Stationary Time Series

Testing and Lag Selection Procedures

N. Bettache[1] C. Butucea[1]

[1]CREST

Half Day Student Talk

Table of Contents

Table of Contents

Setup

Consider X a $p-$dimensional gaussian vector such that $X \sim \mathcal{N}_p(0, \Sigma)$.

- We assume that X is issued from a stationary process.

Setup

Consider X a $p-$dimensional gaussian vector such that $X \sim \mathcal{N}_p(0, \Sigma)$.

- We assume that X is issued from a stationary process.
- Σ has a Toeplitz structure, meaning each descending diagonal from left to right is constant.

Setup

Consider X a $p-$dimensional gaussian vector such that $X \sim \mathcal{N}_p(0, \Sigma)$.

- We assume that X is issued from a stationary process.
- Σ has a Toeplitz structure, meaning each descending diagonal from left to right is constant.
- The covariance matrix Σ has entries $\sigma_{i,j} = \text{Cov}(X^i, X^j) = \sigma_{|i-j|}$.

Observations

Observe repeatedly and independently n samples $(X_1, \dots, X_n)$ of the $\mathbb{R}$-valued time series of length p.
We may consider $n = 1$

Definitions

Consider $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}_p(0, \Sigma)$.

- $\forall i$ we denote by $X_i = (X_i^1, \ldots, X_i^p)$.

Definitions

Consider $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}_p(0, \Sigma)$.

- $\forall i$ we denote by $X_i = (X_i^1, \ldots, X_i^p)$.
- $\Sigma = [\sigma_{|i-j|}]_{1 \le i,j \le p}$.

Definitions

Consider $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}_p(0, \Sigma)$.

- $\forall i$ we denote by $X_i = (X_i^1, \dots, X_i^p)$.
- $\Sigma = [\sigma_{|i-j|}]_{1 \leq i,j \leq p}$.
- Σ belongs to the set of positive definite matrices $\mathcal{S}_p^{++}$.

Toeplitz structure

The information on the Toeplitz matrix is fully contained in the vector $(\sigma_0, \sigma_1, \ldots, \sigma_{p-1})$ of its diagonal values. An empirical estimator can be defined by

- $\forall j \in \{1, \ldots, p-1\}, \; \hat{\sigma}_j = \mathrm{Tr}(A_j \Sigma_n)$ where

CREST
CENTER FOR RESEARCH IN ECONOMICS AND STATISTICS

Toeplitz structure

The information on the Toeplitz matrix is fully contained in the vector $(\sigma_0, \sigma_1, \dots, \sigma_{p-1})$ of its diagonal values. An empirical estimator can be defined by

- $\forall j \in \{1, \dots, p-1\}$, $\hat{\sigma}_j = \mathrm{Tr}(A_j \Sigma_n)$ where
- $\forall (k, l) \in \{1, \dots, p\}^2$, $[A_j]_{k,l} = \frac{1}{2(p-j)} 1[\{|k-l| = j\}]$.
  A_j has 0 entries except on the jth diagonal where it takes the value 1

Toeplitz structure

The information on the Toeplitz matrix is fully contained in the vector $(\sigma_0, \sigma_1, \ldots, \sigma_{p-1})$ of its diagonal values. An empirical estimator can be defined by

- $\forall j \in \{1, \ldots, p-1\}$, $\hat{\sigma}_j = \mathrm{Tr}(A_j \Sigma_n)$ where
- $\forall (k,l) \in \{1, \ldots, p\}^2$, $[A_j]_{k,l} = \frac{1}{2(p-j)} 1[\{|k-l| = j\}]$.
  A_j has 0 entries except on the jth diagonal where it takes the value 1
- $\Sigma_n = \dfrac{1}{n} \sum_{k=1}^{n} X_k X_k^T$ is the empirical covariance matrix

Concentration inequality

Consider $\phi_A : \Sigma \mapsto \mathrm{Tr}(A\Sigma)$ for $A \in \mathcal{S}_p$ and $\Sigma \in \mathcal{M}_p(\mathbb{R})$.

- The random variable $\varphi_A(\Sigma_n - \Sigma)$ is centered and sub-exponential

Concentration inequality

Consider $\phi_A : \Sigma \mapsto \mathrm{Tr}(A\Sigma)$ for $A \in \mathcal{S}_p$ and $\Sigma \in \mathcal{M}_p(\mathbb{R})$.

- The random variable $\varphi_A(\Sigma_n - \Sigma)$ is centered and sub-exponential
- With parameters $\left(\nu^2 = \frac{2||A\Sigma||_F^2}{n(1-K)}, b = \frac{2||A\Sigma||_\infty}{nK}\right)$, for some arbitrary K in $]0,1[$.

Concentration inequality

Consider $\phi_A : \Sigma \mapsto \mathrm{Tr}(A\Sigma)$ for $A \in \mathcal{S}_p$ and $\Sigma \in \mathcal{M}_p(\mathbb{R})$.

- The random variable $\varphi_A(\Sigma_n - \Sigma)$ is centered and sub-exponential
- With parameters $\left(\nu^2 = \frac{2||A\Sigma||_F^2}{n(1-K)}, b = \frac{2||A\Sigma||_\infty}{nK}\right)$, for some arbitrary K in $]0,1[$.
- Consider $t_u = \max\left\{\sqrt{u}\frac{||A\Sigma||_F}{\sqrt{n(1-K)}}, u\frac{||A\Sigma||_\infty}{nK}\right\}$, then:

Concentration inequality

Consider $\phi_A : \Sigma \mapsto \mathrm{Tr}(A\Sigma)$ for $A \in \mathcal{S}_p$ and $\Sigma \in \mathcal{M}_p(\mathbb{R})$.

- The random variable $\varphi_A(\Sigma_n - \Sigma)$ is centered and sub-exponential
- With parameters $\left(\nu^2 = \frac{2||A\Sigma||_F^2}{n(1-K)}, b = \frac{2||A\Sigma||_\infty}{nK}\right)$, for some arbitrary K in $]0,1[$.
- Consider $t_u = \max\left\{\sqrt{u}\frac{||A\Sigma||_F}{\sqrt{n(1-K)}}, u\frac{||A\Sigma||_\infty}{nK}\right\}$, then:
- $$\mathbb{P}[\varphi_A(\Sigma_n - \Sigma) \geq t_u] \leq \exp\left(-\frac{u}{4}\right), \quad u > 0.$$

CREST
CENTER FOR RESEARCH IN ECONOMICS AND STATISTICS

Non parametric testing

- The objective is to test $H_0 : \Sigma = I_p$ against a set of one-sided or two-sided sparse alternatives.

Non parametric testing

- The objective is to test $H_0 : \Sigma = I_p$ against a set of one-sided or two-sided sparse alternatives.
- In the time series setting this hypotheses testing allows to test whether a residual can be considered as a white noise or not.

Non parametric testing

- The objective is to test $H_0 : \Sigma = I_p$ against a set of one-sided or two-sided sparse alternatives.
- In the time series setting this hypotheses testing allows to test whether a residual can be considered as a white noise or not.
- Recall that a test procedure Δ_n is a binary valued random variable $\Delta_n : (\mathbb{R}^p)^{\otimes n} \to \{0, 1\}$.

Two cases

The test procedure needs to be very sensitive both to:

- Moderately sparse case: a relatively large number of very small but significant covariance values.

Two cases

The test procedure needs to be very sensitive both to:

- Moderately sparse case: a relatively large number of very small but significant covariance values.
- Highly sparse case: a small number of significant covariance values.

Test problem

- The test problem definition is

$$H_0 : \Sigma = I_p, \quad \text{vs. } H_1 : \Sigma \in \mathcal{F}_+(s, S, \sigma).$$

Test problem

- The test problem definition is

$$H_0 : \Sigma = I_p, \quad \text{vs. } H_1 : \Sigma \in \mathcal{F}_+(s, S, \sigma).$$

- $\mathcal{F}_+(s, S, \sigma)$ is defined, for $\sigma > 0$ real number and $s \leq S$ integer numbers between 1 and $p - 1$, as the set of sparse Toeplitz covariance matrices Σ such that there are s significantly positive covariance elements with lags no larger than S.

CREST
CENTER FOR RESEARCH IN ECONOMICS AND STATISTICS

Test problem

- The test problem definition is

$$H_0 : \Sigma = I_p, \quad \text{vs. } H_1 : \Sigma \in \mathcal{F}_+(s, S, \sigma).$$

- $\mathcal{F}_+(s, S, \sigma)$ is defined, for $\sigma > 0$ real number and $s \leq S$ integer numbers between 1 and $p - 1$, as the set of sparse Toeplitz covariance matrices Σ such that there are s significantly positive covariance elements with lags no larger than S.
- $$\mathcal{F}_+(s,S,\sigma) = \left\{ \Sigma \in \mathcal{S}_p^{++} \cap \mathcal{T}_p \text{ and there exists } \mathcal{C} \subseteq \{1,\ldots,S\},\ |\mathcal{C}| = s,\ \forall j \in \{1, p-1\},\ \begin{array}{ll} \sigma_j \geq \sigma > 0, & j \in \mathcal{C}, \\ \sigma_j = 0, & j \notin \mathcal{C} \end{array} \right\}.$$

Test procedure

For testing over $\mathcal{F}_+(s, S, \sigma)$, consider for an arbitrary set $\mathcal{C} \subseteq \{1, \ldots, S\}$,

$$Sum_{\mathcal{C}}(\Sigma_n) := \sum_{j \in \mathcal{C}} \mathrm{Tr}(A_j \Sigma_n) = \sum_{j \in \mathcal{C}} \hat{\sigma}_j.$$

For two-sided alternatives it is sufficient to consider $|\sigma_j|$ and $|\hat{\sigma}_j|$ instead of $\hat{\sigma}_j$ in the test statistics.

Moderately sparse case

We consider for some threshold $t_{n,p}^{MS+}$ the test statistic

$$\Delta_n^{MS+} = I\left(Sum_{\{1:S\}}(\Sigma_n - I_p) \geq t_{n,p}^{MS+}\right). \tag{1}$$

Highly sparse case

We consider now for some threshold $t_{n,p}^{HS+}$ the test statistic

$$\Delta_n^{HS+} = \max_{\mathcal{C}\subseteq\{1,\ldots,S\},\#\mathcal{C}=s} I\left(Sum_{\mathcal{C}}(\Sigma_n - I_p) \geq t_{n,p}^{HS+}\right). \qquad (2)$$

The test Δ_n^{HS+} successively tries all possible sets $\mathcal{C}$ of s diagonals among the first S diagonal values. If any of these tests decides to reject H_0, then Δ_n^{HS+} also rejects H_0, otherwise Δ_n^{HS+} doesn't reject the null hypothesis H_0.

Lag selection setup

The objective here is to properly select non-null correlation coefficients. The aim is to find a selector $\widehat{\eta}$ with $\widehat{\eta}_j = 1(|\hat{\sigma}_j| > \tau_n)$ that is consistent in the sense that the risk R^{LS} stays bounded where

$$R^{LS}(\widehat{\eta}, \mathcal{F}) = \sum_{j=1}^{S} \mathbb{E}_{\Sigma}[|\widehat{\eta}_j - \eta_j|].$$

Table of Contents

Non parametric testing

- It separates the set of possible outcomes of some random event in two contiguous sets, we decide to reject H_0 whenever $\Delta_n = 1$ and to accept H_0 whenever $\Delta_n = 0$.

Non parametric testing

- It separates the set of possible outcomes of some random event in two contiguous sets, we decide to reject H_0 whenever $\Delta_n = 1$ and to accept H_0 whenever $\Delta_n = 0$.
- The maximal testing risk is defined as

$$R(\Delta_n, \mathcal{F}_+) = \mathbb{P}_{I_p}(\Delta_n = 1) + \sup_{\Sigma \in \mathcal{F}_+} \mathbb{P}_\Sigma(\Delta_n = 0),$$

Non parametric testing

- It separates the set of possible outcomes of some random event in two contiguous sets, we decide to reject H_0 whenever $\Delta_n = 1$ and to accept H_0 whenever $\Delta_n = 0$.
- The maximal testing risk is defined as

$$R(\Delta_n, \mathcal{F}_+) = \mathbb{P}_{I_p}(\Delta_n = 1) + \sup_{\Sigma \in \mathcal{F}_+} \mathbb{P}_\Sigma(\Delta_n = 0),$$

- A separation rate is the least possible value for $\sigma > 0$ such that the maximal testing risk stays below some prescribed value.

Moderately sparse case

- When the alternative hypothesis is $\mathcal{F}_+(s, S, \sigma)$, we consider for some $t_{n,p}^{MS+}$ the test procedure

$$\Delta_n^{MS+} = I\left(\varphi_{A_{1:S}}(\Sigma_n - I_p) \geq t_{n,p}^{MS+}\right).$$

Moderately sparse case

- When the alternative hypothesis is $\mathcal{F}_+(s, S, \sigma)$, we consider for some $t_{n,p}^{MS+}$ the test procedure

$$\Delta_n^{MS+} = I\left(\varphi_{A_{1:S}}(\Sigma_n - I_p) \geq t_{n,p}^{MS+}\right).$$

- The test Δ_n^{MS+}, with

$$t_{n,p}^{MS+} = \max\left\{\sqrt{\frac{u \cdot S}{n(p-S)}}, \frac{2u \cdot S}{n(p-S)}\right\}$$

for $u > 0$ is such that

$$R(\Delta_n^{MS+}, \mathcal{F}_+) \leq 2\exp\left(-\frac{u}{4}\right)$$

provided that $\sigma \geq \frac{2(s+1)}{s} t_{n,p}^{MS+}$.

Highly sparse case

- Let us consider now for some threshold $t_{n,p}^{HS+}$ the test procedure

$$\Delta_n^{HS+} = \max_{\mathsf{S} \subseteq \{1,\ldots,S\}, \#\mathsf{S}=s} I\left(\varphi_{\mathcal{A}_{\mathsf{S}}}(\Sigma_n - I_p) \geq t_{n,p}^{HS+}\right).$$

Highly sparse case

- Let us consider now for some threshold $t_{n,p}^{HS+}$ the test procedure

$$\Delta_n^{HS+} = \max_{\mathrm{S} \subseteq \{1,\ldots,S\}, \#\mathrm{S}=s} I\left(\varphi_{\mathcal{A}_\mathrm{S}}(\Sigma_n - I_p) \geq t_{n,p}^{HS+}\right).$$

- The test Δ_n^{HS+} successively tries all possible sets S of s diagonals among the first S diagonal values. If any of these tests decides to reject H_0, then Δ_n^{HS+} also rejects H_0, otherwise Δ_n^{HS+} accepts the null hypothesis H_0.

Highly sparse case

- Let us consider now for some threshold $t_{n,p}^{HS+}$ the test procedure

$$\Delta_n^{HS+} = \max_{S \subseteq \{1,\dots,S\}, \# S = s} I\left(\varphi_{\mathcal{A}_S}(\Sigma_n - I_p) \geq t_{n,p}^{HS+}\right).$$

- The test Δ_n^{HS+} successively tries all possible sets S of s diagonals among the first S diagonal values. If any of these tests decides to reject H_0, then Δ_n^{HS+} also rejects H_0, otherwise Δ_n^{HS+} accepts the null hypothesis H_0.
- The test Δ_n^{HS+}, with $t_{n,p}^{HS+} = \max\left\{\sqrt{\frac{4u \cdot s \log\binom{S}{s}}{n(p-S)}}, \frac{8u \cdot s \log\binom{S}{s}}{n(p-S)}\right\}$ for $u > 1$ is such that
$R(\Delta_n^{HS+}, \mathcal{F}^+) \leq \exp\left(-(u-1)\log\binom{S}{s}\right) + \exp\left(-\frac{u}{4}\right)$ provided that
$\sigma \geq \frac{1}{s}\left(t_{n,p}^{HS+} + (2s+1)\max\left\{\sqrt{\frac{u \cdot s}{n(p-S)}}, \frac{2u \cdot s}{n(p-S)}\right\}\right)$

Highly sparse case - computation remark

- When the separation is measured by $\max_C \sum_{j \in C} \sigma_j$, its estimator is known as the scan statistic.

Highly sparse case - computation remark

- When the separation is measured by $\max_C \sum_{j \in C} \sigma_j$, its estimator is known as the scan statistic.
- Note that the computations are not very involved.

Highly sparse case - computation remark

- When the separation is measured by $\max_C \sum_{j \in C} \sigma_j$, its estimator is known as the scan statistic.
- Note that the computations are not very involved.
- After computing $\xi_1 = \varphi_{A_1}(\Sigma_n - I_p), ..., \xi_S = \varphi_{A_S}(\Sigma_n - I_p)$, we sort these values in decreasing order : $\xi_{(1)} \geq \xi_{(2)} \geq ... \geq \xi_{(S)}$

Highly sparse case - computation remark

- When the separation is measured by $\max_C \sum_{j\in C} \sigma_j$, its estimator is known as the scan statistic.
- Note that the computations are not very involved.
- After computing $\xi_1 = \varphi_{A_1}(\Sigma_n - I_p), \ldots, \xi_S = \varphi_{A_S}(\Sigma_n - I_p)$, we sort these values in decreasing order : $\xi_{(1)} \geq \xi_{(2)} \geq \ldots \geq \xi_{(S)}$
- Then

$$\max_{C\subseteq\{1,\ldots,S\},\#C=s} \sum_{j\in C} \varphi_{A_j}(\Sigma_n - I_p) = \xi_{(1)} + \ldots + \xi_{(s)}$$

Lag selection theorem

If Σ belongs to $\mathcal{F}(s, S, \sigma)$, with $\sigma \geq 2\tau_n$, the selector $\hat{\eta}$ with
$a = \left(\sqrt{\log(s)} + \sqrt{\log(S-s)}\right)\sqrt{u\frac{2s+1}{n(p-S)}}$, $b = 2u\log(s(S-s))\frac{2s+1}{n(p-S)}$,

$$\tau_n = \max\{a, b\}$$

for $u > 1$ is such that

$$R_{LS}(\hat{\eta}, \mathcal{F}) \leq 2\exp\left(-(u-1)\frac{\log(s)}{4}\right) + 2\exp\left(-(u-1)\frac{\log(S-s)}{4}\right).$$

Table of Contents

Vector-valued time series

- Consider X a generic $d \times p-$dimensional matrix having a matrix normal distribution $X \sim \mathcal{MN}_{d\times p}(0, \Sigma_L, \Sigma_R)$.

- Consider X a generic $d \times p-$dimensional matrix having a matrix normal distribution $X \sim \mathcal{MN}_{d \times p}(0, \Sigma_L, \Sigma_R)$.
- This is equivalent to saying that $vec(X)$ has multivariate normal distribution $\mathcal{N}_{dp}(0, \Sigma_R \otimes \Sigma_L)$.

Vector-valued time series

- Consider X a generic $d \times p-$dimensional matrix having a matrix normal distribution $X \sim \mathcal{MN}_{d\times p}(0, \Sigma_L, \Sigma_R)$.
- This is equivalent to saying that $vec(X)$ has multivariate normal distribution $\mathcal{N}_{dp}(0, \Sigma_R \otimes \Sigma_L)$.
- Then we have a column covariance matrix $\mathbb{E}[XX^T] = \mathrm{Tr}(\Sigma_R)\Sigma_L$ and a row covariance matrix $\mathbb{E}[X^TX] = \mathrm{Tr}(\Sigma_L)\Sigma_R$.

We assume we observe repeatedly and independently n samples $(X_1, \dots, X_n)$ of the $\mathbb{R}^d$-valued time series of length p.
The goal is to derive similar results as in the real-valued case.