Causal Vector Autoregression and Dynamic Factor Analysis

Marianna Bolla
Rényi Institute of Mathematics
bolla@renyi.hu
Budapest

January 11, 2023.

- *Past: Regular and VARMA Processes*
- *Present: Causal VAR Models*
- *Future: Dynamic Factor Analysis*

References

- 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 (2022).

Regular and VARMA Processes

Bolla M., Szabados T., Multidimensional Stationary Time Series. Dimension Reduction and Prediction. CRC Press, Taylor & Francis Group, a Chapman & Hall book (2021).

Preliminaries: VARMA(p, q) model

$\{\mathbf{X}_t\}$ is d-dimensional, weakly stationary, full rank process of $\mathbf{0}$ expectation.

$$\mathbf{X}_t = \boldsymbol{\alpha}_1 \mathbf{X}_{t-1} + \cdots + \boldsymbol{\alpha}_p \mathbf{X}_{t-p} + \boldsymbol{\varepsilon}_t + \boldsymbol{\beta}_1 \boldsymbol{\varepsilon}_{t-1} + \ldots \boldsymbol{\beta}_q \boldsymbol{\varepsilon}_{t-q}$$

- $\{\boldsymbol{\varepsilon}_t\}$: white noise with covariance matrix $\boldsymbol{\Sigma} > 0$
- $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p$ and $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_q$: $d \times d$ real matrices
- $\boldsymbol{\alpha}(z) = \mathbf{I}_d - \boldsymbol{\alpha}_1 z - \cdots - \boldsymbol{\alpha}_p z^p$ *AR polynomial* that satisfies the stability condition: $|\boldsymbol{\alpha}(z)| \neq 0, \forall\ |z| \leq 1$
- $\boldsymbol{\beta}(z) = \mathbf{I}_d + \boldsymbol{\beta}_1 z + \cdots + \boldsymbol{\beta}_q z^q$ *MA polynomial* that satisfies the inverse stability (strict miniphase) condition: $|\boldsymbol{\beta}(z)| \neq 0, \forall\ |z| \leq 1$

Wold representation

VARMA(p,q) model in short: $\alpha(z)\mathbf{X}_t = \beta(z)\varepsilon_t$

$$\mathbf{X}_t = \alpha^{-1}(z)\beta(z)\varepsilon_t = \mathbf{k}(z)\varepsilon_t = \sum_{j=0}^{\infty} \mathbf{k}_j \varepsilon_{t-j} = \sum_{j=0}^{\infty} \tilde{\mathbf{k}}_j \tilde{\varepsilon}_{t-j} \qquad (1)$$

where $\mathbf{k}_0 = \mathbf{I}_d$, ε_t: innovation process,
$\mathbf{k}(z) = \sum_{j=0}^{\infty} \mathbf{k}_j z^j$: transfer function ($z \leq 1$).
$\tilde{\varepsilon}_t$: fundamental process, orthonormal white noise (its components span the same innovation subspace as the components of ε_t).
$\tilde{\mathbf{k}}_j$ and $\tilde{\varepsilon}_t$ are unique up to unitary transformation $\mathbf{Q}$.
VARMA(p,q) is a special regular process (with rational spectral density) this is why it has the Wold representation (1).
VAR(p): $\mathbf{X}_t = \alpha_1 \mathbf{X}_{t-1} + \cdots + \alpha_p \mathbf{X}_{t-p} + \varepsilon_t$, where the blue part is the best linear prediction of $\mathbf{X}_t$ by its infinite past (which is a p-lag long past now) and the red part is the orthogonal component of the projection (innovation).

General regular processes of rank $r \leq d$

Time domain:
Wold representation: $\mathbf{X}_t = \sum_{j=0}^{\infty} \mathbf{a}_j \boldsymbol{\eta}_{t-j} = \sum_{j=0}^{\infty} \mathbf{b}_j \boldsymbol{\xi}_{t-j}$
$\mathbf{a}_0 = \mathbf{I}_d$, $\mathbf{a}_j$: $d \times d$
$\{\boldsymbol{\eta}_t\} \sim WN(\boldsymbol{\Sigma})$: innovation process
$\mathbf{b}_j$: $d \times r$
$\{\boldsymbol{\xi}_t\} \sim WN(\mathbf{I}_r)$: fundamental process

The components of $\boldsymbol{\xi}_t$ span the same r-dimensional innovation subspace as the components of $\boldsymbol{\eta}_t$. $\boldsymbol{\xi}_t$ is unique up to $r \times r$ unitary transformation $\mathbf{Q}$, akin to the matrices $\mathbf{b}_j$s.

r = rank of the innovation subspaces = (constant) rank of the spectral density matrix

Spectral factors

Frequency domain:

$$\mathbf{f}(\omega) = \frac{1}{2\pi}\boldsymbol{\phi}(\omega)\boldsymbol{\phi}^*(\omega) \quad \omega \in [0, 2\pi]$$

where $\boldsymbol{\phi}(\omega)$: minimum phase spectral factor, it has the Wold coefficients $\mathbf{b}_j$ in its one-sided Fourier expansion (it is an analytic function, in a Hardy-2 space): $\boldsymbol{\phi}(\omega) = \sum_{j=0}^{\infty} \mathbf{b}_j e^{-i\omega j}$ (unique up to unitary transformation $\mathbf{Q}$).
$\boldsymbol{\phi}$ can be analytically extended to the unit disc:

$$\boldsymbol{\phi}(\omega) = \boldsymbol{\Phi}(e^{-i\omega}), \quad \boldsymbol{\Phi}(z) = \sum_{j=0}^{\infty} \mathbf{b}_j z^j, \quad |z| < 1.$$

$\boldsymbol{\Phi}(z)$: maximal analytic matrix ($\boldsymbol{\phi}$ and $\boldsymbol{\Phi}$ are $d \times r$ matrices).

Maximal analytic matrix

$$\mathbf{f}(\omega) = \frac{1}{2\pi}\tilde{\phi}(\omega)\tilde{\phi}^*(\omega), \quad \omega \in [0, 2\pi]$$

may give another admissible spectral factor of $\mathbf{f}$ with a one-sided Fourier series, which is not minimum phase.

$$\tilde{\phi}(\omega) = \sum_{j=0}^{\infty} \tilde{\mathbf{b}}(j) e^{-ij\omega}$$

with the sequence of matrices $\{\tilde{\mathbf{b}}(j)\}_{j=0}^{\infty}$ that is (entrywise) square summable.

A minimum phase spectral factor is not easy to find, however in the 1D case, there are constructions for it (see Lamperti).

An admissible spectral factor can be extended to the open unit disc as well:

$$\tilde{\mathbf{\Phi}}(e^{-i\omega}) := \tilde{\phi}(\omega), \quad \tilde{\mathbf{\Phi}}(z) := \sum_{j=0}^{\infty} \tilde{\mathbf{b}}(j) z^j. \quad |z| < 1.$$

A d-dimensional stationary time series $\{\mathbf{X}_t\}$ is regular if and only if it possesses any of the following properties.

- In the frequency domain: it has an absolutely continuous spectral measure with spectral density matrix of a.e. constant rank $r \leq d$ that has an admissible spectral factor.
- In the time domain: $H_t^-(\mathbf{X}) = H_t^-(\boldsymbol{\eta}) = H_t^-(\boldsymbol{\xi})$ for all $t \in \mathbb{Z}$, where $\{\boldsymbol{\eta}_t\}$ is the unnormalized and $\{\boldsymbol{\xi}_t\}$ is the normalized innovation process of $\{\mathbf{X}_t\}$. Further, in the representation $\mathbf{X}_t = \sum_{j=0}^{\infty} \mathbf{a}(j)\boldsymbol{\eta}_{t-j}$, $\mathbf{a}(0) = \mathbf{I}_d$.
- Also in the time domain: $\mathbf{X}_t$ can be written as a causal MA(∞) process

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \tilde{\mathbf{b}}(j)\tilde{\boldsymbol{\xi}}_{t-j}, \quad t \in \mathbb{Z}, \tag{2}$$

where $\{\tilde{\boldsymbol{\xi}}_t\}$ is a WN($\mathbf{I}_r$) sequence and $\sum_{j=0}^{\infty} \|\tilde{\mathbf{b}}(j)\|_F^2 < \infty$.

$$H_t^-(\mathbf{X}) \subset H_t^-(\tilde{\boldsymbol{\xi}}), \quad t \in \mathbb{Z}$$

and consequently, $H_{-\infty}(\mathbf{X}) = \cap_t H_t^-(\mathbf{X}) = \{\mathbf{0}\}$ too. So the process $\{\mathbf{X}_t\}$ has no remote past, and therefore, it is regular.
However, here $\mathbf{X}_t$ is expanded not with its own innovations, and the MA series is not capable for forecasting.

$$\boldsymbol{\Phi}(0)\,\boldsymbol{\Phi}^*(0) \geq \tilde{\boldsymbol{\Phi}}(0)\,\tilde{\boldsymbol{\Phi}}^*(0)$$

in the sense that the difference between the left and right sides is a positive semidefinite matrix.
This is why the matrix $\boldsymbol{\Phi}$ belonging to the minimum phase spectral factor ϕ is called *maximal analytic matrix*.

Inner and outer functions

Relation between $\mathbf{\Phi}$ and $\tilde{\mathbf{\Phi}}$: $\tilde{\mathbf{\Phi}}(z) = \mathbf{\Phi}(z)\mathbf{Q}(z)$, where the $r \times r$ $\mathbf{Q}(z)$ is analytic in the open unit disc D, belongs to the Hardy class $H^2_{r\times r}$, it is a.e. unitary on the unit circle, and $\mathbf{I}_r - \mathbf{Q}(0)\mathbf{Q}^*(0) \geq 0$.
$\mathbf{Q}(z)$: is an *inner matrix function*,
$\mathbf{\Phi}(z)$: is an outer function,
$\tilde{\mathbf{\Phi}}(z)$: is an admissible spectral factor.

The $d \times r$ matrix $\mathbf{\Phi}(z)$, corresponding to the minimum phase spectral factor, has the same rank r for any $z \in D$.

Processes with a rational spectral density are of the following kind:

- VARMA (with rational spectral density and minimum phase spectral factor is also rational),
- Obey a state space model.
- Obey a matrix fractional description.

Causal VAR Models

Bolla, M., Ye, D., Wang, H., Ma, R., Frappier, V., Thompson, D., Donner, C., Baranyi, M., Abdelkhalek, F., Causal Vector Autoregression Enhanced with Covariance and Order Selection (arXiv: 2211.14203).

VAR(p) models ($p \geq 1$ is fixed integer)

- **Reduced form** VAR(p) model:

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

where $\mathbf{V}_t$ is white noise, it is uncorrelated with $\mathbf{X}_{t-1}, \ldots, \mathbf{X}_{t-p}$, it has zero expectation and covariance matrix Σ (not necessarily diagonal, but positive definite), and the matrices $\mathbf{M}_j$ satisfy the stability conditions. $\mathbf{V}_t$: *innovation*.

- **Structural form** SVAR(p) model:

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

where the *structural shock* $\mathbf{U}_t$ is uncorrelated with $\mathbf{X}_{t-1}, \ldots, \mathbf{X}_{t-p}$, it has zero expectation with uncorrelated components, i.e. with positive definite, diagonal covariance matrix Δ. $\mathbf{A}$ is $d \times d$ upper triangular matrix with 1s along its main diagonal; $\mathbf{B}_1, \ldots, \mathbf{B}_p$ are $d \times d$ matrices (Kilian, Lütkepohl). One-to-one correspondence between them.

VAR(p) models (continued)

- **Causal** CVAR(p) **unrestricted** model: same equation as the Structural, 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.
  The causal effects are meant contemporaneously, and reflected in the upper triangular structure of the matrix **A**.
- **Causal** CVAR(p) **restricted** model: here an incomplete DAG is built, based on partial correlations.
  First we build an undirected graph: not connect i and j if the partial correlation coefficient of X_i and X_j, eliminating the effect of the other variables is 0 (theoretically), or less than a threshold (practically). Then we use covariance selection to ensure 0s in **A** in the no-edge positions.

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.

Partial correlations (Gaussian case)

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

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

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 $\mathbf{X}_{V \setminus \{i,j\}}$. Further,

$$\sigma^{ii} = 1/(\mathrm{Var}(X_i | \mathbf{X}_{V \setminus \{i\}}), \qquad i = 1, \ldots, d$$

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

Partial regression coefficients (Gaussian case)

Consider the regression plane

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

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

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

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

Undirected (Gaussian) graphical model

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

In practice: draw an edge if we can reject

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

The following test uses the empirical partial correlation coefficient:

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

where $\mathbf{S}$ is the sample size (n) times the empirical covariance matrix of the variables in the subscript (product-moments). We reject H_0 for large values of $|t|$, where under H_0, the test statistic

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

is distributed as Student's t with $n - d$ degrees of freedom.

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 $i = j + 1, \ldots, d$:

$$l_{ij} = \left(k_{ij} - \sum_{h=1}^{j-1} l_{ih} \delta_h^{-1} l_{jh} \right) \delta_j \tag{3}$$

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 **M** be a symmetric or an upper triangular matrix of real entries. We say that **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 $j \rightarrow h \leftarrow i$ 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 (3).
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 (subsets) 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 $\mathbf{A}$.

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

Daily relative 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.

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).

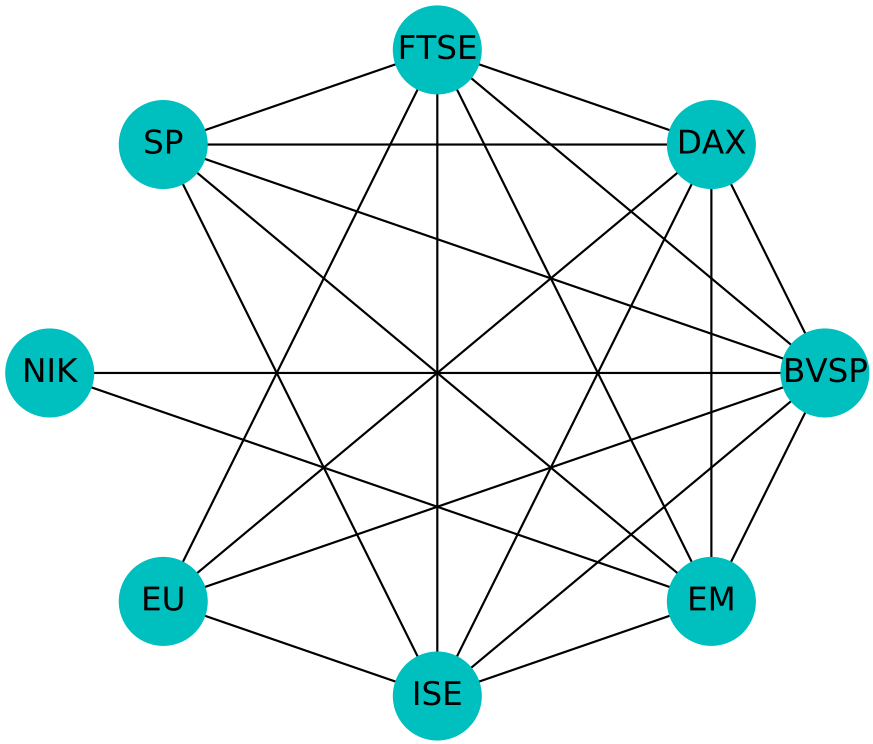
Relative returns

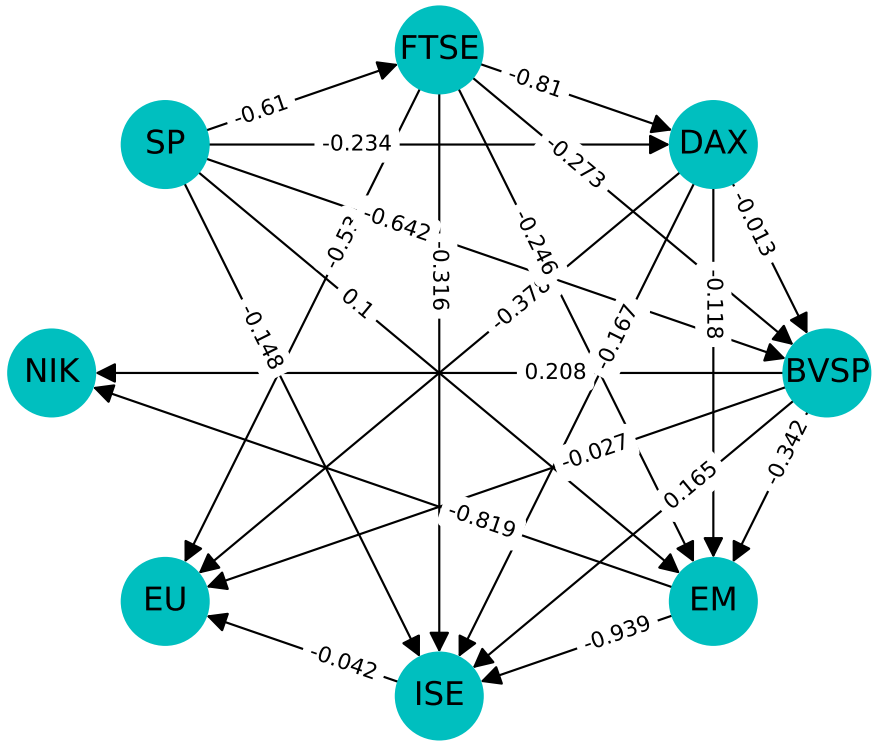
Log asset price $X_t = \log P_t$.
Differenced log asset price:

$$Z_t = X_t - X_{t-1} = \log \frac{P_t}{P_{t-1}} = \log \left(1 + \frac{P_t - P_{t-1}}{P_{t-1}}\right) \approx \frac{P_t - P_{t-1}}{P_{t-1}}$$

which is called *log return (or simply return) for day* t. This is close to the *relative return* $\frac{P_t - P_{t-1}}{P_{t-1}}$ if the price does not change much from one day to the next one, relatively to the previous price.

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.

Covariance selection with lag 1

The lag 1 variables $X_{t-1,1}, \ldots, X_{t-1,d}$ are included too. The new cliques and separators are

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

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

The estimate of the $2d \times 2d$ $\mathbf{K}$, inverse of $\mathfrak{C}_2$ is:

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

where the matrix $\mathbf{S}_{C'}$ is the product-moment estimate based on the $n-1$ element serially correlated sample and $[\mathbf{M}_{C'}]^{2d}$ denotes the $2d \times 2d$ matrix comprising the entries of the larger $2d \times 2d$ matrix $\mathbf{M}$ 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 $\mathbf{A}$.

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

Dynamic Factor Analysis

- Forni, M. and Lippi, M., The general dynamic factor model: one-sided representation results, *Journal of Econometrics* (2011).
- Bolla, M. 2017. Factor Analysis, Dynamic. Wiley StatsRef: Statistics Reference Online: 1-15. John Wiley&Sons, Ltd.
- Lippi, M., Deistler, M., Anderson, B., High-Dimensional Dynamic Factor Models: A Selective Survey and Lines of Future Research (arXiv: 2202.07745).

Time domain

$\{\mathbf{X}_t\}$: d-dimensional, weakly stationary time series with real components and autocovariance matrices $\mathbf{C}(h)$, $\mathbf{C}(-h) = \mathbf{C}^T(h)$, $h \in \mathbb{Z}$.
$\mathfrak{C}_n$: covariance matrix of $[\mathbf{X}_1^T, \ldots, \mathbf{X}_n^T]^T \in \mathbb{R}^{nd}$:

$$\mathfrak{C}_n := \begin{bmatrix} \mathbf{C}(0) & \mathbf{C}(1) & \mathbf{C}(2) & \cdots & \mathbf{C}(n-1) \\ \mathbf{C}^T(1) & \mathbf{C}(0) & \mathbf{C}(1) & \cdots & \mathbf{C}(n-2) \\ \mathbf{C}^T(2) & \mathbf{C}^T(1) & \mathbf{C}(0) & \cdots & \mathbf{C}(n-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}^T(n-1) & \mathbf{C}^T(n-2) & \mathbf{C}^T(n-3) & \cdots & \mathbf{C}(0) \end{bmatrix}.$$

This is a symmetric, positive semidefinite block Toeplitz matrix, the (i, j) block of which is $\mathbf{C}(j - i)$.

Frequency domain

Denoting by $\mathbf{C}(h) = [c_{ij}(h)]$ the $d \times d$ autocovariance matrices ($\mathbf{C}(-h) = \mathbf{C}^T(h)$, $h \in \mathbb{Z}$) in the time domain, assume that their entries are absolutely summable, i.e., $\sum_{h=0}^{\infty} |c_{pq}(h)| < \infty$ for $p, q = 1, \ldots, d$.
Then, the self-adjoint, positive semidefinite spectral density matrix $\mathbf{f}(\omega)$ exists in the frequency domain, and it is defined by

$$\mathbf{f}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \mathbf{C}(h) e^{-ih\omega}, \quad \omega \in [0, 2\pi].$$

We consider it at the Fourier frequencies.

Theorem

Let $\{\mathbf{X}_t\}$ be d-dimensional weakly stationary time series of real components. For odd $n = 2k + 1$, consider $\mathbf{X}_1, \dots, \mathbf{X}_n$ with the block Toeplitz matrix $\mathfrak{C}_n$; further, the Fourier frequencies $\omega_j = \frac{2\pi j}{n}$ for $j = 0, \dots, n-1$. Let

$$\mathbf{D}_n = \operatorname{diag}(\operatorname{Spec} \mathbf{f}(0), \operatorname{Spec} \mathbf{f}(\omega_1), \dots, \operatorname{Spec} \mathbf{f}(\omega_k), \operatorname{Spec} \mathbf{f}(\omega_k), \dots, \operatorname{Spec} \mathbf{f}(\omega_1)).$$

Here Spec contains the eigenvalues of the affected matrix in non-increasing order if not otherwise stated. (The duplication is due to the fact that $\mathbf{f}(\omega_j) = \overline{\mathbf{f}(\omega_{n-j})}$, $j = 1, \dots, k$, for real time series). Then, with the modal matrix $\mathbf{U}$ of the „close" block circulant matrix,

$$\mathbf{U}^T \mathfrak{C}_n \mathbf{U} - 2\pi \mathbf{D}_n \to \mathbf{O}, \quad n \to \infty,$$

i.e., the entries of the matrix $\mathbf{U}^T \mathfrak{C}_n \mathbf{U} - 2\pi \mathbf{D}_n$ tend to 0 uniformly as $n \to \infty$.

Consequence: Complex PCA

$\mathbf{W}$: $nd \times nd$ matrix containing the orthonormal eigenvectors $\mathbf{w}_j$ (of complex coordinates) of the block circulant matrix $\mathfrak{C}_n^{(s)}$ in its columns.
Let $\mathbf{Z} = (\mathbf{Z}_1^T, \dots, \mathbf{Z}_n^T)^T$ denote the random vector obtained by

$$\mathbf{Z} = \mathbf{W}^* \mathbf{X}.$$

Its (complex) components are also uncorrelated and $\mathbb{E}\mathbf{Z}\mathbf{Z}^* \sim 2\pi \mathbf{D}_n$ again. Instead, we consider the blocks $\mathbf{Z}_j$s of it, and perform a „partial principal component transformation" (in d-dimension) of them. Let $\mathbf{w}_{1j}, \dots, \mathbf{w}_{dj}$ be the columns of $\mathbf{W}$ corresponding to the coordinates of $\mathbf{Z}_j$. Then by the block nature of the eigenvectors:

$$\mathbf{Z}_j = \frac{1}{\sqrt{n}} (\mathbf{V}_j^* \otimes \mathbf{r}^*) \mathbf{X},$$

where $\mathbf{r}^* = (1, \rho_j^{-1}, \rho_j^{-2}, \dots, \rho_j^{-(n-1)})$ and $\mathbf{V}_j$ is the $d \times d$ unitary matrix in the spectral decomposition $\mathbf{M}_j = \mathbf{V}_j \mathbf{\Lambda}_j \mathbf{V}_j^*$.

Inverse Discrete Fourier Transform

The main Theorem implies that

$$\mathbb{E}(\mathbf{V}_j\mathbf{Z}_j)(\mathbf{V}_j\mathbf{Z}_j)^* = \mathbf{V}_j\mathbf{\Lambda}_j\mathbf{V}_j^* = \mathbf{M}_j.$$

At the same time,

$$\mathbf{V}_j\mathbf{Z}_j = \frac{1}{\sqrt{n}}\mathbf{V}_j(\mathbf{V}_j^* \otimes \mathbf{r}^*)\mathbf{X} = \frac{1}{\sqrt{n}}(\mathbf{I}_d \otimes \mathbf{r}^*)\mathbf{X}$$

$$= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbf{X}_t e^{-it\omega_j}, \quad j = 1, \ldots, n.$$

This is the finite DFT of $\mathbf{X}_1, \ldots, \mathbf{X}_n$. It is also in accord with the definition of **the orthogonal increment process** $\{\mathbf{Z}_\omega\}$ of which $\mathbf{V}_j\mathbf{Z}_j \sim \mathbf{Z}_{\omega_j}$ is the discrete analogue. Also, $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ are asymptotically pairwise orthogonal akin to $\mathbf{V}_1\mathbf{Z}_1, \ldots, \mathbf{V}_n\mathbf{Z}_n$. Further,

$$\mathbb{E}(\mathbf{V}_j\mathbf{Z}_j)(\mathbf{V}_j\mathbf{Z}_j)^* \sim 2\pi\mathbf{f}(\omega_j),$$

and it is in accord with the fact that

$$\mathbb{E}\mathbf{Z}_j\mathbf{Z}_j^* \sim 2\pi \operatorname{diag}\operatorname{Spec}\mathbf{f}(\omega_j), \quad j = 1, \ldots, n.$$

Low rank approximation of the process

To find the best k-rank approximation of the weakly stationary, regular d-dimensional process, the d-dimensional vectors $\mathbf{V}_j\mathbf{Z}_j$s, obtained by DFT, should be projected onto the subspace spanned by the k leading eigenvectors of $\mathbf{V}_j$, $k \leq d$, denoted by $\mathbf{V}_j^{(k)}$. Assume that the eigenvalues in $\mathbf{\Lambda}_j$ are in non-increasing order. Let us denote the k leading eigenvectors by $\mathbf{v}_{j1}, \ldots, \mathbf{v}_{jk}$. Then the best rank k approximation of $\mathbf{V}_j\mathbf{Z}_j$:

$$(\mathbf{V}_j\mathbf{Z}_j)^{(k)} = \mathrm{Proj}_{\mathrm{Span}\,\{\mathbf{v}_{j1},\ldots,\mathbf{v}_{jk}\}}\mathbf{V}_j\mathbf{Z}_j = \left[\mathbf{V}_j^{(k)}(\mathbf{V}_j^{(k)})^*\right]\mathbf{V}_j\mathbf{Z}_j$$
$$= \sum_{\ell=1}^{k}(\mathbf{v}_{j\ell}^*\mathbf{V}_j\mathbf{Z}_j)\mathbf{v}_{j\ell} = \sum_{\ell=1}^{k} Z_{j\ell}\mathbf{v}_{j\ell},$$

where $Z_{j\ell}$ denotes the ℓth coordinate of $\mathbf{Z}_j$.

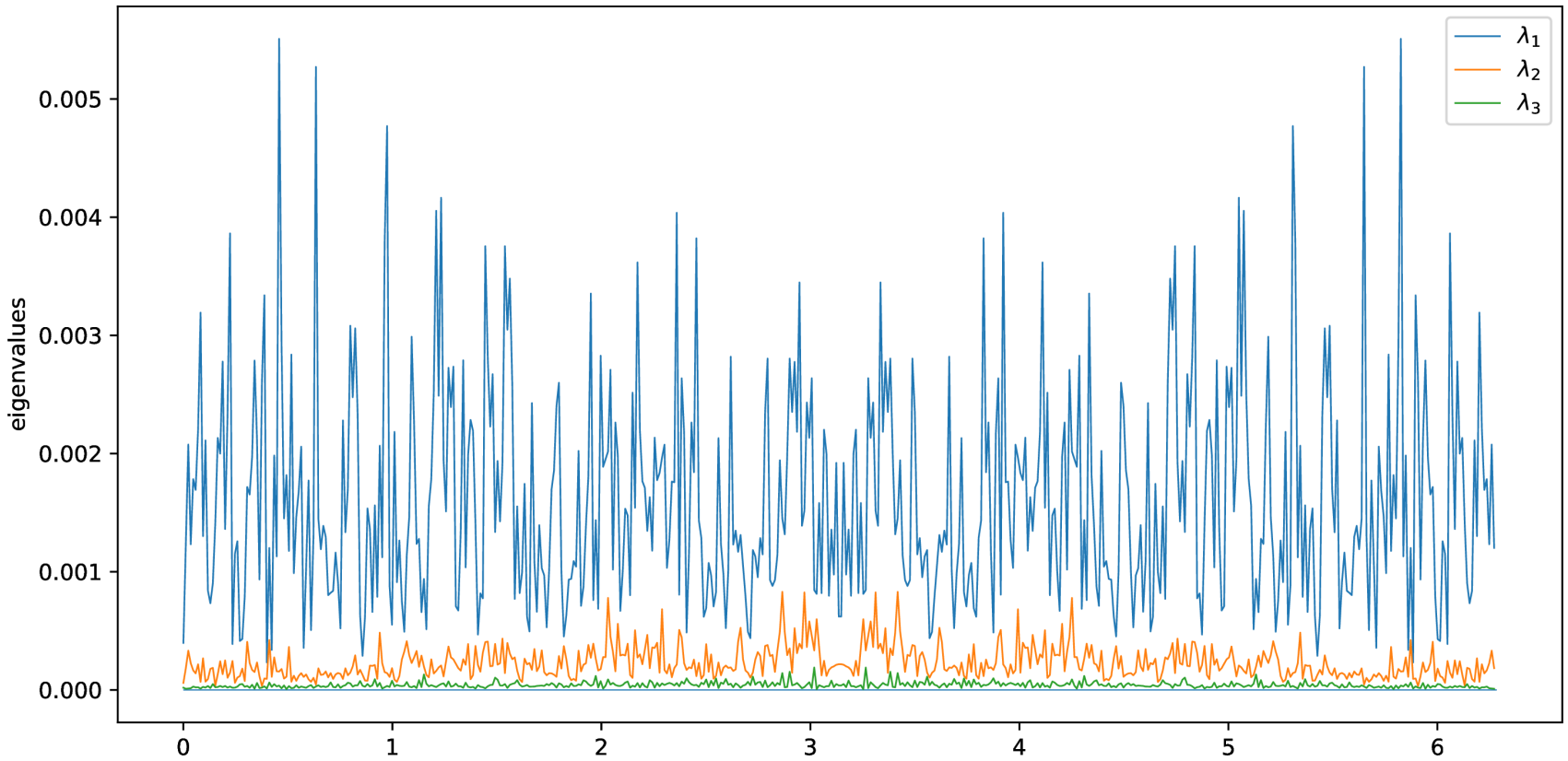
Frequency domain to time domain

This transformation gives rise to **rank** reduction in the frequency domain, then via DFT (due to $\mathbf{X} = \mathbf{WZ}$), in the time domain too.
The best rank k approximation of $\mathbf{X}_t$ via IDFT:

$$(\mathbf{X}_t)^{(k)} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (\mathbf{V}_j \mathbf{Z}_j)^{(k)} e^{it\omega_j}$$
$$= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(\sum_{\ell=1}^{k} Z_{j\ell} \mathbf{v}_{j\ell} \right) e^{it\omega_j},$$

for $t = 1, \ldots, n$.
We can show that is is also a d-dimensional real time series, but its spectral density matrix is of rank $k \leq d$.
We can start the DFA with it.

Application to stock returns (Akbilgic, O. et al.)



The raw data were used.

The spectra shows 3 leading eigenvalues, the size of the gap after the leading eigenvalues depends on the spectral density estimation method.

Singular Autoregression

In a d-dimensional, weakly stationary time series with zero expectation, we linearly predict $\mathbf{X}_n$ based on past values $\mathbf{X}_1, \ldots, \mathbf{X}_{n-1}$. Let $\hat{\mathbf{X}}_1 := 0$, and denote by $\hat{\mathbf{X}}_n$ the best one-step ahead linear prediction (based on $(n-1)$-long past) that minimizes the mean square error

$$\mathbb{E}(\mathbf{X}_n - \hat{\mathbf{X}}_n)^2 = \|\mathbf{X}_n - \hat{\mathbf{X}}_n\|^2, \quad n = 1, 2, \ldots$$

which is the instance of simultaneous linear regressions. $\mathbf{X}_t$ can be expanded in terms of the now d-dimensional innovations, i.e. the prediction error terms

$$\boldsymbol{\eta}_n := \mathbf{X}_n - \hat{\mathbf{X}}_n,$$

with error covariance matrix $\mathbf{E}_n = \mathbb{E}\boldsymbol{\eta}_n \boldsymbol{\eta}_n^T$.

Block LDL decomposition with pseudo-inverses

Consider the first n steps, i.e. the recursive equations

$$\mathbf{X}_j = \sum_{k=1}^{j-1} \mathbf{B}_{jk}\boldsymbol{\eta}_k + \boldsymbol{\eta}_j, \quad j = 1, 2, \ldots, n$$

in the case when the observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are available. If our process is stationary, the coefficient matrices are irrespective of the choice of the starting time, and in the regular case, they approach the one-step ahead projection based on the infinite past. It can be that the error covariance matrices are not zeros, but they are of reduced rank or better and better approach a rank q innovation covariance matrix, with decreasing ranks ($q \leq d$).
Multiplying the above equations by $\mathbf{X}_j^T$ from the right, and taking expectation, the solution for the matrices $\mathbf{B}_{jk}$ and $\mathbf{E}_j$ ($j = 1, \ldots, n$; $k = 1, \ldots, j-1$) can be obtained via the block LDL (variant of the block Cholesky) decomposition:

$$\mathfrak{C}_n = \mathbf{L}_n \mathbf{D}_n \mathbf{L}_n^T.$$

Hhere $\mathfrak{C}_n$ is $nd \times nd$ positive semidefinite block Toeplitz matrix of general block entry $\mathbf{C}(i-j)$,

$$\mathbf{L}_n = \begin{bmatrix} \mathbf{I} & \mathbf{O} & \dots & \mathbf{O} & \mathbf{O} \\ \mathbf{B}_{21} & \mathbf{I} & \dots & \mathbf{O} & \mathbf{O} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{B}_{n1} & \mathbf{B}_{n2} & \dots & \mathbf{B}_{n,n-1} & \mathbf{I} \end{bmatrix}, \quad \mathbf{D}_n = \begin{bmatrix} \mathbf{E}_1 & \mathbf{O} & \dots & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{E}_2 & \dots & \mathbf{O} & \mathbf{O} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{O} & \mathbf{O} & \dots & \mathbf{O} & \mathbf{E}_n \end{bmatrix}.$$

To find the block LDL (Cholesky) decomposition, the following recursion is used: for $j = 1, \dots, n$

$$\mathbf{E}_j := \mathbf{C}(0) - \sum_{k=1}^{j-1} \mathbf{B}_{jk} \mathbf{E}_k \mathbf{B}_{jk}^T, \quad j = 1, \dots, n$$

and for $i = j+1, \dots, n$

$$\mathbf{B}_{ij} := \left(\mathbf{C}(i-j) - \sum_{k=1}^{j-1} \mathbf{B}_{ik} \mathbf{E}_k \mathbf{B}_{ik}^T \right) \mathbf{E}_j^+,$$

where we take the Moore–Penrose inverse (denoted by $+$ in the superscript) if necessary and we do not enter into the summation if $j = 1$

Remarks

- The innovation algorithm (variant of the Durbin–Levinson) also does it, provided $\mathfrak{C}_n$ is non-singular.
- Because of

$$|\mathfrak{C}_n| = |\mathbf{D}_n| = \prod_{j=1}^{n} |\mathbf{E}_j|,$$

  if $|\mathfrak{C}_n| = 0$, then $|\mathbf{E}_j|$ becomes 0 (at least from a certain index j), but we can treat this situation with the pseudoinverse.
- Since $|\mathfrak{C}_n|$ is the product of the eigenvalues of $\mathfrak{C}_n$, which asymptotically comprise the union of the spectra of $\mathbf{f}(\omega)$ ($d \times d$ spectral density matrix) at the Fourier frequencies, singular prediction error matrices indicate reduced rank spectral density.
- $\mathbf{E}_1 = \mathbf{C}(0)$, $\mathrm{rank}\mathbf{C}(0) = r$, and $\mathbf{E}_j$s are the one-step ahead prediction (based on $j-1$ long past) error covariance matrices with non-increasing ranks.

By the multi-dimensional Wold decomposition, $\mathbf{E}_n \to \mathbf{\Sigma}$ in L^2-norm, where $\mathbf{\Sigma}$ is the error covariance matrix of the one-step ahead prediction based on the infinite past, $\mathrm{rank}\mathbf{\Sigma} = q \leq r$.
If the prediction is based on the infinite past, then with $n \to \infty$ this procedure (which is a nested one) extends to the multidimensional Wold decomposition.
If $n \to \infty$, the matrix $\mathbf{L}_n$ better and better approaches a block Toeplitz one, and the matrices $\mathbf{E}_1, \ldots, \mathbf{E}_n$ are closer and closer to $\mathbf{\Sigma}$, the covariance matrix of the innovation process. Since $\|\mathbf{E}_n - \mathbf{\Sigma}\|^2 \to 0$ as $n \to \infty$, $\mathbf{B}_{nj} \to \mathbf{B}_j$ as $n \to \infty$ too, as it continuously depends on $\mathbf{E}_j$s.
As $\mathbf{E}_j$ is Cauchy sequence and we stop at a j ($j < n$, n is „large") where it does not change „much", then the jth block-row of $\mathbf{L}_n$ can be considered that it contains the effective coefficient matrices $\mathbf{B}_{jk}$s ($k = 1, \ldots, j-1$) in a finite segment of the Wold decomposition. So a singular VAR(j) process is obtained if $q < r$.

Perturbation of eigenvalues

If there is a gap in the spectrum of $\mathbf{\Sigma}$ (at $k < r$), like

$$\lambda_1 \geq \cdots \geq \lambda_k \geq \Delta \gg \varepsilon \geq \lambda_{k+1} \geq \cdots \geq \lambda_d,$$

then there is a gap in the spectrum of $\mathbf{E}_n$ too. Indeed, to any $\delta > 0$ there is an N such that for $n \geq N$: $\|\mathbf{E}_n - \mathbf{\Sigma}\| < \delta$. Then for the eigenvalues of $\mathbf{E}_n$,

$$\lambda_1^{(n)} \geq \cdots \geq \lambda_k^{(n)} \geq \Delta - \delta \gg \varepsilon + \delta \geq \lambda_{k+1}^{(n)} \geq \cdots \geq \lambda_d^{(n)}.$$

Consequently, for the best rank k approximations (with Gram-decompositions):

$$\|\mathbf{\Sigma} - \mathbf{\Sigma}^k\| \leq \varepsilon \quad \text{and} \quad \|\mathbf{E}_n - \mathbf{E}_n^k\| \leq \delta + \varepsilon$$

holds by the Weyl perturbation theorem. Therefore,

$$\|\mathbf{\Sigma}^k - \mathbf{E}_n^k\| \leq \|\mathbf{\Sigma}^k - \mathbf{\Sigma}\| + \|\mathbf{\Sigma} - \mathbf{E}_n\| + \|\mathbf{E}_n - \mathbf{E}_n^k\| \leq \varepsilon + \delta + (\delta + \varepsilon) = 2(\delta + \varepsilon)$$

that can be arbitrarily close to 2ε.

Perturbation of spectral subspaces

At the same time, the projections onto the subspaces spanned by the eigenvectors of the k structural eigenvalues of these matrices are close to each other, in the sense of the Davis–Kahan theorem. Let $S_1 := [\Delta - \delta, \lambda_1 + \delta]$ and $S_2 := [\lambda_d + \delta, \varepsilon + \delta]$. Then for $n > N$:

$$\begin{aligned}
\|\mathbb{P}_{\mathbf{\Sigma}}(S_1) - \mathbb{P}_{\mathbf{E}_n}(S_1)\|_F^2 &= \|\mathbb{P}_{\mathbf{\Sigma}}(S_1)\|_F^2 + \|\mathbb{P}_{\mathbf{E}_n}(S_1)\|_F^2 - 2\operatorname{tr}[\mathbb{P}_{\mathbf{\Sigma}}(S_1)\mathbb{P}_{\mathbf{E}_n}^T(S_1)] \\
&= 2r - 2\operatorname{tr}[\mathbb{P}_{\mathbf{\Sigma}}(S_1)(\mathbf{I}_d - \mathbb{P}_{\mathbf{E}_n}^T(S_2))] \\
&= 2r - 2\operatorname{tr}[\mathbb{P}_{\mathbf{\Sigma}}(S_1) - \mathbb{P}_{\mathbf{\Sigma}}(S_1)\mathbb{P}_{\mathbf{E}_n}^T(S_2)] \\
&= 2r - 2r + 2\operatorname{tr}[\mathbb{P}_{\mathbf{\Sigma}}(S_1)\mathbb{P}_{\mathbf{E}_n}^T(S_2)] \\
&\leq 2d\|\mathbb{P}_{\mathbf{\Sigma}}(S_1)\mathbb{P}_{\mathbf{E}_n}^T(S_2)\| \\
&\leq 2d\frac{c}{\Delta - \delta - \varepsilon}\|\mathbf{\Sigma} - \mathbf{E}_n\| \leq 2d\frac{c\delta}{\Delta - \delta - \varepsilon}
\end{aligned}$$

that can be arbitrarily „small" is δ is arbitrarily „small". Here $\mathbb{P}_{\mathbf{\Sigma}}(S)$ denotes the projection onto the subspace spanned by the eigenvectors of $\mathbf{\Sigma}$ corresponding to its eigenvalues in S. It is effective is there is a „large" gap in the spectrum ($O(n)$ vers. $o(n)$).

The Dynamic Factor Model

Assume that $\{\mathbf{X}_t\}$ is *weakly stationary* with an absolutely continuous spectral distribution, i.e. it has the $d \times d$ spectral density matrix $\mathbf{f}_{\mathbf{X}}$. With the integer $1 \leq k < d$, the *dynamic k-factor model* for $\mathbf{X}_t$:

$$\mathbf{X}_t = \boldsymbol{\mu} + \mathbf{B}(L)\mathbf{Z}_t + \mathbf{e}_t = \boldsymbol{\mu} + \boldsymbol{\chi}_t + \mathbf{e}_t$$

or with components,

$$X_t^i = \mu^i + b_{i1}(L)Z_t^1 + \cdots + b_{ik}(L)Z_t^k + e_t^i$$

where the k-dimensional stochastic process $\mathbf{Z}_t = (Z_t^1, \ldots Z_t^k)^T$ is the *dynamic factor*, $\boldsymbol{\chi}_t$ is called *common component*, the d-dimensional stochastic process $\mathbf{e}_t = (e_t^1, \ldots, e_t^d)^T$ is called *idiosyncratic noise*, and the $d \times k$ matrix $\mathbf{B}(L) = (b_{ij}(L))$, $i = 1, \ldots, d$, $j = 1, \ldots, k$, is the *transfer function*. Here L is the *lag operator* (backward shift) and $b_{ij}(L)$ is a square-summable one-sided filter, i.e. $b_{ij}(L) = b_{ij}(0) + b_{ij}(1)L + b_{ij}(2)L^2 + \ldots$, with $\sum_{\ell=0}^{\infty} b_{ij}^2(\ell) < \infty$.

Componentwise conditions

$$\mathbb{E}(\mathbf{Z}_t) = \mathbf{0}, \quad \mathbb{E}(\mathbf{e}_t) = \mathbf{0}, \quad t \in \mathbb{Z}$$
$$\mathrm{Cov}(e_t^i, Z_s^j) = 0, \quad i = 1, \ldots, d, \quad j = 1, \ldots, k, \quad t, s \in \mathbb{Z},\ s \leq t.$$
$$\mathrm{Cov}(e_t^i, e_s^j) = 0, \quad i, j = 1, \ldots d, \quad i \neq j, \quad t, s \in \mathbb{Z},\ s < t.$$

If $\boldsymbol{\chi}_t$ and $\mathbf{e}_t$ are also weakly stationary and they have rational spectral densities $\mathbf{f}_{\boldsymbol{\chi}}$ and $\mathbf{f}_{\mathbf{e}}$, the model equation extends to the spectral density matrices:

$$\mathbf{f}_{\mathbf{X}}(\omega) = \mathbf{f}_{\boldsymbol{\chi}}(\omega) + \mathbf{f}_{\mathbf{e}}(\omega) = \mathbf{B}(e^{-i\omega})\mathbf{f}_{\mathbf{Z}}(\omega)\mathbf{B}(e^{-i\omega})^* + \mathbf{f}_{\mathbf{e}}(\omega), \quad \omega \in [-\pi, \pi].$$

Very frequently, $\mathbf{Z}_t$ is assumed to be orthonormal WN($\mathbf{I}_k$) process. Then the above equation simplifies to

$$\mathbf{f}_{\mathbf{X}}(\omega) = \frac{1}{2\pi}\mathbf{B}(e^{-i\omega})\mathbf{B}(e^{-i\omega})^* + \mathbf{f}_{\mathbf{e}}(\omega).$$

General Dynamic Factor Model (GDFM)

Let $\mathbf{X}_t$ be a *weakly stationary* time series ($t = 1, 2, \dots$) with an absolutely continuous spectral measure and the positive semidefinite *spectral density* matrix $\mathbf{f}_\mathbf{X}$.
Assume that $\mathbf{f}_\mathbf{X}(\omega)$ has *constant rank* r for a.e. $\omega \in [-\pi, \pi]$. If $\mathbf{X}_t$ is also regular (it always holds if $\mathbf{f}_\mathbf{X}$ is a rational spectral density matrix), then the multidimensional Wold decomposition is able to make it a one-sided $VMA(\infty)$ process. It is important that the dimension of the *innovation subspaces* is also r.
With the integer $1 \leq q \leq r$, the *q-factor GDFM*:

$$\mathbf{X}_t = \chi_t + \mathbf{e}_t, \quad t = 1, 2, \dots$$

where now χ_t denotes the common component, $\mathbf{e}_t$ is the *idiosyncratic noise*, and all the expectations are zeros, for simplicity. Here χ_t is subordinated to $\mathbf{X}_t$, but has spectral density matrix of rank $q \leq r$. For example, there are q uncorrelated signals (given by q distinct sources) detected by r sensors.

Nested sequence

Forni, Lippi, and Deistler gave necessary and sufficient conditions for the existence of an underlying GDFM in terms of the expanding sequence of $n \times n$ spectral density matrices $\mathbf{f}_{\mathbf{X}}^{n}(\omega)$, $n \in \mathbb{N}$.

Theorem

The nested sequence $\{\mathbf{X}_t^n : n \in \mathbb{N}, t = 1, 2, \dots\}$ *can be represented by a sequence of q-factor GDFMs if and only if*

- *the q largest eigenvalues,* $\lambda_{\mathbf{X},1}^{n}(\omega) \geq \cdots \geq \lambda_{\mathbf{X},q}^{n}(\omega)$ *of* $\mathbf{f}_{\mathbf{X}}^{n}(\omega)$ *diverge almost everywhere in* $[-\pi, \pi]$ *as* $n \to \infty$*;*
- *the* $(q+1)$*-th largest eigenvalue* $\lambda_{\mathbf{X},q+1}^{n}(\omega)$ *of* $\mathbf{f}_{\mathbf{X}}^{n}(\omega)$ *is uniformly bounded for* $\omega \in [-\pi, \pi]$ *(almost everywhere) and for all* $n \in \mathbb{N}$*.*

The theorem is rather theoretical; its message is that for large n and T (T is not necessarily larger than n) we can conclude for q from the spectral gap of the constant rank spectral density matrix. The estimate χ_t^n is consistent if $n, T \to \infty$.

Assumptions

Assumptions are imposed on the processes $\boldsymbol{\chi}_t$ and $\mathbf{e}_t$:

1. For all n, $\boldsymbol{\chi}_t^n$ has a rational spectral density, so constant rank a.e. on $[-\pi, \pi]$.
2. We suppose that there exists $n_0 \geq q$ s.t. $\operatorname{rank}(\boldsymbol{\chi}_t^n) = q$, that is independent of n, $n \geq n_0$.
3. *Weak cross-sectional dependence of* $\mathbf{e}_t^n$: the eigenvalues of $f_{\mathbf{e}}^n(\omega)$ are uniformly bounded for all n and a.e. ω.
4. *strong cross-sectional dependence of* $\boldsymbol{\chi}_t^n$: the first q eigenvalues of $f_{\boldsymbol{\chi}}^n(\omega)$ diverge as $n \to \infty$ for a.e. ω.

The idiosyncratic noise is less and less important when $n, T \to \infty$, and it may have slightly correlated components. A stationary process with a not full rank spectral density matrix may have some singular components. All these parts are included in the weakly dependent idiosyncratic noise.

Dynamic factors and Wold representation

$\{\boldsymbol{\chi}_t\}$ is a VARMA process: under the stability and strict miniphase conditions,

$$\boldsymbol{\chi}_t = \boldsymbol{\beta}^{-1}(z)\boldsymbol{\alpha}(z)\mathbf{v}_t = \mathbf{k}(z)\mathbf{v}_t = \sum_{j=0}^{\infty} \mathbf{k}_j \mathbf{v}_{t-j},$$

where the q-dimensional $\{\mathbf{v}_t\}$ is the dynamic factor process. Under the stability and strict miniphase conditions, by the Smith–McMillan form, $\mathbf{k}(z)$ has a (not unique) left inverse (generalized inverse) $\mathbf{k}^{-}(z)$. With it,

$$\mathbf{v}_t = \mathbf{k}^{-}(z)\boldsymbol{\chi}_t.$$

Further assumptions, guaranteeing uniqueness can be made (in the $q < r$ case).

Singular VAR

$\{\mathbf{f}_t\}$: r-dimensional static factor process of $\{\mathbf{X}_t^{(k)}\}$, $r = \mathrm{rank}\mathbf{C}(0)$ (of $\mathbf{X}_t^{(k)}$), $r \geq k$.
If $r > q$, then the VARMA process for the static factors

$$\boldsymbol{\alpha}(z)_{r\times r}\mathbf{f}_t = \boldsymbol{\beta}(z)_{r\times q}\mathbf{v}_t,$$

where $\mathbf{v}_t$ is the q-dimensional *dynamic factor process*, can be substituted by a singular VAR process:

$$\mathbf{a}(z)_{r\times r}\mathbf{f}_t = \mathbf{b}_{r\times q}\mathbf{v}_t,$$

where $\mathbf{a}(z)$ is a new VAR polynomial of order $\tilde{p}$, and $\mathbf{b} = \boldsymbol{\beta}(0)$.
Order selection for $\tilde{p}$.
For the singular VAR, we can use the block LDL decomposition with possibly singular innovation covariance matrix (Moore-Penrose inverses are used in the iteration).