AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Methods

© 2021 American Psychological Association
ISSN: 1082-989X

2021, Vol. 26, No. 6, 635–659
https://doi.org/10.1037/met0000401

Informing VAR(1) With Qualitative Dynamical Features Improves Predictive Accuracy

Tim Loossens, Egon Dejonckheere, Francis Tuerlinckx, and Stijn Verdonck
Department of Quantitative Psychology and Individual Differences, University of Leuven (KU Leuven)

Abstract

The AR(1) model has been shown to outperform the general VAR(1) model on typical affective time series. Even in combination with a lasso penalty, the reduced VAR(1) model (VAR-lasso) is generally outperformed. A reason for the AR dominance is that the VAR-lasso selects models that are still too complex—the space of all possible VAR models includes simpler models but these are hard to select with a traditional lasso penalty. In this article, we propose a reparametrization of the VAR model by decomposing its transition matrix into a symmetric and antisymmetric component (denoted as SAD), allowing us to construct a hierarchy of meaningful signposts in the VAR model space ranging from simple to complex. The decomposition enables the lasso procedure to pick up qualitatively distinct dynamical features in a more targeted way (like relaxation, shearing, and oscillations); this procedure is called SAD-lasso. This leads to a more intuitive interpretation of the reduced models. By removing the antisymmetric component altogether, we obtain a subclass of symmetric VAR models that form a natural extension of the AR model with the same simple relaxation dynamics but allowing for interactions between the system components. We apply these reparametrized and constrained VAR models to 1,391 psychological time series of affect, and compare their predictive accuracy. This analysis indicates that the SAD-lasso is a better regularization technique than the VAR-lasso. Additionally, the results of an extensive simulation study suggest the existence of symmetric interactions for almost half of the time series considered in this article.

Translational Abstract

The inclusion of lagged interactions between the variables of interest in an autoregressive model has been shown typically not to improve the description of psychological time series. To many researchers in the field, this finding is counterintuitive because psychological systems, like the affect system, are generally conceived as complicated systems with constituents that are (at least partly) interacting. However, simultaneously including all possible interactions leads to an overly complex model that produces worse predictions. Regrettably, the commonly used method to systematically limit the number of possible interactions, lasso, does not significantly improve predictions. In this article, a taxonomy is provided of all possible dynamical behaviors encompassed by the autoregressive model (exponential relaxation, shearing, and oscillations). By adjusting the lasso procedure to directly target each of these behaviors, the prediction can be significantly improved and lagged interactions can better be detected. Additionally, we find that exponential relaxation dynamics is the most prominent dynamical behavior in time series of affect.

Keywords: cross-validation, network psychometrics, regularization, vector-autoregressive modeling, within-person dynamics

Supplemental materials: https://doi.org/10.1037/met0000401.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Multivariate time series analysis is a commonly used method in psychology and related fields (e.g., neuroscience) to chart the dynamical relations between a set of variables (Lütkepohl, 2005). One of the most popular models is the vector autoregressive (VAR) model of order 1, also denoted as VAR(1). This discrete-time stochastic model tries to capture the linear interdependencies between the constituents of a set of time dependent variables.

This article was published Online First September 16, 2021.

Tim Loossens https://orcid.org/0000-0002-7587-1922
Egon Dejonckheere https://orcid.org/0000-0002-7219-7506
Francis Tuerlinckx https://orcid.org/0000-0002-1775-7654

Tim Loossens developed the models and methods described in the article and implemented them. He also wrote the article. Egon Dejonckheere provided data and a description of the data. He also revised the article. Francis Tuerlinckx wrote the article together with Tim Loossens. Stijn Verdonck contributed to the development of the models and methods described in the article. He also revised the article.

Correspondence concerning this article should be addressed to Tim Loossens, Department of Quantitative Psychology and Individual Differences, University of Leuven (KU Leuven), Tiensestraat 102, 3000 Leuven, Belgium. Email: tim.loossens@kuleuven.be

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

In a general VAR model, the evolution of each variable depends on its own past values, on the other variables at previous time points (all weighted by regression parameters), and on some stochastic fluctuations. In the VAR(1) model, only observations at the previous time point are taken into account. In other words, this model assumes that the current vector of observations is a linear combination of the observations at the previous time point, plus a noise component. The weights of this linear combination are organized in a square matrix that is called the *transition matrix* and quantifies the dynamical relations (or interactions) between the variables.

Recently, the application of the VAR(1) model in psychology has been criticized because it may be overly complex. Bulteel et al. (2018) have shown that the predictive accuracy of the AR(1) model (a VAR(1) model without any cross-regressive effects between the variables) often outperforms that of the VAR(1) model in typical psychological applications. Thus, seemingly there is no evidence for the additional complexity offered by the VAR(1) model. Stated differently, interactions between the different psychological constructs across time cannot be determined with high confidence.

The result that AR(1) outperforms VAR(1) is somewhat counterintuitive because one would expect that there exists at least some degree of interaction between different variables of a psychological system across time. It is conceivable that there exist models more complex than the AR(1) model, but simpler than the overly complex VAR(1) model that do a better job at describing observed and predicting future data.

From this perspective, the crucial question is not: "Is AR(1) or VAR(1) the better model?," but rather: "How does one construct an intermediate model between the AR(1) and the full VAR(1) model that is able to pick up on the relevant interactions between the variables of the system?" Regularization and variable selection techniques could provide the answer. The lasso (least absolute shrinkage and selection operator; Hastie et al., 2009); for instance, puts a penalty on each of the coefficients of the transition matrix, shrinking some of them to zero. It eliminates degrees of freedom of the VAR(1) model by sparsening the transition matrix. The end result is a simpler model. Unfortunately, for the data sets studied by Bulteel et al. (2018), a straightforward implementation of the VAR(1)-lasso did not seem to improve the prediction.

A possible reason for the failure of the VAR(1)-lasso to outperform the AR(1) model is that the penalty of the traditional lasso either eliminates an individual coefficient or retains it (together with some shrinkage). This leads to a less complex model (in terms of number of parameters), but, as we will show below, the dynamics are still very complicated because of the asymmetry in the VAR(1) transition matrix. However, if the major source of information in the data is a symmetric correlation between variables, then it does not make sense to target individual coefficients. Yet, this is what the direct implementation of a lasso does.

The general VAR(1) model is able to generate complicated dynamics. The dynamics of the simple AR(1) model, on the other hand, are relatively simple. By varying the penalty, VAR(1)-lasso traces out a path between VAR(1) and AR(1) that runs predominantly through a model space of complicated dynamics (i.e., all models on the path are VAR-like). However, it may be very unlikely that such dynamics are present in the data. In this article, a hierarchy of meaningful signposts is provided on the path between the simple AR(1) and the complicated VAR(1) model. These signposts mark models that have a clear dynamical interpretation and that are of intermediary complexity. To do so, the VAR(1) transition matrix is decomposed into a symmetric and an antisymmetric part and a taxonomy is provided of qualitatively different features and dynamics that are encompassed by the VAR(1) model. It is then argued that specific dynamical features can more easily be turned on and off by applying the lasso to this reparametrized VAR(1) model. This lasso variant will be referred to as the SAD-lasso (symmetric-antisymmetric decomposition-lasso).

The predictive performance of the SAD-lasso on typical psychological data on affect is examined and compared to that of the AR(1), the VAR(1) and the VAR(1)-lasso models using cross-validation. Because of the special properties related to symmetric transition matrices, special attention is given to symmetric VAR(1) models (Sym-VAR(1) models) as well. Additionally, an extensive simulation study is done to corroborate the results obtained from the cross-validation. Moreover, the results of the simulation study allow us to make inferences about the interactions present in the data and what models best describe them.

The VAR(1) Model

Autoregressive models are typically applied to time series data, referring to data obtained when a set of observables $\mathbf{y} = \{y_1, y_2, \ldots, y_d\}$ is tracked across time:

$$\mathbf{y}_t = \begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{dt} \end{pmatrix}, \quad t = 0, 1, 2, \ldots, N$$

The index t labels the occasions in time when the variables are observed. $\mathbf{y}_t$ is a vector in $\mathbb{R}^d$ containing the values of the d observables at *time t*. Because it describes the state of the system that is being studied at *time t*, it will be referred to as the *state vector*.

Given a series of observations $\{\mathbf{y}_t \in \mathbb{R}^d \mid t = 0, \ldots, N\}$. The VAR(1) model relates the state vector $\mathbf{y}_t$ at *time t* to the previous state vector $\mathbf{y}_{t-1}$ at *time t* – 1 as follows:

$$\mathbf{y}_t = \boldsymbol{c} + A\,\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t. \tag{1}$$

The parameter $\boldsymbol{c} \in \mathbb{R}^d$ denotes the intercept, $A \in \mathbb{R}^{d \times d}$ is the transition matrix, and $\boldsymbol{\varepsilon}_t \in \mathbb{R}^d$ is a vector containing the stochastic fluctuations (innovations) at *time t*.

The innovations $\boldsymbol{\varepsilon}_t$ are assumed to be Gaussian distributed with mean zero. They are assumed to be uncorrelated across time. Hence, Equation 1 is nothing but a system of multivariate linear regression equations, modeling each variable as a linear function of the variables at the previous measurement.

Generally, only stable VAR(1) processes are considered. A VAR(1) process is stable when all the eigenvalues of A have a modulus[1] strictly smaller than 1. If this condition is not met, the process described by Equation 1 is still well-defined, but the state

[1] The modulus extends the notion of absolute values to the complex numbers. For a real number, the modulus is the same as the absolute value. For a complex number $z = x + yi$, the modulus is defined as $|z| = \sqrt{x^2 + y^2}$.

vector $\mathbf{y}_t$ would diverge to infinity. A stable VAR(1) model is stationary, meaning that the first moment, the second moment and the autocovariance of the unconditional (joint) distribution of the state vector $\mathbf{y}_t$ are time invariant (Lütkepohl, 2005).

In this article, only VAR models as in Equation 1 are considered, that is, first order VAR models or VAR(1) models. A general pth order VAR model, denoted as VAR(p), does not only include the first lag $\mathbf{y}_{t-1}$ of the state vector, but also the lags $\mathbf{y}_{t-2}, \ldots, \mathbf{y}_{t-p}$. Because these will not be considered, the lag will not be explicitly written in the remainder of the text.

Lasso: Scanning the VAR Space

Traditional Lasso

Because Equation 1 is just a system of multivariate regression equations, its parameters can be estimated using the least-squares approach. Defining

$$\mathbf{Y}_j = \begin{pmatrix} y_{1,j} \\ y_{2,j} \\ y_{3,j} \\ \vdots \\ y_{N,j} \end{pmatrix}, \quad \boldsymbol{A}_j = \begin{pmatrix} A_{j1} \\ A_{j2} \\ \vdots \\ A_{jd} \end{pmatrix}, \quad \boldsymbol{\epsilon}_j = \begin{pmatrix} \varepsilon_{1,j} \\ \varepsilon_{2,j} \\ \varepsilon_{3,j} \\ \vdots \\ \varepsilon_{N,j} \end{pmatrix}, \tag{2}$$

for all $j = 1, \ldots, d$, and

$$X = \begin{pmatrix} y_{0,1} & y_{0,2} & \cdots & y_{0,d} \\ y_{1,1} & y_{1,2} & \cdots & y_{1,d} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N-1,1} & y_{N-1,2} & \cdots & y_{N-1,d} \end{pmatrix}, \tag{3}$$

the model in Equation 1 can equivalently be written as

$$\mathbf{Y}_j = c_j J_{N,1} + X\boldsymbol{A}_j + \boldsymbol{\epsilon}_j. \tag{4}$$

Here, $J_{N,1}$ denotes a $N \times 1$ all-ones matrix. The vector $\mathbf{Y}_j$ contains all observations across time, except the first, of the jth component of the state vector $\mathbf{y}$; the vector $\mathbf{A}_j$ holds all coefficients of the jth equation in Equation 1; and the vector $\boldsymbol{\epsilon}_j$ contains the innovations of the jth component across time.

The ordinary least-squares function $\mathscr{L}$ associated to Equation 4 is

$$\mathscr{L}(\mathbf{c}, \boldsymbol{A}_1, \ldots, \boldsymbol{A}_d) = \frac{1}{2} \sum_{j=1}^{d} (\mathbf{Y}_j - c_j J_{N,1} - X\boldsymbol{A}_j)^T \times (\mathbf{Y}_j - c_j J_{N,1} - X\boldsymbol{A}_j). \tag{5}$$

The lasso (Tibshirani, 1996) introduces a penalty in the least-squares function (Equation 5) equal to the sum of the absolute values of the VAR coefficients weighted by a tuning parameter $\gamma \in \mathbb{R}^+$,

$$\mathscr{L}_{\text{lasso}}(\mathbf{c}, \boldsymbol{A}_1, \ldots, \boldsymbol{A}_d) = \mathscr{L}(\mathbf{c}, \boldsymbol{A}_1, \ldots, \boldsymbol{A}_d) + \gamma \sum_{j=1}^{d} \sum_{k=1}^{d} |A_{jk}|. \tag{6}$$

When the tuning parameter γ is zero, the lasso least-squares function $\mathscr{L}_{\text{lasso}}$ reduces to the ordinary least-squares function $\mathscr{L}$ of the VAR model. The larger it becomes, the larger the penalty for the coefficients to be different from zero. The absolute value ensures that coefficients can be shrunk to zero entirely. As such, the lasso serves as a variable selection tool.

Although the lasso considers the impact of each of the coefficients separately, many features and dynamics of VAR models depend on a combination of coefficients (see below). Therefore, shrinking a single coefficient generally has no straightforward link to any of these features. This could explain why the traditional VAR-lasso shows little improvement in comparison to the AR model.

Symmetric-Antisymmetric Decomposition

By reparametrizing the VAR model, it is possible to create a lasso variant that is more in tune with the specific structural and dynamical features of VAR models. The reparametrization considered in this article is one in which the transition matrix A is decomposed in a symmetric and an antisymmetric (skew-symmetric) component:

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T),$$

where A^T denotes the transpose of the matrix A.

In the following section, it will be shown that purely symmetric transition matrices form a natural extension of AR models (which, in fact, also have symmetric matrices). Because of its properties, the symmetric VAR or SymVAR model will be considered as a model in its own right. Purely antisymmetric matrices, on the other hand, incorporate cyclic motion (see next section for more information about this feature). Asymmetric matrices, consisting of both a symmetric and an antisymmetric part, also include cyclic motions or shearing-like dynamics depending on whether the eigenvalues and eigenvectors are complex or real (for more information about these features, see the next section).

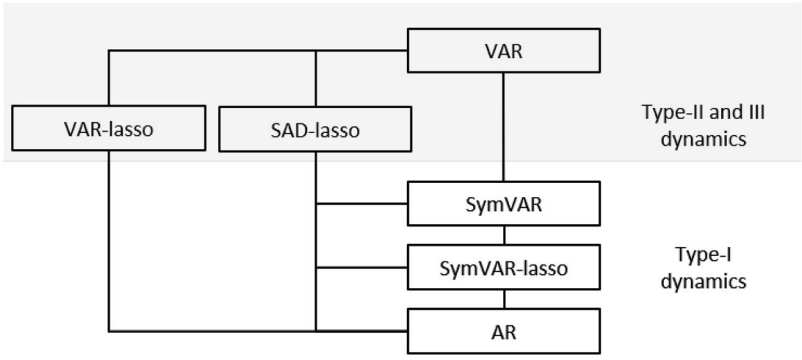
Due to the decomposition of A in a symmetric and an antisymmetric part, the lasso can be made to penalize the symmetric and the antisymmetric coefficients separately. This SAD-lasso has as many degrees of freedom as the traditional lasso (they can create the same models), but the probability distributions of selecting certain models is different for the two approaches.

Shrinking the antisymmetric component, a symmetric interaction is retained. Shrinking the symmetric component, an antisymmetric interaction is obtained. By shrinking both the symmetric and antisymmetric component, the interaction is entirely removed. Keeping both components results in an asymmetric interaction with associated complex features and dynamics (shearing and rotating).

Figure 1 illustrates the hierarchy of the different VAR models that will be considered in this article. The AR and the VAR models form the extremes of the continuum. The AR model prohibits

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 1
Illustration of the Hierarchy of the Different VAR Models Discussed in This Article



Note. The AR and VAR models form the extremes of the continuum. The SymVAR model lies in between the AR model and the VAR model. The SymVAR-lasso searches the space between the SymVAR model and the AR model. The VAR-lasso and the SAD-lasso both search the space in between the AR model and the VAR model. In contrast to the VAR-lasso, however, the SAD-lasso has a natural connection with the SymVAR and the SymVAR-lasso model. The unshaded area only includes Type-I dynamics, whereas the shaded area encompasses the more complicated Type-II and Type-III dynamics. VAR = vector autoregressive; SymVar = symmetric VAR; lasso = least absolute shrinkage and selection operator. See the online article for the color version of this figure.

any kind of interaction, while the VAR model permits all kinds of linear interactions. The SymVAR is a model that is obtained by removing the antisymmetric component of the transition matrix. Together with the AR model, this model incorporates a specific type of dynamics to which will be referred as Type-I dynamics (see next section). As a regression model, the SymVAR can also be combined with the traditional lasso to retain only a select number of symmetric interactions. The targets of the traditional VAR-lasso are the extremes, the AR and the VAR model. The SymVAR-lasso targets the AR model and the SymVAR model. The SAD-lasso targets all of the aforementioned models. It also targets antisymmetric matrices, but because these do not form a class of their own with their own specific dynamics, they are not considered as a separate model.

Dynamics Implied by the Various Models

Given an observation $\mathbf{y}_0$ at *time* $t = 0$, the expected trajectory of the state vector $\mathbf{y}$ at *time* t is

$$\langle \mathbf{y}_t \,|\, \mathbf{y}_0 \rangle = \left(\sum_{i=0}^{t-1} A^i \right) \mathbf{c} + A^t \mathbf{y}_0, \tag{7}$$

where $\langle \mathbf{y}_t \,|\, \mathbf{y}_0 \rangle$ is the conditional expectation of $\mathbf{y}_t$ given $\mathbf{y}_0$.

Because the transition matrix A is bounded for stable VAR processes, the Neumann series $\sum_{i=0}^{t-1} A^i$ converges for t going to infinity:

$$\lim_{t \to \infty} \sum_{i=0}^{t-1} A^i = (1 - A)^{-1}.$$

In addition, for stable VAR processes the term $A^t \mathbf{y}_0$ vanishes as t goes to infinity (see Appendix A). As a consequence, the expected trajectory (Equation 7; the deterministic trajectory) evolves toward the center

$$\boldsymbol{\mu} = (1 - A)^{-1} \mathbf{c} \tag{8}$$

as time moves on. This center is a fixed point on the trajectory. In other words, if $\mathbf{y}_{t-1} = \boldsymbol{\mu}$, then $\langle \mathbf{y}_t \,|\, \mathbf{y}_{t-1} \rangle = \mathbf{y}_{t-1} = \boldsymbol{\mu}$. This is what is implied by stability: there is an attractive force that keeps or brings the state vector $\boldsymbol{y}$ close to the center $\boldsymbol{\mu}$ as time moves on. For that reason, the center $\boldsymbol{\mu}$ is sometimes referred to as the *attractor*. The attractor is equal to the first moment (the mean) of the unconditional distribution of the state vector $\boldsymbol{y}$. A stable VAR process is therefore a mean-reverting process.

The transition matrix A contains the interactions that drive the state vector $\mathbf{y}$. It can be represented as a vector field on the space in which the state vector $\mathbf{y}$ lives. Depending on the form of the transition matrix, several phase portraits can be distinguished (see Figure 2). The term *phase portrait* is used to refer to the expected or deterministic trajectory of the state vector contingent on an observation $\mathbf{y}_0$ at *time* $t = 0$ as described by Equation 7 (depicted in lighter grey in the panels in the first column of Figure 2). An actual instance of a trajectory resembles such a phase portrait, but exhibits additional stochastic fluctuations, thus obscuring the deterministic motion (depicted in black in the panels of the first column of Figure 2). In this article, these phase portraits are briefly discussed; for a more elaborate discussion, see, for example, Strogatz (2000). Many of the dynamical properties of a transition matrix can be derived from its eigenvalues and eigenvectors.

Type-I Dynamics: Simple Exponential Relaxation

In the simplest case, the transition matrix is symmetric (i.e., $\frac{1}{2}(A - A^T) = 0$). The eigenvalues and eigenvectors of symmetric matrices are real (see Appendix B). Furthermore,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

their eigenvectors are orthogonal (see Appendix B). Such a situation is sketched in first two rows of Figure 2. The state vector $\mathbf{y}$ relaxes exponentially toward the center $\boldsymbol{\mu}$ in the directions determined by the eigenvectors. The relaxation rate is determined by the magnitude of the eigenvalues. Smaller eigenvalues (i.e., smaller moduli) imply larger attractive forces; the state vector $\mathbf{y}$ relaxes quicker in the directions spanned by eigenvectors corresponding to smaller eigenvalues.[2] As a consequence, the state vector initially moves in a direction parallel to the eigenvector corresponding to the smallest eigenvalue (the *fast eigenvector*). Then, as the state vector approaches the center $\boldsymbol{\mu}$, it increasingly moves in a direction parallel to the eigenvector corresponding to the largest eigenvalue (the *slow eigenvector*).

An AR model is a special case of symmetric VAR (SymVAR) model for which all of the off-diagonal elements are zero. The eigenvectors then coincide with the natural basis of the space $\mathbb{R}^d$ in which the state vector $\mathbf{y}$ lives. Therefore, the components of the state vector are independent from one another. The diagonal elements coincide with the eigenvalues which determine the rate of relaxation or, equivalently, the autocorrelation of every component.

The AR model thus corresponds to a special situation in which the observables of the system are completely independent from one another. SymVAR models also have perpendicular eigenvectors, meaning that there still exists a frame of reference in which the observables are independent, but this is not the frame of reference in which they are measured. The frame of reference of the measurements does not coincide with the natural basis, but it is a rotated version thereof. Essentially, the dynamics of a SymVAR model are a simple rotation of the dynamics of the AR model.

A consequence of this rotation is that the components of the state vector in the measurement frame of reference interact with one another. These interactions are symmetrical; the effect A_{ij} of the component y_j on the component y_i is the same as the effect A_{ji} of the component y_i on the component y_j. In other words, $A_{ij} = A_{ji}$ or $A = A^T$.

Type-II Dynamics: Shearing

When the transition matrix is asymmetric, the eigenvalues and eigenvectors can still be real. If that is the case, they will not be perpendicular like in the symmetric case. An example of this is portrayed in the third row of Figure 2 There does no longer exist an orthogonal frame of reference in which the components of the state vector $\mathbf{y}$ are independent. The state vector nonetheless still relaxes in the direction of the eigenvectors with a pace determined by the eigenvalues. However, due to the obliqueness of the eigenvectors, this motion appears to be more lateral, shearing-like. In the extreme case where the eigenvectors are aligned, the trajectories are truly lateral on opposite sides of the axis determined by the eigenvectors. Such motion is referred to as pure shearing. Thus, shearing or shearing-like dynamics are caused by the asymmetrical relation between the components y_i and y_j; $A_{ij} \neq A_{ji}$.

Type-III Dynamics: Spiraling

Asymmetric transition matrices do not always have real eigenvalues. In general, the eigenvalues are complex and so are the eigenvectors. In that case, the trajectories are exponentially decaying circles or spirals, as can be seen in the bottom row of Figure 2.

Complex eigenvalues of a real square matrix always appear in pairs of complex conjugates $(\lambda, \overline{\lambda})$ (as can be deduced from Equations B.1 and B.2. Such a pair and their corresponding eigenvectors (which are also complex conjugates of one another, as can also be deduced from Equations B.1 and B.2 describe an orbital motion in a plane. If the moduli of the eigenvalues are equal to 1, the expected trajectory is a closed, elliptic orbit. Closed orbits do not fall under the stable VAR models, because the state vector $\mathbf{y}$ could drift away due to the stochastic fluctuations. When the eigenvalues are complex and their moduli smaller than one, the state vectors spirals toward the center. An example of such a stable, inward spiraling motion with a small angular frequency is shown in the bottom row of Figure 2.

Model Complexity of the Various Models

The general VAR model encompasses a wide spectrum of dynamics. This means that a lot of unknown parameters have to be estimated when applying such a model to data; a general d dimensional VAR model has $d + d^2$ parameters. Hence, the complexity of the VAR model grows quadratically with the number of dimensions, as shown in Figure 3 (red), making it vulnerable to overfitting.

The simple AR model only has $2d$ parameters and therefore its complexity merely grows linearly with the number of dimensions. It is therefore fairly limited regarding the dynamics it includes; it only encompasses Type-I dynamics without any interactions between the system components. This makes it more vulnerable to underfitting.

The SymVAR model (having a symmetric transition matrix) has $d + \frac{d(d-1)}{2}$ parameters and thus the number of parameters grows slower with increasing dimensionality compared with a general VAR model. Its complexity lies right in between that of the AR model and that of the VAR model. Like the AR model, it also only encompasses Type-I dynamics, but it allows interactions between variables to exist.

Both the VAR-lasso and the SAD-lasso have a complexity in between that of the VAR model and that of the AR model. Due to their difference, they give different weights to certain types of dynamics. The space of stable VAR models is predominantly populated by Type-II and Type-III models, as illustrated in Figure 4. Due to the focus of the traditional lasso on individual coefficients, the probability of it retaining pure symmetric or antisymmetric interactions is small. Hence, it is unlikely that a VAR-lasso will

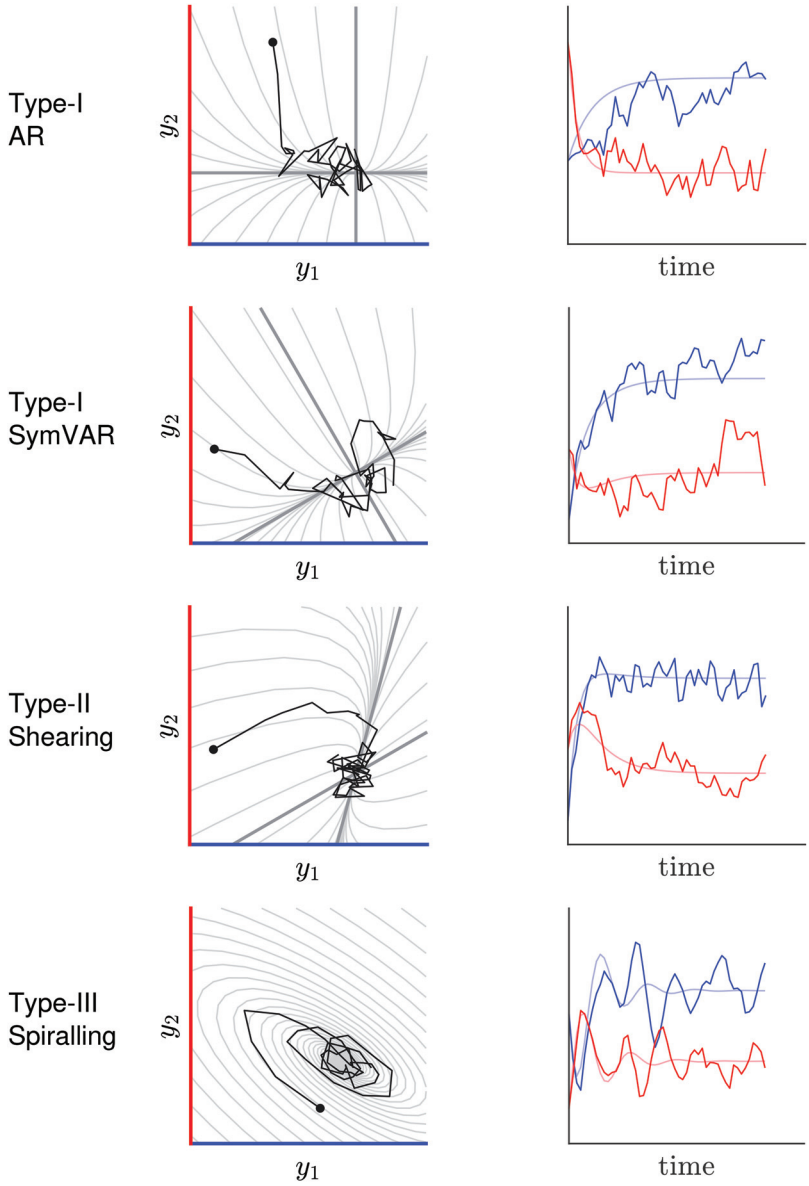
[2] The fact that the smallest eigenvalue is associated with the fast eigenvector may come as a surprise to readers familiar with continuous-time dynamics. Indeed, in Strogatz (2000), it is mentioned that the state vector relaxes quicker in the directions spanned by eigenvectors corresponding to *larger* eigenvalues. In contrast to Strogatz (2000) who describes the evolution of linear systems that are continuously evolving in time, we describe linear systems that evolve discretely in time (which is typical for autoregressive models). If θ is an eigenvalue of a continuous, linear transition matrix, then the associated discrete eigenvalue α is related to the continuous eigenvalue as

$$\alpha \propto \exp(-\theta).$$

Hence, larger θ values are associated with smaller α values and vice versa.

Figure 2

Examples of Phase Portraits and Time Series Demonstrating the Three Fundamental Types (I–III) of Dynamics Incorporated by the VAR Model

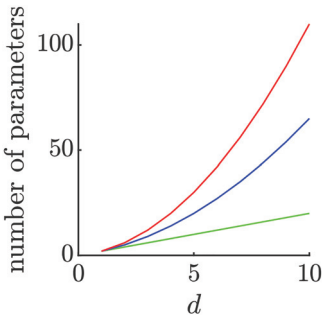


Note. The panels in each row are related to one another. The panel on the left provides illustrations of deterministic trajectories (light grey) for a specific VAR model belonging to the class indicated by the row label. As time moves forward, the trajectories close in on the center situated in the bottom on the right (at the intersection of the two eigenvectors). When applicable, the directions of the eigenvectors are indicated in darker grey. In black is shown a specific stochastic trajectory starting from the black dot. In the right-hand panel is shown the time series corresponding to the black trajectory. Blue corresponds to the first dimension and red to the second dimension in the observer's frame of reference. VAR = vector autoregressive; SymVar = symmetric VAR. See the online article for the color version of this figure.

roam the space of SymVAR models. It will mainly roam in that part of the space populated by Type-II and Type-III models, unless it removes all interactions at which point it becomes an AR model. The SAD-lasso is adapted to be able to scan the subspace of SymVAR models on top of being able to roam in the space of Type-II and Type-III dynamics. It has also more affinity with purely antisymmetric components which is useful for detecting spiraling dynamics.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 3
Complexity of the Different VAR Models



Note. The complexity of the different VAR models are illustrated in function of the dimensionality *d*. The complexity of the AR model is depicted in green, the complexity of the SymVAR model is depicted in blue, and that of the VAR model is depicted in red. VAR = vector autoregressive. See the online article for the color version of this figure.

The lasso can also be made to only scan the subspace of symmetric VAR models—this way, a SymVAR-lasso is obtained. This lasso can be used to fit a SymVAR model which only retains the most important symmetric interactions. This model has a complexity in between that of the full SymVAR model and the AR model.

An Empirical and a Simulation Study

As mentioned above, Bulteel et al. (2018) showed that the AR model often outperforms the VAR model in terms of predictive accuracy for typical psychological data. They also found that the traditional lasso (denoted here as VAR-lasso) does not really improve the predictive accuracy of VAR models. We argued that the reason for this could be that the lasso primordially scans regions of the VAR space where dynamics are rather complicated in comparison to the Type-I dynamics of AR-like models. We have discussed a taxonomy of VAR dynamics and we have related these dynamics to structures of transition matrices (i.e., a series of signposts in the model space). In addition, we have proposed a lasso method (the SAD-lasso) which penalizes the transition coefficients of a reparametrized VAR model, where the transition matrix has been split in a symmetric and an antisymmetric component.

Using cross-validation, we will investigate the predictive performance of the SAD-lasso model in comparison to the other VAR models, and in particular in comparison to that of the traditional lasso. We will also consider the predictive performance of the other intermediate models that have been put forward as signposts in the VAR model space (i.e., the SymVAR and the SymVAR-lasso model). For this analysis, a large data set of affect time series will be considered.

Cross-validation informs us about which model has the best predictive accuracy. This best model does not necessarily coincide with the true model. Given the limited sample sizes of the considered time series, small interaction effects can emerge spuriously. As a consequence, the AR model can be outperformed by a more complex model in terms of predictive accuracy even when it is the data generating mechanism. Besides the cross-validation analysis, we will perform a simulation study to investigate whether the AR model (a model without interactions between the different affect dimensions) could be the true data generating mechanism for the time series considered in this article. The simulation study will also be used to draw conclusions about which other models are required to better understand the data, if the AR model turns out not to be the data generating mechanism.

Data Description

Similar to Bulteel et al. (2018), emotion time series data will be used to compare the predictive accuracy of each of the different VAR models. Emotions are highly dynamic phenomena which are thought to mutually influence one another over time (Pe & Kuppens, 2012). As such, these time series make an excellent application for the VAR regularization techniques proposed in this article.

To arrive at robust conclusions, the total data set considered in this article includes a reanalysis of 10 different previously published studies (Dejonckheere et al., 2019; Geschwind et al., 2011; Heininga et al., 2019; Houben et al., 2016; Koval et al., 2013; Pe et al., 2016; Schmiedek, 2010; Provenzano, 2019; Sels et al., 2017, 2019). This yields an initial sample of 1,552 participants. However, applying the additional selection criterion that the time series should at least consist of 40 measurement occasions in order for a participant's affective time series to comprise the final data set, only 1,391 participants will be retained. This cut-off ensures sufficient emotion assessments per person to fit the various VAR models.

In the different studies, emotion time series were obtained via experience sampling (ESM; Csikszentmihalyi & Larson, 1987) or daily diary methods (Bolger et al., 2003). In all studies, participants were instructed to carry around a mobile device (typically a smartphone or palmtop) in their daily life. Throughout the day, or once at the end of the day, they had to fill out several surveys about their momentary emotion levels. Participants rated the intensity of a fixed set of discrete emotion items, usually on a Likert scale or a continuous scale ranging from *not at all* to *very much*. Capturing people's emotions (almost) in the moment has the strength that it not only significantly reduces memory biases, the ambulatory nature of these assessments also yields emotional time series that are high in ecological validity (Dejonckheere et al., 2019).

The individual study details, the main summary statistics of each (final) time series data set and the specific positive and negative emotions that were assessed at each measurement occasion are given in the Supplementary Table S.1. All studies were ethically approved and all participants provided informed consent.

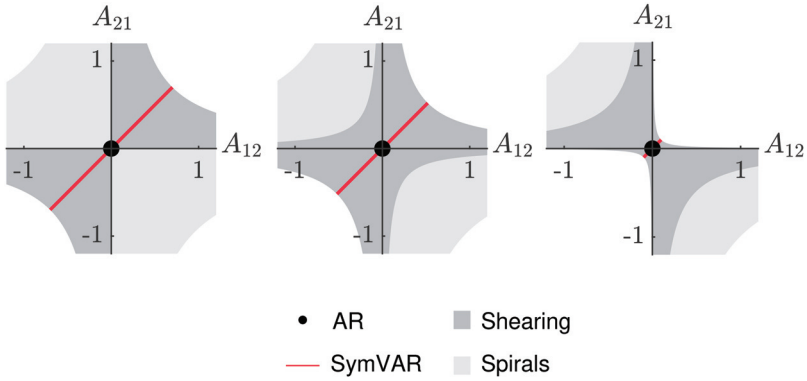
Parameter Estimation

The VAR models are fitted using either the standard regression method, minimizing the least-squares function (Equation 5), or the lasso-regression method, minimizing a least-squares function (Equation 6) with an additional penalty. The VAR and AR model have well-known analytic solutions for the estimates of the intercept and the transition

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 4

Two-Dimensional Phase Diagrams Depicting the Regions Corresponding the Distinct Types of Stable Dynamics Included in the VAR Model



Note. From left to right, the diagonal elements (A_{11}, A_{22}) of the 2×2 transition matrix A are (0.3, 0.3), (0.1, 0.7), and (0.01, 0.99) respectively. The different regions are demarcated in function of the off-diagonal elements (A_{12}, A_{21}). The black dot corresponds to the AR model, the red line corresponds to the SymVAR model (Type-I dynamics), the darker grey regions correspond to Type-II dynamics (shearing), and the lighter grey regions correspond to type-III dynamics (spiraling orbits). VAR = vector autoregressive; SymVar = symmetric VAR. See the online article for the color version of this figure.

matrix which, using the definitions (Equation 2) and (Equation 3), are given by

$$\hat{\mathbf{c}} = \frac{1}{N}(J_{1,N}Y - J_{1,N}XA)^T$$
$$\hat{A} = (X^TX)^{-1}X^TY.$$

$\frac{1}{N}J_{1,N}X \in \mathbb{R}^{1\times d}$ is a vector that contains the averages of the matrix X along its columns. $Y \in \mathbb{R}^{N\times d}$ is a matrix whose columns consists of the vectors $\mathbf{Y}_j$,

$$Y = \begin{pmatrix} \mathbf{Y}_1 & \mathbf{Y}_2 & \cdots & \mathbf{Y}_d \end{pmatrix}.$$

The vector $\frac{1}{N}J_{1,N}Y \in \mathbb{R}^{1\times d}$ contains the averages of the matrix Y along its columns.

Because of the symmetry constraints, the estimate $\hat{A}$ of the SymVAR model is different. Imposing the required symmetry conditions, it can be shown that the estimate of the transition matrix of the SymVAR model is given by (see Appendix C)

$$\mathbf{vec}(\hat{A}) = (Id_d \otimes \tilde{X}^TX + \tilde{X}^TX \otimes Id_d)^{-1}\mathbf{vec}(\tilde{X}^TY + Y^T\tilde{X}),$$

with

$$\tilde{X} \equiv \left(Id_N - \frac{1}{N}J_{N,N}\right)X. \tag{9}$$

The operator $\otimes$ denotes the Kronecker product and Id_N denotes the $N \times N$ identity matrix.

The lasso estimates are obtained using the coordinate descent algorithm (Friedman et al., 2007). When combining the lasso with SymVAR, one has to be careful with imposing the symmetry constraints. In Appendix C, an illustration is given of how this can be done. The tuning parameter γ of the lasso is determined using K-fold cross-validation with $K = 10$ (see below).

When doing lasso-regression, the data are often standardized. For the analyses in this article, it did not matter whether the data were standardized or not; the results largely remained the same. Therefore, we only report the results without standardization.

For parameter estimation in general, only consecutive measurements are considered. Measurement pairs with a gap in between are not included in the regression analysis. Moreover, measurement pairs that are separated by a night are also ignored. That way, all time intervals between measurements have approximately the same order of magnitude.

Cross-Validation: K-Folds and Blocked K-Folds

K-fold cross-validation (Geisser, 1975) is a technique to investigate the predictive accuracy of a model within a data sample. The standard procedure is to randomly divide a data sample (a time series) $\{\mathbf{y}_t \in \mathbb{R}^d \mid t = 0, \ldots, N\}$ into K (more or less) equal subsamples $\mathcal{S}_k$, $k = 1, \ldots, K$. Then, one after another, such a subset is taken to be the test set; the remaining $K - 1$ subsets form the training set. Each time, the model is fitted on the training set and the obtained fit is used to compute the mean-square-error (MSE) between the data points $\mathbf{y}_t$ in the test set and their predicted values $\hat{\mathbf{y}}_t$,

$$\mathrm{MSE}_k = \frac{1}{d\,|\mathrm{S}_k|}\sum_{t\in \mathrm{S}_k}(\mathbf{y}_t - \hat{\mathbf{y}}_t)^T(\mathbf{y}_t - \hat{\mathbf{y}}_t).$$

$t \in \mathrm{S}_k$ implies that only observations in the test set S_k are included in the sum and $|\mathrm{S}_k|$ denotes the number of points in that test set.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Finally, the predictive accuracy of the model on the data sample is defined as the mean of the *MSE* across the k different test sets,

$$\text{MSE} = \frac{1}{K}\sum_{k=1}^{K} \text{MSE}_k$$

The model with the smaller cross-validation *MSE* has the best predictive accuracy. Typically, when a model is not taking into account dynamical patterns present in the data (i.e., the model is underfitting), or when it has too many degrees of freedom and is introducing patterns that are actually not present (i.e., the model is overfitting), the predictive accuracy will become worse—resulting in a larger cross-validation *MSE*.

Because the data format in this article is that of a time series with possibly dependent observations, a variant of this traditional *K*-fold cross-validation is used for model selection: blocked *K*-fold cross-validation (Bergmeir & Benítez, 2012; Bulteel et al., 2018). In blocked *K*-fold cross-validation each of the K subsamples S_k only contains consecutive measurements. For the analyses, we use $K = 10$. For the lasso, the cross-validation procedure consists of two layers because not only the predictive accuracy of the model has to be assessed, the tuning parameter has to be determined as well. Given a choice of the tuning parameter, the predictive accuracy of the model is evaluated using blocked *K*-fold cross-validation. To determine the lasso tuning parameter γ, the standard *K*-fold cross-validation procedure is used (where we also use the value of $K = 10$). The training sample is divided into K subsamples and the average predicted *MSE* is computed for various values of γ. The γ for which this average is minimal is then used as tuning parameter. This means that there are two different layers in the cross-validation of a lasso model. On the level of the blocked *K*-fold cross-validation, the data are divided into K subsamples S_k. For each of these subsamples, the data are again divided into K' subsamples (not necessarily blocked). These K' subsamples are used to determine the optimal tuning parameters γ_k for the subsamples S_k. These γ_k's are subsequently used to computed the predicted mean-square-errors MSE_k of the model on the subsamples S_k. This is the same approach as the one that was used by Bulteel et al. (2018).

A Simulation Study

Besides comparing the models' predictive accuracies, we also conduct a realistic simulation study (i.e., based on the properties of the real data in the empirical study). The goal of the simulation study is to investigate whether the cross-validation results of the real data can be obtained with an AR model as data generating mechanism. In other words, we want to test whether the cross-validation results as obtained with the empirical data can be mimicked in a realistic simulation study assuming there are no interactions between the different affect dimensions. In the same vein, the same simulation study will be used to test whether some of the more complex models (which include interactions) are required to explain the cross-validation results on the substantive data.

The model that has the best predictive accuracy and the true model underlying the data are not necessarily the same. Even if, for instance, the observed data were generated by an AR model, it is possible for a more complex model to outperform the AR model in terms of predictive accuracy. The models we consider are nested (the AR model is nested in the SymVAR-lasso model, which is nested in the SymVAR model, and so on). Hence, there are regions in the model space where different models can generate the same data patterns. Because of the finite sample sizes, spurious interactions between affect dimensions can emerge. As such, an AR model can occasionally be outperformed by a VAR-lasso model or a SymVAR model, even when it actually was the data generating mechanism. This effect can also occur in the other direction. Say, for example, that the VAR model is truly the data generating mechanism but that the interactions between the affect dimensions are small. Then, simpler models can outperform the VAR model.

In a simulation study, we can freely decide which model to use as data generating mechanism. We can pick a model, simulate data, and repeat the cross-validation analysis. If the model we picked corresponds to the data generating mechanism of the true data, then the simulated data will closely resemble the actual data. The cross-validation analysis on the simulated data will therefore give results similar to those obtained for the real data. If, on the other hand, the model we choose deviates from the data generating mechanism, significant differences are expected between the cross-validation results of the simulation study and the actual results. Specifically, we expect an increase in the predictive accuracy of the model that was used to simulate the data.

In order for this reasoning to hold, the features of the simulated time series should maximally resemble those of the actual data. For every time series in the actual data, we therefore first search for the least-squares estimates of the model that we picked. Using these estimates, we then simulate data with the same sample size as the corresponding time series. To increase the resemblance with the actual data, we also condition on the actual observation $\mathbf{y}_{t-1}$ to obtain a simulated observation at measurement occasion t. In other words, the simulated data $\tilde{\mathbf{y}}_t$ are obtained as

$$\tilde{\mathbf{y}}_t = \hat{A}\,\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

where $\mathbf{y}_{t-1}$ denotes the actual observations, $\hat{A}$ denotes the estimated transition matrix and $\boldsymbol{\varepsilon}_t$ is normally distributed with mean zero and covariance $\hat{\Sigma}$. The covariance $\hat{\Sigma}$ is derived from the residuals of the actual data (given the estimate $\hat{A}$).

Using the simulation study, we also want to test whether the AR model (a model without interactions) could have been the data generating mechanism for the time series considered in this article, or if other (more complex) models are required to better explain the results of the substantive data. Instead of considering all of the models independently, we conduct a single test for all of them. For the actual data, the cross-validation analysis gives us a proportion for each model being the best predicting model (i.e., the AR model outperforms all other models in x% of the cases, the VAR model outperforms all other models in y% of the cases, and so on). Each simulation study will result in a similar list of percentages. These simulated lists of proportions give us an indication of what results are to be expected if all data were truly generated by either one of the models. Using these expected proportions,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

we can compute how much we require of each model in order to minimize the error between the simulated list of proportions and the actual list of proportions (i.e., the linear combination of simulated lists that most closely matches the actual list of proportions). We can then evaluate whether the mixing ratio of the AR model (the number of required AR models as generating mechanisms) is different from 1 or not.

Results and Discussion

Predictive Accuracy

Figure 5 gives an overview of the results of the 10-fold cross-validation. The bar graph in the bottom left of the figure summarizes all of the results. It depicts the percentages of each of the models having the smallest predicted *MSE*. The histograms in Figure 5 are organized as an upper triangular matrix showing the pairwise comparisons of the different models. The *x*-axis denotes the difference ΔMSE in the predicted *MSE* of the model indicated by the row label and that of the model indicated by the column label. Data samples for which the row (column) model has the smallest predicted *MSE* fall to the left (right) of the origin. The percentage of data samples in favor of the column model is shown in the upper right corner of every histogram.

From the bar graph, we can immediately conclude that the predictive performance of the VAR model is indeed very limited in comparison to the other models. This result is strengthened by the pairwise comparisons in the last column of the matrix of histograms. In this column, the predictive accuracy of the VAR model is compared to that of all of the other models. The VAR model is hardly ever selected as the best predicting model and all of the histograms are heavily skewed toward the least complex model. These results suggest that the VAR model is too complex for the affect data considered in this article. Because the VAR model performs so poorly, we will not discuss it any further.

The VAR-lasso model is also outperformed by all other models (except the VAR model) in terms of predictive accuracy. All histograms in which the VAR-lasso model is compared to another model are skewed in favor of the other model (see second to last column in Figure 5). We are particularly interested in the comparison of the VAR-lasso model with the SAD-lasso model because we introduced the SAD-lasso as an alternative regularization method; the VAR-lasso and the SAD-lasso models have the same degrees of freedom and can, theoretically speaking, construct exactly the same models. In the pairwise comparison between the VAR-lasso model and the SAD-lasso model, we see that the SAD-lasso model has a better predictive performance for 76% of the time series. The histogram is also heavily skewed in favor of the SAD-lasso model. So, although both models can in principle construct exactly the same models, this result shows that different models are obtained depending on the parametrization. For the affect data analyzed in this article, it is favorable to use the SAD-lasso model instead of the VAR-lasso model for a better predictive accuracy. Hence, when interested in Type-II and Type-III dynamics, the VAR, VAR-lasso, or SAD-lasso model may be considered, but out of these three, the SAD-lasso has the best predictive performance.

Despite the SAD-lasso having the best predictive performance of the three models that encompass more complicated dynamics (Type-II and Type-III dynamics), it is the Type-I models (AR, SymVAR, and SymVAR-lasso) that have the best predictive performance for the large majority of the time series (77%). The SAD-lasso model outperforms the SymVAR model in terms of predictive accuracy, but is outperformed by both the AR and the SymVAR-lasso model. The bar graph shows that on the whole the AR model is still the most dominant. Nevertheless, the AR model is outperformed by one of the intermediate models in terms of predictive accuracy in 62% of the cases. The histogram comparing the AR model with the SymVAR-lasso model also indicates that sparse symmetric interactions lead to a better predictive accuracy in 45% of the cases.

Simulation Study

A summary of the results of the cross-validation simulation study is depicted in Figure 6. For a full overview of the simulation results, including the pairwise comparisons between the models, we refer to Appendix D. Every bar graph in Figure 6 depicts the results that are obtained when data are simulated with the model indicated by the title label. The top panel depicts the results that were obtained for the original data and is the same bar graph as the one in the lower left corner of Figure 5. The top panel has been added for comparison purposes.

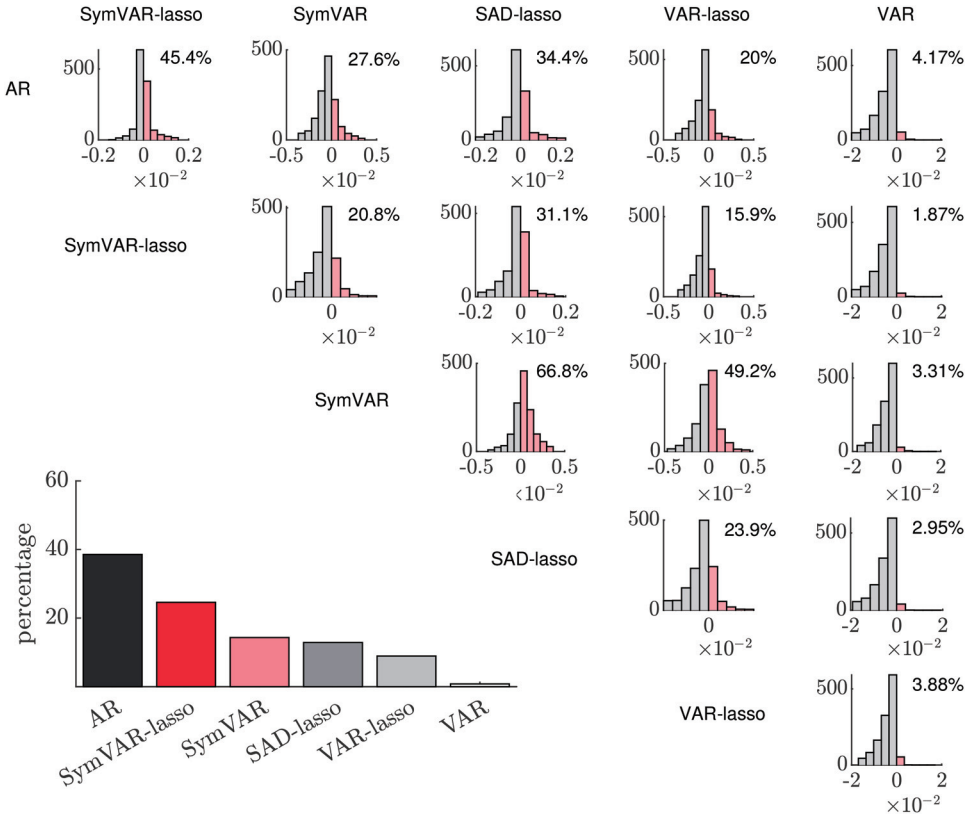
In Figure 6, we see that none of the simulated bar graphs exactly matches the bar graph of the actual data. When data are simulated with the AR model, for instance, we see a substantial increase in the proportion of time series for which the AR model has the best predictive accuracy. In general, we see a boost in predictive performance in the model that was used as data generating mechanism. Furthermore, the SymVAR model is majorly underrepresented (except when the SymVAR model is used for the simulations).

Table 1 depicts the optimal mixing ratios of the different models based on the results in Figure 6. Because the optimal mixing solution is not unique, the results that are depicted are the means and the 95% confidence intervals of the distributions of mixing ratios that have been obtained by means of a nonparametric bootstrap with 10,000 replications. The bar graph corresponding to the optimal mixing solutions is shown in Figure 7.

Based on the 95% confidence interval of [43%, 61%] for the mixing ratio of the AR model, it is unlikely that all time series have been generated by an AR model. In fact, the confidence intervals indicate that none of the models can on itself reproduce the actual cross-validation results. On average, about 53% of the time series could have been produced by an AR model. The second most prevalent model, with

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 5
Results of the 10-Fold Cross-Validation



Note. The histograms depict the distribution of the ΔMSE between the model indicated by the row label and the model indicated by the column label. Points falling to the left (right) of the origin are in favor of the model indicated by the row (column) label. The percentage of points falling strictly to the right of the origin is shown as well. The bar graph depicts the percentage of each model having the best predictive accuracy. VAR = vector autoregressive; SymVar = symmetric VAR; lasso = least absolute shrinkage and selection operator. See the online article for the color version of this figure.

43% [35%, 50%] is the SymVAR model. The VAR model and the lasso models are to a lesser extent required.

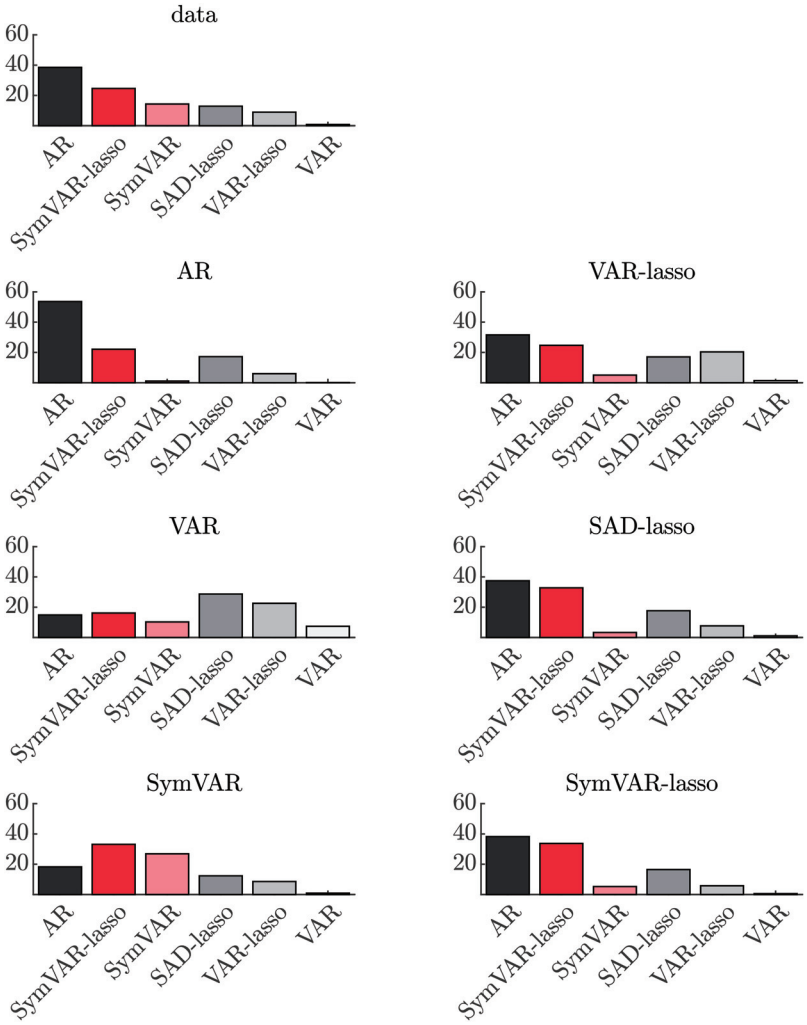
According to the bar graph in Figure 5 (see also the left panel of Figure 7), a model with symmetric interactions (either the SymVAR model or the SymVAR-lasso model) has the best predictive performance in 39% of the cases. A pairwise comparison of the SymVAR-lasso with the AR model also indicates that sparse symmetric interactions are beneficial for prediction in about 45% of the cases. These results agree with the observation that about 43% of the time series could have been generated by a SymVAR model. The reason why the SymVAR-lasso model outperforms the SymVAR model in terms of predictive accuracy is most likely that a significant number of the symmetric interactions are small and difficult to detect given the typical sample size of the time series considered in this article. This idea is supported by the fact that even when data is simulated with the SymVAR model, the SymVAR-lasso model still has the best predictive performance in the majority of the cases (see Figure 6, panel SymVAR).

Conclusion

In this article, we revisited the problem of choosing between full VAR models that are likely to be too complex, and AR models that are possibly too simple in the context of typical psychological applications. The VAR space was partitioned into subclasses with their specific qualitatively different dynamical features. One of these, the class of symmetric VAR (SymVAR) models, can be considered as a natural extension of the simple multivariate AR model. Together, the AR and the SymVAR model encompass simple exponential relaxation dynamics (Type-I dynamics). Both of these models always have an orthogonal eigensystem. Exponential relaxation dynamics can also occur in oblique eigensystems. This type of dynamics (Type-II dynamics), which is shearing-like, belongs to another class of VAR models. The last class of VAR models encompasses orbital (spiraling) motions (Type-III dynamics). Models belonging to this class are characterized by complex eigensystems. The complexity of the VAR model in terms of the number of parameters that has to be estimated grows when going from Type-I dynamics to Type-II dynamics and ultimately to Type-III dynamics. The starting point of the article was that some

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 6
Summary of the Results of the Cross-Validation Simulation Study



Note. Each bar in the bar graphs represents the percentage of cases the corresponding model has the best predictive accuracy when data are simulated using the model indicated by the title label. The top panel is the same bar graph as in Figure 5 and depicts the results that were obtained for the original data. VAR = vector autoregressive; SymVar = symmetric VAR; lasso = least absolute shrinkage and selection operator. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

of these "intermediate" models may be able to pick up on relevant interactions in the data, something which the full VAR model is unable to do because it is overly complex and the sample sizes are generally too small.

Although lasso regularization already enabled us to construct models with a complexity in between that of the VAR model and that of the AR model, it had been shown that the simple AR model in general still performs better in terms of predictive accuracy when compared to models obtained through lasso regularization (Bulteel et al., 2018). We anticipated that the predictive accuracy of models obtained through lasso regularization could be improved by applying the lasso to a reparametrized VAR model. We argued that because the traditional lasso penalizes each off-diagonal element of the transition matrix independently, it primordially scans regions of the VAR space where there are Type-II and Type-III dynamics. Type-I dynamics, which are characterized by symmetric transition matrices, can more easily be retrieved when the transition matrix of the VAR model is split into a symmetric and an antisymmetric component, and when the symmetric and antisymmetric components are penalized instead. This variant of the lasso method is called the SAD (symmetric-antisymmetric decomposition) lasso. Using 10-fold cross-validation, it was shown that the SAD-lasso generally outperforms both the VAR model (97% of the cases) and the VAR-lasso model (76% of the cases).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

In the article, we also investigated whether all time series could have been generated by an AR model (a model without interactions). For this test, we simulated data using the AR model and repeated the 10-fold cross-validation analysis on the simulated data. We then compared the results of the simulation study with the results of the actual data and found significant differences between the results. As an extension, we repeated this simulation study for every model considered in this article. Then, the optimal mixing ratios of the models was computed, the mixing of models which gives us cross-validation results that most closely matches with the actual results. Although the AR model turns out to be the most prominent model (53%), it was shown that it is highly unlikely that all the time series considered in this article were generated from it. The analysis indicates that, on average, up to 43% of the time series could have been generated from a SymVAR model, a simple extension of an AR model with symmetric interactions. This finding is supported by the cross-validation results for the actual data which indicate that models with sparse symmetric interactions (SymVAR-lasso models) outperform the AR model in terms of predictive accuracy in 45% of the time series that are considered. The SymVAR-lasso model outperforming the AR model is in other words not merely a consequence of spurious interactions which can arise because of limited sample sizes. The exact mixing ratios should nonetheless by taken with a pinch of salt, because it is plausible that we did not consider all relevant models in this article.

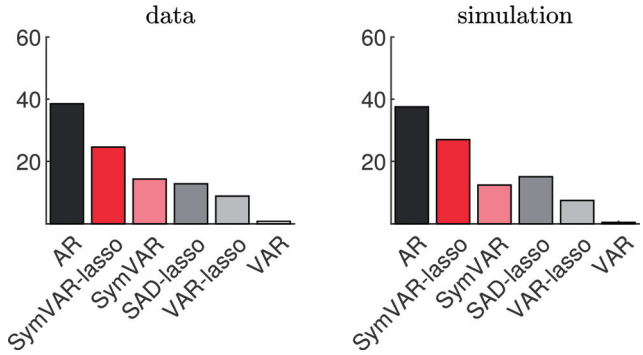
The approach described in this article and the findings can be of particular relevance for the domain of dynamical networks (Borsboom & Cramer, 2013; Bringmann et al., 2016, 2013) and related research domains, where similar data is analyzed. In the dynamical network paradigm, time series data and VAR models are used to construct networks of directed relations. Because of the particular interest in the interactions between different system components, typically unrestricted VAR models are used to build the networks, or VAR models constrained through lasso regularization or other regularization techniques. However, restricting attention to undirected dynamics and allowing only an occasional directed interaction may yield more sound networks. Undirected networks can be constructed using SymVAR models, and if one does not want to exclude the possibility of Type-II and Type-III interactions, the SAD-lasso is a viable alternative to search for both symmetries and asymmetries.

Table 1
Optimal Mixing Ratios of the Models

Model	*M*	95% bootstrap confidence interval
AR	5.3×10^{-1}	$[4.3 \times 10^{-1}, 6.1 \times 10^{-1}]$
SymVAR	4.3×10^{-1}	$[3.5 \times 10^{-1}, 5.0 \times 10^{-1}]$
VAR-lasso	3.4×10^{-2}	$[1.7 \times 10^{-19}, 1.5 \times 10^{-1}]$
VAR	1.1×10^{-2}	$[8.2 \times 10^{-21}, 7.6 \times 10^{-2}]$
SymVAR-lasso	2.4×10^{-4}	$[2.7 \times 10^{-22}, 5.1 \times 10^{-10}]$
SAD-lasso	1.1×10^{-4}	$[3.7 \times 10^{-22}, 6.3 \times 10^{-10}]$

Note. VAR = vector autoregressive; SymVar = symmetric VAR; lasso = least absolute shrinkage and selection operator.

Figure 7
Comparison of the Cross-Validation Results for the Actual Data and the Optimally Mixed Simulated Data



Note. The actual bar graph of the real data (same as in Figure 5) is shown on the left and the simulated bar graph obtained with ideal mixing ratios is shown on the right. VAR = vector autoregressive; SymVar = symmetric VAR; lasso = least absolute shrinkage and selection operator. See the online article for the color version of this figure.

References

Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, *191*, 192–213. https://doi.org/10.1016/j.ins.2011.12.028

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, *54*(1), 579–616. https://doi.org/10.1146/annurev.psych.54.101601.145030

Borsboom, D., & Cramer, A. O. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*(1), 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Bringmann, L. F., Pe, M. L., Vissers, N., Ceulemans, E., Borsboom, D., Vanpaemel, W., Tuerlinckx, F., & Kuppens, P. (2016). Assessing temporal emotion dynamics using networks. *Assessment*, *23*(4), 425–435. https://doi.org/10.1177/1073191116645909

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., Borsboom, D., & Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLoS ONE*, *8*(4), e60188. https://doi.org/10.1371/journal.pone.0060188

Bulteel, K., Mestdagh, M., Tuerlinckx, F., & Ceulemans, E. (2018). VAR (1) based models do not always outpredict AR(1) models in typical psychological applications. *Psychological Methods*, *23*(4), 740–756. https://doi.org/10.1037/met0000178

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *Journal of Nervous and Mental Disease*, *175*(9), 526–536. https://doi.org/10.1097/00005053-198709000-00004

Dejonckheere, E., Kalokerinos, E. K., Bastian, B., & Kuppens, P. (2019). Poor emotion regulation ability mediates the link between depressive symptoms and affective bipolarity. *Cognition and Emotion*, *33*(5), 1076–1083. https://doi.org/10.1080/02699931.2018.1524747

Dejonckheere, E., Mestdagh, M., Houben, M., Rutten, I., Sels, L., Kuppens, P., & Tuerlinckx, F. (2019). Complex affect dynamics add limited information to the prediction of psychological well-being. *Nature Human Behaviour*, *3*(5), 478–491. https://doi.org/10.1038/s41562-019-0555-0

Friedman, J., Hastie, T., Höfling, H., & Tibshirani, R. (2007). Pathwise coordinate optimization. *The Annals of Applied Statistics*, *1*(2), 302–332. https://doi.org/10.1214/07-AOAS131

Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American Statistical Association*, *70*(350), 320–328. https://doi.org/10.2307/2285815

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. https://doi.org/10.1037/a0024595

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer-Verlag.

Heininga, V. E., Dejonckheere, E., Houben, M., Obbels, J., Sienaert, P., Leroy, B., van Roy, J., & Kuppens, P. (2019). The dynamical signature of anhedonia in major depressive disorder: positive emotion dynamics, reactivity, and recovery. *BMC Psychiatry*. Advance online publication. https://doi.org/10.1186/s12888-018-1983-5

Houben, M., Vansteelandt, K., Claes, L., Sienaert, P., Berens, A., Sleuwaegen, E., & Kuppens, P. (2016). Emotional switching in borderline personality disorder: A daily life study. *Personality Disorders: Theory, Research, and Treatment*, *7*(1), 50–60. https://doi.org/10.1037/per0000126

Koval, P., Pe, M. L., Meers, K., & Kuppens, P. (2013). Affect dynamics in relation to depressive symptoms: Variable, unstable or inert? *Emotion*, *13*(6), 1132–1141. https://doi.org/10.1037/a0033579

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer.

Pe, M. L., Brose, A., Gotlib, I. H., & Kuppens, P. (2016). Affective updating ability and stressful events interact to prospectively predict increases in depressive symptoms over time. *Emotion*, *16*(1), 73–82. https://doi.org/10.1037/emo0000097

Pe, M. L., & Kuppens, P. (2012). The dynamic interplay between emotions in daily life: Augmentation, blunting, and the role of appraisal overlap. *Emotion*, *12*(6), 1320–1328. https://doi.org/10.1037/a0028262

Provenzano, J. (2019). *The neural basis of the temporal dynamics of emotions*. KU Leuven.

Schmiedek, F. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: findings from the COGITO study. *Frontiers in Aging Neuroscience*. Advance online publication. https://doi.org/10.3389/fnagi.2010.00027

Sels, L., Ceulemans, E., & Kuppens, P. (2017). Partner-expected affect: How you feel now is predicted by how your partner thought you felt before. *Emotion*, *17*(7), 1066–1077. https://doi.org/10.1037/emo0000304

Sels, L., Ceulemans, E., & Kuppens, P. (2019). All's well that ends well? A test of the peak-end rule in couples' conflict discussions. *European Journal of Social Psychology*, *49*(4), 794–806. https://doi.org/10.1002/ejsp.2547

Strogatz, S. H. (2000). *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering* (1st ed.). CRC Press.

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix A

Dynamics of the VAR(1) Model in the Long-Time Limit

Given a stable VAR model as in Equation 1. The stability implies that the moduli of the eigenvalues of the transition matrix A are strictly smaller than 1. Therefore, A is a bounded operator. Given an observation $\mathbf{y}_0$ at *time* $t = 0$, the expected trajectory of the state vector is given by Equation 7. Here, it is shown that for any initial condition $\mathbf{y}_0 \in \mathbb{R}^d$ and for any consistent norm $||\cdot||$

$$\lim_{t\to\infty} ||A^t\,\mathbf{y}_0|| = 0,$$

so that

$$\lim_{t\to\infty} \langle\, \mathbf{y}_t \,|\, \mathbf{y}_0 \,\rangle = \lim_{t\to\infty} \left(\sum_{i=0}^{t-1} A^i\right) \mathbf{c}$$

$$= (1-A)^{-1}\,\mathbf{c},$$

where the second equality follows from $\sum_{i=0}^{t-1} A^i$ being a Neumann series which converges toward $(1-A)^{-1}$ for any bounded operator A in the limit for t going to infinity.

From the eigen decomposition theorem, it follows that the matrix A can be decomposed as

$$A = Q\,\Lambda\,Q^{-1},$$

where $\Lambda \in \mathbb{C}^{d\times d}$ is a diagonal matrix with the eigenvalues on the diagonal and where $Q \in \mathbb{C}^{d\times d}$ is a matrix with the corresponding eigenvectors in the columns. As a consequence,

$$||A^t\,\mathbf{y}_0|| = ||\,Q\,\Lambda^t\,Q^{-1}\,\mathbf{y}_0\,||.$$

From the consistency of the norm $||\cdot||$ it follows that

$$||\,Q\,\Lambda^t\,Q^{-1}\,\mathbf{y}_0\,|| \leq ||\,Q\,||\,||\,\Lambda^t\,Q^{-1}\,\mathbf{y}_0\,||.$$

Using the standard Euclidean norm, the norm $||\,\Lambda^t\,Q^{-1}\,\mathbf{y}_0\,||$ on the right-hand side can be written as

$$||\,\Lambda^t\,Q^{-1}\,\mathbf{y}_0\,|| = \left[\sum_{i=1}^{d} (\Lambda^t)_{ii}^2\,(Q^{-1}\,\mathbf{y}_0)_i^2\right]^{1/2} = \left[\sum_{i=1}^{d} (\Lambda)_{ii}^{2t}\,(Q^{-1}\,\mathbf{y}_0)_i^2\right]^{1/2}.$$

Because

$$\lim_{t\to\infty} |\,\Lambda_{ii}^t\,| = 0,$$

it follows that

$$\lim_{t\to\infty} \left[\sum_{i=1}^{d} (\Lambda)_{ii}^{2t}\,(Q^{-1}\,\mathbf{y}_0)_i^2\right]^{1/2} = 0.$$

Thus,

$$\lim_{t\to\infty} ||A^t\,\mathbf{y}_0|| = 0.$$

This implies that the vector $A^t\,\mathbf{y}_0$ converges toward the null vector as t goes to infinity.

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix B

Eigenvalues and Eigenvectors of Real, Symmetric Matrices

B.1 Eigenvalues

Theorem

Let $A \in \mathbb{R}^{d\times d}$ be a real, symmetric matrix: $A = A^{\mathrm{T}}$. Then the eigenvalues of A are real.

Proof

Given a $d \times d$ real, symmetric matrix A. Let $\lambda \in \mathbb{C}$ be an eigenvalue of the matrix A with corresponding eigenvector $\mathbf{v}$,

$$A\,\mathbf{v} = \lambda\,\mathbf{v}. \tag{B.1}$$

The complex conjugate of this expression is

$$A\,\overline{\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}}. \tag{B.2}$$

From the Equation B.1 it follows that

$$\overline{\mathbf{v}}^T A\,\mathbf{v} = \lambda\,(\overline{\mathbf{v}}^T\,\mathbf{v}). \tag{B.3}$$

From Equation B.2, on the other hand, it follows that

$$\begin{aligned}\overline{\mathbf{v}}^T A\,\mathbf{v} &= (A\overline{\mathbf{v}})^T\,\mathbf{v}\\ &= \overline{\lambda}\,(\overline{\mathbf{v}}^T\,\mathbf{v}).\end{aligned} \tag{B.4}$$

By equating Equations B.3 and B.4, the equation

$$\lambda\,(\overline{\mathbf{v}}^T\,\mathbf{v}) = \overline{\lambda}\,(\overline{\mathbf{v}}^T\,\mathbf{v}) \tag{B.5}$$

is obtained. Because $\mathbf{v}$ is an eigenvector, it is not a null vector. Therefore, $\overline{\mathbf{v}}^T\,\mathbf{v} > 0$ ($\overline{\mathbf{v}}^T\mathbf{v}$ is a squared norm) and thus

$$\lambda = \overline{\lambda}.$$

In other words, the eigenvalue λ is real.

B.2 Real Eigenvectors

Theorem

The eigenvectors of a real symmetric matrix corresponding to nonzero eigenvalues are real.

Proof

Given a $d \times d$ real, symmetric matrix A. Let $\lambda \in \mathbb{R}\backslash\{0\}$ be an eigenvalue of the matrix A with corresponding eigenvector $\mathbf{v}$. Because λ is real, Equation B.2 states that

$$A\,\overline{\mathbf{v}} = \lambda\overline{\mathbf{v}}.$$

Subtracting Equation B.1 from this equation, the following equality is obtained:

$$\lambda\,(\overline{\mathbf{v}} - \mathbf{v}) = 0.$$

Because $\lambda \neq 0$ and $\mathbf{v} \neq 0$, the equality

$$\overline{\mathbf{v}} = \mathbf{v}$$

holds. Hence, the eigenvector $\mathbf{v}$ is real.

B.3 Orthogonal Eigenvectors

Theorem

Let $A \in \mathbb{R}^{d\times d}$ be a real, symmetric matrix and let $\mathbf{v}_\lambda$ and $\mathbf{v}_{\lambda^*}$ be two eigenvectors corresponding to the eigenvalues λ and λ^* with $\lambda \neq \lambda^*$. Then the eigenvectors $\mathbf{v}_\lambda$ and $\mathbf{v}_{\lambda^*}$ are orthogonal: $\mathbf{v}_\lambda^T\,\mathbf{v}_{\lambda^*} = 0$.

Proof

Given a $d \times d$ real, symmetric matrix A. Let $\lambda, \lambda^* \in \mathbb{R}$ be two eigenvalues of the matrix A for with $\lambda \neq \lambda^*$. Let $\mathbf{v}_\lambda$ and $\mathbf{v}_{\lambda^*}$ denote the corresponding eigenvectors,

$$\begin{aligned}A\,\mathbf{v}_\lambda &= \lambda\,\mathbf{v}_\lambda\\ A\,\mathbf{v}_{\lambda^*} &= \lambda^*\,\mathbf{v}_{\lambda^*}.\end{aligned}$$

On the one hand,

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

$$\mathbf{v}_{\lambda^*}^T A\,\mathbf{v}_\lambda = \lambda\,\mathbf{v}_{\lambda^*}^T\,\mathbf{v}_\lambda.$$

On the other hand,

$$\begin{aligned}\mathbf{v}_{\lambda^*}^T A\,\mathbf{v}_\lambda &= \mathbf{v}_\lambda^T A\,\mathbf{v}_{\lambda^*}\\ &= \lambda^*\,\mathbf{v}_{\lambda^*}^T\,\mathbf{v}_\lambda.\end{aligned}$$

Hence,

$$(\lambda - \lambda^*)\,\mathbf{v}_{\lambda^*}^T\,\mathbf{v}_\lambda = 0.$$

Because $\lambda \neq \lambda^*$ and given that neither of the eigenvectors $\mathbf{v}_\lambda$ and $\mathbf{v}_{\lambda^*}$ is equal to the null vector, it must be true that

$$\mathbf{v}_{\lambda^*}^T\,\mathbf{v}_\lambda = 0.$$

Hence, the eigenvectors $\mathbf{v}_\lambda$ and $\mathbf{v}_{\lambda^*}$ are orthogonal.

Note that the eigenvectors of a degenerate subspace can also be chosen in such a way that they are perpendicular. As a consequence, all eigenvectors of a symmetric, real matrix are either orthogonal or can be made to be orthogonal.

Appendix C

SymVAR

Least-Squares Regression

For any $k = 1, \ldots, d$, the derivative of the least-squares function (5) with respect to c_k is

$$\begin{aligned}\partial_{c_k}\mathscr{L} =& \frac{1}{2}\sum_{j=1}^{d}\Big[\partial_{c_k}(\mathbf{Y}_j - c_j J_{N,1} - XA_j)^T(\mathbf{Y}_j - c_j J_{N,1} - XA_j)\\ &+ (\mathbf{Y}_j - c_j J_{N,1} - XA_j)^T\partial_{c_k}(\mathbf{Y}_j - c_j J_{N,1} - XA_j)\Big]\\ =& -\frac{1}{2}\Big[J_{N,1}^T(\mathbf{Y}_k - c_k J_{N,1} - XA_k) + (\mathbf{Y}_k - c_k J_{N,1} - XA_k)TJ_{N,1}\Big]\\ =& -J_{1,N}(\mathbf{Y}_k - c_k J_{N,1} - XA_k)\\ =& -(J_{1,N}\mathbf{Y}_k - Nc_k - J_{1,N}XA_k).\end{aligned}$$

Equating this expression to zero for all $k = 1, \ldots, d$, and noting that

$$\begin{aligned}A &= \begin{pmatrix} A_1 & A_2 & \cdots & A_d\end{pmatrix},\\ Y &= \begin{pmatrix} Y_1 & Y_2 & \cdots & Y_d\end{pmatrix},\end{aligned}$$

the intercept $\mathbf{c}$ is found to be

$$\hat{\mathbf{c}} = \frac{1}{N}(J_{1,N}Y - J_{1,N}XA)^T. \tag{C.6}$$

The derivative of the least-squares function (5) with respect to the interaction coefficients A_{kl} is bit more intricate because of the symmetry conditions $A_{kl} = A_{lk}$. First, note that for $k \neq l$

$$\begin{aligned}\partial_{A_{kl}}(\mathbf{Y}_j - c_j J_{N,1} - XA_j) &= -\partial_{A_{kl}}\left(\sum_{m=1}^{d}\mathbf{X}_m A_{jm}\right)\\ &= -(\delta_{jk}\,\delta_{l\alpha} + \delta_{jl}\,\delta_{km})\,\mathbf{X}_m\\ &= -(\delta_{jk}\,\mathbf{X}_l + \delta_{jl}\,\mathbf{X}_k),\end{aligned}$$

where

$$\mathbf{X}_m = \begin{pmatrix} y_{0,m}\\ y_{1,m}\\ \vdots\\ y_{N-1,m}\end{pmatrix}$$

refers to the mth column of the matrix X. Consequently,

$$\begin{aligned}\partial_{A_{kl}}\mathscr{L} =& \frac{1}{2}\sum_{j=1}^{d}\Big[\partial_{A_{kl}}(\mathbf{Y}_j - c_j J_{N,1} - XA_j)^T(\mathbf{Y}_j - c_j J_{N,1} - XA_j)\\ &+ (\mathbf{Y}_j - c_j J_{N,1} - XA_j)^T\partial_{A_{kl}}(\mathbf{Y}_j - c_j J_{N,1} - XA_j)\Big]\\ =& -\frac{1}{2}\Big[(\delta_{jk}X_l - \delta_{jl}X_k)^T + (\mathbf{Y}_j - c_j J_{N,1} - XA_j)\\ &+ (\mathbf{Y}_j - c_j J_{N,1} - XA_j)^T(\delta_{jk}X_l - \delta_{jl}X_k)\Big]\\ =& -(\delta_{jk}X_l + \delta_{jl}X_k)^T + (\mathbf{Y}_j - c_j J_{N,1} - XA_j)\\ =& -X_l^T(\mathbf{Y}_{k-c_k}J_{N,1} - XA_k) - X_k^T(\mathbf{Y}_{l-c_l}J_{N,1} - XA_l).\end{aligned}$$

The derivatives with respect to the self-interactions A_{kk} are simply given by

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

$$\partial_{A_{kk}}\mathscr{L} = -\mathbf{X}_k^T(\mathbf{Y}_k - c_k J_{N,1} - XA_k).$$

Equating all of the derivatives with respect to the interactions A_{kl} to zero, the following set of $d \times d$ equations can be obtained:

$$\begin{aligned} 0 &= (Y - J_{N,1}\,\mathbf{c}^T - XA)TX + X^T(Y - J_{N,1}\,\mathbf{c}^T - XA) \\ &= -[X^T(Id_N - \frac{1}{N}J_{N,N})XA + AX^T(Id_N - \frac{1}{N}J_{N,N})X] \\ &\quad + [X^T(Id_N - \frac{1}{N}J_{N,N})Y + Y^T(Id_N - \frac{1}{N}J_{N,N})X] \end{aligned}$$

(note that $\mathbf{c}$ is a column vector and $\mathbf{c}^T$ a row vector). In the second equality, we have use Equation C.6 and rearranged the terms. Using the definition (9), we can rearrange the obtained system of equations to obtain the following Lyapunov equation:

$$\tilde{X^T}XA + AX^T\tilde{X} = \tilde{X}^TY + Y^T\tilde{X}.$$

Using the Kronecker product $\otimes$ and the vec-operator $\mathbf{vec}(A)$ which stacks the columns of any matrix A, this Lyapunov equation can be rewritten as

$$(Id_d \otimes \tilde{X^T}X + \tilde{X^T}X \otimes Id_d)\mathbf{vec}(A) = \mathbf{vec}(\tilde{X}^TY + Y^T\tilde{X}),$$

and thus the estimate for A is found to be

$$\mathbf{vec}(A) = (Id_d \otimes \tilde{X^T}X + \tilde{X^T}X \otimes Id_d)^{-1}\mathbf{vec}(\tilde{X}^TY + Y^T\tilde{X}).$$

In summary, the least-square estimates of the SymVAR(1) model are given by

$$\begin{aligned} \hat{\mathbf{c}} &= \frac{1}{N}(J_{1,N}Y - J_{1,N}XA)T, \\ \mathbf{vec}(\hat{A}) &= (Id_d \otimes \tilde{X^T}X + \tilde{X^T}X \otimes Id_d)^{-1}\mathbf{vec}(\tilde{X}^TY + Y^T\tilde{X}). \end{aligned}$$

Lasso Regression

Equation 4 can equivalently be written as

$$\mathrm{vec}(Y) = \mathbf{c} \otimes J_{N,1} + (Id_d \otimes X)\,\mathbf{vec}(A^T) + \boldsymbol{\epsilon} \otimes \mathrm{J}_{\mathrm{N},1}.$$

For $d = 3$, for example, this system of equations is given by

$$\begin{aligned} \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \mathbf{Y}_3 \end{pmatrix} &= \begin{pmatrix} J_{N,1} & 0 & 0 \\ 0 & J_{N,1} & 0 \\ 0 & 0 & J_{N,1} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \\ &\quad + \begin{pmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{pmatrix} \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \mathbf{A}_3 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_1 J_{N,1} \\ \boldsymbol{\epsilon}_2 J_{N,1} \\ \boldsymbol{\epsilon}_3 J_{N,1} \end{pmatrix}. \end{aligned}$$

For the SymVAR, there are coefficients that occur twice in the vector $\mathbf{vec}(A)$. To see this, write

$$\begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \mathbf{X}_3 & 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & \mathbf{X}_1 & \mathbf{X}_2 & \mathbf{X}_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{X}_1 & \mathbf{X}_2 & \mathbf{X}_3 \end{pmatrix} \begin{pmatrix} A_{11} \\ A_{12} \\ A_{13} \\ A_{21} \\ A_{22} \\ A_{23} \\ A_{31} \\ A_{32} \\ A_{33} \end{pmatrix}.$$

Imposing the symmetry constraint $A = A^T$, this product can equivalently be written as

$$\begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \mathbf{X}_3 & 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & \mathbf{X}_1 & \mathbf{X}_2 & \mathbf{X}_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{X}_1 & \mathbf{X}_2 & \mathbf{X}_3 \end{pmatrix} \begin{pmatrix} A_{11} \\ A_{12} \\ A_{13} \\ A_{12} \\ A_{22} \\ A_{23} \\ A_{13} \\ A_{23} \\ A_{33} \end{pmatrix}.$$

During a lasso regression, the algorithm does not necessarily know that there are coefficients that appear twice. Therefore, these duplicates should be removed. The first coefficient that occurs for the second time is the coefficient A_{12}. This coefficient is associated with the second and fourth column of the matrix $(Id_d \otimes X)$. Adding the fourth column to the second and removing it afterward together with the fourth interaction coefficient does not alter the system of equations. In other words, the matrix product above is equivalent with

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

$$\begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \mathbf{X}_3 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{X}_1 & 0 & \mathbf{X}_2 & \mathbf{X}_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{X}_1 & \mathbf{X}_2 & \mathbf{X}_3 \end{pmatrix} \begin{pmatrix} A_{11} \\ A_{12} \\ A_{13} \\ A_{22} \\ A_{23} \\ A_{13} \\ A_{23} \\ A_{33} \end{pmatrix}.$$

$$(Id_3 \otimes X)\,\mathbf{vec}(A^T) = \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \mathbf{X}_3 & 0 & 0 & 0 \\ 0 & \mathbf{X}_1 & 0 & \mathbf{X}_2 & \mathbf{X}_3 & 0 \\ 0 & 0 & \mathbf{X}_1 & \mathbf{X}_2 & 0 & \mathbf{X}_3 \end{pmatrix} \begin{pmatrix} A_{11} \\ A_{12} \\ A_{13} \\ A_{22} \\ A_{23} \\ A_{33} \end{pmatrix} \tag{C.7}$$

The same reasoning applies to the coefficients A_{13} and A_{23}. Following the same methodology, the equivalence can be obtained. Note that every coefficient in Equation C.7 appears only once, as required.

Appendix D

Results of the Simulation Study

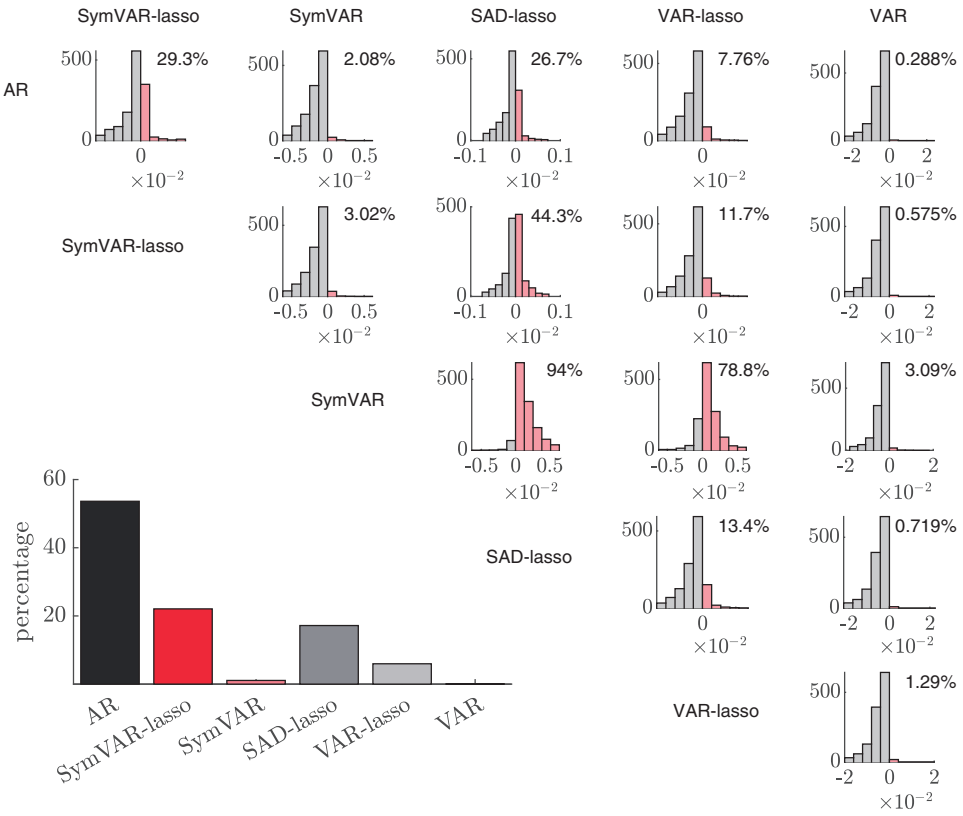
This appendix includes all the results of the simulation study. The results are depicted in the Figures D1–D6. Each of these figures is similar to Figure 5 which exhibits the results of the cross-validation analysis on the true data. An overview of the bar graph appearing in Figures D1–D6 is shown in Figure 6.

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure D1

Results of the 10-Fold Cross-Validation on Data Simulated With the AR Model



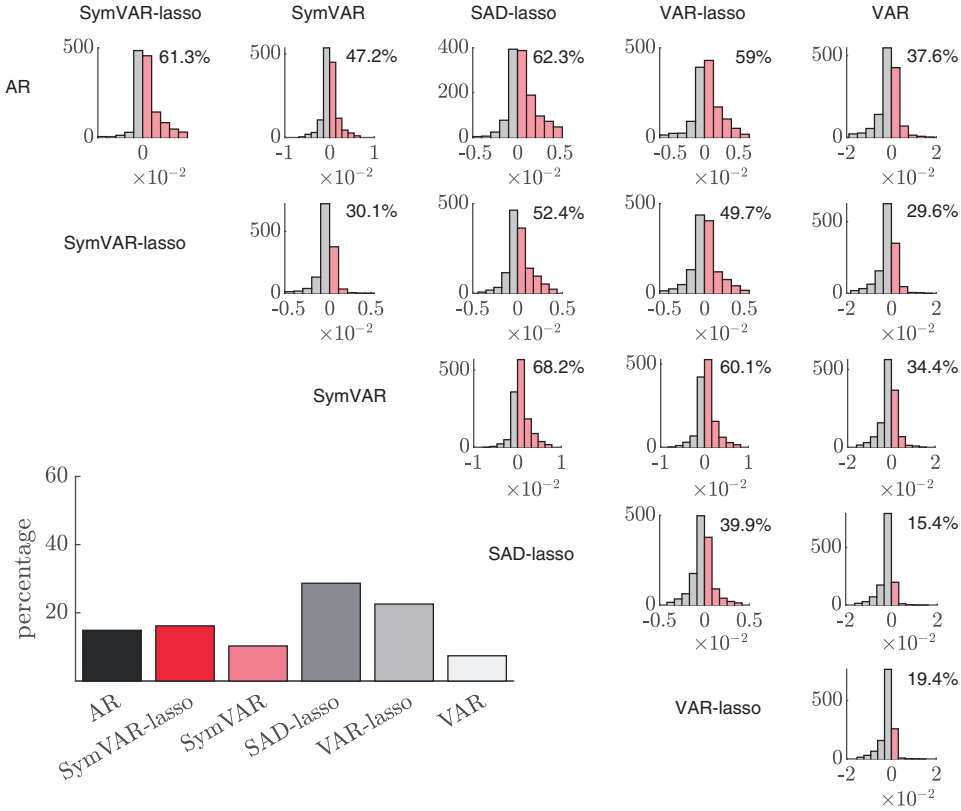
Note. The histograms depict the distribution of the Δ*MSE* between the model indicated by the row label and the model indicated by the column label. Points falling to the left (right) of the origin are in favor of the model indicated by the row (column) label. The percentage of points falling strictly to the right of the origin is shown as well. The bar graph depicts the percentage of each model having the best predictive accuracy. VAR = vector autoregressive; SymVar = symmetric VAR; lasso = least absolute shrinkage and selection operator. See the online article for the color version of this figure.

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure D2

Results of the 10-Fold Cross-Validation on Data Simulated With the VAR Model



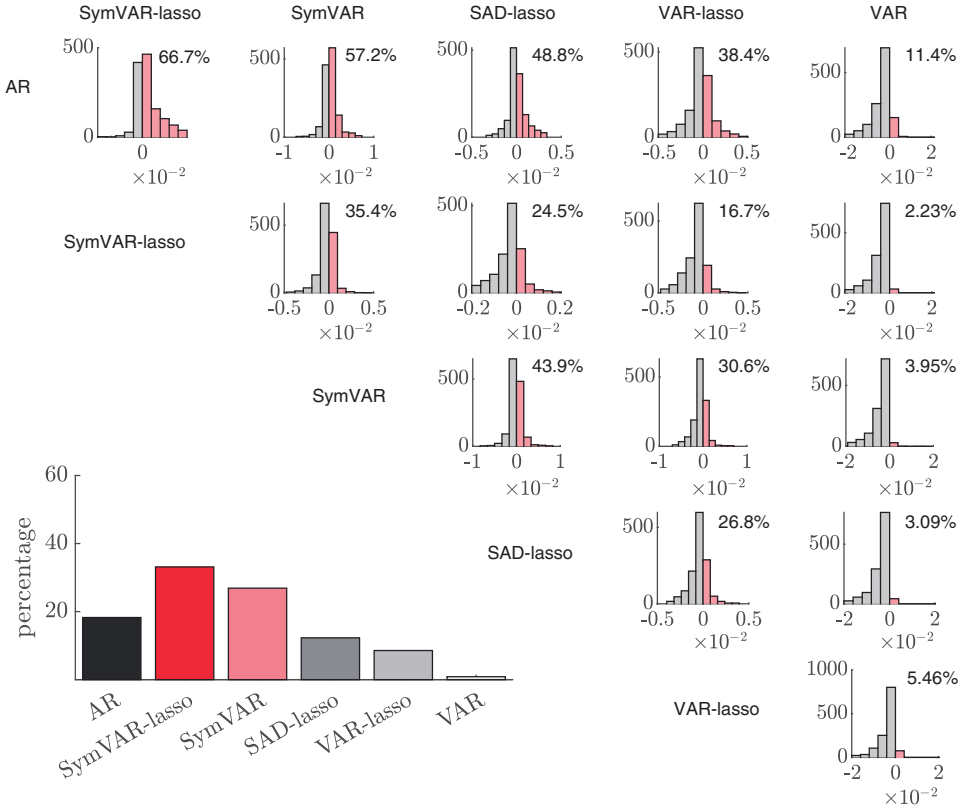
Note. The histograms depict the distribution of the Δ*MSE* between the model indicated by the row label and the model indicated by the column label. Points falling to the left (right) of the origin are in favor of the model indicated by the row (column) label. The percentage of points falling strictly to the right of the origin is shown as well. The bar graph depicts the percentage of each model having the best predictive accuracy. VAR = vector autoregressive; SymVar = symmetric VAR; lasso = least absolute shrinkage and selection operator. See the online article for the color version of this figure.

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure D3

Results of the 10-Fold Cross-Validation on Data Simulated With the SymVAR Model



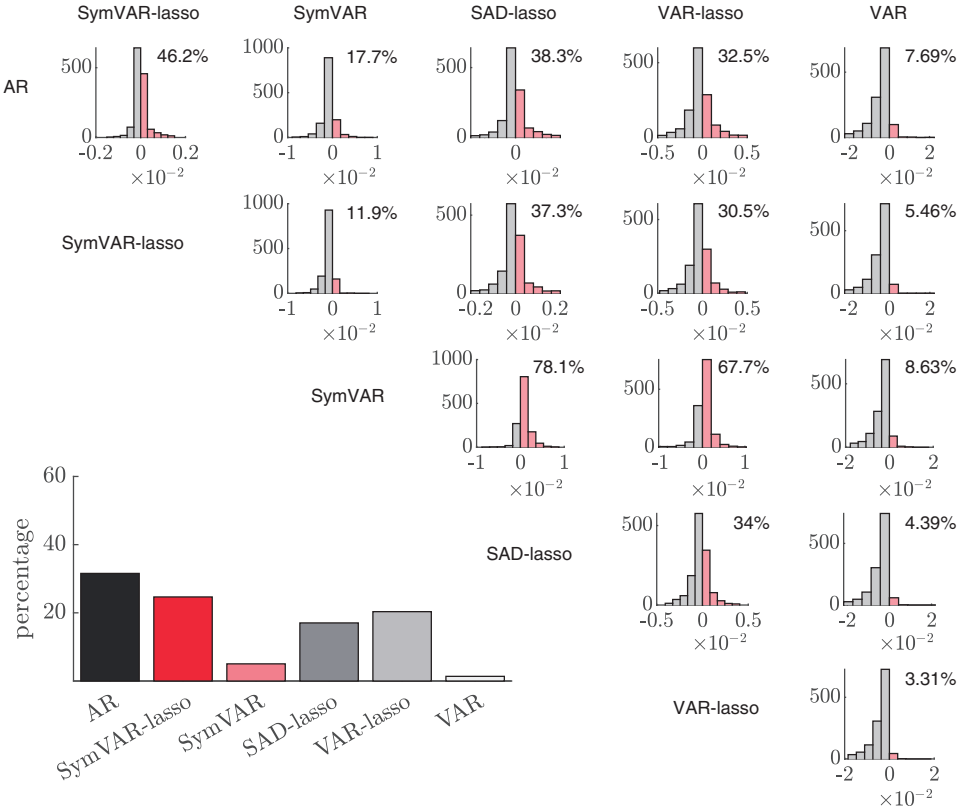
Note. The histograms depict the distribution of the Δ*MSE* between the model indicated by the row label and the model indicated by the column label. Points falling to the left (right) of the origin are in favor of the model indicated by the row (column) label. The percentage of points falling strictly to the right of the origin is shown as well. The bar graph depicts the percentage of each model having the best predictive accuracy. VAR = vector autoregressive; SymVar = symmetric VAR; lasso = least absolute shrinkage and selection operator. See the online article for the color version of this figure.

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure D4

Results of the 10-Fold Cross-Validation on Data Simulated With the VAR-Lasso Model



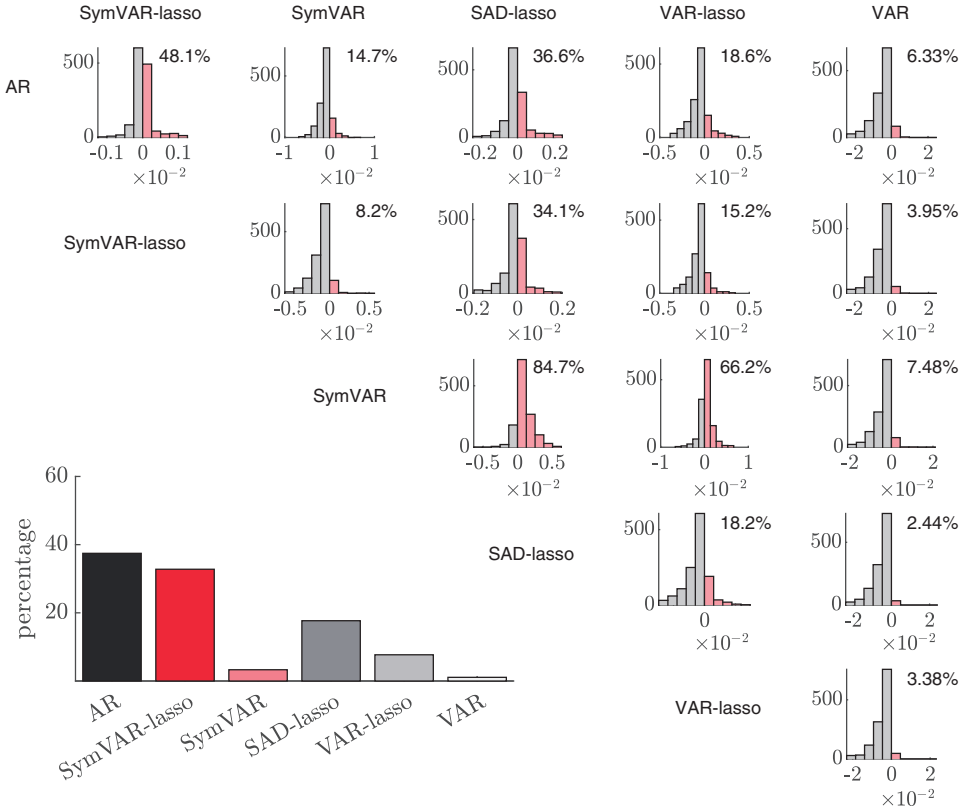
Note. The histograms depict the distribution of the Δ*MSE* between the model indicated by the row label and the model indicated by the column label. Points falling to the left (right) of the origin are in favor of the model indicated by the row (column) label. The percentage of points falling strictly to the right of the origin is shown as well. The bar graph depicts the percentage of each model having the best predictive accuracy. VAR = vector autoregressive; SymVar = symmetric VAR; lasso = least absolute shrinkage and selection operator. See the online article for the color version of this figure.

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure D5

Results of the 10-Fold Cross-Validation on Data Simulated With the SAD-Lasso Model



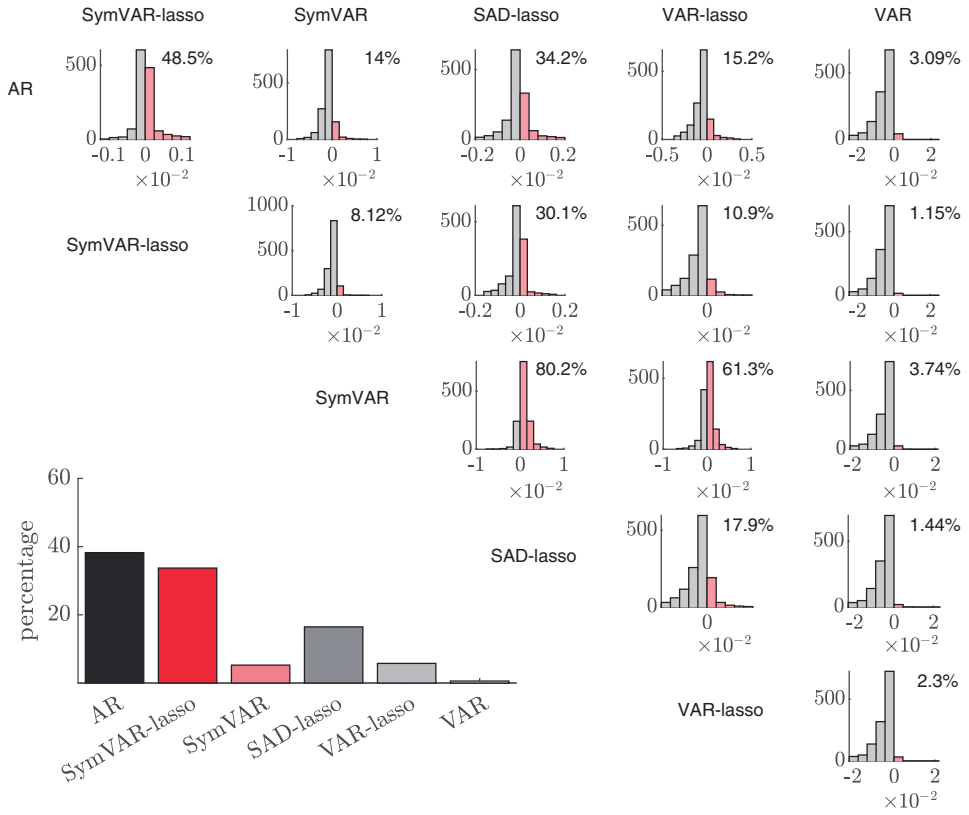
Note. The histograms depict the distribution of the Δ*MSE* between the model indicated by the row label and the model indicated by the column label. Points falling to the left (right) of the origin are in favor of the model indicated by the row (column) label. The percentage of points falling strictly to the right of the origin is shown as well. The bar graph depicts the percentage of each model having the best predictive accuracy. VAR = vector autoregressive; SymVar = symmetric VAR; lasso = least absolute shrinkage and selection operator. See the online article for the color version of this figure.

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure D6

Results of the 10-Fold Cross-Validation on Data Simulated With the SymVAR-Lasso Model



Note. The histograms depict the distribution of the Δ*MSE* between the model indicated by the row label and the model indicated by the column label. Points falling to the left (right) of the origin are in favor of the model indicated by the row (column) label. The percentage of points falling strictly to the right of the origin is shown as well. The bar graph depicts the percentage of each model having the best predictive accuracy. VAR = vector autoregressive; SymVar = symmetric VAR; lasso = least absolute shrinkage and selection operator. See the online article for the color version of this figure.

Received July 9, 2019
Revision received February 24, 2021
Accepted February 25, 2021 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.