymbol{\xi}) = f\left(\boldsymbol{y}_T \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{T_p-1}, \boldsymbol{\xi}\right) \cdots f(\boldsymbol{y}_3 \mid \boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{\xi})\, f(\boldsymbol{y}_2 \mid \boldsymbol{y}_1, \boldsymbol{\xi})\, f(\boldsymbol{y}_1 \mid \boldsymbol{\xi}).$$

The model above, although fully characterizing the joint likelihood, is not estimable without stringent assumptions, so we make the following assumptions.

1. The joint probability distribution can be factorized according to a graph.
2. The conditional probability distributions are stable and independent of t.
3. The first measurements are treated as exogenous and not modeled.
4. The conditional distributions are multivariate normal.

The first assumption is that the time series follows some graph structure such that the factorization of the joint probability distribution can be made easier. Figure 6.3 shows three such potential graph structures. The first panel shows the *Lag 0* factorization, in which each observation is assumed to be independent of others. As described above, although this is a sparse representation, the Lag 0 model is not plausible in most time-series psychological datasets. As such, we could use the graph factorization of the second panel of Figure 6.3 instead, denoting the *Lag 1* factorization

$$f(\boldsymbol{y} \mid \boldsymbol{\xi}) = f\left(\boldsymbol{y}_T \mid \boldsymbol{y}_{T_p-1}, \boldsymbol{\xi}\right) \cdots f(\boldsymbol{y}_2 \mid \boldsymbol{y}_1, \boldsymbol{\xi})\, f(\boldsymbol{y}_1 \mid \boldsymbol{\xi}).$$

This is a powerful factorization because it does not assume that measurement are independent of one another. For example, the Lag 1 factorization does not

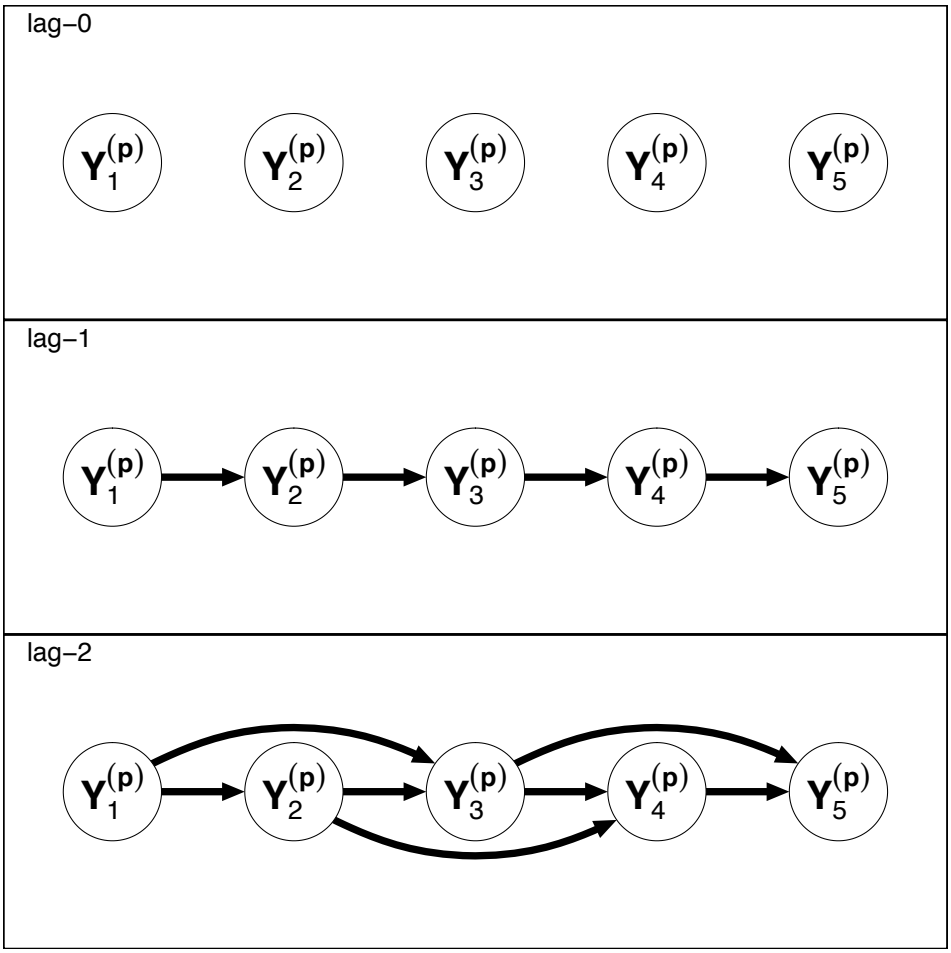

Figure 6.3: Here are three possible graph factorizations of the within-subject likelihood of subject p. Each node represents a vector of measurements at some time point. The top panel shows the Lag 0 factorization, indicating that cases are independent. Because this is usually not a plausible assumption in psychology, we can instead use another factorization. The middle panel shows the Lag 1 factorization, indicating that cases are independent given only the previous case. The bottom panel shows the Lag 2 factorization, indicating that cases are independent given the past two cases.

assume there are no *Lag 2* correlations—correlations between $\boldsymbol{y}_{t-2}$ and $\boldsymbol{y}_t$ for any t—but instead that these Lag 2 correlations can fully be explained by the Lag 1 interactions, $\boldsymbol{y}_{t-2} \perp\!\!\!\perp \boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}$ for all t. However, more flexible models can be specified as well, such as the Lag 2 model shown in the last panel of Figure 6.3. Because the number of observations in ESM studies is often relatively low (people cannot be expected to fill out tens of questionnaires daily over the course of several days), adding complexity to the model requires more observations to reliably estimate parameter values. Therefore, we will only describe the Lag 1 factorization in the remainder of this chapter because this factorization is the simplest that also controls for the most obvious violations of independence between consecutive cases.

The second assumption is that the conditional probability distributions do not depend on t and are thus stable over time. This is called the assumption of *stationarity*. Using this assumption, the time series of a single subject now features multiple observations of the same relationship (e.g., the Lag 1 relationship), making the model estimable. Combining this with the third assumption of first measurements being treated as exogenous, and thereby not modeled, renders the probability distribution simple and straightforward. For example, the Lag 1 factorization then becomes

$$f\left(\boldsymbol{y} \mid \boldsymbol{y}_1, \boldsymbol{\xi}\right)_{\text{lag}-1} = \prod_t f\left(\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}, \boldsymbol{\xi}\right).$$

The assumption of stationarity is not trivial because people can develop over time. In a typical ESM study, data are gathered in a relatively short time span (e.g., a few weeks). Assuming a person stays relatively stable in such a short interval is much more plausible. It is therefore important to note that the assumption of stability does not assume a person never changes, merely that the person's scores are distributed similarly in a relatively short time span (Fleeson, 2001).

Finally, we assume that these conditional distributions are multivariate normal. Using the Schur complement, these distributions can be shown to be equivalent to a linear regression model with correlated multivariate normal residuals. We can, without loss of information, center the lagged predictors such that we obtain

$$\boldsymbol{Y}_t \mid \boldsymbol{y}_{t-1} \sim N\left(\boldsymbol{\mu} + \boldsymbol{B}\left(\boldsymbol{y}_{t-1} - \boldsymbol{\mu}\right), \boldsymbol{\Theta}\right),$$

in which $\boldsymbol{B}$ denotes an $I \times I$ matrix of *temporal* effects, $\boldsymbol{\mu}$ denotes the I length vector of stationary means, and $\boldsymbol{\Theta}$ denotes the $I \times I$ variance–covariance matrix conditional on the previous time point, which we will term *contemporaneous* effects. This model is also known as the VAR because it can be seen as a multivariate multiple regression on the previous time point. VAR has become popular in psychology because $\boldsymbol{B}$ encodes temporal prediction: Element $\beta_{ij}^{(p)}$ being nonzero means that Y_{ti} is predicted by $Y_{t-1,j}$. Such a temporal prediction is termed *Granger causality* in the economic literature (Eichler, 2007; Granger, 1969) because it satisfies at least the temporal requirement for causation (i.e., cause must precede the effect).

This model implies a stationary distribution of cases:

$$\boldsymbol{y}_t \sim N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right),$$

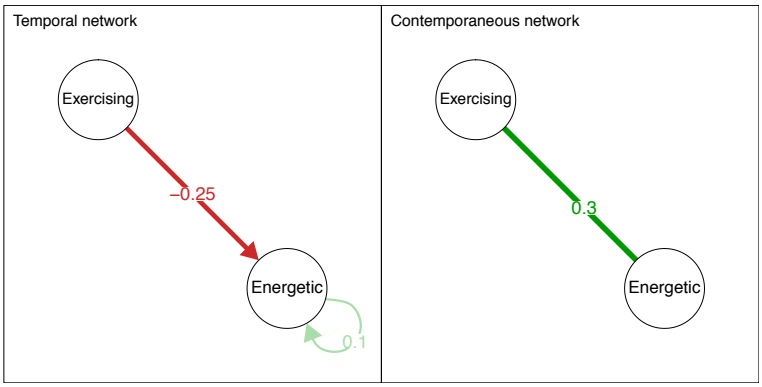

Figure 6.4: A hypothetical example of two network structures obtained from a vector-autoregression model. The network on the left indicates the temporal network, demonstrating that a variable predicts another variable at the next time point. The network on the right indicates the contemporaneous network, demonstrating that two variables predict each other in the same time point.

in which $\boldsymbol{\Sigma}$ can be obtained from $\boldsymbol{B}$ and $\boldsymbol{\Theta}$, making use of the vectorization operator Vec and the Kronecker product $\otimes$:

$$\text{Vec}\left(\boldsymbol{\Sigma}\right) = \left(\boldsymbol{I} - \boldsymbol{B} \otimes \boldsymbol{B}\right)^{-1} \text{Vec}\left(\boldsymbol{\Theta}\right). \tag{6.3}$$

A proof of Equation (6.3) is beyond the scope of this chapter and can be requested from the author. It is important to note that in addition to the estimation of temporal effects, the VAR model can be used to obtain the GGM (the inverse of the variance–covariance matrix described above) for nonindependent cases. We can further note that if cases are independent, $\boldsymbol{B} = \boldsymbol{O}$ and subsequently $\boldsymbol{\Sigma} = \boldsymbol{\Theta}$. Thus, the GGM is a special case of the VAR model. 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.

Contemporaneous Causation

In order to disentangle the temporal and contemporaneous relationship, it is best not to combine them into a single GGM model but rather to investigate them separately. Following Wild et al. (2010), the inverse of $\boldsymbol{\Theta}$ can be standardized to a GGM model encoding residual partial correlation coefficients, which can be drawn in a network model. As a result, the VAR model returns two network models: the *temporal network*, a directed network indicating temporal prediction or Granger causality, and the *contemporaneous network*, a partial-correlation network of effects in the same window of measurement. Both network structures can highlight potential causal pathways. In psychology, there will likely be many causal relationships that occur much faster than the lag interval in 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.

Figure 6.4 shows a hypothetical example of the two network structures obtained in a VAR analysis. The left panel shows the temporal network (a graphical representation of $\boldsymbol{B}$). This network 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 as well. The contemporaneous network in the right panel (a graphical representation of the GGM based on $\boldsymbol{\Theta}$) shows a plausible reverse relationship: Whenever this person exercised, he or she felt more energetic in the same measurement occasion.

Estimation

We can estimate the VAR model by specifying it as a regression model. Without the loss of information, we can center the variables to have a mean of zero. The corresponding multivariate regression model then becomes

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{B}\boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t \\ \boldsymbol{\varepsilon}_t &\sim N(\boldsymbol{0}, \boldsymbol{\Theta}). \end{aligned}$$

Alternatively, the VAR model can be estimated in steps using separate *univariate* models for every variable:

$$\begin{aligned} y_{ti} &= \boldsymbol{\beta}_i \boldsymbol{y}_{t-1} + \varepsilon_{ti} \\ \varepsilon_{ti} &\sim N(0, \sqrt{\theta_{ii}}), \end{aligned}$$

in which $\boldsymbol{\beta}_i$ denotes the ith row of $\boldsymbol{B}$. Figure 6.5 shows the difference between univariate and multivariate estimation. In univariate estimation, every model contains a different subset of the parameters of interest. In addition, the contemporaneous covariance θ_{12} is not obtained in any of the models and needs to be estimated post hoc by correlating the residuals of both regression models.

Abegaz and Wit (2013) proposed to apply LASSO estimation of $\boldsymbol{B}$ and $\boldsymbol{\Theta}$ using the multivariate regression with the covariance estimation (MRCE) algorithm described by Rothman et al. (2010). MRCE involves iteratively optimizing $\boldsymbol{B}$ using cyclical-coordinate descent and $\boldsymbol{K}$ using the glasso algorithm (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, 2015), which was designed to estimate the model on the psychological data of a single subject. The *graphicalVAR* package also allows for unregularized multivariate estimation.

6.5 When Cases Are Not Independent: $n > 1$

When multiple subjects are measured, we need to characterize the likelihood for every subject. Using the assumptions described above, we can model the time

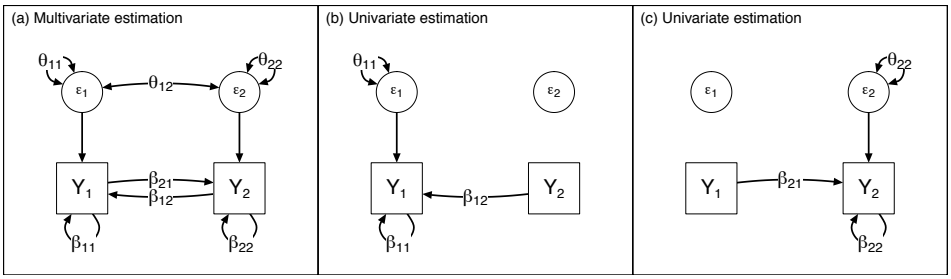

Figure 6.5: A multivariate and univariate estimation of a VAR model with two observed variables. Intercepts are not shown to improve clarity. Panel A shows that in multivariate estimation, the entire model is estimated at once whereas Panels B and C show that in sequential estimation two separate models are estimated.

series of each subject with a subject-specific VAR model:

$$\boldsymbol{y}_t^{(p)} \mid \boldsymbol{y}_{t-1}^{(p)} \sim N\left(\boldsymbol{\mu}^{(p)} + \boldsymbol{B}^{(p)}\left(\boldsymbol{y}_{t-1}^{(p)} - \boldsymbol{\mu}^{(p)}\right), \boldsymbol{\Theta}^{(p)}\right).$$

Often, however, researchers are not interested in the dynamics of a single participant but rather in the generalizability of dynamics over multiple subjects. To this end, researchers may want to estimate the average effects and interindividual differences of such intraindividual dynamics. We can model these by using the language of multilevel modeling (Bringmann et al., 2013). For each parameter, we denote the average effect as the *fixed effects*, $\boldsymbol{f}$, and the person-level deviance from this mean as the *random effects*, $\boldsymbol{R}^{(p)}$, with the realization $\boldsymbol{r}^{(p)}$. Using this notation, the parameter vector of a single subject becomes

$$\begin{bmatrix} \boldsymbol{\mu}^{(p)} \\ \text{Vec}\left(\boldsymbol{B}^{(p)}\right) \\ \text{Vech}\left(\boldsymbol{\Theta}^{(p)}\right) \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}_{\boldsymbol{\mu}}^{(p)} \\ \boldsymbol{f}_{\boldsymbol{B}}^{(p)} \\ \boldsymbol{f}_{\boldsymbol{\Theta}}^{(p)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{r}_{\boldsymbol{\mu}}^{(p)} \\ \boldsymbol{r}_{\boldsymbol{B}}^{(p)} \\ \boldsymbol{r}_{\boldsymbol{\Theta}}^{(p)} \end{bmatrix},$$

in which Vec stacks the columns of a matrix, and Vech does the same but only takes the upper-triangular elements including the diagonal. The random effects are centered on zero:

$$\mathcal{E}\left(\begin{bmatrix} \boldsymbol{R}_{\boldsymbol{\mu}} \\ \boldsymbol{R}_{\boldsymbol{B}} \\ \boldsymbol{R}_{\boldsymbol{\Theta}} \end{bmatrix}\right) = \boldsymbol{0},$$

such that the fixed effects reflect the population means of the parameters. The variance of the random effects can be interpreted as the *individual differences*.

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.

Modeling and estimating a random distribution for the contemporaneous variance–covariance matrix is still a topic for future research and not readily implemented in open-source software. This is mainly because these matrices must be positive definite. We cannot simply assign normal distributions to elements of the contemporaneous (partial) variance–covariance matrix because doing so might lead to nonzero probability of matrices that are not positive definite. Therefore, we do not define this distribution here and merely state that there is some population mean for its elements, $\boldsymbol{f_\Theta}$. We assume the means and lagged regression parameters to be normally distributed:

$$\begin{bmatrix} \boldsymbol{R_\mu} \\ \boldsymbol{R_B} \end{bmatrix} \sim N\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{\Omega_\mu} & \boldsymbol{\Omega_{\mu B}} \\ \boldsymbol{\Omega_{B\mu}} & \boldsymbol{\Omega_B} \end{bmatrix}\right).$$

To summarize, the multilevel VAR model makes use of the following parameters for all subjects:

- $\boldsymbol{f_B}$: The average within-person temporal relationships between consecutive time points.
- $\boldsymbol{f_\Theta}$: The average within-person contemporaneous relationships.
- $\boldsymbol{\Omega_\mu}$: The between-person relationships between observed variables.
- $\boldsymbol{\Omega_{\mu B}}$ and $\boldsymbol{\Omega_B}$: Individual differences between the temporal relationships and other temporal relationships or the means. Of particular interest is $\sqrt{\text{Diag}\left(\boldsymbol{\Omega_B}\right)}$, which shows the individual differences of each temporal relationship (Bringmann et al., 2013).

For any researcher interested in investigating results of particular subjects, the subject-specific structures are also of interest:

- $\boldsymbol{\mu}^{(p)}$: The stationary means of subject p.
- $\boldsymbol{B}^{(p)}$: The within-person temporal relationships of subject p.
- $\boldsymbol{\Theta}^{(p)}$: The within-person contemporaneous relationships of subject p.

A New Look at Cross-Sectional Analysis

Such variance decomposition exposes a major limitation of cross-sectional analyses. In cross-sectional data, each subject is only measured once: $T_1 = T_2 = \ldots = T_P = 1$. This can be seen as a special case of the multilevel VAR model in which the Lag 0 factorization is used to model the single response of a subject. This single response can then be written as the stationary mean of person p and the deviation from that mean:

$$\begin{aligned}\boldsymbol{y}_1^{(p)} &= \boldsymbol{\mu}^{(p)} + \boldsymbol{\varepsilon}_1^{(p)} \\ \boldsymbol{\varepsilon}_1^{(p)} &\sim N(\mathbf{0}, \boldsymbol{\Theta}^{(p)}).\end{aligned}$$

It is immediately clear that $\boldsymbol{\Theta}^{(p)}$ cannot be estimated from a single set of responses. Moreover, even if we assume that within-person contemporaneous effects are equal across people and drop the superscript (p), this still leaves us without an estimable model because $\boldsymbol{R_\mu}$ is also assumed to be normally distributed. Therefore, we get

$$\begin{aligned}\boldsymbol{\varepsilon}_1^{(p)} &\sim N(\mathbf{0}, \boldsymbol{\Theta}) \\ \boldsymbol{R_\mu} &\sim N(\mathbf{0}, \boldsymbol{\Omega_\mu}).\end{aligned}$$

In no way do we know if deviations from the grand mean are due to the within-person variance in $\boldsymbol{\Theta}$ or the between-person variance in $\boldsymbol{\Omega_\mu}$. Thus, in cross-sectional analysis, within- and between-subject variances are not distinguishable. We can estimate $\boldsymbol{\Theta}$ by assuming $\boldsymbol{\Omega_\mu} = \boldsymbol{O}$, or we can estimate $\boldsymbol{\Omega_\mu}$ by assuming $\boldsymbol{\Theta} = \boldsymbol{O}$. Both assumptions lead to the exact same estimates. This does not mean that cross-sectional analysis is unusable by default because the obtained structure can highlight potential causal relationships between variables; however, it cannot disentangle between-subject relationships from short-term, within-subject relationships (Hamaker, 2012).

Between-Subjects Causation

The variance–covariance matrix $\boldsymbol{\Omega_\mu}$ encodes how variables relate to one another across subjects and can be modeled using a GGM network of partial correlation coefficients. As such, the multilevel VAR model returns three types of network structures describing relationships between observed variables. In addition to the temporal and contemporaneous network fixed effects (the average temporal and contemporaneous network) and random effects (the personal deviations from these averages), the multilevel VAR model also returns a *between-subjects network*—the network structure between stationary means of subjects based on $\boldsymbol{\Omega_\mu}$.

Hamaker (2012) described an example of how within- and between-person effects can strongly differ from each other. Suppose we let people write several texts, and we measure the number of spelling errors they made and the number of words per minute they typed (typing speed). We would expect the seemingly paradoxical three network structures shown in Figure 6.6. First, we would not expect the temporal network to show any relationships. There is no logical reason to assume that observing someone type a text faster than his or her average has any influence on the number of spelling errors in the next text. Second, we expect

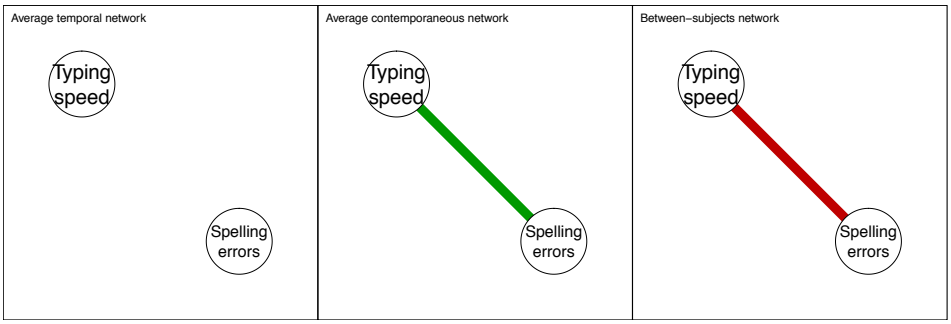

Figure 6.6: A hypothetical example three network structures obtained from a multilevel vector-autoregression model. The network on the left indicates the temporal network, showing that a variable predicts another variable at the next time point. The network in the middle indicates the contemporaneous network, showing that personal deviations from the means predict each other at the same time point. Finally, the network on the right indicates the between-subjects network, showing how the *means* of different subjects relate to one another

a positive relationship in the contemporaneous network. When a person types faster than his or her average typing speed, that person will make more spelling errors. Finally, we expect a negative relationship in the between-person 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 court room stenographer) and therefore are less likely to make a lot of spelling errors, compared to someone who types infrequently.

The different ways of thinking about the effects of manipulations in time-series models can be organized in terms of recently developed 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 would lead 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 $\text{Do}\left(X = x\right)$, and the causal effect on Y is formally expressed as $\mathbb{E}\left(Y \mid \text{Do}\left(X = x\right)\right)$. Pearl distinguished this from the classical statistical association, in which no intervention is present, and we get the ordinary regression $\mathbb{E}\left(Y \mid \text{See}\left(X = x\right)\right)$. This is a useful notation because it immediately raises the important point that there is a difference between doing and seeing, which of course parallels the classic distinction between experimental and correlational research (Cronbach & Meehl, 1955).

Cashing out causal effects in terms of interventions is useful to get a grip on the causal information in the different GGMs defined in this chapter. In a time-series model, interventions on variables can be conceptualized in different ways. In particular, consider the intervention $\text{Do}\left(X = x\right)$. We can think of this in terms of a random shock to the system, which sets X to value x on a particular time point and evaluates the effect on another variable Y on the next time point (or series of time points as in continuous time models). If we want to gauge this

type of causal relationship, we may look at the within-subjects VAR model. To take Hamaker's (2014) example, say we want to know what would happen to the number of typing errors: If a researcher forced a given person to type very fast, that researcher would need to evaluate the within-person data, which would show a negative association between typing speed and typing errors. Between-subject data would be misleading because individual differences would probably yield a positive correlation between speed and accuracy—faster typists are likely to be more accurate typists.

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, we should consider the behavior of the system on the changes of the intercept of X. Clearly, in order to evaluate this type of intervention, the within-subject time-series model is useless (as per stationarity). However, the between-subjects model may contain important clues because it contains the relationships between the long-term averages across people. Thus, if we want to gauge the effect of a long-term change (most plausibly conceptualized as a change in intercept), the between-subjects model is a better guide. In terms of Hamaker's (2014) example, if we are interested in the effect of changing someone's typing speed structurally (e.g., by training a person systematically), our preferred source of causal information would likely lie in the between-subjects model because the parameters of the within-subjects model would undoubtedly lead to the wrong conclusion.

Estimation

A straightforward technique is to estimate separate VAR models for each subject. Afterwards, fixed effects (i.e., average effects in the population) can be estimated by pooling the parameters and averaging them. This estimation technique is relatively fast even for large models, but it requires a high number of observations per person. As described above, an alternative is to use multilevel modeling (Hamaker, 2012). The benefit of the latter approach is that instead of estimating the VAR model in each subject, only the fixed effects and variance–covariance of the random effects need to be estimated. This can be done by integrating over the distribution of the random effects or by specifying the model using hierarchical Bayesian Monte-Carlo sampling methods (Gelman & Hill, 2006; Schuurman, Grasman, & Hamaker, 2016). Here, we propose a novel two-step, multilevel estimation procedure that estimates the fixed effects for the temporal, between-subjects, and contemporaneous networks as well as the random effects for the temporal and contemporaneous networks. The contemporaneous networks are estimated in a second step, by analyzing the residuals of the first step.

Temporal network. Although multivariate multilevel estimation is possible in theory, it is computationally expensive in practice. For example, when we want to explore potential dynamics in medium-sized ESM datasets on around 10 to 20 variables, multivariate multilevel estimation becomes very slow in both MLE

and Bayesian estimation. Therefore, we only describe univariate estimation procedures (Bringmann et al., 2013). Because the joint conditional distribution of $\boldsymbol{y}_t^{(p)} \mid \boldsymbol{y}_{t-1}^{(p)}$ 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:

$$y_{ti}^{(p)} \mid \boldsymbol{y}_{t-1}^{(p)} \sim N\left(\mu_i^{(p)} + \boldsymbol{\beta}_i^{(p)}\left(\boldsymbol{y}_{t-1}^{(p)} - \boldsymbol{\mu}^{(p)}\right), \theta_{ii}^{(p)}\right),$$

in which $\boldsymbol{\beta}_i$ indicates the row vector of the ith row of $\boldsymbol{B}^{(p)}$. 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, $\bar{\boldsymbol{y}}^{(p)}$, can be taken as a substitute for $\boldsymbol{\mu}^{(p)}$ (Hamaker & Grasman, 2014). The model then becomes a univariate multilevel regression model with within-subject centered predictors, estimable by functions such as the `lmer` in `lme4` (Bates, Mächler, Bolker, & Walker, 2015). The Level 1 model becomes

$$y_{ti}^{(p)} = \mu_i^{(p)} + \boldsymbol{\beta}_i^{(p)}\left(\boldsymbol{y}_{t-1}^{(p)} - \bar{\boldsymbol{y}}^{(p)}\right) + \varepsilon_{ti}^{(p)} \tag{6.4}$$

$$\varepsilon_{ti}^{(p)} \sim N(0, \theta_{ii}^{(p)}), \tag{6.5}$$

and the Level 2 model becomes

$$\begin{bmatrix} \mu_i^{(p)} \\ \boldsymbol{\beta}_i^{(p)} \end{bmatrix} = \begin{bmatrix} f_{\mu_i} \\ \boldsymbol{f}_{\boldsymbol{\beta}_i} \end{bmatrix} + \begin{bmatrix} r_{\mu_i}^{(p)} \\ \boldsymbol{r}_{\boldsymbol{\beta}_i}^{(p)} \end{bmatrix}$$
$$\begin{bmatrix} R_{\mu_i} \\ \boldsymbol{R}_{\boldsymbol{\beta}_i} \end{bmatrix} \sim N\left(\boldsymbol{0}, \begin{bmatrix} \omega_{\mu_i} & \boldsymbol{\Omega}_{\boldsymbol{\mu}\boldsymbol{\beta}_i} \\ \boldsymbol{\Omega}_{\boldsymbol{\beta}_i\boldsymbol{\mu}} & \boldsymbol{\Omega}_{\boldsymbol{\beta}_i} \end{bmatrix}\right).$$

Estimation of such univariate models only requires the numeric approximation of an $I + 1$ dimensional integral, which is much easier to compute. Therefore, 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, off-diagonal elements of $\boldsymbol{\Theta}^{(p)}$ and $\boldsymbol{\Omega}_{\boldsymbol{\mu}}$ as well as certain elements of $\boldsymbol{\Omega}_{\boldsymbol{\mu B}}$ and $\boldsymbol{\Omega}_{\boldsymbol{B}}$ are not obtained. 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*. The performance of orthogonal estimation, although the random effects are in reality correlated, is assessed in the simulation study below.

Between-subjects network. To obtain estimates of between-subject effects, Hamaker and Grasman (2014) suggest that the sample means of every subject, $\bar{\boldsymbol{y}}^{(p)}$ in Equation (6.4), can be included as predictors at the subject level. With this extension, the Level 2 model for the person-specific mean of the ith variable now becomes

$$\mu_i^{(p)} = f_{\mu_i} + \boldsymbol{\gamma}_{\mu,i}\bar{\boldsymbol{y}}_{-(i)}^{(p)} + r_{\mu_i}^{(p)}, \tag{6.6}$$

in which we use $\boldsymbol{\gamma}_{\mu,i}$ to denote the ith row (without the diagonal element i) of a $I \times I$ matrix $\boldsymbol{\Gamma}_{\mu}$, and $\bar{\boldsymbol{y}}^{(p)}_{-(i)}$ denotes the vector $\bar{\boldsymbol{y}}^{(p)}$ without the i-th element. Because $\bar{y}^{(p)}_i$ is itself an estimate of $\mu^{(p)}_i$, Equation (6.6) seems to take the form of Equation (6.1). As such, these estimates can be used, as seen in Equation (6.2), to estimate a GGM between the means (Lauritzen, 1996; Meinshausen & Bühlmann, 2006)—the between-subjects network. Due to the estimation in a multilevel framework, the resulting matrix will not be perfectly symmetric and must be made symmetric by averaging lower and upper triangular elements. Thus, each edge (i.e., partial correlation) in the between-subjects network is estimated by standardizing and averaging two regression parameters: the parameter denoting how well mean A predicts mean B and the regression parameter denoting how well mean B predicts mean A.

Contemporaneous network. 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 $\hat{\varepsilon_{ti}}^{(p)}$ denote the estimated residual of variable i at time point t of person p, and let $\hat{\boldsymbol{\varepsilon}}^{(p)}_{t,-(i)}$ denote the vector of residuals of all other variables at this time point. The Level 1 model then becomes

$$\hat{\varepsilon_{ti}}^{(p)} = \tau^{(p)}_i + \boldsymbol{\gamma}^{(p)}_{\varepsilon,i}\hat{\boldsymbol{\varepsilon}}^{(p)}_{t,-(i)} + \zeta^{(p)}_{ti}, \tag{6.7}$$

in which $\boldsymbol{\gamma}^{(p)}_{\varepsilon,i}$ represents the i-th row (without the diagonal element i) of a $I \times I$ matrix, $\boldsymbol{\Gamma}^{(p)}_{\mu}$, $\tau^{(p)}_i$ represents some intercept, and $\zeta^{(p)}_{ti}$ represents a residual. In the Level 2 model, we again assign a multivariate normal distribution to parameters $\tau^{(p)}_i$ and $\boldsymbol{\gamma}^{(p)}_{\varepsilon,i}$. It can be seen that Equation (6.7) also takes the form of Equation (6.1). Thus, this model can again be seen as the node-wise GGM estimation procedure. Estimates of both the person-specific and fixed-effects contemporaneous networks can be obtained by using Equation (6.2), where again the matrices need to be made symmetric by averaging upper and lower triangle elements. 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 this thresholding to be done. For the temporal networks, each edge is represented by one parameter and thus by one p value. The contemporaneous and between-subjects networks, however, are a function of two parameters that are standardized and averaged: a regression parameter for the multiple regression model of the first node and a regression parameter for the multiple regression model of the second node. As

such, for every edge, two p values are obtained. We can choose to retain edges of which at least one of the two p values is significant, termed the OR-rule, or we can choose to retain edges in which both p values are significant, termed the AND-rule (Barber, Drton, & Others, 2015).

Summary. In sum, the above described two-step estimation method proposes to estimate a multilevel model per variable, using within-person centered lagged variables as within-subject predictors and the sample means as between-subject predictors. These models can be used to obtain estimates for the temporal network and between-subjects network. In a second step, the contemporaneous networks can be estimated by estimating a second multilevel on the residuals of the first multilevel model. The *mlVAR* R package implements these methods (Epskamp, Deserno, & Bringmann, 2016). 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 "unique" (all residuals are used to obtain one GGM), "correlated" (second step multilevel model with correlated random effects), "orthogonal" (second step multilevel model with uncorrelated random effects), or "unique" (residuals are used to obtain a GGM per subject). The *mlVAR* package can also be used to plot the estimated networks, in which significance thresholding is used by default with a significance level of $\alpha = 0.05$.

6.6 Empirical Example

To provide an empirical example of the multilevel VAR methods described above, we reanalyzed the data of Mõttus et al. (2016). This data consists of two independent ESM samples, in which items tapping three of the five Five-Factor Model (McCrae & John, 1992) domains (neuroticism, extraversion, and conscientiousness) 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 2 hr. For more information about the samples and the specific questions asked, we refer readers to Mõttus et al. (2016).

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 lead 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 *mlVAR* package. 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.

Figure 6.7 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 6.4: 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 6.8 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 instance, we used orthogonal random effects to estimate the model. Figure 6.9 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.

6.7 Simulation Study

In this section, we present a simulation study to assess the performance of *ml-VAR* and 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; Foygel & Drton, 2010; see also Chapter 2). For this study, we simulated datasets of 10 variables, in which the fixed-effect temporal, contemporaneous, and between-subjects networks were simulated to be 50% sparse (i.e., containing only half the possible edges). A more detailed description of how the models were simulated can be read in the Appendix. We varied the number of subjects (50, 100, and 250) and the number of measurements per subject (25, 50, 75, and 100) and replicated

Figure 6.7: 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.

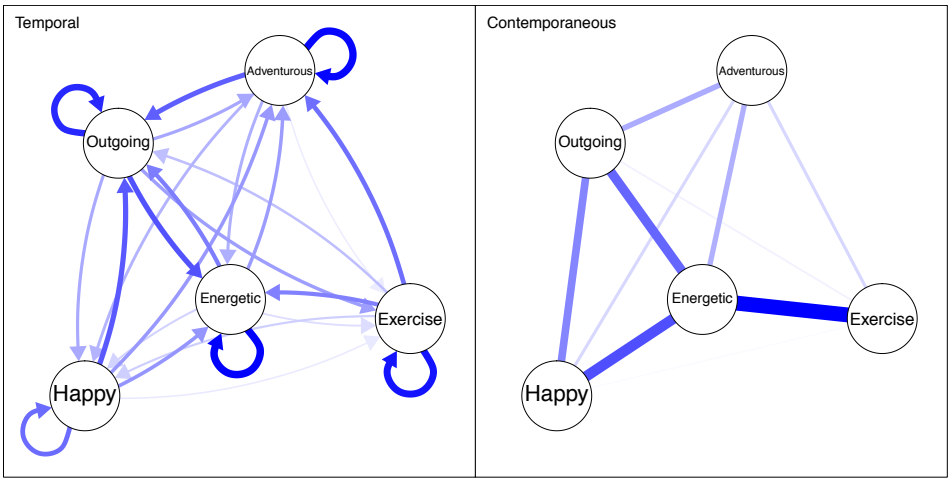

Figure 6.8: The networks showing the standard deviation of random effects in the temporal and contemporaneous networks. Due to scale differences, networks are plotted using different maximum values.

each condition 100 times. This led to a total number of 1,200 simulated datasets. In each dataset, we estimated a multilevel VAR model using orthogonal random effects.

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 random effects of the temporal and contemporaneous networks—because the between-subjects network does not have random effects. In addition, we assessed the performance of using significance in thresholding the network. We used the OR-rule in thresholding the fixed-effects contemporaneous and between-subjects network and removed in all the networks all edges not significant at $\alpha = 0.05$. In line with other studies on assessing how well a method retrieves the structure of a network (e.g., van Borkulo et al., 2014), we computed the *sensitivity* and *specificity*. The sensitivity (also termed *true positive rate*) is high when the method retains edges that are in the true network, and the specificity (also termed *true negative rate*) is high when the method does not retain edges that are not in the true model (i.e., models without edges that are, in reality, zero).

Figure 6.10 shows the results of the simulation study. It can be seen that performance was generally good. 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). Random-effect estimation was poorer but steeply increased with more measurements per subject. The between-subjects network was better estimated with more people. At low sample-sizes, the method lacked power to detect true edges (low sensitivity) but did not estimate false edges

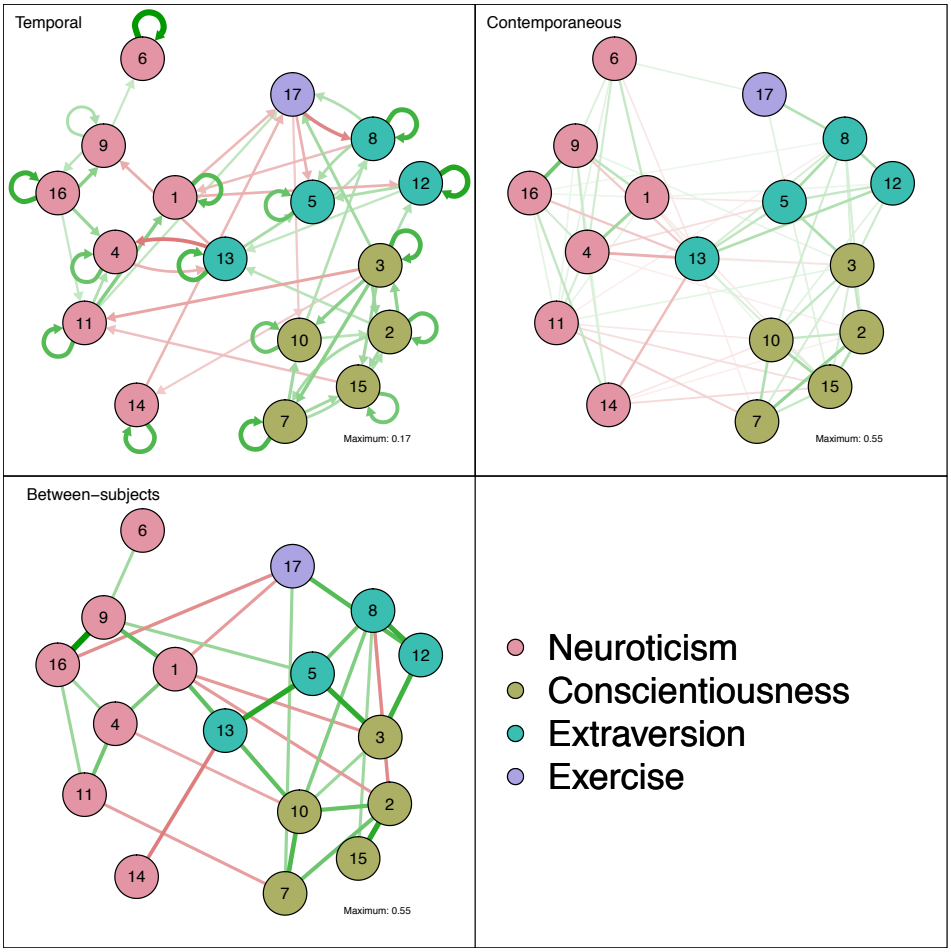

Figure 6.9: The estimated fixed effects of the three network structures based on all 17 variables administered. Only significant edges are shown. Legend: 1 = "Worried"; 2 = "Organized"; 3 = "Ambitious"; 4 = "Depressed"; 5 = "Outgoing"; 6 = "Self-Conscious"; 7 = "Self-Disciplined"; 8 = "Energetic"; 9 = "Frustrated"; 10 = "Focused"; 11 = "Guilty"; 12 = "Adventurous"; 13 = "Happy"; 14 = "Control"; 15 = "Achieved"; 16 = "Angry"; 17 = "Exercise."

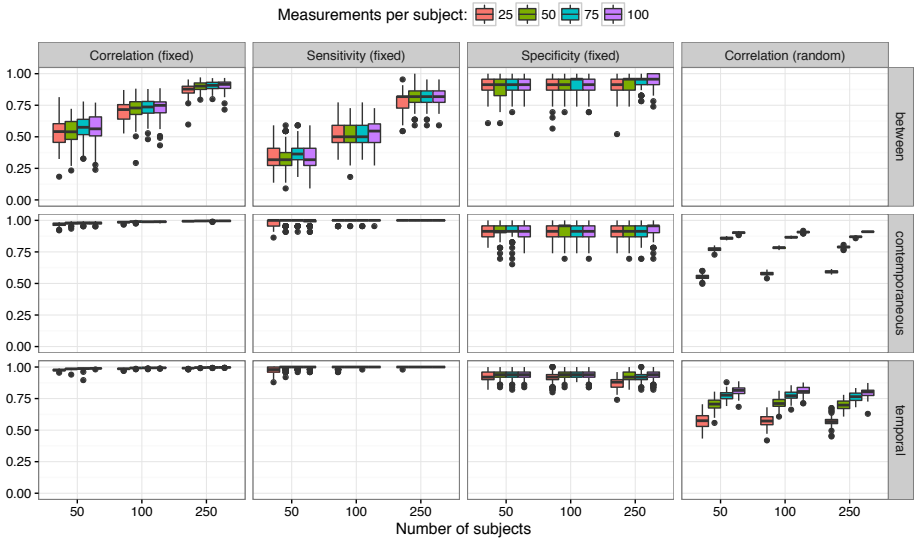

Figure 6.10: Here are the results of the simulation study. Boxplots indicate the distribution of the measures over all the 100 simulated datasets per condition. From left to right is shown: the correlation between true and estimated fixed effects, the sensitivity (i.e., the ability to detect true edges), the specificity (i.e., the ability to remove false edges), and the correlation between true and estimated random effects.

(high specificity).

6.8 Conclusion

In this chapter, we presented an overview of statistical methods that estimate network models—both cross-sectional and time-series—of multivariate Gaussian data. In our cross-sectional data analysis, we described the GGM, which takes the form of a network of partial correlation coefficients. In time-series data, 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. We argued that both can generate causal hypotheses. When multiple subjects were measured, 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 people. We argued that this network can also show potential causal relationships but in a different way than the temporal and contemporaneous networks. We proposed a two-step, multilevel estimation procedure to estimate temporal, contemporaneous, and between-subjects networks, which we implemented in the open-source R package, *mlVAR*. We presented a

simulation study showing that *mlVAR* closely estimates the true network structure and presented an empirical example showcasing the three network structures described above.

The outlined methodology in this chapter is not the only possible methodology for obtaining network structures from multivariate Gaussian data. A detailed description of these methods was beyond the scope of this chapter, as this chapter focussed on the GGM and its natural generalizations in time-series data. In particular, much work has been done on the estimation of directed acyclic graphs (DAG; Kalisch & Bühlmann, 2007; Pearl, 2000) which aim to model causal effects. When cases can be assumed to be independent, such DAGs can be fitted in standard structural equation (SEM) modeling software (see Chapter 7). Several software packages exist that aim to find such a DAG (e.g., *pcalg*, Kalisch et al., 2012; *bnlearn*, Scutari, 2010; *BDgraph*, Mohammadi & Wit, 2015). In time-series data, one can use structural VAR (Chen et al., 2011; also termed unified SEM, Gates, Molenaar, Hillary, Ram, & Rovine, 2010) to fit contemporaneous effects in a directed network. Structural VAR can be shown to be equivalent to the VAR model discussed in this chapter, and can under strict assumptions be interpreted as a causal model. A promising 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 in the *gimme* package (Lane, Gates, Molenaar, Hallquist, & Pike, 2016).

The presented methods are not without problems and have several limitations. First, multivariate estimation of the multilevel VAR model is not yet feasible for larger datasets. As such, we only focused on combining univariate models. Doing so, however, means that not all parameters are in the same model. It is important to note that 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 the 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, MLE software. In this case, orthogonal random effects can be used. Although the simulation study showed that the networks are still attainable when using orthogonal random effects (even though random effects were correlated in the true model), using orthogonal estimation enforces parsimony on the model that may not be plausible. Finally, even when orthogonal estimation was used, the analysis ran very slowly in models with more than 20 variables. As such, multilevel VAR analysis of high-dimensional datasets is not yet feasible. LASSO estimation as used in $n = 1$ models can also be used with multiple subjects, but it does not take individual differences into account (Abegaz & Wit, 2013). LASSO estimation methods that combine the strengths of high-dimensional network estimation in $n = 1$ models, with the ability to use information of other subjects, could be promising in this regard, but they have not yet been worked out in detail.

It should further be noted that all network structures only generate hypotheses and are in no way confirmatory of causal relationships. The analyses showcased in

this chapter are therefore 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, in the thresholding of edges, no correction for multiple testing was applied 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. This means that care should be taken in substantively interpreting the selected edges of the networks.

One of the main innovations in this chapter comes in the substantial interpretation of between-subjects effects being potentially causal. The function of between-subjects differences in causal models has been argued to be problematic (Borsboom, Mellenbergh, & Van Heerden, 2003; Markus & Borsboom, 2013a). In order to make inferences on causal processes based on how people differ from each other, we must place very strong assumptions on the homogeneity of causal structures across individuals. In essence, we must assume that individuals are independent realizations of the same causal model (see also Hamaker, 2012). It is rarely acknowledged, however, that a similar problem holds for intraindividual data. As in the between-subjects case, the inference from a statistical association in the data to a causal model, operative at the level of the individual, is dependent on the strength of the research design and does not follow from the statistical associations themselves. In addition, if time series do not contain actual manipulations, the generalization in question can be equally problematic as in between-subjects designs.

Suppose we found a robust association between X and Y together with the temporal precedence of X (e.g., as in Granger causality) in a time-series analysis; we still would not know whether interventions on X would actually lead to changes in Y. Associations in within-subjects models can be subject to third-variable issues, such as Simpson's paradox, just as well as between-subjects models can. Correlations remain correlations, whether they come from individual differences or from time series, and rather than categorically preferring one type of data over another, it appears more sensible to let our judgment on optimal inferences depend on the substantive context.

In sum, this chapter provided methodological tools that can be used to gain insight in the potential dynamics present in psychological data. The described software packages and estimation methods present the current state-of-the-art in a field that is growing rapidly. These methods provide new ways to look at data—both literally, through the use of networks to visualize the results, and figuratively, by investigating contemporaneous and between-subjects effects in combination with temporal effects.

6.9 Appendix A: Simulating Multi-level VAR Models and Data

This algorithm generates data for P subjects on T measurement occasions of I items, using a lag-1 factorization. The fixed effect temporal, contemporaneous, and between-subjects networks all have a sparsity of 50%. This algorithm is implemented in the `mlVARsim` function. Given parameters are

- $\text{DF}_{\boldsymbol{\Theta}}$: The degrees of freedom (sampling variation) in sampling contemporaneous covariance matrices (default: $2I$)
- $S_{\boldsymbol{f_\beta}}$: The shrinkage factor of the temporal fixed effects (default: 0.9)
- $S_{\boldsymbol{\sigma^2_\beta}}$: The shrinkage factor of the temporal random effects (default: 0.9)
- $\boldsymbol{V_B}$: Vector of variances of the temporal effects (default uniformly drawn between 0.01 and 1)

1. Generate the following structures:
   - Inverse $I \times I$ variance–covariance matrices $\boldsymbol{\Theta}^{-1}$, $\boldsymbol{\Omega}_{\boldsymbol{\mu}}^{-1}$, and $I^2 \times I^2$ variance–covariance matrix $\boldsymbol{\Omega}_{\boldsymbol{B}}^{-1}$. All with 50% sparsity and 50% negative edges, using the methodology described by Yin and Li (2011) with a constant of 1.1 and a parameter range of 0.5 to 1. Standardize these matrices such that the diagonals of $\boldsymbol{\Theta}$ and $\boldsymbol{\Omega}_{\boldsymbol{\mu}}$ are equal to ones and the diagonal of $\boldsymbol{\Omega}_{\boldsymbol{B}}$ is equal to $\boldsymbol{V_B}$.
   - I length vector $\boldsymbol{f_\mu} \sim N(\boldsymbol{0}, \boldsymbol{I})$
   - I^2 length vector $\boldsymbol{f_\beta} \sim N(\boldsymbol{0}, \boldsymbol{I})$. Subsequently, set 50% lowest absolute values to zero.
2. Generate P covariance matrices $\boldsymbol{\Theta}^{(p)} \sim \text{Wishart}^{-1}(\boldsymbol{\Theta}/\text{DF}_{\boldsymbol{\Theta}}, \text{DF}_{\boldsymbol{\Theta}})$
3. Generate P parameter sets $\boldsymbol{\mu}^{(p)} \sim N(\boldsymbol{f_\mu}, \boldsymbol{\Omega_\mu})$ and $\text{Vec}(\boldsymbol{B}^{(p)}) \sim N(\boldsymbol{f_\beta}, \boldsymbol{\Omega_B})$
4. Compute eigenvalues of $\boldsymbol{B}^{(p)}$: $\lambda_1^{(1)} \dots \lambda_I^{(P)}$
5. If $\max\left(\text{Re}(\lambda_i^{(p)})^2 + \text{Im}(\lambda_i^{(p)})^2 > 1\right)$
   a) Set $\boldsymbol{f_\beta} \leftarrow S_{\boldsymbol{f_\beta}} \boldsymbol{f_\beta}$
   b) Scale $\boldsymbol{\Omega_B}$ such that $\text{diag}(\boldsymbol{\Omega_B}) \leftarrow S_{\boldsymbol{\sigma^2_\beta}} \text{diag}(\boldsymbol{\Omega_B})$
   c) Go to 3
6. For each p, set $\boldsymbol{y}_{-100}^{(p)} = \boldsymbol{\mu}^{(p)}$
7. For each p, generate for $t = -99, -98, \dots, T$ the scores $\boldsymbol{y}_t^{(p)}$
8. Discard all scores with $t < 1$
9. If any $|y_{ti}^{(p)}| > 100$, go to 5a