

Article

# Causal Vector Autoregression Enhanced with Covariance and Order Selection

Marianna Bolla <sup>1,\*</sup> , Dongze Ye <sup>2</sup> , Haoyu Wang <sup>3</sup> , Renyuan Ma <sup>4</sup>, Valentin Frappier <sup>5</sup>, William Thompson <sup>6</sup>, Catherine Donner <sup>7</sup>, Máté Baranyi <sup>1</sup>  and Fatma Abdelkhalek <sup>8</sup> 

- <sup>1</sup> Department of Stochastics, Budapest University of Technology and Economics, 1111 Budapest, Hungary  
<sup>2</sup> Department of Computer Science, University of Southern California, Los Angeles, CA 90007, USA  
<sup>3</sup> Committee on Computational and Applied Mathematics, University of Chicago, Chicago, IL 60637, USA  
<sup>4</sup> Department of Statistics, Yale University, New Haven, CT 06520, USA  
<sup>5</sup> UFR Sciences and Techniques, Nantes University, 44035 Nantes, France  
<sup>6</sup> Lindner College of Business, University of Cincinnati, Cincinnati, OH 45221, USA  
<sup>7</sup> Data Science and Analytics Institute, University of Oklahoma, Norman, OK 73019, USA  
<sup>8</sup> Department of Statistics, Mathematics, and Insurance, Faculty of Commerce, Assiut University, Assiut Governorate 71515, Egypt  
\* Correspondence: marib@math.bme.hu; Tel.: +36-1-2000646

**Abstract:** A causal vector autoregressive (CVAR) model is introduced for weakly stationary multivariate processes, combining a recursive directed graphical model for the contemporaneous components and a vector autoregressive model longitudinally. Block Cholesky decomposition with varying block sizes is used to solve the model equations and estimate the path coefficients along a directed acyclic graph (DAG). If the DAG is decomposable, i.e., the zeros form a reducible zero pattern (RZP) in its adjacency matrix, then covariance selection is applied that assigns zeros to the corresponding path coefficients. Real-life applications are also considered, where for the optimal order  $p \geq 1$  of the fitted CVAR( $p$ ) model, order selection is performed with various information criteria.



**Citation:** Bolla, Marianna, Dongze Ye, Haoyu Wang, Renyuan Ma, Valentin Frappier, William Thompson, Catherine Donner, Máté Baranyi, and Fatma Abdelkhalek. 2023. Causal Vector Autoregression Enhanced with Covariance and Order Selection. *Econometrics* 11: 7. <https://doi.org/10.3390/econometrics11010007>

Academic Editors: Benedikt Poetscher, Leopold Sögner and Martin Wagner

Received: 22 August 2022  
Revised: 17 February 2023  
Accepted: 20 February 2023  
Published: 24 February 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** structural vector autoregression; causality along a DAG; block Cholesky decomposition; covariance selection; order selection

**MSC:** 15B05; 62M10; 65F99; 65F30

## 1. Introduction

The purpose of the present paper is to connect graphical modeling tools and time series models together via path coefficient estimation. In statistics, the path analysis was established by the geneticist Wright (1934) about a century ago, but he used complicated entrywise calculations with partial correlations. Taking these partial correlations of a pair of variables in a multidimensional data set conditioned on another set of variables makes things overtly complicated, as the conditioning set changes in the steps. A bit later, in econometrics, structural equation modeling (SEM) was developed; the prominent author Haavelmo (1943) obtained the Nobel price for it later. The maximum likelihood estimation (MLE) of the parameters in the Gaussian case was elaborated by Joreskog (1977). At the same time, Ref. Wold (1985), the inventor of partial least squares regression (PLS), used matrix calculations, and Ref. Kiiveri et al. (1984) already used block matrix decompositions when dividing their variables into endogenous and exogenous ones. However, none of these authors applied steadily algorithms of block LDL (variant of the Cholesky) decomposition alone, without using partial correlations. Furthermore, they did not consider time series.

Here, we give a rigorous block matrix approach of these problems that originated in statistics and time series analysis. Furthermore, we enhance the usual and structural vector autoregressive (VAR and SVAR) models, discussed e.g., in Deistler and Scherrer

(2019) and [Deistler and Scherrer \(2022\)](#), with a causal component that has an effect between the coordinates contemporaneously. Therefore, we call our causal vector autoregressive model CVAR. Joint effects between the contemporaneous components are also considered in SVAR models of [Keating \(1996\)](#); [Lütkepohl \(2005\)](#); and [Kilian and Lütkepohl \(2017\)](#), but just recursive ordering of the variables, and no specific structure of the underlying directed graph is investigated. Though in [Wold \(1960\)](#) a causal chain model is introduced with an exogenous and a lagged endogenous variable, Gaussian Markov processes and usual regression estimates are used in the context of econometric problems. This research is also inspired by the paper [Wermuth \(1980\)](#), where recursive ordering of the variables is crucial, without using any time component.

Ref. [Eichler \(2006\)](#) introduces causality as a fundamental tool for the empirical investigations of dynamic interactions in multivariate time series. He also discusses the differences between the structural and Granger causality. The former one appears in the SVAR models (see [Geweke \(1984\)](#)), whereas the latter one first appears in [Wiener \(1956\)](#), then in [Granger \(1969\)](#), and is sometimes called Wiener–Granger causality. Without causality between the contemporaneous components, our model in the Gaussian case also resembles the one of [Eichler \(2012\)](#), where the error term (shock) can have correlated components. Our higher order recursive VAR model can be transformed into a model like this, but the price is losing the recursive structure. The VAR model of [Brillinger \(1996\)](#) has a similar structure as ours with uncorrelated error terms; but no further benefits of recursive models, such as RZP, induced by the underlying DAG, are discussed. Ref. [Sims \(1980\)](#) investigates the use of different types of structural equations and autoregressive models in macroeconomics, without suggesting numerical algorithms. However, historically, this survey paper was among the first ones which pointed out the difference between the existing macroeconomic models so far and distinguished endogenous and exogenous variables. The method of the most recent paper [Bazinas and Nielsen \(2022\)](#) is based on the reduced form system and is constructed by the conditional distribution of two endogenous variables, given a catalyst or multiple catalysts; lagged effects are assessed, without having a longer time series, and stationarity is not assumed.

Throughout the paper, second order processes are considered that can be assumed to asymptotically follow multivariate Gaussian distribution. In Section 2, the different types of VAR models are compared, and a novel CVAR model is introduced, combining a recursive graphical model contemporaneously and a VAR( $p$ ) model longitudinally. In Section 3, the models are described in details, together with introducing algorithms for the parameter estimation. In Section 3.1, the unrestricted CVAR( $p$ ) model is introduced, while in Section 3.2, the restricted cases are treated, with some prescribed zeros in the path coefficients. Relation to covariance selection and decomposability is discussed too. In Section 4, application to real life data is presented together with information criteria for order selection (optimal choice of  $p$ ). The results and estimation schemes are summarized in Section 5; finally, in Section 6, conclusions and further perspectives are posed. The proofs of the theorems and the detailed description of the algorithms are to be found in Appendix A, while the pseudocodes in Appendix B. To illustrate the CVAR model and related algorithms, there are supporting Python files and notebooks uploaded, together with some additional tables and figures. These are included in the Supplementary Material.

## 2. Materials and Methods

First, the different purpose VAR( $p$ ) models for the  $d$ -dimensional, weakly stationary process  $\{\mathbf{X}_t\}$  are enlisted and compared. The first two models are known in the literature, whereas the last two are our contributions, for which block matrix decomposition based algorithms are introduced in Section 3, and they are illustrated in Section 4 on real life data.

- **Reduced form** VAR( $p$ ) model: for given integer  $p \geq 1$ , it is

$$\mathbf{X}_t + M_1\mathbf{X}_{t-1} + \dots + M_p\mathbf{X}_{t-p} = \mathbf{V}_t, \quad t = p + 1, p + 2, \dots, \quad (1)$$

where  $\mathbf{V}_t$  is white noise, it is uncorrelated with  $\mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-p}$ , it has zero expectation and covariance matrix  $\Sigma$  (not necessarily diagonal, but positive definite), and the matrices  $M_j$  satisfy the stability conditions (see [Deistler and Scherrer \(2019\)](#)). (Sometimes,  $\mathbf{X}_t$  is isolated on the left-hand side.)  $\mathbf{V}_t$  is called *innovation*, i.e., the error term of the (added value to the) best one-step ahead linear prediction of  $\mathbf{X}_t$  with its past, which (in the case of a VAR( $p$ ) model) can be carried out with the  $p$ -lag long past  $\mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-p}$ . Here, the ordering of the components of  $\mathbf{X}_t$  does not matter: if it is changed (with some permutation of  $\{1, \dots, d\}$ ), then clearly the rows of the matrices  $M_j$ s and, furthermore, the rows and columns of  $\Sigma$  are permuted accordingly.

- **Structural form SVAR( $p$ ) model:** for given integer  $p \geq 1$ , it is

$$A\mathbf{X}_t + B_1\mathbf{X}_{t-1} + \dots + B_p\mathbf{X}_{t-p} = \mathbf{U}_t, \quad t = p + 1, p + 2, \dots, \quad (2)$$

where the white noise term  $\mathbf{U}_t$  is uncorrelated with  $\mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-p}$ , and it has zero expectation with uncorrelated components, i.e., with positive definite, diagonal covariance matrix  $\Delta$ .  $A$  is a  $d \times d$  upper triangular matrix with 1s along its main diagonal, whereas  $B_1, \dots, B_p$  are  $d \times d$  matrices; see also [Lütkepohl \(2005\)](#). The components of  $\mathbf{U}_t$  are called *structural shocks*, and they are mutually uncorrelated and assigned to the individual variables.

Here, the ordering of the components of  $\mathbf{X}_t$  does matter: if it is changed (with some permutation of  $\{1, \dots, d\}$ ), then the matrices  $A$ ,  $B_j$  and  $\Delta$  cannot be obtained in a simple way; they profoundly change under the given permutation.

However, there is a one-to-one correspondence between the reduced and structural model; since  $A$  is invertible, from Equation (2), Equation (1) can be obtained (and vice versa):

$$\mathbf{X}_t + A^{-1}B_1\mathbf{X}_{t-1} + \dots + A^{-1}B_p\mathbf{X}_{t-p} = A^{-1}\mathbf{U}_t, \quad t = p + 1, p + 2, \dots,$$

where  $M_j = A^{-1}B_j$ ,  $\mathbf{V}_t = A^{-1}\mathbf{U}_t$ , and  $\Sigma = A^{-1}\Delta A^{T-1}$ ; further,  $|\Sigma| = |\Delta|$  as  $|A| = 1$ .

- **Causal CVAR( $p$ ) unrestricted model:** it also obeys Equation (2), but here the ordering of the components follows a causal ordering, given e.g., by an expert's knowledge. This is a recursive ordering along a "complete" DAG, where the permutation (labeling) of the graph nodes (assigned to the components of  $\mathbf{X}_t$ ) is such that  $X_{t,i}$  can be caused by  $X_{t,j}$  whenever  $i < j$ , which means a  $j \rightarrow i$  directed edge. Here, the causal effects are meant contemporaneously, and reflected by the upper triangular structure of the matrix  $A$ .

It is important that, in any ordering of the jointly Gaussian variables, a Bayesian network (in other words, a Gaussian directed graphical model) can be constructed, in which every node (variable) is regressed linearly with the variables corresponding to higher label nodes. The partial regression coefficients behave like path coefficients, also used in SEM. If the DAG is complete, then there are no zero constraints imposed on the partial regression coefficients. Here, building the DAG just aims at finding a sensible ordering of the variables.

- **Causal CVAR( $p$ ) restricted model:** here, an incomplete DAG is built, based on partial correlations.

First, we build an undirected graph: do not connect  $i$  and  $j$  if the partial correlation coefficients of  $X_i$  and  $X_j$ , eliminating the effect of the other variables is 0 (theoretically), or less than a threshold (practically). Such an undirected graphical model is called Markov random field (MRF). It is known (see [Rao \(1973\)](#) and [Lauritzen \(2004\)](#)) that partial correlations can be calculated from the concentration matrix (inverse of the covariance matrix). However, here the upper left block of the inverse of the large block matrix, containing the first  $p$  autocovariance matrices, is used. If this undirected graph is triangulated, then in a convenient (so-called perfect) ordering of the nodes, the zeros of the adjacency matrix form an RZP. We can find such a (not necessarily unique) ordering of the nodes with the maximal cardinality search (MCS) algorithm, together

with cliques and separators of a so-called junction tree (JT); see Lauritzen (2004), Koller and Friedman (2009), and Bolla et al. (2019). In this ordering (labeling) of the nodes, a DAG can also be constructed, which is Markov equivalent to the undirected one (it has no so-called sink V configuration); for further details, see Section 3.2.

Having an RZP in the CVAR restricted model, we use the incomplete DAG for estimation. With the covariance selection method of Dempster (1972), the starting concentration matrix is re-estimated by imposing zero constraints for its entries in the RZP positions (symmetrically). By the theory (see, e.g., Bolla et al. (2019)), this will result in zero entries of  $A$  in the no directed edge positions.

Note that the unrestricted CVAR model can use an incomplete DAG as well, where the labeling of its nodes follows the perfect labeling of the undirected graph; still, the parameter matrices  $A$  and  $B$ 's are "full" in the sense that no zeros of  $A$  are guaranteed in the no-edge positions of the graph. Their entries are just considered as path coefficients of the contemporaneous and lagged effects, respectively. On the contrary, in the restricted CVAR model, action is carried out for introducing zero entries in  $A$  in the no-edge positions. If the desired zeros form an RZP, the covariance selection has a closed form (see Lauritzen (2004)). In the lack of an RZP, the covariance selection still works, but it needs an infinite (convergent) iteration, called iterative proportional scaling (IPS), see Lauritzen (2004) and Bolla et al. (2019). Other possibility is to moralize the DAG (connect parents that are not connected and thus eliminate the sink Vs), and work with the so-obtained undirected graph.

Then, both in the unrestricted and restricted CVAR( $p$ ) models, an order selection is initiated to choose the optimal  $p$ , based on information criteria, such as AIC, BIC, AICC, and HQ, where only the number of parameters differs in the two cases. Actually, in the restricted case, the product-moments are calculated only within the cliques, and since separators are subsets of them, we can reduce the computational complexity of our algorithm that is spectacular when the number of nodes is "large".

### 3. Results

#### 3.1. The Unrestricted Causal VAR( $p$ ) Model

The directed Gaussian graphical model of Wermuth (1980) does not consider time development; it is, in fact, a CVAR(0) model. In addition, note that at this point, the ordering of the jointly Gaussian variables is not relevant, since in any recursive ordering of them (encoded in  $A$ ), a Gaussian directed graphical model (in other words, a Gaussian Bayesian network) can be constructed, where every variable is regressed linearly with the higher label ones.

To illustrate the  $p > 0$  case, first we introduce the unrestricted CVAR(1) model. This has a special interest, as can be used for longitudinal data spanning a short time interval or adapted to the situation when  $\mathbf{X}_{t-1}$  represents the exogenous, and  $\mathbf{X}_t$  the endogenous variables in their components.

Let  $\{\mathbf{X}_t\}$  be a  $d$ -dimensional, weakly stationary process with real valued components of zero expectation and covariance matrix function  $C(h)$ ,  $h = 0, \pm 1, \pm 2, \dots$ ;  $C(-h) = C^T(h)$ . All deterministic and random vectors are column vectors and so  $C(h) = \mathbb{E}\mathbf{X}_t\mathbf{X}_{t+h}^T$  does not depend on  $t$ , by weak stationarity. The CVAR(1) model equation is

$$A\mathbf{X}_t + B\mathbf{X}_{t-1} = \mathbf{U}_t, \quad t = 1, 2, \dots, \quad (3)$$

where  $A$  is a  $d \times d$  upper triangular matrix with 1s along its main diagonal,  $B$  is a  $d \times d$  matrix; furthermore, the white noise random vector  $\mathbf{U}_t$  is uncorrelated with (in the Gaussian case, independent of)  $\mathbf{X}_{t-1}$ , has zero expectation, and covariance matrix  $\Delta = \text{diag}(\delta_1, \dots, \delta_d)$ .

Let  $\mathcal{C}_2$  denote the covariance matrix of the stacked random vector  $(\mathbf{X}_t^T, \mathbf{X}_{t-1}^T)^T$  which, in block matrix form, is as follows:

$$\mathcal{C}_2 = \begin{pmatrix} C(0) & C^T(1) \\ C(1) & C(0) \end{pmatrix}. \quad (4)$$

It is symmetric and positive definite if the process is of full rank regular (which means that its spectral density matrix is of full rank, see [Bolla and Szabados \(2021\)](#)) that is assumed in the sequel. It is well known that the inverse of  $\mathfrak{C}_2$ , the so-called concentration matrix  $\mathbf{K}$ , has the block-matrix form

$$\mathbf{K} = \begin{pmatrix} \mathbf{C}^{-1}(1|0) & -\mathbf{C}^{-1}(1|0)\mathbf{C}^T(1)\mathbf{C}^{-1}(0) \\ -\mathbf{C}^{-1}(0)\mathbf{C}(1)\mathbf{C}^{-1}(1|0) & \mathbf{C}^{-1}(0) + \mathbf{C}^{-1}(0)\mathbf{C}(1)\mathbf{C}^{-1}(1|0)\mathbf{C}^T(1)\mathbf{C}^{-1}(0) \end{pmatrix},$$

where  $\mathbf{C}(1|0) = \mathbf{C}(0) - \mathbf{C}^T(1)\mathbf{C}^{-1}(0)\mathbf{C}(1)$  is the conditional covariance matrix  $\mathbf{C}(t|t-1)$  of the distribution of  $\mathbf{X}_t$ , given  $\mathbf{X}_{t-1}$ ; by weak stationarity, it does not depend on  $t$  either; therefore, it is denoted by  $\mathbf{C}(1|0)$ . In addition,  $\mathfrak{C}_2$  is positive definite if and only if both  $\mathbf{C}(0)$  and  $\mathbf{C}(1|0)$  are positive definite.

Observe that  $\mathbf{C}(1|0) = \mathbf{A}^{-1}\mathbf{\Delta}\mathbf{A}^{-1T}$  is the covariance matrix of the innovation  $\mathbf{V}_t = \mathbf{A}^{-1}\mathbf{U}_t$ . Therefore, the left upper block of  $\mathbf{K}$  contains its inverse, which is  $\mathbf{A}^T\mathbf{\Delta}^{-1}\mathbf{A}$ .

**Theorem 1.** *The parameter matrices  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{\Delta}$  of model Equation (3) can be obtained by the block LDL decomposition of the (positive definite) concentration matrix  $\mathbf{K}$  (inverse of the covariance matrix  $\mathfrak{C}_2$  in Equation (4)) of the  $2d$ -dimensional Gaussian random vector  $(\mathbf{X}_t^T, \mathbf{X}_{t-1}^T)^T$ . If  $\mathbf{K} = \mathbf{LDL}^T$  is this (unique) decomposition with block-triangular matrix  $\mathbf{L}$  and block-diagonal matrix  $\mathbf{D}$ , then they have the form*

$$\mathbf{L} = \begin{pmatrix} \mathbf{A}^T & \mathbf{O}_{d \times d} \\ \mathbf{B}^T & \mathbf{I}_{d \times d} \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} \mathbf{\Delta}^{-1} & \mathbf{O}_{d \times d} \\ \mathbf{O}_{d \times d} & \mathbf{C}^{-1}(0) \end{pmatrix}, \tag{5}$$

where the  $d \times d$  upper triangular matrix  $\mathbf{A}$  with 1s along its main diagonal, the  $d \times d$  matrix  $\mathbf{B}$ , and the diagonal matrix  $\mathbf{\Delta}$  of model Equation (3) can be retrieved from them.

The proof of this theorem together with the detailed description of the algorithm is to be found in Appendices A.1 and A.2 of Appendix A.

The above model naturally generalizes to the following recursive CVAR( $p$ ) model: for given integer  $p \geq 1$ ,

$$\mathbf{A}\mathbf{X}_t + \mathbf{B}_1\mathbf{X}_{t-1} + \dots + \mathbf{B}_p\mathbf{X}_{t-p} = \mathbf{U}_t, \quad t = p + 1, p + 2, \dots, \tag{6}$$

where the white noise term  $\mathbf{U}_t$  is uncorrelated with  $\mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-p}$ , it has zero expectation and covariance matrix  $\mathbf{\Delta} = \text{diag}(\delta_1, \dots, \delta_d)$ .  $\mathbf{A}$  is a  $d \times d$  upper triangular matrix with 1s along its main diagonal; whereas,  $\mathbf{B}_1, \dots, \mathbf{B}_p$  are  $d \times d$  matrices.

Here, we have to perform the block Cholesky decomposition of the inverse covariance matrix of  $(\mathbf{X}_t^T, \mathbf{X}_{t-1}^T, \dots, \mathbf{X}_{t-p}^T)^T$ , i.e., the inverse of the matrix

$$\mathfrak{C}_{p+1} = \begin{pmatrix} \mathbf{C}(0) & \mathbf{C}^T(1) & \mathbf{C}^T(2) & \dots & \mathbf{C}^T(p) \\ \mathbf{C}(1) & \mathbf{C}(0) & \mathbf{C}^T(1) & \dots & \mathbf{C}^T(p-1) \\ \mathbf{C}(2) & \mathbf{C}(1) & \mathbf{C}(0) & \dots & \mathbf{C}^T(p-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}(p) & \mathbf{C}(p-1) & \mathbf{C}(p-2) & \dots & \mathbf{C}(0) \end{pmatrix}. \tag{7}$$

This is a symmetric, positive definite block Toeplitz matrix with  $(p + 1) \times (p + 1)$  blocks which are  $d \times d$  matrices. Again,  $\mathbf{C}^T(h) = \mathbf{C}(-h)$ , and it is well known that the inverse matrix  $\mathfrak{C}_{p+1}^{-1}$  has the following block-matrix form:

- Upper left block:  $\mathbf{C}^{-1}(p|0, \dots, p-1)$ ;
- Upper right block:  $-\mathbf{C}^{-1}(p|0, \dots, p-1)\mathbf{C}^T(1, \dots, p)\mathfrak{C}_p^{-1}$ ;
- Lower left block:  $-\mathfrak{C}_p^{-1}\mathbf{C}(1, \dots, p)\mathbf{C}^{-1}(p|0, \dots, p-1)$ ;
- Lower right block:  $\mathfrak{C}_p^{-1} + \mathfrak{C}_p^{-1}\mathbf{C}(1, \dots, p)\mathbf{C}^{-1}(p|0, \dots, p-1)\mathbf{C}^T(1, \dots, p)\mathfrak{C}_p^{-1}$ ,

where  $\mathbf{C}(p|0, \dots, p-1) = \mathbf{C}(0) - \mathbf{C}^T(1, \dots, p)\mathfrak{C}_p^{-1}\mathbf{C}(1, \dots, p)$  is the conditional covariance matrix  $\mathbf{C}(t|t-1, \dots, t-p)$  of the distribution of  $\mathbf{X}_t$  given  $\mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-p}$ ; due to stationarity,

it does not depend on  $t$  either; therefore, it is denoted by  $C(p|0, \dots, p - 1)$ . Furthermore,  $C^T(1, \dots, p) = (C^T(1), \dots, C^T(p))$  is a  $d \times pd$  and  $C(1, \dots, p)$  is a  $pd \times d$  matrix. In addition,  $\mathfrak{C}_{p+1}$  is positive definite if and only if both  $\mathfrak{C}_p$  and  $C(p|0, \dots, p - 1)$  are positive definite.

**Theorem 2.** The parameter matrices  $A, B_1, \dots, B_p$  and  $\Delta$  of model Equation (6) can be obtained by the block LDL decomposition of the (positive definite) concentration matrix  $K$  (inverse of the covariance matrix  $\mathfrak{C}_{p+1}$  in Equation (7)) of the  $(p + 1)d$ -dimensional Gaussian random vector  $(X_t^T, X_{t-1}^T, \dots, X_{t-p}^T)^T$ . If  $K = LDL^T$  is this (unique) decomposition with block-triangular matrix  $L$  and block-diagonal matrix  $D$ , then they have the form

$$L = \begin{pmatrix} A^T & O_{d \times pd} \\ B^T & I_{pd \times pd} \end{pmatrix}, \quad D = \begin{pmatrix} \Delta^{-1} & O_{d \times pd} \\ O_{pd \times d} & \mathfrak{C}_p^{-1} \end{pmatrix} \tag{8}$$

where the  $d \times d$  upper triangular matrix  $A$  with 1s along its main diagonal, the  $d \times pd$  matrix  $B = (B_1 \dots B_p)$  (transpose of  $B^T$ , partitioned into blocks) and the diagonal matrix  $\Delta$  of model Equation (6) can be retrieved from them.

The proof of this theorem together with the detailed description of the algorithm is to be found in Appendices A.3 and A.4 of Appendix A.

### 3.2. The Restricted Causal VAR(p) Model

First, we again consider the  $p = 1$  case. Assume that we have a causal ordering of the coordinates  $X_1, \dots, X_d$  of  $X$  such that  $X_j$  can be the cause of  $X_i$  whenever  $i < j$ . We can think of  $X_i$ s as the nodes of a graph in a directed graphical model (Bayesian network) and their labeling corresponds to a topological ordering of the nodes in the underlying DAG. Thus,  $i < j$  can imply a  $j \rightarrow i$  edge, and then we say that  $X_j$  is a parent (cause) of  $X_i$ . ( $X_i$  can have multiple parents, maximum  $d_i$  ones.) For example, when asset prices or relative returns of different assets or currencies (on the same day) influence each other in a certain (recursive) order. Now, restricted cases are analyzed, when only certain arrows (causes) are present, but the DAG is connected. In particular, only certain asset prices influence some others on a DAG contemporaneously, but not all possible directed edges are present. In this case, a covariance selection technique can be initiated to re-estimate the covariance matrix so that the partial regression coefficients in the no-edge positions are zeros.

Our DAG is sometimes given by an expert’s knowledge, but usually it is built from an undirected graph, when we also require that the so-constructed DAG be Markov equivalent to its undirected skeleton. Then, the DAG must not contain a sink V configuration. Under sink V configuration, a triplet  $j \rightarrow h \leftarrow i$  is understood, where  $i$  is not connected by  $j$  ( $h < i < j$ ); see Figure 1.

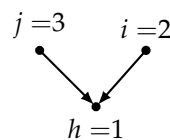


Figure 1. Triplet sink V.

Here, we include a short description on the relation between directed and undirected graphical models, with emphases on the Gaussian case, based on Lauritzen (2004) and Bolla et al. (2019). Directed and undirected models have many properties in common, and under some conditions, there are important correspondences between them.

Let  $X \sim \mathcal{N}_d(\mu, \Sigma)$  be a  $d$ -variate non-degenerate Gaussian random vector with expectation  $\mu$  and positive definite, symmetric  $d \times d$  covariance matrix  $\Sigma$ . The also positive definite, symmetric matrix  $\Sigma^{-1}$  of entries  $\sigma^{ij}$  is called a concentration matrix, it appears in the joint density, and its zero entries indicate conditional independences between two

components of  $\mathbf{X}$ , given the remaining ones. Mostly, the variables are already centered, so  $\boldsymbol{\mu} = \mathbf{0}$  is assumed.

Let us form an undirected graph  $G$  on the node-set  $V = \{1, \dots, d\}$ , where  $V$  corresponds to the components of  $\mathbf{X}$ , and the edges are drawn according to the rule

$$i \sim j \Leftrightarrow \sigma^{ij} \neq 0, \quad i \neq j. \tag{9}$$

This is called an undirected Gaussian graphical model, which is a special Markov Random Field (MRF). To establish conditional independence statements, we use the following facts.

**Proposition 1.** Let  $\mathbf{X} = (X_1, \dots, X_d)^T \sim \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma})$  be a random vector, and let  $V := \{1, \dots, d\}$  denote the index set of the variables,  $d \geq 3$ . Assume that  $\boldsymbol{\Sigma}$  is positive definite. Then,

$$r_{X_i X_j | \mathbf{X}_{V \setminus \{i,j\}}} = \frac{-\sigma^{ij}}{\sqrt{\sigma^{ii} \sigma^{jj}}} \quad i \neq j,$$

where  $r_{X_i X_j | \mathbf{X}_{V \setminus \{i,j\}}}$  denotes the partial correlation coefficient between  $X_i$  and  $X_j$  after eliminating the effect of the remaining variables  $\mathbf{X}_{V \setminus \{i,j\}}$ . Furthermore,

$$\sigma^{ii} = 1 / (\text{Var}(X_i | \mathbf{X}_{V \setminus \{i\}})), \quad i = 1, \dots, d$$

is the reciprocal of the conditional (residual) variance of  $X_i$ , given the other variables  $\mathbf{X}_{V \setminus \{i\}}$ .

**Definition 1.** Let  $\mathbf{X} \sim \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma})$  be a random vector with  $\boldsymbol{\Sigma}$  positive definite. Consider the regression plane

$$\mathbb{E}(X_i | \mathbf{X}_{V \setminus \{i\}} = \mathbf{x}_{V \setminus \{i\}}) = \sum_{j \in V \setminus \{i\}} \beta_{j \cdot V \setminus \{i\}} x_j, \quad j \in V \setminus \{i\},$$

where  $x_j$ 's are the coordinates of  $\mathbf{x}_{V \setminus \{i\}}$ . Then, we call the coefficient  $\beta_{j \cdot V \setminus \{i\}}$  the partial regression coefficient of  $X_j$  when regressing  $X_i$  with  $\mathbf{X}_{V \setminus \{i\}}$ ,  $j \in V \setminus \{i\}$ .

**Proposition 2.**

$$\beta_{j \cdot V \setminus \{i\}} = -\frac{\sigma^{ij}}{\sigma^{ii}}, \quad j \in V \setminus \{i\}.$$

**Corollary 1.** An important consequence of Propositions 1 and 2 is that

$$\beta_{j \cdot V \setminus \{i\}} = r_{X_i X_j | \mathbf{X}_{V \setminus \{i,j\}}} \sqrt{\frac{\sigma^{jj}}{\sigma^{ii}}} = r_{X_i X_j | \mathbf{X}_{V \setminus \{i,j\}}} \sqrt{\frac{\text{Var}(X_i | \mathbf{X}_{V \setminus \{i\}})}{\text{Var}(X_j | \mathbf{X}_{V \setminus \{j\}})}}, \quad j \in V \setminus \{i\}.$$

(The formula is analogous to the one of unconditioned regression.) Thus, only the variables  $X_j$ 's, whose partial correlation with  $X_i$  (after eliminating the effect of the remaining variables) is not 0, enter into the regression of  $X_i$  with the other variables.

To form the edges, instead of Equation (9), for  $i \neq j$ , we have to test the following statistical hypothesis, and draw an edge if we can reject  $H_0$  with a "small enough" significance:

$$H_0 : r_{X_i X_j | \mathbf{X}_{V \setminus \{i,j\}}} = 0,$$

i.e.,  $X_i$  and  $X_j$  are conditionally independent conditioned on the remaining variables. Equivalently,  $H_0$  means that  $\beta_{ij | V \setminus \{i\}} = 0$ ,  $\beta_{ji | V \setminus \{j\}} = 0$ , or simply,  $\sigma^{ij} = \sigma^{ji} = 0$  ( $\boldsymbol{\Sigma} > 0$  is assumed).

To test  $H_0$  in some format, several exact tests are known that are usually based on likelihood ratio tests. The following test uses the empirical partial correlation coefficient, denoted by  $\hat{r}_{X_i X_j | \mathbf{X}_{V \setminus \{i,j\}}}$ , and the following statistic is based on it:

$$B = 1 - (\hat{r}_{X_i X_j | \mathbf{X}_{V \setminus \{i,j\}}})^2 = \frac{|\mathbf{S}_{V \setminus \{i,j\}}| \cdot |\mathbf{S}_V|}{|\mathbf{S}_{V \setminus \{i\}}| \cdot |\mathbf{S}_{V \setminus \{j\}}|},$$

where  $S$  is the sample size ( $n$ ) times the empirical covariance matrix of the variables in the subscript (its entries are the product-moments).

It can be proven that, under  $H_0$ , the test statistic

$$t = \sqrt{n-d} \cdot \sqrt{\frac{1}{B} - 1} = \sqrt{n-d} \cdot \frac{\hat{r}_{X_i X_j | \mathbf{X}_{V \setminus \{i,j\}}}}{\sqrt{1 - (\hat{r}_{X_i X_j | \mathbf{X}_{V \setminus \{i,j\}}})^2}}$$

is distributed as Student's  $t$  with  $n - d$  degrees of freedom. Therefore, we reject  $H_0$  for large values of  $|t|$ , or equivalently, for large values of  $\hat{r}_{X_i X_j | \mathbf{X}_{V \setminus \{i,j\}}}$ .

In the directed model (Bayesian network), the nodes of the graph  $G$  correspond to random variables  $X_1, \dots, X_d$ , whereas the directed edges to *causal* dependences between them. In the case of a DAG  $G$  with node-set  $V = \{1, \dots, d\}$ , there are no directed cycles, and therefore, there exists a recursive ordering (labeling) of the nodes such that, for every directed edge  $j \rightarrow i$ , the relation  $i < j$  holds.

Let  $\text{par}(i) \subset \{i + 1, \dots, d\}$  denote the set of the parents of  $i$  and, for any  $A \subset V$ , we use the notation  $\mathbf{x}_A = \{x_i : i \in A\}$  and  $\mathbf{X}_A = \{X_i : i \in A\}$ . To draw the edges, the *directed pairwise Markov property* is used: for  $i < j$ , there is no  $j \rightarrow i$  directed edge, whenever  $X_i$  and  $X_j$  are conditionally independent, given  $\mathbf{X}_{\text{par}(i)}$ . With notation,

$$X_i \perp\!\!\!\perp X_j | \mathbf{X}_{\text{par}(i)} \quad \text{for } j \in \{i + 1, \dots, d\} \setminus \text{par}(i), \quad i = 1, \dots, d - 1.$$

In the case of a non-degenerate Gaussian distribution, by the Hammersley–Clifford theorem, the following *undirected factorization property* is also equivalent to the undirected pairwise Markov property (9) that defines the graph. It means the factorization of the joint density of the components of  $\mathbf{X}$ , for any state configuration  $\mathbf{x} = (x_1, \dots, x_d)$  as follows:

$$f(\mathbf{x}) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \Psi_C(\mathbf{x}_C),$$

where  $Z > 0$  is a normalizing constant, and the non-negative *compatibility functions*  $\Psi_C$ s are assigned to the cliques of  $G$ . Under *clique*, we understand a maximal complete subgraph. (Note that, in graph theory, it is sometimes called a maximal clique.) The above factorization is far from unique, but in special (so-called decomposable) models, the forthcoming Equation (10) gives an explicit formula for the compatibility functions.

In addition, even if the underlying graph is undirected, a decomposable structure of it gives a (not necessarily unique) so-called *perfect ordering* of the nodes, in which order directed edges can be drawn. Conversely, a decomposable directed graph (with no sink  $V$  configurations) can be made undirected by disregarding the orientation of the edges.

Decomposable graphs have a special interest with regard to exact MLE. There are several equivalent properties of a decomposable graph, based on [Wermuth \(1980\)](#); [Lauritzen \(2004\)](#); [Bolla et al. \(2019\)](#):

- $G$  is *triangulated* (with other words, *chordal*), i.e., every cycle in  $G$  of a length of at least four has a chord.
- $G$  has a *perfect numbering* of its nodes such that, in this labeling,  $\text{ne}(i) \cap \{i + 1, \dots, d\}$  is a complete subgraph, where  $\text{ne}(i)$  is the set of neighbors of  $i$ , for  $i = 1, \dots, d$ . It is also called *single node elimination ordering* (see [Wainwright \(2015\)](#)), and obtainable



with the *maximal cardinality search* (MCS) algorithm of [Tarjan and Yannakakis \(1984\)](#); see also [Koller and Friedman \(2009\)](#).

- $G$  has the following *running intersection property*: we can number the cliques of it to form a so-called *perfect sequence*  $C_1, \dots, C_k$  where each combination of the subgraphs induced by  $H_{j-1} = C_1 \cup \dots \cup C_{j-1}$  and  $C_j$  is a decomposition ( $j = 2, \dots, k$ ), i.e., the necessarily complete subgraph  $S_j = H_{j-1} \cap C_j$  is a separator. More precisely,  $S_j$  is a node cutset between the disjoint node subsets  $H_{j-1} \setminus S_j$  and  $R_j = C_j \setminus S_j = H_j \setminus H_{j-1}$ . This sequence of cliques is also called a *junction tree* (JT). Here, any clique  $C_j$  is the disjoint union of  $R_j$  (called *residual*), the nodes of which are not contained in any  $C_i, i < j$  and of  $S_j$  (called *separator*) with the following property: there is an  $i^* \in \{1, \dots, j - 1\}$  such that

$$S_j = C_j \cap (\cup_{i=1}^{j-1} C_i) = C_j \cap C_{i^*}.$$

This (not necessarily unique)  $C_{i^*}$  is called *parent clique* of  $C_j$ . Here,  $S_1 = \emptyset$  and  $R_1 = C_1$ . Furthermore, if such an ordering is possible, a version may be found in which any prescribed set is the first one. Note that the junction tree is indeed a tree with nodes  $C_1, \dots, C_k$  and one less edge that are the separators  $S_2, \dots, S_k$ .

- There is a labeling of the nodes such that the adjacency matrix contains a *reducible zero pattern* (RZP). It means that there is an index set  $I \subset \{(i, j) : 1 \leq i < j \leq d\}$  which is reducible in the sense that, for each  $(i, j) \in I$  and  $h = 1, \dots, i - 1$ , we have  $(h, i) \in I$  or  $(h, j) \in I$  or both.

Indeed, this convenient labeling is a perfect numbering of the nodes.

- The following *Markov chain* property also holds:  $f(\mathbf{x}_{R_j} | \mathbf{x}_{C_1 \cup \dots \cup C_{j-1}}) = f(\mathbf{x}_{R_j} | \mathbf{x}_{S_j})$ . Therefore, if we have a perfect sequence  $C_1, \dots, C_k$  of the cliques with separators  $S_1 = \emptyset, S_2, \dots, S_k$ , then, for any state configuration  $\mathbf{x}$ , we have the following factorized form of the density:

$$f(\mathbf{x}) = \frac{\prod_{j=1}^k f(\mathbf{x}_{C_j})}{\prod_{j=2}^k f(\mathbf{x}_{S_j})} = \prod_{i=1}^k f(\mathbf{x}_{R_i} | \mathbf{x}_{S_i}). \tag{10}$$

To find the structure, where one of the equivalent criteria of decomposability holds, we can use the MCS method of [Tarjan and Yannakakis \(1984\)](#) and [Koller and Friedman \(2009\)](#). The simple MCS gives label  $d$  to an arbitrary node. Then, the nodes are labeled consecutively, from  $d$  down to 1, choosing as the next to label a node with a maximum number of previously labeled neighbors and breaking ties arbitrarily. (Note that [Lauritzen \(2004\)](#) labels the nodes conversely.) The MCS ordering is far from unique, and this simple version is not always capable of finding the JT structure behind a triangulated graph in one run, but another run is needed. There are also variants of this algorithm which are applicable to a non-triangulated graph too, and capable of triangulating it with adding a minimum number of edges.

In the unrestricted model, no restrictions for the upper-diagonal entries of  $A$  were made. In practice, we have a sample and all the autocovariance matrices are estimated, and the resulting  $A, B$  matrices are calculated with them. Usually, a statistical hypothesis testing advances this procedure, during which it can be found that certain partial correlations (closely related to the entries of  $K$ ) do not significantly differ from zero. Then, we naturally want to introduce zeros for the corresponding entries of  $A$ . For this, the method of covariance selection of [Dempster \(1972\)](#) was elaborated; see also [Lauritzen \(2004\)](#) and [Wermuth \(1980\)](#). First, we give a more general definition of the notion of an RZP.

**Definition 2.** Let  $M$  be a symmetric or an upper triangular matrix of real entries. We say that  $M$  has a *reducible zero pattern* (RZP) with respect to the index set  $I \subset \{(i, j) : 1 \leq i < j \leq d\}$  if, for each  $(i, j) \in I$  and  $h = 1, \dots, i - 1$ , we have  $(h, i) \in I$  or  $(h, j) \in I$  or both.

In view of this, we can find relation between the zeros of  $A$  in the CVAR(1) model and those of the inverse covariance matrix.

**Proposition 3.** *If the upper triangular matrix  $A$  of model Equation (3) has an RZP with respect to the index set  $I$ , then the upper left  $d \times d$  block of  $\mathbf{K} = \mathfrak{C}_2^{-1}$  has an RZP with respect to  $I$ . Conversely, if  $\mathbf{K}$  has an RZP with respect to the index set  $I$ , then it is inherited to  $A$ .*

**Proof.** In the forward direction, the proof follows from Equation (A4) of Appendix A.2 in Appendix A. Indeed, in the presence of an index set  $I$ , giving an RZP in  $A$ , for  $1 \leq j < i \leq d$ : if  $l_{ij} = a_{ji} = 0$ , then  $k_{ij} = 0$ , since either  $\ell_{ih} = a_{hi} = 0$  or  $\ell_{jh} = a_{hj} = 0$  (or both) for  $h = 1, \dots, j - 1$ , which are the intrinsic entries of the summation in Equation (A4).

In the backward direction, if  $k_{ij} = 0$ , then  $l_{ij} = a_{ji} = 0$  too because of the Markov equivalence of the DAG and its undirected skeleton in the decomposable case. The presence of the RZP guarantees decomposability (see the equivalences to decomposability). Furthermore, by the nested structure of the block LDL decomposition, the entry  $l_{ij}$  is a partial regression coefficients that is zero at the same time as the corresponding partial correlation coefficient and the entry  $k_{ij}$  of  $\mathbf{K}$  (see Proposition 1 and Corollary 1).  $\square$

Note that, in both directions, the other matrix ( $\mathbf{K}$  or  $A$ ) may have additional zeros. Consequently, if we have causal relations between the contemporaneous components of  $\mathbf{X}_t$ , and the so-constructed DAG has an RZP, then this RZP is inherited by the left upper block of  $\mathbf{K}$ , which is  $\mathbf{C}^{-1}(1|0)$ . Therefore, we further improve the covariance selection model Dempster (1972), by introducing zero entries into the sample conditional covariance matrix. Actually, fixing the zero entries in the left upper block of  $\mathbf{K}$ , we re-estimate the matrix  $\mathfrak{C}_2$ .

In the possession of a sample, there are exact MLEs developed for this purpose, for an i.i.d. sample (see Bolla et al. (2019), Lauritzen (2004)). Note that here we do not have an i.i.d. sample, but a serially correlated sample. However, by ergodicity, for “large”  $n$ , this method also works and gives an asymptotic MLE, akin to the product-moment estimates.

For estimation purposes, we use the empirical partial correlation coefficients, and based on them, the above exact test to check whether they significantly differ from 0 or not. In Theorem 5.3 of Lauritzen (2004), it is proved that, based on an i.i.d. sample, under the covariance selection model, the MLE of the mean vector is the sample mean  $\bar{\mathbf{X}}$ , and the restricted covariance matrix  $\Sigma^* = (\sigma_{ij}^*)$  can be estimated as follows. The entries in the edge-positions are estimated as in the saturated model (no restrictions):

$$\hat{\sigma}_{ij}^* = \frac{1}{n} s_{ij}, \quad \{i, j\} \in E, \quad (11)$$

where  $S = (s_{ij}) = \sum_{\ell=1}^n (\mathbf{X}_\ell - \bar{\mathbf{X}})(\mathbf{X}_\ell - \bar{\mathbf{X}})^T$  is the usual product-moment estimate. The other entries (in the no-edge positions) of  $\Sigma^*$  are free, but satisfy the model conditions: after taking the inverse  $\mathbf{K}$  of  $\Sigma^*$  with these undetermined entries, we obtain the same number of equations for them from  $k^{ij} = 0$  whenever  $\{i, j\} \notin E$ . To do so, there are numerical algorithms at our disposal, for instance, the iterative proportional scaling (IPS), see Lauritzen (2004), p. 134, where an infinite iteration is needed because, in general, there is no explicit solution for the MLE. However, the fixed point of this iteration gives a unique positive definite matrix  $\hat{\mathbf{K}}$ .

In the decomposable case, there is no need of running the IPS, but an explicit estimate can be given as follows. Recall that, if the Gaussian graphical model is decomposable (its concentration graph  $G$  is decomposable), then the cliques, together with their separators (with possible multiplicities), form a JT structure. Denote  $\mathcal{C}$  as the set of the cliques and  $\mathcal{S}$  the set of the separators in  $G$ . Then, direct density estimates, using (10), are available. In particular, the MLE of  $\mathbf{K}$  can be calculated based on the product-moment estimates applied for subsets of the variables, corresponding to the cliques and separators.

Let  $n$  be the size of the sample for the underlying  $d$ -variate normal distribution, and assume that  $n > d$ . For the clique  $C \in \mathcal{C}$ , let  $[S_C]^V$  denote  $n$  times the empirical covariance matrix corresponding to the variables  $\{X_i : i \in C\}$  complemented with zero entries to have a  $d \times d$  (symmetric, positive semidefinite) matrix. Likewise, for the separator  $S \in \mathcal{S}$ , let  $[S_S]^V$  denote  $n$  times the empirical covariance matrix corresponding to the variables  $\{X_i : i \in S\}$  complemented with zero entries to have an  $d \times d$  (symmetric, positive semidefinite) matrix. Then, the MLE of the mean vector is the sample average (as usual), while that of the concentration matrix is

$$\hat{K} = n \left\{ \sum_{C \in \mathcal{C}} [S_C^{-1}]^V - \sum_{S \in \mathcal{S}} [S_S^{-1}]^V \right\}, \tag{12}$$

see Proposition 5.9 of Lauritzen (2004). This proposition states that the above MLE exists with probability one if and only if  $n$  is greater than the maximum clique size.

However, here we use a serially correlated sample as follows. Assume that the cliques of the node set  $\{1, \dots, d\}$  of  $\mathbf{X}_t$  are  $C_1, \dots, C_k$ , to which a last clique is added, formed by the components  $X_{t-1,1}, \dots, X_{t-1,d}$  of  $\mathbf{X}_{t-1}$ . If  $C_1, \dots, C_k$  form a JT in this ordering, the joint density of  $\mathbf{X}_t$  and  $\mathbf{X}_{t-1}$  factorizes like

$$f(\mathbf{x}_t, \mathbf{x}_{t-1}) = f(\mathbf{x}_{t-1}) \prod_{j=1}^k f(\mathbf{x}_{t,R_j} | \mathbf{x}_{t,S_j}, \mathbf{x}_{t-1}).$$

For covariance selection, we include the lag 1 variables  $X_{t-1,1}, \dots, X_{t-1,d}$  too. Therefore, the new cliques and separators are

$$C'_j := C_j \cup \{X_{t-1,1}, \dots, X_{t-1,d}\}, \quad j = 1, \dots, k$$

and

$$S'_j := S_j \cup \{X_{t-1,1}, \dots, X_{t-1,d}\}, \quad j = 2, \dots, k.$$

Having this, we are able to re-estimate the  $2d \times 2d$   $K$ , inverse of  $\mathcal{C}_2$  in (4), for our VAR(1) model as follows:

$$\hat{K} = (n - 1) \left\{ \sum_{j=1}^k [S_{C'_j}^{-1}]^{2d} - \sum_{j=2}^k [S_{S'_j}^{-1}]^{2d} \right\}, \tag{13}$$

where the matrix  $S_{C'}$  is the product-moment estimate based on the  $n - 1$  element serially correlated sample with the following variables:

$$(X_{t,i} : i \in C \quad \text{and} \quad X_{t-1,1}, \dots, X_{t-1,d}), \quad t = 2, \dots, n;$$

furthermore,  $[M_{C'}]^{2d}$  denotes the  $2d \times 2d$  matrix comprising the entries of the larger  $2d \times 2d$  matrix  $M$  in the  $|C'| \times |C'|$  block corresponding to  $C'$ , and otherwise zeros. By the properties of the LDL decomposition, these zeros go into zeros of  $A$ .

In the financial example of Section 4, the cliques and separators of the forthcoming Equation (14) are used. There the estimate of  $K$  is as follows:



Restricted cases in the  $p > 1$  scenario can be treated similarly. Here, too, the existence of an RZP in the DAG on  $d$  nodes is equivalent to the existence of an RZP in the left upper  $d \times d$  corner of the concentration matrix  $\mathbf{C}_{p+1}^{-1}$ . From the model equations, it is obvious that

$$X_{t,i} = - \sum_{j=i+1}^d a_{ij} X_{t,j} - \sum_{h=1}^p \sum_{j=1}^d b_{h,ij} X_{t-h,j} - U_{t,i},$$

where  $X_{t,i}$  is the  $i$ th coordinate of  $\mathbf{X}_t$ . By weak stationarity, it follows that the entries of the matrices  $\mathbf{A}$  and  $\mathbf{B}_h = (b_{h,ij})_{i,j=1}^d$  are partial regression coefficients as follows:

$$a_{ij} = -\beta_{X_{t,i} X_{t,j} \cdot \{X_{t,i+1}, \dots, X_{t,d}, X_{t-1,1}, \dots, X_{t-1,d}, \dots, X_{t-p,1}, \dots, X_{t-p,d}\}}, \quad 1 \leq i < j \leq d;$$

$$b_{h,ij} = -\beta_{X_{t,i} X_{t+h,j} \cdot \{X_{t,i+1}, \dots, X_{t,d}, X_{t-1,1}, \dots, X_{t-1,d}, \dots, X_{t-p,1}, \dots, X_{t-p,d}\}},$$

$$1 \leq i < j \leq d, \quad h = 1, \dots, p.$$

Since the conditioning set changes from equation to equation, it is easier to use the block LDL decompositions here, without the exact meaning of the coefficients.

Considering the components of  $\mathbf{X}_t, \mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-p}$  as nodes of the expanded graph, the joint density of  $\mathbf{X}_t, \mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-p}$  factorizes like

$$f(\mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-p}) = f(\mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-p}) f(\mathbf{x}_t | \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-p})$$

$$= f(\mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-p}) \cdot \prod_{i=1}^d f(x_{t,i} | \mathbf{x}_{t, \text{par}(i)}, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-p}).$$

Now, assume that the cliques of the node set  $\{1, \dots, d\}$  of  $\mathbf{X}_t$  are  $C_1, \dots, C_k$ , and they form a JT with residuals  $R_1, \dots, R_k$  and separators  $S_1, \dots, S_k$  (with the understanding that  $S_1 = \emptyset$  and  $R_1 = C_1$ ). Enhancing the preceding density with this, we obtain the following factorization:

$$f(\mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-p}) = f(\mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-p}) \cdot \prod_{j=1}^k f(\mathbf{x}_{t,R_j} | \mathbf{x}_{t,S_j}, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-p}).$$

Covariance selection can be carried out similarly as in the  $p = 1$  case, but here zero entries of the left upper  $d \times d$  block of  $\mathbf{C}_{p+1}^{-1}$  provide the zero entries of  $\mathbf{A}$ . For this purpose, the  $n - p$  element sample entries are used with the following coordinates:

$$\left( X_{t,i} : i \in C \quad \text{and} \quad X_{t-1,1}, \dots, X_{t-1,d}, \dots, X_{t-p,1}, \dots, X_{t-p,d} \right),$$

for  $t = p + 1, \dots, n$  when we calculate the product-moment estimate  $S_{C'}$  with  $C' = C \cup \{\mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-p}\}$ . For more details, see the explanation after Equation (13) and Section 4.

Again, here the covariance selection is carried out based only on a serially correlated and not an independent sample. However, when  $n$  is “large”, then ergodicity issues (see, e.g., Bolla and Szabados (2021)) give rise to this relaxation of the original algorithm. In addition, by the theory of Brockwell and Davis (1991) (p. 424), it is guaranteed that the Yule–Walker equations have a stable stationary solution to the VAR( $p$ ) model, whenever the starting covariance matrix of  $(p + 1) \times (p + 1)$  blocks is positive definite. However, we assume this in our theorems. In this case, the empirical versions are also positive definite (almost surely as  $n \rightarrow \infty$ ), and the covariance selection also gives a positive definite estimate. Thus, the estimated parameter matrices provide a stable VAR model in view of the standard theory and ergodicity if  $n$  is “large”.

#### 4. Applications with Order Selection

##### 4.1. Financial Data

We used the data communicated in the paper Akbilgic et al. (2014) on daily relative returns of eight different asset prices, spanning 534 days. The multivariate time series was found to be stationary and nearly Gaussian.

First, we applied the unrestricted  $CVAR(p)$  model. We constructed a DAG by making the undirected graph on eight nodes directed. The undirected graph was constructed by testing statistical hypotheses for the partial correlations of the pairs of the variables conditioned on all the others. As the test statistic is increasing in the absolute value of the partial correlation in question, a threshold 0.04 for the latter one was used that corresponds to significance level  $\alpha = 0.008851$  of the partial correlation test. Table 1 contains the partial correlations based on  $C^{-1}(0)$ .

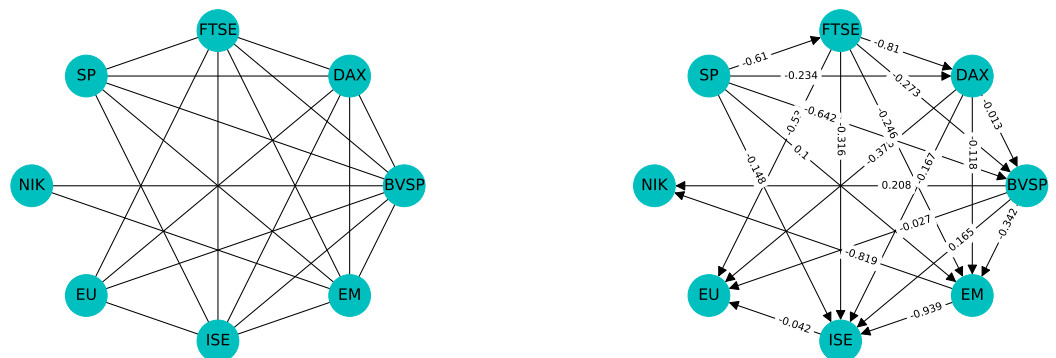
**Table 1.** Partial correlation coefficients from  $C^{-1}(0)$ . Entries marked by asterisk are less than 0.04 in absolute value (i.e., they correspond to no-edge positions in the graph), and the corresponding significance is  $\alpha = 0.008851$ .

	NIK	EU	ISE	EM	BVSP	DAX	FTSE	SP
NIK		0.016 *	0.035 *	0.522	-0.260	-0.019 *	-0.076	0.024 *
EU	0.016 *		0.217	0.034 *	0.067	0.687	0.747	0.018 *
ISE	0.035 *	0.217		0.358	-0.157	-0.077	-0.059	0.034 *
EM	0.522	0.034 *	0.358		0.546	0.048	0.086	-0.184
BVSP	-0.260	0.067	-0.157	0.546		-0.093	-0.045	0.533
DAX	-0.019 *	0.687	-0.077	0.048	-0.093		-0.203	0.191
FTSE	-0.076	0.747	-0.059	0.086	-0.045	-0.203		0.057
SP	0.024 *	0.018 *	0.034 *	-0.184	0.533	0.191	0.057	

Since the graph was triangulated, with the MCS algorithm, we were able to (not necessarily uniquely) label the nodes so that the adjacency matrix of this undirected graph had an RZP:

- 1 : NIK (stock market return index of Japan),
- 2 : EU (MSCI European index),
- 3 : ISE (Istanbul stock exchange national 100 index)
- 4 : EM (MSCI emerging markets index),
- 5 : BVSP (stock market return index of Brazil),
- 6 : DAX (stock market return index of Germany),
- 7 : FTSE (stock market return index of UK),
- 8 : SP (Standard & Poor’s 500 return index).

If this is considered as the topological labeling of the DAG, where directed edges point from a higher label node to a lower label one, then the so-obtained directed graph is Markov equivalent to its undirected skeleton; see Figure 2a,b. However, the RZP is used only in the restricted case; in the unrestricted case, only the DAG ordering of the variables is used.



(a) MRF fitted to the financial dataset.

(b) DAG oriented in the MCS ordering.

**Figure 2.** Graphical models fitted to the financial dataset.

We ran the VAR( $p$ ) algorithm with  $p = 1, 2, 3, 4, 5$  and found that the  $A$  matrices do not change much with increasing  $p$ , akin to  $B_1$ . The  $B_2, \dots, B_5$  matrices have relatively “small” entries. Consequently, contemporaneous effects and one-day lags are the most important. This is also supported by the forthcoming order selection investigations. For the  $p = 1$  and  $p = 2$  cases, see Tables 2–6, respectively. The  $p = 3, 4, 5$  cases are represented by tables in the Supplementary Material.

Then, we considered the restricted CVAR(1) model. Here, we want to introduce structural zeros into the matrix  $A$ . Now, the matrix  $C^{-1}(1|0)$ , the left upper  $8 \times 8$  corner of  $\mathcal{C}_2^{-1}$  is used for covariance selection. Figure 2b shows this DAG with the significant path coefficients above the arrows, based on Table 7.

**Table 2.**  $A$  matrix for the unrestricted Financial VAR(1) model (rounded to 4 decimals).

	NIK	EU	ISE	EM	BVSP	DAX	FTSE	SP
NIK	1	0.0264	0.0042	−0.8902	0.2030	0.0170	0.0781	−0.0336
EU	0	1	−0.0418	−0.0146	−0.0239	−0.3746	−0.5255	−0.0033
ISE	0	0	1	−0.9518	0.1613	−0.1658	−0.3129	−0.1413
EM	0	0	0	1	−0.3507	−0.1182	−0.2464	0.1077
BVSP	0	0	0	0	1	−0.0129	−0.2782	−0.6375
DAX	0	0	0	0	0	1	−0.8102	−0.2336
FTSE	0	0	0	0	0	0	1	−0.6100
SP	0	0	0	0	0	0	0	1

**Table 3.**  $B$  matrix for the unrestricted Financial VAR(1) model (rounded to 4 decimals).

	NIK <sub>−1</sub>	EU <sub>−1</sub>	ISE <sub>−1</sub>	EM <sub>−1</sub>	BVSP <sub>−1</sub>	DAX <sub>−1</sub>	FTSE <sub>−1</sub>	SP <sub>−1</sub>
NIK	0.1845	−0.1685	−0.0874	0.0852	0.0635	0.0205	−0.1236	−0.2798
EU	−0.0131	0.1219	−0.0044	0.0291	−0.0124	−0.0393	−0.0979	0.0011
ISE	0.0677	0.2811	−0.0657	0.2473	−0.2940	−0.0543	0.0098	−0.1442
EM	−0.0016	−0.0569	−0.0159	0.1076	−0.0917	−0.0945	0.0875	−0.1071
BVSP	−0.0140	0.0704	0.0142	−0.1046	0.1397	−0.1497	0.1188	−0.0812
DAX	−0.0034	0.2021	−0.0342	−0.0044	−0.0352	−0.0476	−0.0670	−0.0673
FTSE	0.0293	−0.0168	−0.0109	0.0420	−0.1129	0.2141	0.0805	−0.2641
SP	0.0417	0.2603	−0.0261	0.0112	−0.0026	−0.0709	−0.2850	0.1240

**Table 4.**  $A$  matrix for the unrestricted Financial VAR(2) model (rounded to 4 decimals).

	NIK	EU	ISE	EM	BVSP	DAX	FTSE	SP
NIK	1	−0.0114	0.0103	−0.8822	0.1995	0.0233	0.0856	−0.0214
EU	0	1	−0.0426	−0.0110	−0.0240	−0.3745	−0.5137	−0.0128
ISE	0	0	1	−0.9788	0.1701	−0.1669	−0.3139	−0.1361
EM	0	0	0	1	−0.3450	−0.1154	−0.2375	0.0922
BVSP	0	0	0	0	1	−0.0047	−0.2655	−0.6601
DAX	0	0	0	0	0	1	−0.8120	−0.2339
FTSE	0	0	0	0	0	0	1	−0.6320
SP	0	0	0	0	0	0	0	1

**Table 5.**  $B_1$  matrix for the unrestricted Financial VAR(2) model (rounded to 4 decimals).

	NIK <sub>−1</sub>	EU <sub>−1</sub>	ISE <sub>−1</sub>	EM <sub>−1</sub>	BVSP <sub>−1</sub>	DAX <sub>−1</sub>	FTSE <sub>−1</sub>	SP <sub>−1</sub>
NIK	0.2063	−0.1826	−0.1106	0.1063	0.0731	0.0187	−0.1502	−0.2580
EU	−0.0037	0.1364	−0.0010	0.0232	−0.0150	−0.0371	−0.0996	−0.0107
ISE	0.0409	0.2476	−0.0771	0.2274	−0.2772	−0.0447	0.0331	−0.1284
EM	0.0489	−0.0200	−0.0030	0.1360	−0.1150	−0.0996	0.0468	−0.1162
BVSP	−0.0066	0.0931	0.0261	−0.1091	0.1312	−0.1573	0.1161	−0.0935
DAX	−0.0123	0.2146	−0.0319	0.0073	−0.0406	−0.0536	−0.0727	−0.0694
FTSE	0.0852	0.0019	0.0275	0.0145	−0.1117	0.2377	0.1035	−0.3427
SP	0.0530	0.2759	−0.0565	−0.0033	0.0024	−0.0945	−0.3106	0.1789

The ordering of the variables is the same as in the unrestricted case, but the RZP is a bit different. The decomposable structure has the following cliques and separators:

$$\begin{aligned}
 C_1 &= \{BVSP, DAX, EM, FTSE, ISE, SP\} = \{3, 4, 5, 6, 7, 8\} \\
 C_2 &= \{BVSP, DAX, EU, FTSE, ISE\} = \{2, 3, 5, 6, 7\} \\
 C_3 &= \{BVSP, EM, NIK\} = \{1, 4, 5\} \\
 S_2 &= \{BVSP, DAX, FTSE, ISE\} = \{3, 5, 6, 7\} \\
 S_3 &= \{BVSP, EM\} = \{4, 5\},
 \end{aligned} \tag{14}$$

where the parent clique of both  $C_2$  and  $C_3$  is  $C_1$ . Note that the set of nodes in the second braces is the same, but they follow increasing labels so that they better see the JT structure. The  $16 \times 16$  matrix  $\hat{K}$  is estimated by covariance selection, using the lag 1 variables too; see it in a table in the Supplementary Material.

**Table 6.**  $B_2$  matrix for the unrestricted Financial VAR(2) model (rounded to 4 decimals).

	NIK <sub>-2</sub>	EU <sub>-2</sub>	ISE <sub>-2</sub>	EM <sub>-2</sub>	BVSP <sub>-2</sub>	DAX <sub>-2</sub>	FTSE <sub>-2</sub>	SP <sub>-2</sub>
NIK	-0.0402	-0.1695	-0.0410	0.0156	0.0998	-0.0406	0.1367	-0.0091
EU	0.0017	0.0771	-0.0065	0.0054	0.0037	0.0192	-0.0762	-0.0394
ISE	-0.0142	-0.1725	-0.0276	-0.0088	0.0389	0.1167	0.0826	0.0357
EM	-0.0054	0.0650	-0.0322	0.1155	-0.0695	-0.0959	-0.0162	-0.0270
BVSP	-0.0423	0.0332	-0.0449	0.2878	-0.0717	-0.0221	-0.0381	-0.0120
DAX	-0.0372	0.0177	0.0130	0.0658	-0.0360	-0.0108	-0.0202	0.0059
FTSE	0.0491	0.3107	-0.0820	0.0693	0.0299	0.0153	-0.0840	-0.3038
SP	0.0447	-0.0628	0.0804	-0.1824	0.0785	0.0133	-0.1775	0.1284

The matrices  $A$  and  $B$  were estimated via the algorithm for the LDL decomposition of  $\hat{K}$ . Here, the zeros of the left upper  $8 \times 8$  block of  $\hat{K}$  will necessarily result in the zeros of  $A$  in the same positions. The upper-diagonal entries of  $A$  and the entries of  $B$  are considered as path coefficients which represent the contemporaneous and 1-day lagged effect of the assets to the others, respectively; see Tables 7 and 8.

**Table 7.**  $A$  matrix for the restricted Financial VAR(1) model (rounded to 4 decimal places).

	NIK	EU	ISE	EM	BVSP	DAX	FTSE	SP
NIK	1	0	0	-0.8193	0.2080	0	0	0
EU	0	1	-0.0421	0	-0.0269	-0.3782	-0.5297	0
ISE	0	0	1	-0.9386	0.1653	-0.1675	-0.3161	-0.1477
EM	0	0	0	1	-0.3419	-0.1184	-0.2464	0.0997
BVSP	0	0	0	0	1	-0.0130	-0.2729	-0.6423
DAX	0	0	0	0	0	1	-0.8102	-0.2336
FTSE	0	0	0	0	0	0	1	-0.6104
SP	0	0	0	0	0	0	0	1

**Table 8.**  $B$  matrix for the restricted Financial VAR(1) model (rounded to 4 decimal places).

	NIK <sub>-1</sub>	EU <sub>-1</sub>	ISE <sub>-1</sub>	EM <sub>-1</sub>	BVSP <sub>-1</sub>	DAX <sub>-1</sub>	FTSE <sub>-1</sub>	SP <sub>-1</sub>
NIK	0.1811	-0.1797	-0.0856	0.0842	0.0739	-0.0058	-0.1146	-0.2662
EU	-0.0131	0.1213	-0.0046	0.0304	-0.0130	-0.0415	-0.0969	0.0002
ISE	0.0676	0.2814	-0.0658	0.2483	-0.2941	-0.0567	0.0120	-0.1472
EM	-0.0016	-0.0567	-0.0158	0.1067	-0.0908	-0.0951	0.0890	-0.1085
BVSP	-0.0139	0.0704	0.0142	-0.1041	0.1391	-0.1488	0.1195	-0.0828
DAX	-0.0034	0.2019	-0.0342	-0.0046	-0.0353	-0.0474	-0.0669	-0.0672
FTSE	0.0292	-0.0171	-0.0109	0.0419	-0.1130	0.2142	0.0807	-0.2642
SP	0.0417	0.2608	-0.0261	0.0115	-0.0026	-0.0713	-0.2853	0.1239



In the VAR(2) situation, the graph, constructed by  $C^{-1}(2|1,0)$ , is the same, has the same JT with 3 cliques, and the same RZP as based on  $C^{-1}(1|0)$ . It is in accordance with our former observation that the lag of 2 or more days' effects of the assets to the others is negligible compared to the 1-day lag effect (the forthcoming order selection also supports this).

Here, the  $24 \times 24$  matrix  $\hat{K}$  was estimated by adapting Equation (13) to the  $3d \times 3d$  situation, by using both the lag 1 and lag 2 variables for covariance selection. This is to be found in the Supplementary Material. The estimated  $A$ ,  $B_1$ , and  $B_2$  matrices are shown in Tables 9–11.

**Table 9.**  $A$  matrix for the restricted Financial VAR(2) model (rounded to 4 decimals).

	NIK	EU	ISE	EM	BVSP	DAX	FTSE	SP
NIK	1	0	0	−0.8191	0.2076	0	0	0
EU	0	1	−0.0423	0	−0.0293	−0.3811	−0.5192	0
ISE	0	0	1	−0.9662	0.1790	−0.1713	−0.3112	−0.1470
EM	0	0	0	1	−0.3361	−0.1153	−0.2372	0.0835
BVSP	0	0	0	0	1	−0.0069	−0.2544	−0.6664
DAX	0	0	0	0	0	1	−0.8128	−0.2336
FTSE	0	0	0	0	0	0	1	−0.6319
SP	0	0	0	0	0	0	0	1

**Table 10.**  $B_1$  matrix for the restricted Financial VAR(2) model (rounded to 4 decimals).

	NIK <sub>−1</sub>	EU <sub>−1</sub>	ISE <sub>−1</sub>	EM <sub>−1</sub>	BVSP <sub>−1</sub>	DAX <sub>−1</sub>	FTSE <sub>−1</sub>	SP <sub>−1</sub>
NIK	0.2009	−0.1869	−0.1098	0.1089	0.0824	−0.0079	−0.1493	−0.2428
EU	−0.0038	0.1387	−0.0013	0.0260	−0.0153	−0.0410	−0.1027	−0.0086
ISE	0.0353	0.2865	−0.0750	0.2479	−0.2741	−0.0639	0.0101	−0.1418
EM	0.0494	−0.0218	−0.0027	0.1338	−0.1144	−0.0990	0.0500	−0.1177
BVSP	−0.0107	0.1202	0.0276	−0.0947	0.1327	−0.1674	0.0987	−0.1030
DAX	−0.0110	0.2072	−0.0322	0.0034	−0.0412	−0.0503	−0.0677	−0.0675
FTSE	0.0824	0.0176	0.0281	0.0224	−0.1104	0.2309	0.0928	−0.3463
SP	0.0506	0.2898	−0.0560	0.0040	0.0037	−0.1010	−0.3199	0.1760

**Table 11.**  $B_2$  matrix for the restricted Financial VAR(2) model (rounded to 4 decimals).

	NIK <sub>−2</sub>	EU <sub>−2</sub>	ISE <sub>−2</sub>	EM <sub>−2</sub>	BVSP <sub>−2</sub>	DAX <sub>−2</sub>	FTSE <sub>−2</sub>	SP <sub>−2</sub>
NIK	−0.0455	−0.1847	−0.0391	0.0264	0.0906	−0.0486	0.1427	0.0089
EU	0.0017	0.0755	−0.0058	0.0047	0.0033	0.0179	−0.0765	−0.0370
ISE	−0.0161	−0.1634	−0.0290	−0.0021	0.0352	0.1113	0.0821	0.0313
EM	−0.0056	0.0659	−0.0330	0.1189	−0.0701	−0.0959	−0.0167	−0.0283
BVSP	−0.0430	0.0415	−0.0456	0.2906	−0.0729	−0.0258	−0.0389	−0.0168
DAX	−0.0369	0.0163	0.0130	0.0656	−0.0356	−0.0100	−0.0203	0.0064
FTSE	0.0485	0.3142	−0.0820	0.0716	0.0290	0.0128	−0.0845	−0.3054
SP	0.0442	−0.0606	0.0805	−0.1825	0.0778	0.0117	−0.1773	0.1281

Summarizing, in the  $p = 1$  and  $p = 2$  cases, when we took into consideration the lag 1 and 2 variables, respectively, in the graph building, we obtained the same graph with the same threshold for the partial correlation coefficients as in the CVAR(0) case.

To find the optimal order  $p$ , information criteria are suggested; see e.g., [Box et al. \(2015\)](#); [Brockwell and Davis \(1991\)](#). Here, the following criteria will be used: the AIC (Akaike Information Criterion), the AICC (bias corrected version of the AIC), the BIC (Bayesian information criterion), and the HQ (Hannan and Quinn's criterion). Each criterion can be decomposed into two terms: an *information term* that quantifies the information brought by the model (via the likelihood) and a *penalization term* that penalizes too "large" number of parameters, in order to avoid over-fitting. It can be proven that the AIC has a positive probability of overspecification and the BIC is strongly consistent, but sometimes it underspecifies the true model. The explicit forms of AIC, BIC, and HQ, which are to be minimized with respect to  $p$ , are as follows:

$$\begin{aligned}
 \text{AIC}(p) &= \ln |\hat{\Delta}| + \frac{2 \left[ pd^2 + \binom{d}{2} \right]}{n-p} = \sum_{j=1}^d \ln \hat{\delta}_j + \frac{2 \left[ pd^2 + \binom{d}{2} \right]}{n-p}, \\
 \text{BIC}(p) &= \ln |\hat{\Delta}| + \frac{\left[ pd^2 + \binom{d}{2} \right] \ln(n-p)}{n-p} = \sum_{j=1}^d \ln \hat{\delta}_j + \frac{\left[ pd^2 + \binom{d}{2} \right] \ln(n-p)}{n-p}, \\
 \text{HQ}(p) &= \ln |\hat{\Delta}| + \frac{2 \left[ pd^2 + \binom{d}{2} \right] \ln(\ln(n-p))}{n-p} = \sum_{j=1}^d \ln \hat{\delta}_j + \frac{2 \left[ pd^2 + \binom{d}{2} \right] \ln(\ln(n-p))}{n-p},
 \end{aligned}$$

where  $\hat{\Delta}$  is the estimated error covariance matrix  $\Delta$ .

The AICC (Akaike Information Criterion Corrected) is a bias-corrected version of Akaike’s AIC, which is an estimate of the Kullback–Leibler index of the fitted model relative to the true model and needs further explanation. Here,

$$\text{AICC}(p) = -2 \ln L(\hat{A}, \hat{B}_1, \dots, \hat{B}_p, \hat{\Delta}) + \text{penalty}(p),$$

where the first term is  $-2$  times the log-likelihood function, evaluated at the parameter estimates of Theorems 1 and 2, whereas the second term penalizes the computational complexity. The model parameters  $A, B_1, \dots, B_p$ , and  $\Delta$  are estimated by the block Cholesky decomposition of the estimated inverse covariance matrix  $\mathcal{C}_{p+1}^{-1}$  of the Gaussian random vector  $(\mathbf{X}_t^T, \mathbf{X}_{t-1}^T, \dots, \mathbf{X}_{t-p}^T)^T$ ; see Algorithms of Appendices A.2 and A.4. This is a moment estimation, but since our underlying distribution is multivariate Gaussian, which belongs to the exponential family, asymptotically, it is also an MLE (for “large”  $n$ ) that satisfies the moment matching equations, see Wainwright and Jordan (2008). Of course, the matrices  $\hat{A}, \hat{B}_1, \dots, \hat{B}_p$ , and  $\hat{\Delta}$  also depend on  $p$ , but for simplicity, we do not denote this dependence. More exactly,

$$\begin{aligned}
 L(\hat{A}, \hat{B}_1, \dots, \hat{B}_p, \hat{\Delta}) &= (2\pi)^{-\frac{(n-p)d}{2}} |\hat{\Delta}|^{-\frac{n-p}{2}} e^{-\frac{1}{2} \sum_{t=p+1}^n \mathbf{U}_t^T \hat{\Delta}^{-1} \mathbf{U}_t} \\
 &= (2\pi)^{-\frac{(n-p)d}{2}} \left( \prod_{j=1}^d \hat{\delta}_j \right)^{-\frac{n-p}{2}} e^{-\frac{1}{2} \sum_{t=p+1}^n \sum_{j=1}^d (U_{tj}^2 / \hat{\delta}_j)},
 \end{aligned}$$

where

$$\mathbf{U}_t = \hat{A}(\mathbf{X}_t - \hat{\mathbf{X}}_t),$$

and

$$\hat{\mathbf{X}}_t = -\hat{A}^{-1} \hat{B}_1 \mathbf{X}_{t-1} - \dots - \hat{A}^{-1} \hat{B}_p \mathbf{X}_{t-p},$$

for  $t = p + 1, \dots, n$ .

In the unrestricted model, the complexity term (see Brockwell and Davis (1991)) is

$$\frac{2 \left[ pd^2 + \binom{d}{2} \right] (n-p)d}{(n-p)d - pd^2 - \binom{d}{2} - 1}.$$

Therefore,

$$\begin{aligned}
 \text{AICC}(p) &= (n-p)d \ln(2\pi) + (n-p) \ln |\hat{\Delta}| + \sum_{t=p+1}^n \mathbf{U}_t^T \hat{\Delta}^{-1} \mathbf{U}_t + \frac{2 \left[ pd^2 + \binom{d}{2} \right] (n-p)d}{(n-p)d - pd^2 - \binom{d}{2} - 1} \\
 &= (n-p)d \ln(2\pi) + (n-p) \sum_{j=1}^d \ln \hat{\delta}_j + \sum_{t=p+1}^n \sum_{j=1}^d \frac{U_{tj}^2}{\hat{\delta}_j} + \frac{2 \left[ pd^2 + \binom{d}{2} \right] (n-p)d}{(n-p)d - pd^2 - \binom{d}{2} - 1}.
 \end{aligned}$$

In the restricted model, the penalization term depends on the cardinalities of the cliques  $C_1, \dots, C_k$  that are the same for all  $p$ . The penalization terms for the four criteria are

$$\begin{aligned} \text{penalty}_{\text{AIC}}(p) &= \frac{2 \left[ pd^2 + \sum_{j=1}^k \binom{|C_j|}{2} - \sum_{j=2}^k \binom{|S_j|}{2} \right]}{n - p}, \\ \text{penalty}_{\text{BIC}}(p) &= \frac{\left[ pd^2 + \sum_{j=1}^k \binom{|C_j|}{2} - \sum_{j=2}^k \binom{|S_j|}{2} \right] \ln(n - p)}{n - p}, \\ \text{penalty}_{\text{HQ}}(p) &= \frac{2 \left[ pd^2 + \sum_{j=1}^k \binom{|C_j|}{2} - \sum_{j=2}^k \binom{|S_j|}{2} \right] \ln(\ln(n - p))}{n - p}, \\ \text{penalty}_{\text{AICC}}(p) &= \frac{2 \left[ pd^2 + \sum_{j=1}^k \binom{|C_j|}{2} - \sum_{j=2}^k \binom{|S_j|}{2} \right] (n - p)d}{(n - p)d - pd^2 - \sum_{j=1}^k \binom{|C_j|}{2} - 1}. \end{aligned}$$

The cliques are usually of “small” sizes that can reduce computational complexity, in particular, when the number of variables  $d$  is much “larger” than the clique sizes. Furthermore, the separators are intersections of the cliques, so the number of product-moments calculated within them can be subtracted.

All of these criteria are tested, for both the restricted and unrestricted CVAR( $p$ ) models, using the financial data above for  $p = 1, 2, \dots, 9$ . The results for the unrestricted case are shown in Table 12.

**Table 12.** Order selection criteria for the unrestricted Financial CVAR( $p$ ) model, bold-face values represent minimum of each criterion.

$p$	AIC	AICC	BIC	HQ
1	−76.81	− <b>33,222.68</b>	− <b>76.07</b>	− <b>76.52</b>
2	− <b>76.85</b>	−33,173.98	−75.60	−76.36
3	−76.84	−33,095.75	−75.08	−76.15
4	−76.83	−33,011.98	−74.55	−75.94
5	−76.77	−32,893.23	−73.97	−75.67
6	−76.69	−32,766.33	−73.37	−75.39
7	−76.58	−32,612.38	−72.74	−75.08
8	−76.48	−32,457.38	−72.11	−74.77
9	−76.41	−32,316.33	−71.52	−74.49

Observe that, in the unrestricted case, AIC reaches the minimum for  $p = 2$ , whereas AICC, BIC, and HQ for  $p = 1$ . This is in accordance with our previous experience that the parameter matrices did not change much after the first or second day.

In the restricted case (see Table 13), except for the AIC, every criterion suggests that the best model is obtained with  $p = 1$ . Thus, the parameter matrices did not change much after the first day, except for AIC, which seems to overspecify the model and was the lowest on the 4th day, i.e., the last workday after the first workday. In addition, these criteria showed only a minuscule decrease in the restricted case; probably because the clique sizes were not significantly smaller than  $d$ .

**Table 13.** Order selection criteria for the restricted Financial CVAR( $p$ ) model; bold-face values represent the minimum of each criterion.

$p$	AIC	AICC	BIC	HQ
1	−76.87	− <b>33,239.11</b>	− <b>76.19</b>	− <b>76.60</b>
2	−76.91	−33,190.27	−75.71	−76.44
3	−76.93	−33,129.67	−75.22	−76.26
4	− <b>77.00</b>	−33,084.90	−74.77	−76.13
5	−76.94	−32,969.37	−74.19	−75.86
6	−76.92	−32,869.31	−73.65	−75.64
7	−76.81	−32,718.36	−73.02	−75.33
8	−76.80	−32,612.63	−72.49	−75.11
9	−76.78	−32,495.56	−71.94	−74.88

Componentwise predictions of  $X_t$  with RMSEs and figures are shown in the Supplementary Material.

#### 4.2. IMR (Infant Mortality Rate) Longitudinal Data

Here, we used the longitudinal data of six indicators (components of  $X_t$ ), spanning 21 years (1995–2015) from the World Bank in the case of Egypt:

- 1 : IMR (Infant mortality rate),
  - 2 : MMR (Maternal mortality ratio),
  - 3 : HepB (Hepatitis-B immunization),
  - 4 : GDP (Gross domestic per capita),
  - 5 : OPExp (Out-of-pocket health expenditure as % of HExp),
  - 6 : HExp (Current health expenditure as % of GDP).
- (15)

For more details about these indicators, see [Abdelkhalek and Bolla \(2020\)](#). Through the CVAR( $p$ ) model, we show the contemporaneous and lagged time effects between the components. Since the sample size is small, we investigate only the CVAR(1) model in the unrestricted and restricted situations. Furthermore, the variables are measured on different scales; thus, we use the autocorrelations which are the autocovariances of the standardized variables. We distinguish between two working hypotheses with respect to two different ordering of the variables given by an expert:

- Case 1: {IMR, MMR, HepB, OPExp, HExp, GDP}.
- Case 2: {IMR, MMR, HepB, GDP, OPExp, HExp}.

In the unrestricted CVAR(1) model, both orderings work, but we present only Case 1. (The estimated matrices  $A$  and  $B$  are mostly the same in both cases, but the entries are interchanged with respect to the ordering of the variables.) The entries of matrix  $A$  (see Table 14) represent the contemporaneous effects (path coefficients) between the components at time  $t$ . The MMR has the largest contemporaneous inverse causal effect on the IMR, i.e., an increase in the MMR caused a decrease in the IMR by 1.13. Matrix  $B$  (see Table 15), on the other hand, indicates the path coefficients of the one time lag causal effect of  $X_{t-1}$  on the current  $X_t$  components. An increase in the IMR at a one-year time lag caused an increase in the IMR at the current time by 0.29. All other path coefficients in the matrices  $A, B$  can be explained likewise.

**Table 14.** A matrix for the IMR unrestricted VAR(1) model of Case 1 (rounded to 4 decimals).

	IMR	MMR	HepB	OPExp	HExp	GDP
IMR	1.0	−1.1259	−0.0161	0.0003	0.0176	−0.1348
MMR	0.0	1.0000	0.3594	0.0492	−0.0684	0.7135
HepB	0.0	0.0000	1.0000	−0.1626	0.2510	−0.8196
OPExp	0.0	0.0000	0.0000	1.0000	−0.6876	−0.4229
HExp	0.0	0.0000	0.0000	0.0000	1.0000	0.6749
GDP	0.0	0.0000	0.0000	0.0000	0.0000	1.0000

In the restricted CVAR(1) model, the graph structure is important. We consider only Case 2 that provides the RZP and corresponds to the ordering of (15). Note that the so-obtained DAG is Markov equivalent to its undirected skeleton. The decomposable structure of the JT has two cliques and only one separator as follows:

$$\begin{aligned}
 C_1 &= \{\text{IMR, MMR, HepB, GDP, HExp}\} = \{1, 2, 3, 4, 6\}, \\
 C_2 &= \{\text{OPExp, HExp}\} = \{5, 6\}, \\
 S_2 &= \{\text{HExp}\} = \{6\}.
 \end{aligned}$$

In this case, the  $n - 1$  element sample including lag 1 variables is used to estimate the  $12 \times 12$  matrix  $\hat{K}$  with covariance selection, see Equation (13). Then, the LDL algorithm was applied to the so-obtained  $\hat{K}$  (shown in the Supplementary Material) to estimate the model parameters  $A, B$ . Unlike the unrestricted model, here there are prescribed zero entries in  $\hat{K}$  and  $A$ . Specifically, the zeros of the left upper  $6 \times 6$  corner of  $\hat{K}$  will necessarily result in the zeros of the estimated matrix  $A$  in the same positions. Similarly to the unrestricted situation, the non-zero upper-diagonal entries of  $A$  (see Table 16) represent the path coefficients of the contemporaneous causal effects of  $X_t$ , while the entries of the matrix  $B$  (see Table 17) represent the one time lag causal effects of the  $X_{t-1}$  components on the  $X_t$  components.

**Table 15.**  $B$  matrix for the IMR unrestricted VAR(1) model of Case 1 (rounded to 4 decimals).

	IMR-1	MMR-1	HepB-1	OPExp-1	HExp-1	GDP-1
IMR	0.2986	−0.3589	−0.0076	0.0042	−0.0115	−0.0639
MMR	−0.0149	−0.7469	−0.2358	0.0540	0.0193	−0.5577
HepB	13.0541	−15.7658	−1.1915	−0.3506	0.2170	−1.7902
OPExp	7.1616	−8.0906	−0.2994	−0.1038	−0.0215	−0.7720
HExp	1.6861	−2.9922	−0.4650	−0.1566	−0.0681	−1.0913
GDP	−11.0674	13.2182	0.3254	0.3204	−0.3099	1.2129

**Table 16.**  $A$  matrix for the IMR restricted VAR(1) model of Case 2 (rounded to 4 decimals).

	IMR	MMR	HepB	GDP	OPExp	HExp
IMR	1.0	0.0736	0.0052	0.0111	0.0	0.0040
MMR	0.0	1.0000	0.0237	0.1180	0.0	−0.0116
HepB	0.0	0.0000	1.0000	−0.3418	0.0	0.0236
GDP	0.0	0.0000	0.0000	1.0000	0.0	−0.0035
OPExp	0.0	0.0000	0.0000	0.0000	1.0	−0.7854
HExp	0.0	0.0000	0.0000	0.0000	0.0	1.0000

**Table 17.**  $B$  matrix for the IMR restricted VAR(1) model of Case 2 (rounded to 4 decimals).

	IMR-1	MMR-1	HepB-1	GDP-1	OPExp-1	HExp-1
IMR	−0.9198	−0.0592	−0.0021	0.0053	−0.0013	−0.0049
MMR	−1.0175	0.2511	−0.0002	0.0607	−0.0047	0.0039
HepB	3.6850	−4.4606	−0.9058	−0.4230	−0.0811	−0.0950
GDP	0.5849	−0.4453	0.0640	−0.8573	0.0896	0.1086
OPExp	3.6432	−3.7486	−0.1413	−0.4380	0.0273	−0.0926
HExp	1.9561	−3.4687	−0.5221	−0.6302	−0.2298	−0.1171

## 5. Discussion

The main contribution of our paper is the introduction of causality in VAR models by using graphical modeling tools. SVAR models are known in the literature, but there the inclusion of the upper triangular matrix  $A$  rather facilitates an alternative solution for the Yule–Walker equations, and not the causal ordering of the contemporaneous effects.

Our unrestricted CVAR model does this job, where the recursive ordering of the variables follows a DAG ordering in the directed graphical model contemporaneously, and the entries of  $A$  are treated like path coefficients of SEM. In addition, the white noise process  $U_t$  of structural shocks (see Equation (6)) is obtained from the process  $V_t$  of innovations in the reduced form (see Equation (1)) and has an econometric interpretation. The structural shocks are mutually uncorrelated, and they are assigned to the individual variables. They also represent unanticipated changes in the observed econometric variables. However, they are not just orthogonalized innovations, but here the labeling of the nodes and the graph skeleton behind the matrix  $A$  also matters.

In the unrestricted case, the following estimation scheme is used. The DAG is built partly by expert knowledge and partly by starting with an undirected Gaussian graphical model, using known algorithms (e.g., MCS) to find a triangulated graph and a (not necessarily unique) perfect labeling of the nodes, in which ordering the directed and undirected models are Markov equivalent to each other (there are no sink  $V$  configurations in the DAG). However, here the Markov equivalence is not important: even if the undirected graph is not triangulated, and the DAG contains sink  $V$ s, the DAG ordering (given, e.g., by an expert) can be used to estimate the  $A$  and  $B$  matrices, which are full in the sense that no zero constraints for their entries are assumed at the beginning. After having the DAG ordering, we apply the block LDL decomposition for the estimated block matrix  $\mathcal{C}_2$  or  $\mathcal{C}_{p+1}$ , and retrieve the estimated parameter matrices by Theorem 1 or Theorem 2.

It is the restricted CVAR model, where zero constraints for the entries of  $A$  (in the given DAG ordering) are made. For this purpose, we re-estimate the covariance matrix (the big block matrix, the size of which depends on the order  $p$  of the model) such that the entries in the left upper block of its inverse are zeros in the no-edge positions. For this, there is the method of covariance selection at our disposal, which works for Gaussian variables even if the prescribed zeros in the inverse covariance matrix do not have the RZP (RZP is just the property of decomposable models). In this case, our algorithm first applies algorithms (e.g., MCS) to find the JT structure of the graph (which is equivalent to having an RZP). The estimation scheme is enhanced with covariance selection, for which there are closed form estimates in the decomposable case. Actually, we use an improved version of the covariance selection that needs higher order autocovariances too and relaxes the independence of the sample entries which are only serially correlated. This is supported by ergodicity issues if  $n$  is “large”. Note that, in the lack of an RZP, the covariance selection still works, but it needs the infinite iteration (IPS).

Since the necessary product-moment estimates include only variables belonging to the cliques and separators, and the separators are intersections of the cliques, this fact can reduce the computational complexity of the restricted CVAR model compared to the unrestricted one. The information criteria, applied to select the optimal order  $p$ , also take into consideration the number of relevant parameters to be estimated.

## 6. Conclusions and Further Perspectives

Our algorithm is also applicable to longitudinal data instead of time series. The  $p = 0$  case resolves the problem posed in [Wermuth \(1980\)](#), and the  $p = 1$  case is also applicable to solve a SEM with endogenous and exogenous variables.

As a further perspective, lagged causalities could also be introduced, with some upper triangular matrices  $B$ . For example, if the previous time observations influence the present time ones, and the order of causalities is the same as that of the contemporaneous ones, then  $B_1$  is also upper triangular. This problem can be solved by running the block Cholesky decomposition with  $2d$  singleton blocks and treating only the other blocks “en block”.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/econometrics11010007/s1>. For illustrating the CVAR model and related algorithms, there are supporting Python and notebook files uploaded, together with some additional tables and figures: CVAR.py, CVAR\_example.ipynb, CVARtables.pdf, CVARfigures.pdf.

**Author Contributions:** Conceptualization, M.B. (Marianna Bolla), D.Y.; methodology, M.B. (Marianna Bolla), M.B. (Máté Baranyi), F.A. and V.F.; software, D.Y., H.W., R.M. and V.F.; validation, W.T., C.D.; formal analysis, F.A., V.F.; investigation, D.Y.; resources, M.B. (Máté Baranyi); data curation, M.B. (Máté Baranyi), F.A.; writing—original draft preparation, M.B. (Marianna Bolla); writing—review and editing, M.B. (Marianna Bolla), M.B. (Máté Baranyi); visualization, D.Y., V.F. and F.A.; supervision, M.B. (Marianna Bolla); project administration, M.B. (Marianna Bolla). All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The third-party financial dataset analyzed in the current study is available in the UCI Machine Learning Repository, and was collected by the authors of Akbilgic et al. (2014). The dataset is available in: Dua, D. and Graff, C. (2019). UCI Machine Learning Repository, Irvine, CA: University of California, School of Information and Computer Science, <https://archive.ics.uci.edu/ml/datasets/ISTANBUL+STOCK+EXCHANGE> (accessed on 1 August 2022). The World Bank Data on infant mortality rates are available on <https://data.worldbank.org/indicator> (accessed on 1 August 2022); see also Abdelkhalek and Bolla (2020).

**Acknowledgments:** The research was carried out under the auspices of the Budapest Semesters in Mathematics program, in the framework of an undergraduate online research course in summer 2021, with the participation of US undergraduate students. Two PhD students of the corresponding author also participated. In particular, Fatma Abdelkhalek’s work was funded by a scholarship under the Stipendium Hungaricum program between Egypt and Hungary, whereas Valentin Frappier’s internship by the Erasmus program of the EU.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

VAR	Vector AutoRegression
SVAR	Structural Vector AutoRegression
CVAR	Causal Vector AutoRegression
SEM	Structural Equation Modeling
DAG	Directed Acyclic Graph
JT	Junction Tree
MCS	Maximal Cardinality Search
IPS	Iterative Proportional Scaling
RZP	Reducible Zero Pattern
MRF	Markov Random Field
AIC	Akaike Information Criterion
AICC	Akaike Information Criterion Corrected
BIC	Bayesian Information Criterion
HQ	Hannan and Quinn’s criterion
MLE	Maximum Likelihood Estimate
PLS	Partial Least Squares regression
RMSE	Root Mean Square Error
IMR	Infant Mortality Rate
LDL	variant of the Cholesky decomposition for a symmetric, positive semidefinite matrix as $L$ (lower triangular) $\times D$ (diagonal) $\times L^T$

## Appendix A. Proofs of the Main Theorems

### Appendix A.1. Proof of Theorem 1

First of all, note that the block Cholesky decomposition applies to  $K$  partitioned symmetrically into  $(d + 1) \times (d + 1)$  blocks of sizes  $1, \dots, 1, d$ , where the number of singleton blocks (of size 1) is  $d$ . (In the  $p = 0$  case, all the blocks are singletons, so the standard LDL decomposition is applicable.) Therefore, in the main diagonal of the resulting  $L$ , we have number  $d$  of 1s and  $I_{d \times d}$  (the  $d \times d$  identity matrix), see the forthcoming Equation (A3). In other words, the last  $d$  rows (and columns) are treated “en block”; this is why here indeed the block LDL (variant of the block Cholesky) decomposition is applicable.

Let us compute the inverse of the matrix  $LDL^T$  with block matrices  $L$  and  $D$  partitioned as in Equation (5). For the time being, we only assume that  $A$  is a  $d \times d$  upper triangular matrix with 1s along its main diagonal,  $B$  is  $d \times d$ , and the diagonal matrix  $\Delta$  has positive diagonal entries. We will use the computation rule of the inverse of a symmetrically partitioned block matrix Rózsa (1991), which is applicable due to the fact that  $|A| = 1$ , so the matrix  $A$  is invertible:

$$\begin{aligned}
(\mathbf{LDL}^T)^{-1} &= \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{O}_{d \times d} & \mathbf{I}_{d \times d} \end{pmatrix}^{-1} \begin{pmatrix} \Delta^{-1} & \mathbf{O}_{d \times d} \\ \mathbf{O}_{d \times d} & \mathbf{C}^{-1}(0) \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{A}^T & \mathbf{O}_{d \times d} \\ \mathbf{B}^T & \mathbf{I}_{d \times d} \end{pmatrix}^{-1} \\
&= \begin{pmatrix} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B} \\ \mathbf{O}_{d \times d} & \mathbf{I}_{d \times d} \end{pmatrix} \begin{pmatrix} \Delta & \mathbf{O}_{d \times d} \\ \mathbf{O}_{d \times d} & \mathbf{C}(0) \end{pmatrix} \begin{pmatrix} (\mathbf{A}^T)^{-1} & \mathbf{O}_{d \times d} \\ -\mathbf{B}^T(\mathbf{A}^T)^{-1} & \mathbf{I}_{d \times d} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{A}^{-1}\Delta(\mathbf{A}^{-1})^T + \mathbf{A}^{-1}\mathbf{BC}(0)\mathbf{B}^T(\mathbf{A}^{-1})^T & -\mathbf{A}^{-1}\mathbf{BC}(0) \\ -\mathbf{C}(0)\mathbf{B}^T(\mathbf{A}^T)^{-1} & \mathbf{C}(0) \end{pmatrix}.
\end{aligned}$$

Now, we are going to prove that the above matrix equals  $\mathfrak{C}_2$  if and only if  $\mathbf{A}, \mathbf{B}, \Delta$  satisfy the model equations. Comparing the blocks to those of (4), the right bottom block is  $\mathbf{C}(0)$  in both expressions. Comparing the left bottom blocks, we obtain  $-\mathbf{C}(0)\mathbf{B}^T(\mathbf{A}^T)^{-1} = \mathbf{C}(1)$ , and so  $\mathbf{B}^T = -\mathbf{C}^{-1}(0)\mathbf{C}(1)\mathbf{A}^T$  and  $\mathbf{B} = -\mathbf{AC}^T(1)\mathbf{C}^{-1}(0)$  should hold for  $\mathbf{B}$ . It is in accordance with the model equation. Indeed, (3) is equivalent to

$$\mathbf{B}\mathbf{X}_{t-1} = -\mathbf{A}\mathbf{X}_t + \mathbf{U}_t,$$

which, after multiplying with  $\mathbf{X}_{t-1}^T$  from the right and taking expectations into consideration, yields  $\mathbf{BC}(0) = -\mathbf{AC}^T(1)$  that in turn provides

$$\mathbf{B} = -\mathbf{AC}^T(1)\mathbf{C}^{-1}(0). \quad (\text{A1})$$

By symmetry, the same applies to the right upper block. As for the left upper block,

$$\mathbf{A}^{-1}\Delta(\mathbf{A}^T)^{-1} + \mathbf{A}^{-1}\mathbf{BC}(0)\mathbf{B}^T(\mathbf{A}^T)^{-1} = \mathbf{C}(0)$$

should hold. Multiplying this equation with  $\mathbf{A}$  from the left and with  $\mathbf{A}^T$  from the right, we obtain the equivalent equation

$$\Delta = \mathbf{AC}(0)\mathbf{A}^T - \mathbf{BC}(0)\mathbf{B}^T. \quad (\text{A2})$$

This is in accordance with Equation (3), which implies

$$\mathbb{E}(\mathbf{A}\mathbf{X}_t + \mathbf{B}\mathbf{X}_{t-1})(\mathbf{A}\mathbf{X}_t + \mathbf{B}\mathbf{X}_{t-1})^T = \mathbf{AC}(0)\mathbf{A}^T + \mathbf{AC}^T(1)\mathbf{B}^T + \mathbf{BC}(1)\mathbf{A}^T + \mathbf{BC}(0)\mathbf{B}^T = \Delta.$$

Combining this with Equation (A1), we obtain

$$\begin{aligned}
\Delta &= \mathbf{AC}(0)\mathbf{A}^T + \mathbf{AC}^T(1)\mathbf{B}^T + \mathbf{BC}(1)\mathbf{A}^T + \mathbf{BC}(0)\mathbf{B}^T \\
&= \mathbf{AC}(0)\mathbf{A}^T - \mathbf{AC}^T(1)\mathbf{C}^{-1}(0)\mathbf{C}(1)\mathbf{A}^T - \mathbf{AC}^T(1)\mathbf{C}^{-1}(0)\mathbf{C}(1)\mathbf{A}^T \\
&\quad + \mathbf{AC}^T(1)\mathbf{C}^{-1}(0)\mathbf{C}(0)\mathbf{C}^{-1}(0)\mathbf{C}(1)\mathbf{A}^T \\
&= \mathbf{AC}(0)\mathbf{A}^T - \mathbf{AC}^T(1)\mathbf{C}^{-1}(0)\mathbf{C}(1)\mathbf{A}^T = \mathbf{AC}(0)\mathbf{A}^T - \mathbf{BC}(0)\mathbf{B}^T,
\end{aligned}$$

which also satisfies (A2).

Summarizing, we have proved that, under the model equations,  $(\mathbf{LDL}^T)^{-1} = \mathfrak{C}_2$ , or equivalently,  $\mathbf{LDL}^T = \mathbf{K}$  indeed holds. In view of the uniqueness of the block LDL decomposition (under positive definiteness of the involved matrices), this finishes the proof.

#### Appendix A.2. Algorithm for the Block LDL Decomposition of Appendix A.1

By the preliminary assumptions,  $\mathbf{K}$  and so  $\mathbf{D}$  are positive definite; therefore,  $\Delta$  has positive diagonal entries. To apply the protocol of the block Cholesky decomposition, which gives the theoretically guaranteed unique solution, it is worth writing the above matrices according to the blocks as follows. The matrix  $\mathbf{L}$  has the partitioned form



$$L = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \ell_{21} & 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ \ell_{31} & \ell_{32} & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & 0 & \dots & 0 \\ \ell_{d1} & \ell_{d2} & \dots & \ell_{d,d-1} & 1 & 0 & \dots & 0 \\ \ell_{d+1,1} & \ell_{d+1,2} & \vdots & \ell_{d+1,d-1} & \ell_{d+1,d} & \mathbf{I}_{d \times d} & & \end{pmatrix}, \tag{A3}$$

where the  $2d \times 2d$  lower triangular matrix  $L$  is also lower triangular with respect to its blocks which are partly scalars, partly vectors, and partly matrices as follows:

$$\ell_{ij} = \begin{cases} a_{ji}, & j = 1, \dots, d-1; \quad i = j+1, \dots, d; \\ 1 & i = j = 1, \dots, d; \\ 0 & i = 1, \dots, d; \quad j = i+1, \dots, 2d; \end{cases}$$

Furthermore, the vectors  $\ell_{d+1,j}$  are  $d \times 1$  for  $j = 1, \dots, d$ , and comprise the column vectors of the  $d \times d$  matrix  $B^T$ . The matrix in the bottom right block is the  $d \times d$  identity  $I_{d \times d}$ , and above it, the zero entries can be arranged into the  $d \times d$  zero matrix  $O_{d \times d}$ .

The  $2d \times 2d$  block-diagonal matrix  $D$  in partitioned form is

$$D = \begin{pmatrix} \delta_1^{-1} & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \delta_2^{-1} & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \delta_3^{-1} & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \mathbf{0} & \dots & \mathbf{0} \\ 0 & 0 & 0 & 0 & \delta_d^{-1} & 0 & \dots & 0 \\ \mathbf{0} & \mathbf{0} & \vdots & \mathbf{0} & \mathbf{0} & & C^{-1}(0) & \end{pmatrix},$$

where the  $d \times 1$  vectors  $\mathbf{0}$  comprise  $O_{d \times d}$  in the left bottom, and the entries comprise the inverse of the  $d \times d$  positive definite matrix  $C(0)$  in the right bottom block. We perform the following multiplications of block matrices, also using formulas Golub and Van Loan (2012); Rózsa (1991) for their inverses and the algorithm proposed in Nocedal and Wright (1999) to obtain the recursion of the block LDL decomposition that goes on as follows:

- Outer cycle (column-wise). For  $j = 1, \dots, d$ :  $\delta_j^{-1} = k_{jj} - \sum_{h=1}^{j-1} \ell_{jh} \delta_h^{-1} \ell_{jh}$  (with the reservation that  $\delta_1^{-1} = k_{11}$ );
- Inner cycle (row-wise). For  $i = j+1, \dots, d$ :

$$\ell_{ij} = \left( k_{ij} - \sum_{h=1}^{j-1} \ell_{ih} \delta_h^{-1} \ell_{jh} \right) \delta_j \tag{A4}$$

and

$$\ell_{d+1,j} = \left( \mathbf{k}_{d+1,j} - \sum_{h=1}^{j-1} \ell_{d+1,h} \delta_h^{-1} \ell_{jh} \right) \delta_j$$

(with the reservation that, in the  $j = 1$  case, the summand is zero), where  $\mathbf{k}_{d+1,j}$  for  $j = 1, \dots, d$  is  $d \times 1$  vector in the bottom left block of  $K$ .

Note that the last step of the outer cycle, when  $j = d+1$ , formally would be

$$C^{-1}(0) = \mathbf{K}_{d+1,d+1} - \sum_{h=1}^d \ell_{d+1,h} \delta_h^{-1} \ell_{d+1,h}^T = \mathbf{K}_{d+1,d+1} - \sum_{h=1}^d \delta_h^{-1} \ell_{d+1,h} \ell_{d+1,h}^T,$$

where  $\ell_{d+1,h}$  for  $h = 1, \dots, d$  are  $d \times 1$  vectors, and  $\mathbf{K}_{d+1,d+1}$  is the bottom right  $d \times d$  block of the  $2d \times 2d$  concentration matrix  $K$ , but it need not be performed as it is in accordance with Theorem 1. Then, no inner cycle follows and the recursion ends in one run.

It is obvious that the above decomposition has a nested structure, so, for the first  $d$  rows of  $L$ , only its previous rows or preceding entries in the same row enter into the calculation, as if we performed the standard LDL decomposition of  $K$ . Therefore,  $\ell_{ij} = a_{ji}$  for  $j = 1, \dots, d - 1, i = j + 1, \dots, d$  that are the partial regression coefficients akin to those offered by the standard LDL decomposition  $K = \tilde{L}\tilde{D}\tilde{L}^T$ , so the first  $d$  rows of  $\tilde{L}$  and  $L$  are the same, and the first  $d$  rows of  $\tilde{D}$  and  $D$  are the same too.

When the process terminates after finding the first  $d$  rows of  $L$ , we consider the blocks “en block” and obtain the matrix  $B = (\ell_{d+1,1}, \dots, \ell_{d+1,d})^T$ .

Appendix A.3. Proof of Theorem 2

Note that here the block Cholesky decomposition applies to  $K$  partitioned symmetrically into  $(d + 1) \times (d + 1)$  blocks of sizes  $1, \dots, 1, pd$  with number  $d$  of singleton blocks. (Therefore, in the main diagonal of  $L$ , we have number  $d$  of 1s and  $I_{pd \times pd}$ .) The  $d \times pd$  matrix  $B$ , a transpose of  $B^T$ , will contain the coefficient matrices of Equation (6) in its blocks, like

$$B = (B_1 \dots B_p).$$

The proof goes on similarly as in Appendix A.1. However, for completeness and being able to formulate the algorithm, we discuss it herein. Let us compute the inverse of the matrix  $LDL^T$  with block matrices  $L$  and  $D$  partitioned as in Equation (8). For the time being, we only assume that  $A$  is a  $d \times d$  upper triangular matrix with 1s along its main diagonal,  $B$  is  $d \times pd$ , and the diagonal matrix  $\Delta$  has positive diagonal entries. We can again use the computation rule of the inverse of symmetrically partitioned block matrices, since the matrix  $A$  is invertible.

$$\begin{aligned} (LDL^T)^{-1} &= \begin{pmatrix} A & B \\ O_{pd \times d} & I_{pd \times pd} \end{pmatrix}^{-1} \begin{pmatrix} \Delta^{-1} & O_{d \times pd} \\ O_{pd \times pd} & \mathfrak{C}_p^{-1} \end{pmatrix}^{-1} \begin{pmatrix} A^T & O_{d \times pd} \\ B^T & I_{pd \times pd} \end{pmatrix}^{-1} \\ &= \begin{pmatrix} A^{-1} & -A^{-1}B \\ O_{pd \times d} & I_{pd \times pd} \end{pmatrix} \begin{pmatrix} \Delta & O_{d \times pd} \\ O_{pd \times d} & \mathfrak{C}_p \end{pmatrix} \begin{pmatrix} (A^T)^{-1} & O_{d \times pd} \\ -B^T(A^T)^{-1} & I_{pd \times pd} \end{pmatrix} \\ &= \begin{pmatrix} A^{-1}\Delta(A^{-1})^T + A^{-1}B\mathfrak{C}_pB^T(A^{-1})^T & -A^{-1}B\mathfrak{C}_p \\ -\mathfrak{C}_pB^T(A^T)^{-1} & \mathfrak{C}_p \end{pmatrix}. \end{aligned}$$

Now, we are going to prove that the above matrix equals  $\mathfrak{C}_{p+1}$  if and only if  $A, B, \Delta$  satisfy the model equations. Comparing the blocks to those of (7), the right bottom block is  $\mathfrak{C}_p$  in both expressions. Comparing the left bottom blocks, we obtain  $-\mathfrak{C}_pB^T(A^T)^{-1} = C(1, \dots, p)$ , and so,  $B^T = -\mathfrak{C}_p^{-1}C(1, \dots, p)A^T$  and  $B = -AC^T(1, \dots, p)\mathfrak{C}_p^{-1}$  should hold for  $B$ . It is in accordance with the model equation: indeed, (6) is equivalent to

$$B_1X_{t-1} + \dots + B_pX_{t-p} = -AX_t + U_t,$$

which, after multiplying with  $X_{t-1}^T, \dots, X_{t-p}^T$  from the right and taking expectation, in concise form yields  $B\mathfrak{C}_p = -AC^T(1, \dots, p)$  that in turn provides

$$B = -AC^T(1, \dots, p)\mathfrak{C}_p^{-1}. \tag{A5}$$

By symmetry, it also applies to the right upper block. As for the left upper block,

$$A^{-1}\Delta(A^T)^{-1} + A^{-1}B\mathfrak{C}_pB^T(A^T)^{-1} = C(0)$$

should hold. Multiplying this equation with  $A$  from the left and with  $A^T$  from the right, we obtain the equivalent equation

$$\Delta = AC(0)A^T - B\mathfrak{C}_pB^T. \tag{A6}$$

This is in accordance with Equation (6) that implies

$$\begin{aligned} & \mathbb{E}(\mathbf{A}\mathbf{X}_t + \mathbf{B}_1\mathbf{X}_{t-1} + \dots + \mathbf{B}_p\mathbf{X}_{t-p})(\mathbf{A}\mathbf{X}_t + \mathbf{B}_1\mathbf{X}_{t-1} + \dots + \mathbf{B}_p\mathbf{X}_{t-p})^T \\ & = \mathbf{A}\mathbf{C}(0)\mathbf{A}^T + \mathbf{A}\mathbf{C}^T(1, \dots, p)\mathbf{B}^T + \mathbf{B}\mathbf{C}(1, \dots, p)\mathbf{A}^T + \mathbf{B}\mathbf{C}_p\mathbf{B}^T = \mathbf{\Delta}. \end{aligned}$$

Combining this with Equation (A5), we have

$$\begin{aligned} \mathbf{\Delta} & = \mathbf{A}\mathbf{C}(0)\mathbf{A}^T + \mathbf{A}\mathbf{C}^T(1, \dots, p)\mathbf{B}^T + \mathbf{B}\mathbf{C}(1, \dots, p)\mathbf{A}^T + \mathbf{B}\mathbf{C}_p\mathbf{B}^T \\ & = \mathbf{A}\mathbf{C}(0)\mathbf{A}^T - \mathbf{A}\mathbf{C}^T(1, \dots, p)\mathbf{C}_p^{-1}\mathbf{C}(1, \dots, p)\mathbf{A}^T - \mathbf{A}\mathbf{C}^T(1, \dots, p)\mathbf{C}_p^{-1}\mathbf{C}(1, \dots, p)\mathbf{A}^T \\ & \quad + \mathbf{A}\mathbf{C}^T(1, \dots, p)\mathbf{C}_p^{-1}\mathbf{C}_p\mathbf{C}_p^{-1}\mathbf{C}(1, \dots, p)\mathbf{A}^T \\ & = \mathbf{A}\mathbf{C}(0)\mathbf{A}^T - \mathbf{A}\mathbf{C}^T(1, \dots, p)\mathbf{C}_p^{-1}\mathbf{C}(1, \dots, p)\mathbf{A}^T = \mathbf{A}\mathbf{C}(0)\mathbf{A}^T - \mathbf{B}\mathbf{C}_p\mathbf{B}^T, \end{aligned}$$

which also satisfies (A6).

Summarizing, we have proved that, under the model equations,  $(\mathbf{L}\mathbf{D}\mathbf{L}^T)^{-1} = \mathbf{C}_{p+1}$ , or equivalently,  $\mathbf{L}\mathbf{D}\mathbf{L}^T = \mathbf{K}$  indeed holds. In view of the uniqueness of the block LDL decomposition (under positive definiteness of the involved matrices), this finishes the proof.

Appendix A.4. Algorithm for the Block LDL Decomposition of Appendix A.3

Again, the protocol of the block Cholesky decomposition is applied to the involved matrices in block partitioned form. Here,

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \ell_{21} & 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ \ell_{31} & \ell_{32} & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & 0 & \dots & 0 \\ \ell_{d1} & \ell_{d2} & \dots & \ell_{d,d-1} & 1 & 0 & \dots & 0 \\ \ell_{d+1,1} & \ell_{d+1,2} & \vdots & \ell_{d+1,d-1} & \ell_{d+1,d} & \mathbf{I}_{pd \times pd} & & \end{pmatrix},$$

where the  $(p + 1)d \times (p + 1)d$  lower triangular matrix  $\mathbf{L}$  is also lower triangular with respect to its blocks which are partly scalars, partly vectors, and partly matrices as follows:

$$\ell_{ij} = \begin{cases} a_{ji}, & j = 1, \dots, d - 1; \quad i = j + 1, \dots, d; \\ 1 & i = j = 1, \dots, d; \\ 0 & i = 1, \dots, d; \quad j = i + 1, \dots, (p + 1)d. \end{cases}$$

Furthermore, the vectors  $\ell_{d+1,j}$  are  $pd \times 1$  for  $j = 1, \dots, d$ , and comprise the column vectors of the  $pd \times d$  matrix  $\mathbf{B}^T$ . The matrix in the bottom right block is the  $pd \times pd$  identity, and above it, the zero entries can be arranged into the  $d \times pd$  zero matrix.

The  $(p + 1)d \times (p + 1)d$  block-diagonal matrix  $\mathbf{D}$  in partitioned form is

$$\mathbf{D} = \begin{pmatrix} \delta_1^{-1} & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \delta_2^{-1} & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \delta_3^{-1} & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \mathbf{0} & \dots & \mathbf{0} \\ 0 & 0 & 0 & 0 & \delta_d^{-1} & 0 & \dots & 0 \\ \mathbf{0} & \mathbf{0} & \vdots & \mathbf{0} & \mathbf{0} & \mathbf{C}_p^{-1} & & \end{pmatrix},$$

where the  $pd \times 1$  vectors  $\mathbf{0}$  comprise  $\mathbf{O}_{pd \times d}$  in the left bottom, and the matrix  $\mathbf{C}_p^{-1}$  stands in the right bottom block. With multiplication rules of block matrices and their inverses, the recursion of the block LDL decomposition goes on as follows:

- Outer cycle (column-wise). For  $j = 1, \dots, d$ :  $\delta_j^{-1} = k_{jj} - \sum_{h=1}^{j-1} \ell_{jh}\delta_h^{-1}\ell_{jh}$  (with the reservation that  $\delta_1^{-1} = k_{11}$ );

- Inner cycle (row-wise). For  $i = j + 1, \dots, d$ :

$$\ell_{ij} = \left( k_{ij} - \sum_{h=1}^{j-1} \ell_{ih} \delta_h^{-1} \ell_{jh} \right) \delta_j$$

and

$$\ell_{d+1,j} = \left( \mathbf{k}_{d+1,j} - \sum_{h=1}^{j-1} \ell_{d+1,h} \delta_h^{-1} \ell_{jh} \right) \delta_j$$

(with the reservation that, in the  $j = 1$  case, the summand is zero), where  $\mathbf{k}_{d+1,j}$  for  $j = 1, \dots, d$  are  $pd \times 1$  vectors in the bottom left block of  $\mathbf{K}$ .

The recursion ends in one run.

The above decomposition is again a nested one, so for the first  $d$  rows of  $\mathbf{L}$ , only its previous rows or preceding entries in the same row enter into the calculation, as if we performed the usual LDL decomposition of  $\mathbf{K}$ . Therefore,  $\ell_{ij} = a_{ji}$  for  $j = 1, \dots, d - 1$ ,  $i = j + 1, \dots, d$  that are the negatives of the partial regression coefficients akin to those offered by the standard LDL decomposition  $\mathbf{K} = \tilde{\mathbf{L}} \tilde{\mathbf{D}} \tilde{\mathbf{L}}^T$ , so the first  $d$  rows of  $\tilde{\mathbf{L}}$  and  $\mathbf{L}$  are the same, and the first  $d$  rows of  $\tilde{\mathbf{D}}$  and  $\mathbf{D}$  are the same too. When the process terminates, we consider the blocks “en block” and obtain the  $pd \times d$  matrix  $\mathbf{B}^T = (\ell_{d+1,1}, \dots, \ell_{d+1,d})$ .

### Appendix B. Pseudocodes

In practice, the  $n \times d$  data matrix  $\mathbf{D} = (\mathbf{X}_1, \dots, \mathbf{X}_n)^T$  is given, where  $\mathbf{X}_t$  is a serially correlated sample of the underlying  $d$ -dimensional stacked random vector  $\mathbf{X} = (X_1, \dots, X_d)^T$  at time  $t \in \{1, \dots, n\}$ ,  $n > d$ . To construct a CVAR model, the first step is to construct an undirected graph  $G$  and a causal ordering of its nodes (the  $d$  observed variables). The algorithm below is a general procedure for this step. Notice that we only consider triangulated (thus decomposable) graphs in this section.<sup>1</sup>

Algorithm A1 will fail if the specified threshold  $r^*$  does not lead to a triangulated graph. Therefore, it is recommended that users manually inspect the initial partial correlations (step 3) and repeat the graph construction step (step 4) with various reasonable thresholds. If there is an expert’s advice on the causal structure (in the form of a causal ordering or a junction tree) of the variables in a dataset, the users may also skip Algorithm A1 and build a CVAR model directly using the following algorithms.

---

#### Algorithm A1: Constructing an undirected graph and a causal ordering of variables

---

**Input** :  $\mathbf{D}$ ,  $n \times d$  data matrix  
 $p$ , order of the CVAR model  
 $r^*$ , threshold for the partial correlation statistical test

**Output**: undirected graph  $G$  and its perfect ordering

- 1 Compute the block Toeplitz matrix  $\mathfrak{C}_{p+1}$  as in Equation (7), by using the autocovariance matrices  $C(h)$  for  $h = 0, 1, \dots, p$ .
- 2 Compute the concentration matrix  $\mathbf{K} = \mathfrak{C}_{p+1}^{-1} = (\sigma^{ij})$ .
- 3 Compute the partial correlation coefficient  $r_{ij}$  for  $X_i$  and  $X_j$  conditioned on all other variables up to lag  $p$ . By Proposition 1,

$$r_{ij} = \frac{-\sigma^{ij}}{\sqrt{\sigma^{ii}\sigma^{jj}}} \quad (1 \leq i < j \leq d).$$

- 4 Construct an undirected graph  $G = (V, E)$  based on the partial correlations such that  $V = \{1, \dots, d\}$  and  $E = \{(i, j) : i < j, r_{ij} \geq r^*\}$ .
  - 5 If  $G$  is triangulated, proceed to the next step; otherwise, terminate with an appropriate warning (then, choose another  $r^*$ ).
  - 6 Apply MCS (Maximal Cardinality Search) on  $G$  to obtain a perfect elimination ordering (see Algorithm 9.3 of Koller and Friedman (2009)).
-

Last but not least, to find the optimal order  $p$  for the CVAR model (i.e., order selection), we recommend repeating Algorithm A2 or A3 for different  $ps$  (e.g., for  $p = 1, \dots, 10$ ) and then comparing various information criteria of the resulting models (as illustrated in Section 4).

---

**Algorithm A2: Constructing an unrestricted CVAR model**


---

**Input** :  $D$ :  $n \times d$  data matrix or the existing  $K$  from Algorithm A1  
 $p$ : order of the CVAR model  
 $(i_1, \dots, i_d)$ , causal ordering of the  $d$  observed variables.

**Output**: parameter matrices  $A, B_1, \dots, B_p$

- 1 Reorder the columns of  $D$  according to the causal ordering.
  - 2 Compute the concentration matrix  $K$  for the reordered  $D$  (or, reorder the rows and columns of the existing  $K$  from Algorithm A1).
  - 3 Run Appendix A.4 with  $(K, p)$  to obtain the parameter matrices  $A, B_1, \dots, B_p$  for the unrestricted CVAR( $p$ ) model.
- 

---

**Algorithm A3: Constructing a restricted CVAR model**


---

**Input** :  $D$ :  $n \times d$  data matrix  
 $p$ : order of the CVAR model  
 $G$ , (undirected) chordal graph for observed variables  
 $(i_1, \dots, i_d)$ , causal ordering of observed variables.

**Output**: parameter matrices  $\hat{A}, \hat{B}_1, \dots, \hat{B}_p$

- 1 Re-label the nodes in  $G$  according to the causal ordering.
  - 2 Build a JT (junction tree) based on  $G$  and the causal (i.e., perfect elimination) ordering using a JT algorithm (e.g., `NetworkX.junction_tree(G)` in Hagberg et al. (2008)).
  - 3 Apply covariance selection (as in Equation (13)) to obtain a re-estimated concentration matrix  $\hat{K}$  using the JT from step 2.
  - 4 Run Appendix A.4 with  $(\hat{K}, p)$  to obtain the parameter matrices  $\hat{A}, \hat{B}_1, \dots, \hat{B}_p$  for the restricted CVAR( $p$ ) model.
- 

## Notes

- <sup>1</sup> Please see the main text for suggestions on graphs that are not triangulated when moralization and running the IPS algorithm is needed. The default  $r^*$  threshold is usually set according to a significance level (e.g.,  $\alpha = 0.05$ ) for the partial correlation test. This can be changed based on the sample size and the effect size.

## References

- Abdelkhalik, Fatma, and Marianna Bolla. 2020. Application of Structural Equation Modeling to Infant Mortality Rate in Egypt. In *Demography of Population Health, Aging and Health Expenditures*. Edited by Christos H. Skiadas and Charilaos Skiadas. Cham: Springer, pp. 89–99.
- Akbilgic, Oguz, Hamparsum Bozdogan, and M. Erdal Balaban. 2014. A Novel Hybrid RBF Neural Networks Model as a Forecaster. *Statistics and Computing* 24: 365–75. [CrossRef]
- Bazinas, Vassilios, and Bent Nielsen. 2022. Causal Transmission in Reduced-Form Models. *Econometrics* 10: 14. [CrossRef]
- Bolla, Marianna, Fatma Abdelkhalik, and Máté Baranyi. 2019. Graphical models, regression graphs, and recursive linear regression in a unified way. *Acta Scientiarum Mathematicarum (Szeged)* 85: 9–57. [CrossRef]
- Bolla, Marianna, and Tamás Szabados. 2021. *Multidimensional Stationary Time Series: Dimension Reduction and Prediction*. New York: CRC Press, Taylor and Francis Group.
- Box, George EP, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung. 2015. *Time series Analysis: Forecasting and Control*. New York: Wiley.
- Brillinger, David R. 1996. Remarks concerning graphical models for time series and point processes. *Revista de Econometria* 16: 1–23. [CrossRef]
- Brockwell, Peter J., and Richard A. Davis. 1991. *Time Series: Theory and Methods*. Berlin/Heidelberg: Springer.
- Deistler, Manfred, and Wolfgang Scherrer. 2019. Vector Autoregressive Moving Average Models. In *Handbook of Statistics*. Berlin/Heidelberg: Springer, vol. 41.
- Deistler, Manfred, and Wolfgang Scherrer. 2022. *Time Series Models*. Cham: Springer Nature.

- Dempster, Arthur P. 1972. Covariance selection. *Biometrics* 28: 157–75. [[CrossRef](#)]
- Eichler, Michael. 2006. Graphical modelling of dynamic relationships in multivariate time series. In *Handbook of Time Series Analysis*. Edited by Schelter Björn, Winterhalder Matthias and Timmer Jens. Berlin/Heidelberg: Wiley-VCH Berlin.
- Eichler, Michael. 2012. Graphical modelling of multivariate time series. *Probability Theory Related Fields* 153: 233–68. [[CrossRef](#)]
- Geweke, John. 1984. Inference and causality in economic time series models. In *Handbook of Econometrics*. Amsterdam: Elsevier, vol. 2.
- Golub, Gene H., and Charles F. Van Loan. 2012. *Matrix Computations*. Baltimore: JHU Press.
- Granger, Clive W. J. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37: 424–38. [[CrossRef](#)]
- Haavelmo, Trygve. 1943. The statistical implications of a system of simultaneous equations. *Econometrica* 11: 1–12. [[CrossRef](#)]
- Hagberg, Aric, Pieter Swart, and Daniel S Chult. 2008. Exploring network structure, dynamics, and function using NetworkX. Paper presented at the 7th Python in Science Conference (SciPy2008), Pasadena, CA, USA, August 19–24; Edited by Varoquaux Gäel, Vaught Travis and Millman Jarrod. Los Alamos: Los Alamos National Lab (LANL), pp. 11–15.
- Jöreskog, Karl G. 1977. Structural equation models in the social sciences. Specification, estimation and testing. In *Applications of Statistics*. Edited by Pathak R. Krishnaiah. Amsterdam: North-Holland Publishing Co., pp. 265–87.
- Keating, John W. 1996. Structural information is recursive VAR orderings. *Journal of Econometric Dynamics and Control* 20: 1557–80. [[CrossRef](#)]
- Kiiveri, Harri, Terry P. Speed, and John B. Carlin. 1984. Recursive causal models. *Journal of the Australian Mathematical Society* 36: 30–52. [[CrossRef](#)]
- Kilian, Lutz, and Helmut Lütkepohl. 2017. *Structural Vector Autoregressive Analysis*. Cambridge: Cambridge University Press.
- Koller, Daphne, and Nir Friedman. 2009. *Probabilistic Graphical Models. Principles and Techniques*. Cambridge: MIT Press.
- Lauritzen, Steffen L. 2004. *Graphical Models*. Oxford Statistical Science Series; Oxford: Clarendon Press, Oxford University Press, reprint with corr. edition.
- Lütkepohl, Helmut. 2005. *New Introduction to Multiple Time Series Analysis*. Berlin/Heidelberg: Springer.
- Nocedal, Jorge, and Stephen J. Wright. 1999. *Numerical Optimization*. Berlin/Heidelberg: Springer.
- Rao, Calyampudi Radhakrishna. 1973. *Linear Statistical Inference and its Applications*. New York: Wiley.
- Rózsa, Pál. 1991. *Linear Algebra and Its Applications*. Budapest: Műszaki Kiadó. (In Hungarian)
- Sims, Christopher A. 1980. Macroeconomics and reality. *Econometrica* 48: 1–48. [[CrossRef](#)]
- Tarjan, Robert E., and Mihalis Yannakakis. 1984. Simple Linear-Time Algorithms to Test Chordality of Graphs, Test Acyclicity of Hypergraphs, and Selectively Reduce Acyclic Hypergraphs. *SIAM Journal on Computing* 13: 566–79. [[CrossRef](#)]
- Wainwright, Martin J., and Michael I. Jordan. 2008. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning* 1: 1–305. [[CrossRef](#)]
- Wainwright, Martin J. 2015. Graphical Models and Message-Passing Algorithms: Some Introductory Lectures. In *Mathematical Foundations of Complex Networked Information Systems*. Lecture Notes in Mathematics 2141. Edited by Fagnani Fabio, Sophie M. Fosson and Ravazzi Chiara. Cham: Springer.
- Wermuth, Nanny. 1980. Recursive equations, covariance selection, and path analysis. *Journal of the American Statistical Association* 75: 963–72. [[CrossRef](#)]
- Wiener, Norbert. 1956. The theory of prediction. In *Modern Mathematics for Engineers*. Edited by E. F. Beckenback. New York: McGraw–Hill.
- Wold, Herman O. A. 1960. A generalization of causal chain models. *Econometrica* 28: 444–63. [[CrossRef](#)]
- Wold, Herman O. A. 1985. Partial least squares. In *Encyclopedia of Statistical Sciences*. Edited by Samuel Kotz, Norman L. Johnson and C. R. Read. New York: Wiley.
- Wright, Sewall. 1934. The method of path coefficients. *The Annals of Mathematical Statistics* 5: 161–215. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.