Causal Vector Autoregression

Marianna Bolla
Math. Inst., Budapest University of Technology and Economics
marib@math.bme.hu

High Dimensional Time Series Workshop, Vienna

June 9, 2022.

Preliminaries

- Haavelmo, T., The statistical implications of a system of simultaneos equations, *Econometrica* **11** (1943). Nobel prize.
- Wiener, N., The theory of prediction. *Modern Mathematics for Engineers*, McGraw–Hill, New York (1956).
- Wold, H. O. A., A generalizarion of causal chain models, *Econometrica* **28** (1960).
- Granger, C., Investigating causal relations by econometric models and cross-spectral methods, *Econometrica* **37**(1969).
- Wermuth, N., Recursive equations, covariance selection, and path analysis, *J. Amer. Stat. Assoc.* **75** (1980).
- Wold, H. O. A., Partial least squares (PLS). In *Encyclopedia of Statistical Sciences*, Wiley (1985).
- Kilian, L., Lütkepohl, H., *Structural Vector Autoregressive Analysis*. Cambridge University Press (2017).
- Deistler, M., Scherrer, W., *Time series models*, Springer

The unrestricted causal VAR(0) model

Recursive linear equations (Wermuth):
$\mathbf{X} \sim \mathcal{N}_d(\mathbf{0}, \mathbf{\Sigma})$ is d-dimensional Gaussian random vector.

$$\mathbf{AX} = \mathbf{U} \quad \text{with} \quad \mathbf{U} = (U_1, \ldots, U_d)^T \sim \mathcal{N}_d(\mathbf{0}, \mathbf{\Delta}),$$

where $\mathbf{A}$ is a $d \times d$ upper triangular matrix with 1s along its main diagonal, otherwise it contains the negatives of the partial regression coefficients a_{ji}'s, when X_j is the target of a multivariate linear regression with predictors $\{X_i : i > j\}$.
a_{ji}: path coefficient, statistical tests for its significance.
$\mathbf{\Delta} = \mathrm{diag}(\delta_1, \ldots, \delta_d)$: diagonal matrix with positive diagonal entries, covariance matrix of the error term $\mathbf{U}$ (residual variances).
Taking the covariance matrix, we get

$$\mathbb{E}[(\mathbf{AX})(\mathbf{AX})^T] = \mathbf{A\Sigma A}^T = \mathbf{\Delta}.$$

By the standard LDL (variant of the simple Cholesky) decomposition of $\mathbf{\Sigma}^{-1}$ (Bolla et al., Acta Sci. Math. (Szeged) 85, 2019):

$$\mathbf{\Sigma}^{-1} = \mathbf{L}\mathbf{\Delta}^{-1}\mathbf{L}^T.$$

This decomposition of the positive definite matrix $\mathbf{\Sigma}^{-1}$ is unique, where $\mathbf{L} = \mathbf{A}^T$ is lower triangular of entries 1s along its main diagonal and $\mathbf{\Delta}^{-1}$ is a diagonal matrix of entries all positive along its main diagonal. By uniqueness, this $\mathbf{A} = \mathbf{L}^T$ and $\mathbf{\Delta}$ give the solution to the original problem.
At this point, the ordering of the jointly Gaussian variables is not relevant, since in any recursive ordering of them (encoded in $\mathbf{A}$) a Gaussian directed graphical model (in other words, a Gaussian Bayesian network) can be constructed, where every variable is regressed linearly with the higher index ones.

The unrestricted causal VAR(1) model

$\{\mathbf{X}_t\}$: d-dimensional, weakly stationary process with real valued components and covariance matrix function $\mathbf{C}(h)$,
$h = 0, \pm 1, \pm 2, \ldots$; $\mathbf{C}(-h) = \mathbf{C}^T(h)$. $\mathbb{E}\mathbf{X}_t = \mathbf{0}$,
$\mathbf{C}(h) = \mathbb{E}\mathbf{X}_t\mathbf{X}_{t+h}^T$ does not depend on t.
Recursive VAR(1) model:

$$\mathbf{A}\mathbf{X}_t + \mathbf{B}\mathbf{X}_{t-1} = \mathbf{U}_t, \quad t = 1, 2, \ldots,$$

where the white noise random vector $\mathbf{U}_t$ is uncorrelated with $\mathbf{X}_{t-1}$, has zero expectation and covariance matrix $\mathbf{\Delta} = \mathrm{diag}(\delta_1, \ldots, \delta_d)$.
The covariance matrix of $(\mathbf{X}_t^T, \mathbf{X}_{t-1}^T)^T$:

$$\mathfrak{C}_2 = \begin{pmatrix} \mathbf{C}(0) & \mathbf{C}^T(1) \\ \mathbf{C}(1) & \mathbf{C}(0) \end{pmatrix}.$$

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) 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

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

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 $\mathbf{X}_t$ conditioned on $\mathbf{X}_{t-1}$; by weak stationarity, it does not depend on t either, therefore it is denoted by $\mathbf{C}(1|0)$. Also, $\mathfrak{C}_2$ is positive definite if and only if both $\mathbf{C}(0)$ and $\mathbf{C}(1|0)$ are positive definite.

Theorem

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

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

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 the model equation can be retrieved from them.*

Algorithm (recursion)

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

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

and

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

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

$\mathbf{B}^T := (\mathbf{l}_{d+1,1}, \ldots, \mathbf{l}_{d+1,d})$ is $d \times d$ as $\mathbf{l}_{d+1,j}$ is $d \times 1$.

Remark: nested structure

It is obvious that the above decomposition has a nested structure, 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 standard LDL decomposition of $\mathbf{K}$.
Therefore, $l_{ij} = a_{ji}$ for $j = 1, \ldots, d-1$, $i = j+1, \ldots, d$ that are the partial regression coefficients akin to those offered by the standard LDL decomposition $\mathbf{K} = \tilde{\mathbf{L}}\tilde{\mathbf{D}}\tilde{\mathbf{L}}^T$;

Consequently, 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 after finding the first d rows of $\mathbf{L}$, we consider the blocks „en block" and get the matrix
$\mathbf{B}^T = (\mathbf{I}_{d+1,1}, \ldots, \mathbf{I}_{d+1,d})$.

The restricted causal VAR(1) model

Assume that we have a causal ordering of the coordinates $X_1, \ldots, X_d$ of $\mathbf{X}$ such that X_i can be the cause of X_j whenever $j < i$. We can think of X_is as the nodes of a graph in a directed graphical model (Bayesian network) and their labeling corresponds to a topological ordering of the nodes of the underlying directed acyclic graph (DAG).

For example, when asset prices or log-returns of different assets or currencies (on the same day) influence each other in a certain (recursive) order, contemporaneously.

In the restricted cases, only certain asset prices influence some others on a DAG, 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 be zeros.

Decomposable graph, triangulated (chordal), junction tree, RZP (equivalent)

Definition

Let $\mathbf{M}$ be a symmetric or an upper triangular matrix of real entries. We say that $\mathbf{M}$ has a **reducible zero pattern (RZP)** if $m_{ji} = 0$ $(j < i)$ implies that for each $h = 1, \ldots, j-1$: either $m_{hj} = 0$ or $m_{hi} = 0$ holds (or both hold).

In the adjacency matrix of a DAG, an RZP is present if and only if there is no sink V configuration in the topological ordering of the DAG. Under sink V configuration a triplet $j \to h \leftarrow i$ is understood, where i is not connected to j $(h < j < i)$.
Indeed, in this case the DAG has a triplet $h < j < i$ with $m_{hi} \neq 0$, $m_{hj} \neq 0$, but $m_{ij} = 0$, in contrast to the Definition.

If we also require that the so constructed DAG be Markov equivalent to its undirected skeleton, then the DAG must not contain sink V configuration. In this case, the positions of the zero entries of $\mathbf{\Sigma}^{-1}$ are identical to the positions of the zero entries of $\mathbf{A}$ in the VAR(0) model.
If such an ordering exists, it gives a special DAG which defines a decomposable graph, the skeleton also triangulated, labeling compatible with the so-called maximal cardinality search (MCS) ordering, defines a possible (not necessarily unique) causal ordering of the variables, and junction tree (JT) structure of the cliques (maximal complete subgraphs).

Junction Tree

The cliques $C_1, \ldots, C_k$ can be numbered so that C_j is the disjoint union of R_j (residual) and of S_j (separator) with the following property: there is an $i^* \in \{1, \ldots, j-1\}$ 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, where $S_1 = \emptyset$ and $R_1 = C_1$. Furthermore, if such an ordering is possible, a version can be found in which any prescribed clique is the first one. Also equivalently, any path between C_i and C_j $(i \neq j)$ contains $C_i \cap C_j$. Note that the junction tree is indeed a tree with nodes $C_1, \ldots, C_k$ and one less edges, that are the separators $S_2, \ldots, S_k$.

Also note, that if an undirected graph is triangulated, then in a convenient labeling of its nodes, its adjacency matrix has an RZP. This so-called perfect ordering is obtainable, e.g. by the maximal cardinality search (MCS) algorithm. In this ordering of the nodes, a DAG can be constructed, the adjacency matrix of which has the same RZP. In this way, a decomposable undirected graph can be made directed.

At the beginning, no restrictions for the upper-diagonal entries of $\mathbf{A}$ are made. In practice, we have a sample and all the autocovariance matrices are estimated, consequently the resulting $\mathbf{A}, \mathbf{B}$ matrices are also estimated. Usually a statistical hypothesis testing advances this procedure, during which it can be found that certain partial correlations (closely related to the entries of $\mathbf{K}$) do not significantly differ from zero.

Covariance Selection

Then we naturally want to introduce zeros for the corresponding entries of **A**. The covariance selection (Dempster) is improved:

Proposition

The upper triangular matrix **A** *of the VAR(1) model has an RZP if and only if the upper left* $d \times d$ *block of* $\mathbf{K} = \mathfrak{C}_2^{-1}$ *has an RZP. Moreover, the zero entries of* **A** *are exactly in the same positions as the zero entries of the upper diagonal part of the upper left block of* **K**.

The proof follows from Equation (1).
Fixing the zero entries in the left upper block of **K**, we re-estimate the matrix $\mathfrak{C}_2$ that means a product-moment estimate (MLE in the RZP model).
Product moments are calculated only for the cliques and separators that decreases computational complexity.

Higher order causal VAR models

The above model is further generalized to the VAR(p) model ($p \geq 1$):

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

where the white noise term $\mathbf{U}_t$ is uncorrelated with $\mathbf{X}_{t-1}, \ldots, \mathbf{X}_{t-p}$, it has zero expectation and covariance matrix $\mathbf{\Delta} = \mathrm{diag}(\delta_1, \ldots, \delta_d)$. $\mathbf{A}$ is $d \times d$ upper triangular matrix with 1s along its main diagonal; whereas, $\mathbf{B}_1, \ldots, \mathbf{B}_p$ are $d \times d$ matrices. Here we have to perform the block Cholesky decomposition of the inverse covariance matrix (concentration matrix) $\mathbf{K}$ of $\mathfrak{C}_{p+1}$, covariance matrix of the stacked vector $(\mathbf{X}_t^T, \mathbf{X}_{t-1}^T, \ldots, \mathbf{X}_{t-p}^T)^T$.

Estimation of the parameter matrices

Theorem

The parameter matrices $\mathbf{A}, \mathbf{B}_1, \ldots, \mathbf{B}_p$ *and* $\mathbf{\Delta}$ *of the model equation can be obtained by the block LDL decomposition of the (positive definite) concentration matrix* $\mathbf{K}$*. If* $\mathbf{K} = \mathbf{L}\mathbf{D}\mathbf{L}^T$ *is this (unique) decomposition with block-triangular matrix* $\mathbf{L}$ *and block-diagonal matrix* $\mathbf{D}$*, then they have the form*

$$\mathbf{L} = \begin{pmatrix} \mathbf{A}^T & \mathbf{O}_{d\times pd} \\ \mathbf{B}^T & \mathbf{I}_{pd\times pd} \end{pmatrix}, \qquad \mathbf{D} = \begin{pmatrix} \mathbf{\Delta}^{-1} & \mathbf{O}_{d\times pd} \\ \mathbf{O}_{pd\times d} & \mathfrak{C}_p^{-1} \end{pmatrix},$$

where the $d \times d$ *upper triangular matrix* $\mathbf{A}$ *with 1s along its main diagonal, the* $d \times pd$ *matrix* $\mathbf{B} = (\mathbf{B}_1 \ldots \mathbf{B}_p)$ *(transpose of* $\mathbf{B}^T$*, partitioned into blocks) and the diagonal matrix* $\mathbf{\Delta}$ *of the model equation can be retrieved from them.*

Restricted causal VAR(p) models

Restricted cases can be treated similarly as in the $p = 1$ case. Here too, the existence of an RZP in the DAG on p nodes is equivalent to the existence of an RZP in the left upper $d \times d$ corner of the concentration matrix $\mathbf{K} = \mathfrak{C}_{p+1}^{-1}$. The selection of p is an issue in the usual (not causal) VAR models too. However, this problem needs statistical hypothesis testing, akin to the test of the partial correlations.

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.

Covariance selection can be done similarly, but here zero entries of the left upper $d \times d$ block of $\mathfrak{C}_{p+1}^{-1}$ provide the zero entries of $\mathbf{A}$.

Conclusions

- The model equation has the equivalent reduced form:

  $$\mathbf{X}_t+\mathbf{A}^{-1}\mathbf{B}_1\mathbf{X}_{t-1}+\cdots+\mathbf{A}^{-1}\mathbf{B}_p\mathbf{X}_{t-p}=\mathbf{V}_t, \quad t=p+1,p+2,\ldots,$$

  where $\mathbf{V}_t$ is white noise with covariance matrix $\mathbf{\Sigma}=\mathbf{A}^{-1}\mathbf{\Delta}\mathbf{A}^{-1\,T}=(\mathbf{A}^T\mathbf{\Delta}^{-1}\mathbf{A})^{-1}$, usually not diagonal, but positive definite. So we have a solution for the Yule–Walker equations, an alternative of the Durbin–Levinson algorithm (essentially encodes the block Cholesky decomposition).
- The process $\mathbf{U}_t=\mathbf{A}\mathbf{V}_t$ of structural shocks is obtained from the process $\mathbf{V}_t$ of innovations in the reduced form and have an econometric interpretation. The structural shocks represent unanticipated changes in the observed econometric variables and they are mutually uncorrelated. In fact, they are the orthogonalized innovations, but here the structure of the matrix $\mathbf{A}$ plays an important role.

Application to financial data (Akbilgic, O. et al.)

Daily log-returns of 8 different asset prices, spanning 534 trading days (nearly stationary and Gaussian).
A DAG was constructed by making the undirected graph on 8 nodes directed. The undirected graph was made 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.

RZP

Since the graph was triangulated, with the MCS algorithm, we were able to 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),

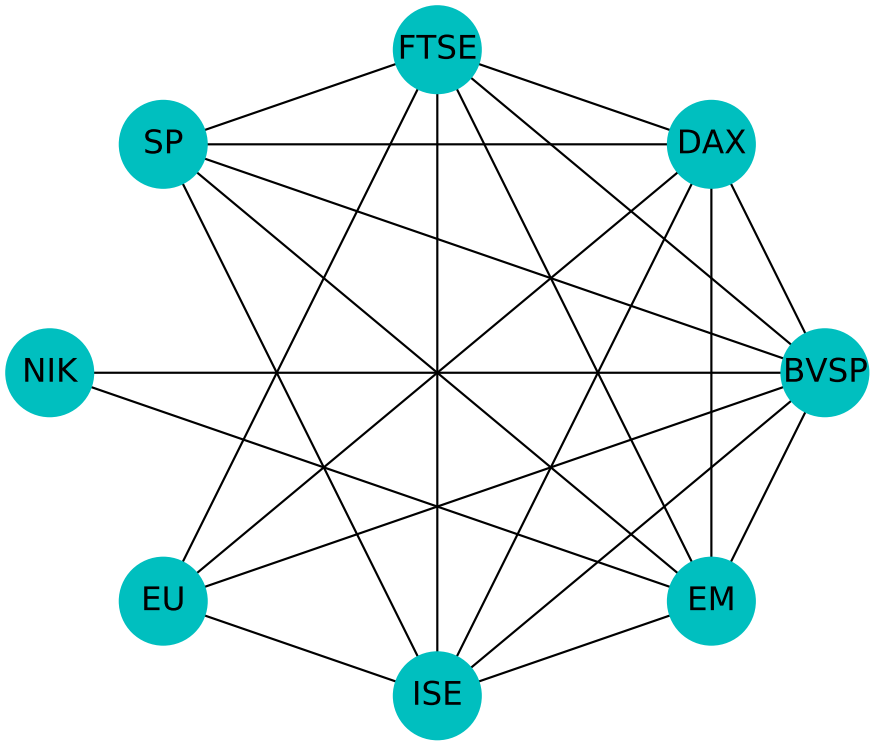
5 : BVSP (stock market return index of Brazil),

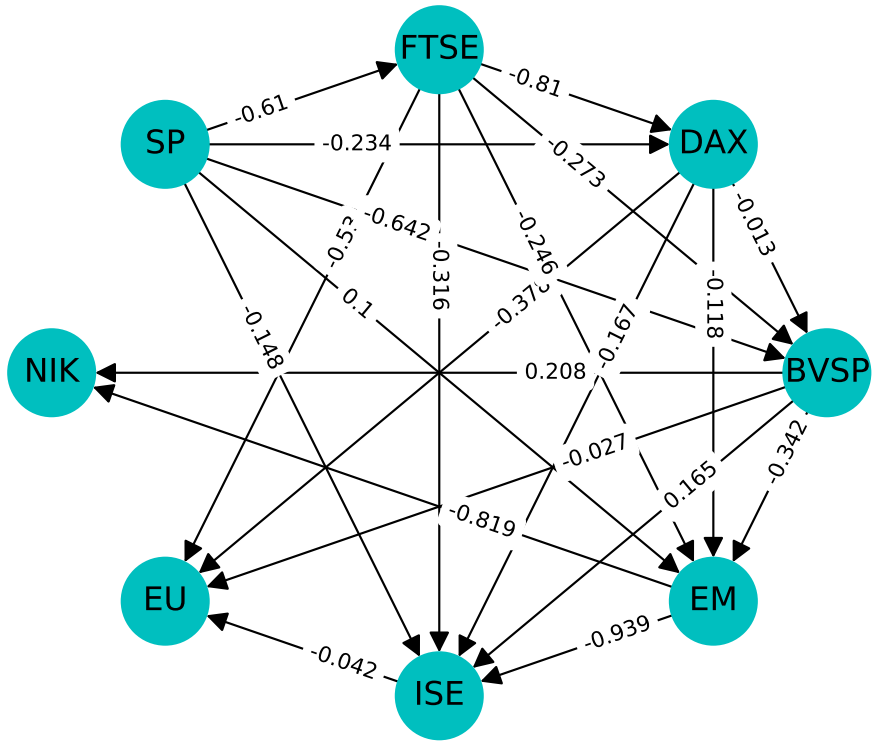
6 : DAX (stock market return index of Germany),

7 : FTSE (stock market return index of UK),

8 : SP (Standard & poor's 500 return index).

Undirected and directed graphs (path coefficients)





The unrestricted VAR(p) model, $p = 1, 2, 3, 4, 5$

First we run the unrestricted VAR(p) algorithm with $p = 1, 2, 3, 4, 5$ and found that the $\mathbf{A}$ matrices do not change much with increasing p, akin to $\mathbf{B}_1$. The $\mathbf{B}_2, .., \mathbf{B}_5$ matrices have relatively „small" entries.
Consequently, contemporaneous effects and one-day lags are the most important.

The restricted causal VAR(1) model

We want to introduce structural zeros into the matrix **A**. Now the matrix $C^{-1}(1|0)$, the left upper 8×8 corner of $\mathfrak{C}_2^{-1}$ is used for covariance selection.
The JT structure has the following cliques and separators:

$$C_1 = \{\text{BVSP,DAX,EM,FTSE,ISE,SP}\}$$
$$C_2 = \{\text{BVSP,DAX,EU,FTSE,ISE}\}$$
$$C_3 = \{\text{BVSP,EM,NIK}\}$$
$$S_2 = \{\text{BVSP,DAX,FTSE,ISE}\}$$
$$S_3 = \{\text{BVSP,EM}\},$$

where the parent clique of both C_2 and C_3 is C_1.

A matrix of the restricted causal VAR(1) model

	NIK	EU	ISE	EM	BVSP	DAX	FTSE	SP
NIK	1	0	0	-0.819	0.208	0	0	0
EU	0	1	-0.042	0	-0.027	-0.378	-0.529	0
ISE	0	0	1	-0.939	0.165	-0.168	-0.316	-0.148
EM	0	0	0	1	-0.342	-0.118	-0.246	0.099
BVSP	0	0	0	0	1	-0.013	-0.273	-0.642
DAX	0	0	0	0	0	1	-0.810	-0.234
FTSE	0	0	0	0	0	0	1	-0.610
SP	0	0	0	0	0	0	0	1

B matrix of the restricted causal VAR(1) model

	NIK_{-1}	EU_{-1}	ISE_{-1}	EM_{-1}	$BVSP_{-1}$	DAX_{-1}	$FTSE_{-1}$	SP_{-1}
NIK	0.181	-0.179	-0.086	0.084	0.074	-0.006	-0.115	-0.266
EU	-0.013	0.121	-0.005	0.0304	-0.013	-0.042	-0.097	0.000
ISE	0.068	0.281	-0.066	0.248	-0.294	-0.057	0.012	-0.147
EM	-0.002	-0.057	-0.017	0.107	-0.091	-0.095	0.089	-0.109
BVSP	-0.014	0.070	0.014	-0.104	0.139	-0.149	0.119	-0.083
DAX	-0.003	0.202	-0.034	-0.005	-0.035	-0.047	-0.067	-0.067
FTSE	0.029	-0.017	-0.011	0.042	-0.113	0.214	0.081	-0.264
SP	0.042	0.261	-0.026	0.012	-0.003	-0.071	-0.285	0.124