Simultaneous Statistical Inference in Dynamic Factor Models

Estimation, Simulation and Application

Thorsten Dickhaus, Marcus Pauly and Natalia Sirotko-Sibirskaya

University of Bremen
Faculty of Mathematics
Institute for Statistics

April 29, 2016

Motivation

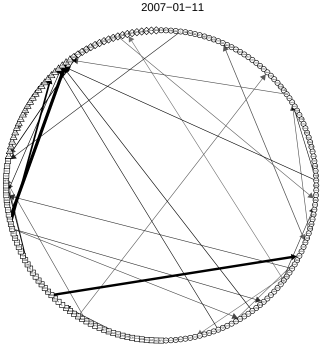

Figure : An example of a financial network of 200 U. S. financial companies based on semi-parametric single-index quantile regression with variable selection, Härdle, Wang and Sirotko-Sibirskaya 2014 SFB 649 Discussion paper.

Dynamic Factor Model

$$\mathbf{X}(t) = \sum_{s=-\infty}^{\infty} \Lambda(s)\mathbf{f}(t-s) + \varepsilon(t), \quad 1 \leq t \leq T,$$

- $\mathbf{X} = \Big(\mathbf{X}(t) : 1 \leq t \leq T\Big)$, a p-dimensional, covariance-stationary stochastic process with mean zero,
- $\mathbf{f}(t-s) = \Big(f_1(t-s), \cdots, f_k(t-s)\Big)^{\mathsf{T}}$, a k-dimensional vector of "common factors", $k < p$,
- $\varepsilon(t) = \Big(\varepsilon_1(t), \cdots, \varepsilon_p(t)\Big)^{\mathsf{T}}$, a p-dimensional vector of "specific factors",
- $f \perp \varepsilon$.

Dynamic Factor Model

Example: s = 0, 1, i. e. AR (1)-type DFM.

$$\underset{p\times 1}{x} = \sum_{s=0}^{1} \underset{p\times k}{\Lambda(s)} \underset{k\times 1}{f(t-s)} + \underset{p\times 1}{\varepsilon} = \underset{p\times k}{\Lambda(0)} \underset{k\times 1}{f(t)} + \underset{p\times k}{\Lambda(1)} \underset{k\times 1}{f(t-1)} + \underset{p\times 1}{\varepsilon}$$

$$\begin{bmatrix} x_{1,1} \\ x_{2,1} \\ \vdots \\ x_{p,1} \end{bmatrix} = \begin{bmatrix} \lambda_{1,1}^{(0)} & \cdots & \lambda_{1,k}^{(0)} \\ \lambda_{2,1}^{(0)} & \cdots & \lambda_{2,k}^{(0)} \\ \vdots & & \vdots \\ \lambda_{p,1}^{(0)} & \cdots & \lambda_{p,k}^{(0)} \end{bmatrix} \begin{bmatrix} f_{1,t} \\ f_{2,t} \\ \vdots \\ f_{k,t} \end{bmatrix} + \begin{bmatrix} \lambda_{1,1}^{(1)} & \cdots & \lambda_{1,k}^{(1)} \\ \lambda_{2,1}^{(1)} & \cdots & \lambda_{2,k}^{(1)} \\ \vdots & & \vdots \\ \lambda_{p,1}^{(1)} & \cdots & \lambda_{p,k}^{(1)} \end{bmatrix} \begin{bmatrix} f_{1,(t-1)} \\ f_{2,(t-1)} \\ \vdots \\ f_{k,(t-1)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,1} \\ \varepsilon_{2,1} \\ \vdots \\ \varepsilon_{p,1} \end{bmatrix}$$

$$x_{1,1} = \lambda_{1,1}^{(0)} f_{1,t} + \cdots + \lambda_{1,k}^{(0)} f_{k,t} + \lambda_{1,1}^{(1)} f_{1,(t-1)} + \cdots + \lambda_{1,k}^{(1)} f_{k,(t-1)} + \varepsilon_{1,1}$$

Dynamic Factor Model

$$\mathbf{X}(t) = \sum_{s=-\infty}^{\infty} \Lambda(s)\mathbf{f}(t-s) + \varepsilon(t), \quad 1 \leq t \leq T,$$

where

- $\Sigma_\varepsilon = diag$ (here), or $\Sigma_\varepsilon \neq diag$ (future work),
- $\Sigma_f = diag$, or $\Sigma_f \neq diag$,
- $\varepsilon \neq WN$ and $\Gamma_f \neq diag$.

Testing in Dynamic Factor Model

Dickhaus and Pauly 2015

Example 1: Which of the common factors have a lagged influence on X?

Example 2: Which of the specific factors have a non-trivial autocorrelation structure?

Overview

Testing Approach

Definition 1 [Empirical Fourier Transform]

$$\tilde{\mathbf{X}}(\omega_j) = (2\pi T)^{-1/2} \sum_{t=1}^{T} X(t) \exp(it\omega_j),$$

where $\omega_j = 2j\pi/T \quad \text{and} \quad -T/2 < j < T/2.$

Testing Approach

Definition 2 [Spectral Density Matrix]

$$S_{\mathbf{X}}(\omega) = \tilde{\Lambda}(\omega) S_{\mathbf{f}}(\omega) \tilde{\Lambda}^{\mathsf{T}}(\omega) + S_{\varepsilon}(\omega),$$

where $\omega_j = 2j\pi/T \quad \text{and} \quad -T/2 < j < T/2.$

$$\begin{aligned} S_{\mathbf{X}}(\omega) &= (2\pi)^{-1} \sum_{u=-\infty}^{\infty} \Gamma_{\mathbf{X}}(u) \exp(-i\omega u) \\ \Gamma_{\mathbf{X}}(u) &= \mathbb{E}[\mathbf{X}(t)\mathbf{X}(t+u)^{\mathsf{T}}] \\ &= \sum_{s=-\infty}^{\infty} \Lambda(s) \sum_{v=-\infty}^{\infty} \Gamma_f(u+s-v)\Lambda(v)^{\mathsf{T}} + \Gamma_{\varepsilon}(u) \end{aligned}$$

Testing Approach

Piecewise Constant Spectrum Assumption

$\exists$ B disjoint frequency bands $\Omega_1, \cdots, \Omega_B$ centred at $w^{(b)} \notin \{0, \pi\}$, $1 \leq b \leq B$, s.t. S_X can be assumed approximately constant and different from zero within each of these bands.

Let $n_b = n_b(T)$ be a number of harmonic frequencies $(\omega_{j,b})_{1 \leq j \leq n_b}$ of the form $2\pi j_u / T$ which are as near as possible to $\omega^{(b)}$ with j_u, $1 \leq u \leq n_b$.

Testing Approach

Theorem 1 [Convergence of Multivariate Fourier Transforms]

Under Assumptions (1) to (3) (See Appendix)

$$((\tilde{\mathbf{X}}(\omega_{j,b}))_{1\le j\le n_b}, 0_{\mathbb{N}}) \to (Z_{j,b})_{j\in\mathbb{N}}, \quad \min(n_b(T), T) \to \infty,$$

where $(Z_{j,b})_{j\in\mathbb{N}} \sim N(0, S_{\mathbf{X}}(\omega^{(b)}))$.

Proof. **Dickhaus and Pauly 2015**

Testing Approach

Definition 3 [Complex Normal Likelihood]

$$l_b(\theta_b, x) = \pi^{-p \times n_b} |ivech(\theta_b)|^{-n_b} \exp\left(-\sum_{j=1}^{n_b} \tilde{x}(\omega_{j,b})^{\mathsf{T}} ivech(\theta_b)^{-1} \tilde{x}(\omega_{j,b})\right),$$

where

- $\theta_b = ivech(S_{\mathbf{X}}(\omega^{(b)}))$,
- $w^{(b)}$ is the center of one of the b bands, $1 \leq b \leq B$,
- n_b is a number of frequencies in a respective band b, $1 \leq j_u \leq n_b$.

Testing Approach

Asymptotic Normality of MLE and Theorem 1 imply

$$\sqrt{n_b}(\widehat{\theta}_b - \theta_b) \to T_b \sim N_d(0, V_b) \quad \text{as} \quad \min(n_b(T), T) \to \infty,$$

where

- V_b, a covariance matrix of θ_b, is estimated by $\widehat{V}_b$,
- T_b's are independent for $1 \leq b \leq B$.

Testing Approach

Let

$$H : C\theta = \xi,$$

where $C \in \mathbb{R}^{r \times Bd}$ is a contrast matrix and $d = 2pk + p^2 + p$ is the number of parameters to estimate, $\xi \in \mathbb{R}^r$, $\theta = (\theta_1, \cdots, \theta_B)$.

Then

$$W = N(C\widehat{\theta} - \xi)^T (C\widehat{V}C)^+ (C\widehat{\theta} - \xi),$$

where

- $N = \sum_{b=1}^{B} n_b$,
- $\widehat{V}$ is a block-diagonal matrix consisting of $N\widehat{V}_b/n_b$, $1 \leq b \leq B$,
- A^+ is a Moore-Penrose inverse of a matrix A.

Testing Approach

Theorem 2 [Distribution of Wald Statistic]

Under Assumptions (1) to (3) (See Appendix) W has multivariate χ^2-distribution with $rank(C)$ degrees of freedom under the null hypothesis H provided that V is positive-definite and $N/n_b \leq K \leq \infty$ for all $1 \leq b \leq B$.

Proof. **Dickhaus and Pauly 2015**

Estimation

0 (relevant for empirical application) Pre-whitening of data
1 Imposition of restrictions and identification scheme*
2 Change-point estimation**
3 Estimation of the model by Fletchell-Powell algorithm (1963)***

References:

Fletcher and Powell (1963),
($*$) Geweke and Singleton 1981,
($**$) Laviere and Ludena 2000, Reschnehofer 2008,
($***$) Geweke 1977, Geweke and Singleton 1981, Joereskog 1966, Joereskog 1967, Joereskog 1969.

Identification

Let $M(\omega)$ be any nonsingular $p \times p$ matrix, then

$$\tilde{\Lambda}(\omega) S_f(\omega) \tilde{\Lambda}(\omega)^{\mathsf{T}} = \Lambda^* S_f(\omega)^* \Lambda^{\mathsf{T}},$$

where $\Lambda^*(\omega) = \tilde{\Lambda}(\omega) M(\omega)$ and $S_f^*(\omega) = M(\omega)^{-1} S_f(\omega) (M(\omega)^{-1})^{\mathsf{T}}$.

Identication Schemes: **Geweke and Singleton 1981**

Change-Point Estimation

Let $\mathcal{T}_Q$ be the set of configurations of change-points and Γ_Q the space of parameters.

$$\mathcal{T}_Q = \{t = (t_0, t_1, \ldots, t_Q), \quad t_0 = 0 < t_1 < t_2 < \ldots < t_Q = n\},$$
$$\Gamma_Q = \{\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_Q), \quad \gamma_q \in \Gamma\},$$

where Q is the number of change-points and n is the number of frequency points at which spectrum is evaluated.

Change-Point Estimation

The negative Whittle log likelihood of $(\tilde{\mathbf{x}}, t \in T_q)$ evaluated at γ_q is

$$W_n(T_q, \gamma_q) = \int_{(0,\pi)} \Big(\log f(\omega, \gamma_q) + \frac{I_n(\omega, T_q)}{f(\omega, \gamma_q)} \Big) d\omega$$

where

- $I_n(T_q, \omega) = \frac{1}{n_q} | \sum_{q \in Tq} x_t \exp^{-it\omega} |$ is the periodogram computed over the window T_q,
- $f(\omega, \gamma_q) =$ $a(b_1 \mathbb{1}_{[t_0,t_1)} + b_2 \mathbb{1}_{[t_1,t_2)} + \ldots + b_{Q-1} \mathbb{1}_{[t_{Q-2},t_{Q-1})} + \mathbb{1}_{[t_{Q-1},t_Q]})$ is the piecewise linear function with $\boldsymbol{\gamma}_q = (a, b_1, \ldots, b_{Q-1})$.

Change-Point Estimation

If the number of change-points is known in advance, the estimates for the parameters of the piecewise linear function are obtained by minimising

$$J_n(\mathbf{t}, \gamma) = \frac{1}{n} \sum_{q=1}^{Q} n_q W_n(T_q, \gamma_q).$$

Change-Point Estimation

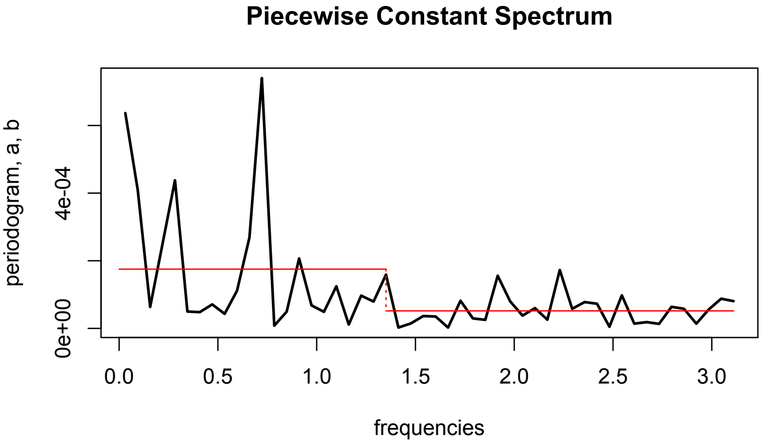

Figure : Piecewise constant spectrum estimation with one change-point for U. S. GDP 1947 - 2014, Reschenhofer 2008.

Change-Point Estimation

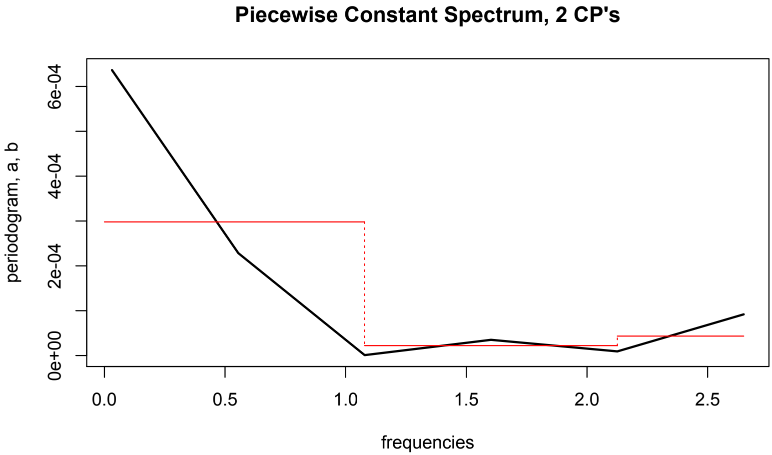

Figure : Piecewise constant spectrum estimation with 2 change-points for U. S. GDP 1947 - 2014, Reschenhofer 2008.

Change-Point Estimation

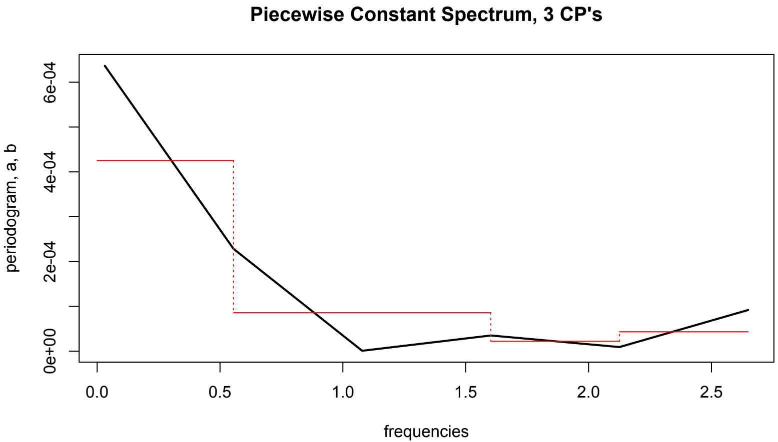

Figure : Piecewise constant spectrum estimation with 3 change-points for U. S. GDP 1947 - 2014, Reschenhofer 2008.

Model Estimation

Objective function:

$$l_b(\theta_b, x) = \pi^{-p\times n_b}|ivech(\theta_b)|^{-n_b} \exp\Big(-\sum_{j=1}^{n_b} \tilde{x}(\omega_{j,b})^{\mathsf{T}} ivech(\theta_b)^{-1}\tilde{x}(\omega_{j,b})\Big)$$

$$\ln l_b(\theta_b, x) = n_b\Big(-p\ln\pi - \ln|S_{\mathbf{X}}| - tr(SS_{\mathbf{X}}^{-1})\Big),$$

$$f(\theta_b, x) = -\ln l_b(\theta_b, x) = \ln|S_{\mathbf{X}}| + tr(SS_{\mathbf{X}}^{-1}) - \ln|S| - p$$

$$\widehat{\theta}_b = \arg\min f(\theta_b, x)$$

where

- n_b are frequencies in a respective band b, $1 \leq b \leq B$,
- $S = (n_b)^{-1}\sum_{j=1}^{n_b} \tilde{x}(\omega_{j,b})\tilde{x}(\omega_{j,b})^{\mathsf{T}}$.

Model Estimation

Normal equations:

$$\frac{\partial f}{\partial \Lambda} = 2S_{\mathbf{X}}^{-1}(S_{\mathbf{X}} - S)S_{\mathbf{X}}^{-1}\tilde{\Lambda}S_f = 0,$$
$$\frac{\partial f}{\partial S_f} = 2\tilde{\Lambda}^{\mathsf{T}}S_{\mathbf{X}}^{-1}(S_{\mathbf{X}} - S)S_{\mathbf{X}}^{-1}\tilde{\Lambda} = 0,$$
$$\frac{\partial f}{\partial S_\varepsilon} = diag(S_{\mathbf{X}}^{-1}(S_{\mathbf{X}} - S)S_{\mathbf{X}}^{-1}) = 0,$$

where

- $S = (n_b)^{-1}\sum_{j=1}^{n_b}\tilde{x}(\omega_{j,b})\tilde{x}(\omega_{j,b})^{\mathsf{T}}$,
- $S_{\mathbf{X}} = S_{\mathbf{X}}(\omega^{(b)})$, $\tilde{\Lambda} = \tilde{\Lambda}(\omega^{(b)})$ and $S_f = S_f(\omega^{(b)})$.

Algorithm

Fletcher and Powell (1963)

0 Choose initial values $\widehat{\gamma_0}^{\mathsf{T}} = [\theta_1^{\mathsf{T}} \theta_2^{\mathsf{T}}]$, where r_1 is dimension of a real part, θ_1, and r_2 is dimension of imaginary part, θ_2.

1 Compute a $(r_1 + r_2)$-dimensional gradient vector g at $\widehat{\gamma_0}$.

2 Update $\widehat{\gamma_0}$ to $\widehat{\gamma_1}$ as follows

$$\widehat{\gamma}_{j+1} = \widehat{\gamma}_j - d_j E_j \mathbf{g}_j,$$

where d is a step-size and E is an approximation to an inverse of the matrix of second derivatives with dimensions $(r_1 + r_2) \times (r_1 + r_2)$.

Simulation Set-Up

0 Choose Σ_f, Σ_ε and $\Lambda \sim U(0,1)$'s.

1 Let $f \sim N(0, \Sigma_f)$ with $\Sigma_f = diag$ and $\varepsilon \sim N(0, \Sigma_\varepsilon)$ with $\Sigma_\varepsilon = 0.01\mathbb{1}_{p\times p}$, simulate time series f and ε with $T = 100$.

2 Check identification scheme.

3 Compute Γ_f, Γ_ε and $S_X(\omega)$.

4 Estimate the change-point(s) t_k.

5 Estimate the model and compare $\widehat{S}_X$ to S_X in each segment.

Simulation Results

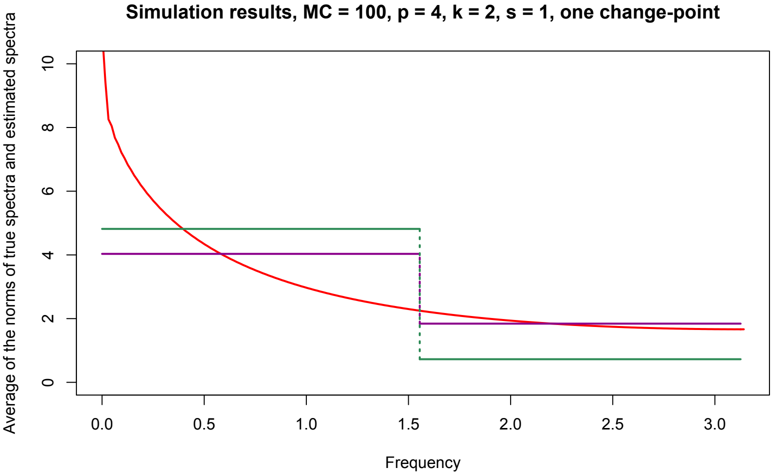

Figure : Averages of the norms of the true (green) and estimated spectra (magenta), true spectrum - red.

Simulation Results

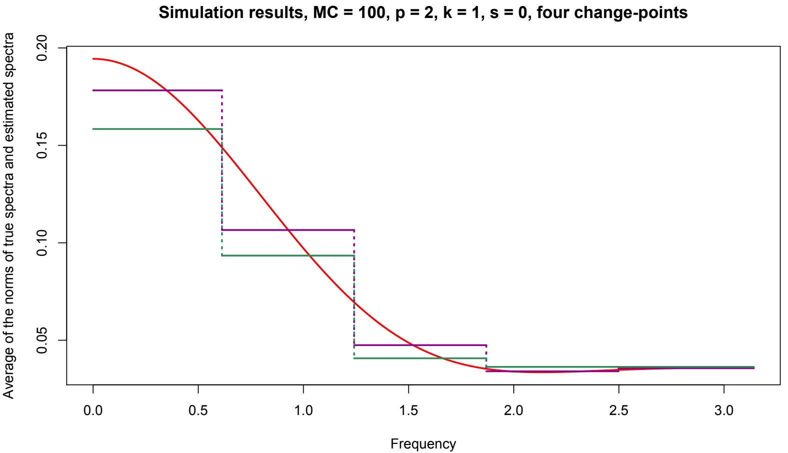

Figure : Averages of the norms of the true (green) and estimated spectra (magenta), true spectrum - red.

Simulation Results

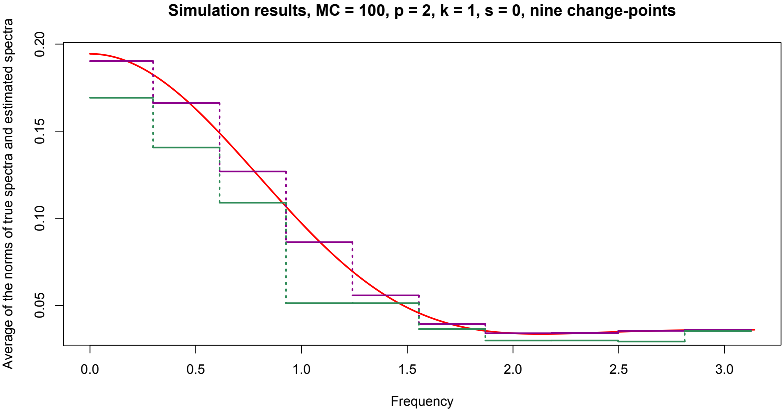

Figure : Averages of the norms of the true (green) and estimated spectra (magenta), true spectrum - red.

Test 1

Example 1: Which of the common factors have a lagged influence on X?

$$W_{ij} = (C_{Dunett}|\tilde{\Lambda_{ij}}|^2)^\top [C_{Dunett}\hat{V}_{|\tilde{\Lambda_{ij}}|^2} C_{Dunett}^\top]^+ (C_{Dunett}|\tilde{\Lambda_{ij}}|^2),$$

where C_{Dunett} is a $(B-1)\times B$ contrast matrix and $\mathbf{W} = (W_{ij} : 1 \leq i \leq p, 1 \leq j \leq k) \sim \chi^2_{B-1}$.

$$H_{ij} : C_{Dunett}|\tilde{\Lambda_{ij}}|^2 = 0.$$

Test 2

Example 2: Which of the specific factors have a non-trivial autocorrelation structure?

$$W_i = (C_{Dunett}\hat{s}_{\varepsilon_i})^{\mathsf{T}}[C_{Dunett}\hat{V}_{\varepsilon_i}C_{Dunett}^{\mathsf{T}}]^{+}(C_{Dunett}\hat{s}_{\varepsilon_i}),$$

where C_{Dunett} is a $(B-1)\times B$ contrast matrix and $\mathbf{W} = (W_i : 1 \leq i \leq p) \sim \chi^2_{B-1}$.

$$H_i : C_{Dunett}\varepsilon_i = 0.$$

Bootstrap

1. Given the data $\mathbf{X} = \mathbf{x}$, calculate in each band Ω_b the quantities $\hat{\vartheta}_b$ and $\hat{V}_b$.
2. For all $1 \leq b \leq B$, generate (pseudo) random numbers which behave like realizations of independent random vectors $Z^*_{1,b}, \ldots, Z^*_{n_b,b} \overset{i.i.d.}{\sim} \mathcal{N}_d(\hat{\vartheta}_b, \hat{V}_b)$.
3. For all $1 \leq b \leq B$, calculate $\hat{\vartheta}^*_b = n_b^{-1} \sum_{j=1}^{n_b} Z^*_{j,b}$ and $\hat{V}^*_b = n_b^{-1} \sum_{j=1}^{n_b} (Z^*_{j,b} - \hat{\vartheta}^*_b)(Z^*_{j,b} - \hat{\vartheta}^*_b)^\top$.
4. Calculate $W^* = N(\hat{\vartheta}^* - \hat{\vartheta})^\top C^\top (C \hat{V}^* C^\top)^+ C(\hat{\vartheta}^* - \hat{\vartheta})$, where $\hat{\vartheta}^*$ and $\hat{V}^*$ are constructed in analogy to $\hat{\vartheta}$ and $\hat{V}$.
5. Repeat steps 2. - 4. M times to obtain M pseudo replicates of W^* and approximate the distribution of W by the empirical distribution of these pseudo replicates.

Summary

1. Numerical comparison of bootstrap procedure with multiplicity-corrected methods.
2. Modelling dynamics of a certain process, i.e. a financial network, with DFM.

References

Dickhaus and Pauly, 2015, Simultaneous Statistical Inference in Dynamic Factor Models, Proceedings ITISE 2015, Granada 1-3 July, 2015.

Fletcher and Powell, 1963, A Rapidly Convergent Descent Method for Minimization, Computer Journal, 6, 163 - 168.

Geweke and Singleton, 1981, Maximum Likelihood "Confirmatory" Factor Analysis of Economic Time Series, International Economic Review, 22/1, 1991, 37 - 54.

References

Geweke, 1977, The Dynamic Factor Analysis of Economic Time-Series Models, in Contributions to Economic Analysis 'Latent Variables in Socio-Economic Models', ed. Aigner and Goldberger, 365 - 383.

Joereskog, 1967, Some Contributions to Maximum Likelihood Factor Analysis, Psychometrika, 32/4, 443 - 482.

Joereskog, 1969, A General Approach to Confirmatory Maximum Likelihood Factor Analysis, Psychometrika, 34/2, 183 - 202.

Lavielle and Ludena, 2000, Bernouilli 6(5), 845 - 869.

References

Reschenhofer, 2008, Frequency Domain Modelling with Piecewise Constant Spectra, Journal of Modern Applied Statistical Methods, 7/2, 467 - 470.

Appendix: Identification

1 $\Sigma_f = \text{diag}$:

Let $\Lambda^{(j)}(\omega) = [\Lambda_1^{(j)} : 0 : \Lambda_2^{(j)}]$ be the $p_j \times k$ submatrix of $\tilde{\Lambda}$ consisting of those rows of $\tilde{\Lambda}$ that have zero elements in the j-th column, where p_j is the number of zeros in the j-th column of $\Lambda(\omega)$. If the $p_j \times (j-1)$ matrices $\Lambda_1^{(j)}(\omega)$ have rank equal to $(j-1)$, for all ω, then $\tilde{\Lambda}$ is identified.

2 $\Sigma_f \neq \text{diag}$:

If the $p_j \times k$ matrices $\Lambda^{(j)}(\omega)$ have rank equal to $(p-1)$ for all ω and j, then $\tilde{\Lambda}(\omega)$ and $S_f(\omega)$ are identified.