18th International Joint Conference CFE-CMStatistics Computational and Methodological Statistics (CMStatistics 2024) Computational and Financial Econometrics (CFE 2024)

14-16 December 2024, King's College London, UK



BLOCK TESTING IN COVARIANCE AND PRECISION MATRICES FOR FUNCTIONAL DATA ANALYSIS

Marie Morvan¹, **Alessia Pini**², Madison Giacofci³, Valerie Monbet¹

¹ Dept. of Mathematics, Université Rennes 1

² Dept. of Statistical Sciences, Università Cattolica del Sacro Cuore, Milan, Italy

³ Dept. of Mathematics, Université Rennes 2











Valerie Monbet, Marie Morvan Université de Rennes 1



Madison Giacofci Université de Rennes 2

RENNES 2



- **Infrared spectroscopy**: a technique used to study the composition of a mixture.
- Data are infrared spectra, which can be seen as continuous functions of the wavelength.
- The different parts of the signal are related to the mixture components.
- We are particularly interested in the dependency structure of data along the domain, to understand if the different components are related between each other.





MOTIVATION

• Aim: to test (conditional) independence between different parts of the functions' domain.

Example: fruit purees spectra





MOTIVATION

- Aim: to test (conditional) independence between different parts of the functions' domain.
- Assumption 1: the domain can be partitioned into regions of interest.

Example: fruit purees spectra





- Aim: to test (conditional) independence between different parts of the functions' domain.
- Assumption 1: the domain can be partitioned into regions of interest.
- Assumption 2: data can be described with a B-splines basis expansion.
 - Coefficients are directly related to the parts of the domain.
 - A significant dependence between basis coefficients translates into a significant dependence between parts of the domain.

Example: fruit purees spectra



Coefficients of the basis expansion



- Aim: to test (conditional) independence between different parts of the functions' domain.
- Assumption 1: the domain can be partitioned into regions of interest.
- Assumption 2: data can be described with a B-splines basis expansion.
 - Coefficients are directly related to the parts of the domain.
 - A significant dependence between basis coefficients translates into a significant dependence between parts of the domain.
- We look at the block structure of the covariance matrix (precision matrix) of the basis coefficients.

Example: fruit purees spectra



Coefficients of the basis expansion



INTUITION



Covariance matrix

1:150



INTUITION



Precision matrix



100

80

120

140

1e+08

8e+07

6e+07

- 4e+07

2e+07

0e+00

-2e+07

-4e+07



THE MODEL

Functional data. We assume to observe functional data in $L^2(D)$ with $D \subset \mathbb{R}$

$$\mathbf{X}_1(t), \dots, \mathbf{X}_n(t)$$
 $t \in D, \mathbf{X}_i \in L^2(D)$

Basis expansion.

$$\mathbf{X}_{i}(t) = \sum_{j=1}^{p} \phi_{j}(t) C_{ij} \qquad (\mathbf{X}_{1}, \dots, \mathbf{X}_{n}) \Rightarrow (\mathbf{C}_{1}, \dots, \mathbf{C}_{n})$$

Blocks. The coefficients $\{C_{ij}\}$ are partitioned into M blocks, associated to different RoI:

$$\mathcal{J}_1,\ldots,\mathcal{J}_M\subset\{1,\ldots,p\}$$



THE MODEL

Functional data. We assume to observe functional data in $L^2(D)$ with $D \subset \mathbb{R}$

$$\mathbf{X}_1(t), \dots, \mathbf{X}_n(t)$$
 $t \in D, \mathbf{X}_i \in L^2(D)$

Basis expansion.

$$\mathbf{X}_{i}(t) = \sum_{j=1}^{p} \phi_{j}(t) C_{ij} \qquad (\mathbf{X}_{1}, \dots, \mathbf{X}_{n}) \Rightarrow (\mathbf{C}_{1}, \dots, \mathbf{C}_{n})$$

Blocks. The coefficients $\{C_{ij}\}$ are partitioned into M blocks, associated to different RoI:

$$\mathcal{J}_1,\ldots,\mathcal{J}_M\subset\{1,\ldots,p\}$$



Permutation test of linear independence between coefficients in blocks \mathcal{J}_m and $\mathcal{J}_{m'}$:

 $H_{0,m,m'}: \mathbf{C}_{\mathcal{J}'_m}, \mathbf{C}_{\mathcal{J}'_m}$ are independent

 $H_{1,m,m'}: \Sigma_{\mathcal{J}_m \times \mathcal{J}_m'} \neq 0$



Permutation test of linear independence between coefficients in blocks \mathcal{J}_m and $\mathcal{J}_{m'}$:

 $H_{0,m,m'}: \mathbf{C}_{\mathcal{J}'_m}, \mathbf{C}_{\mathcal{J}'_m}$ are independent

$$H_{1,m,m'}: \Sigma_{\mathcal{J}_m \times \mathcal{J}_m'} \neq 0$$

• Test statistic:

$$T_{mm'} = \sum_{j \in \mathcal{J}_m, j' \in \mathcal{J}_{m'}} \left(\frac{\widehat{\Sigma}_{jj'}}{\sqrt{\widehat{\Sigma}_{jj}\widehat{\Sigma}_{j'j'}}} \right)^2$$



Permutation test of linear independence between coefficients in blocks \mathcal{J}_m and $\mathcal{J}_{m'}$:

- $H_{0,m,m'}: \mathbf{C}_{\mathcal{J}'_m}, \mathbf{C}_{\mathcal{J}'_m}$ are independent $H_{1,m,m'}: \Sigma_{\mathcal{J}_m \times \mathcal{J}'_m} \neq 0$
- Test statistic:

$$T_{mm'} = \sum_{j \in \mathcal{J}_m, j' \in \mathcal{J}_{m'}} \left(\frac{\widehat{\Sigma}_{jj'}}{\sqrt{\widehat{\Sigma}_{jj} \widehat{\Sigma}_{j'j'}}} \right)^2$$

• **Permutations:** n! permutations of the coefficients C_{ij} $j \in \mathcal{J}_m$ keeping C_{ij} $j \in \mathcal{J}_{m'}$ fixed, (or vice-versa).





Permutation test of linear independence between coefficients in blocks \mathcal{J}_m and $\mathcal{J}_{m'}$:

 $H_{0,m,m'}: \mathbf{C}_{\mathcal{J}'_m}, \mathbf{C}_{\mathcal{J}'_m}$ are independent $H_{1,m,m'}: \Sigma_{\mathcal{J}_m \times \mathcal{J}'_m} \neq 0$

Test statistic:

$$T_{mm'} = \sum_{j \in \mathcal{J}_m, j' \in \mathcal{J}_{m'}} \left(\frac{\widehat{\Sigma}_{jj'}}{\sqrt{\widehat{\Sigma}_{jj} \widehat{\Sigma}_{j'j'}}} \right)^2$$

- **Permutations:** n! permutations of the coefficients C_{ij} $j \in \mathcal{J}_m$ keeping $C_{ij} \ j \in \mathcal{J}_{m'}$ fixed, (or vice-versa).
- p-value:

$$p_{m,m'} = \frac{\#\{T^*_{mm'} \ge T_{mm'}\}}{n!} \qquad \text{Test statistic evaluated on original data}$$

Test statistic evaluated on permuted data

data



THEORETICAL PROPERTIES: TEST ON COVARIANCE MATRIX

• The test of linear independence is **exact**, but it might be **not consistent under general alternatives** (e.g., when the dependence is not linear):

$$\mathbb{P}_{H_{0,m,m'}}[p_{m,m'} \leq \alpha] = \alpha$$
$$\lim_{n \to \infty} \mathbb{P}_{H_{1,m,m'}}[p_{m,m'} \leq \alpha] = 1$$
$$\lim_{n \to \infty} \mathbb{P}_{H_{0,m,m'}}[p_{m,m'} \leq \alpha] \leq 1$$

• For Gaussian data: the test of linear independence is a test for independence. It is **exact** for any sample size and **consistent** under every alternative.

$$\mathbb{P}_{H_{0,m,m'}}[p_{m,m'} \leq \alpha] = \alpha$$
$$\lim_{n \to \infty} \mathbb{P}_{H_{1,m,m'}}[p_{m,m'} \leq \alpha] = 1$$
$$\lim_{n \to \infty} \mathbb{P}_{H_{0,m,m'}}[p_{m,m'} \leq \alpha] = 1$$



Permutation test of conditional linear independence between coefficients in blocks \mathcal{J}_m and $\mathcal{J}_{m'}$:

$$H_{0,m,m'}: \begin{cases} \mathbf{C}_{\mathcal{J}_m} = \mathbf{C}_{\{1,\dots,p\} \setminus (\mathcal{J}_m \cup \mathcal{J}_{m'})} \mathbf{A} + \boldsymbol{\varepsilon}_{\mathcal{J}_m} \\ \mathbf{C}_{\mathcal{J}_{m'}} = \mathbf{C}_{\{1,\dots,p\} \setminus (\mathcal{J}_m \cup \mathcal{J}_{m'})} \mathbf{A}' + \boldsymbol{\varepsilon}_{\mathcal{J}_{m'}} \\ \boldsymbol{\varepsilon}_{\mathcal{J}_m}, \boldsymbol{\varepsilon}_{\mathcal{J}_{m'}} & \text{independent} \end{cases} \qquad H_{1,m,m'} = \boldsymbol{\Omega}_{\mathcal{J}_m \times \mathcal{J}_{m'}} \neq 0$$



Permutation test of conditional linear independence between coefficients in blocks \mathcal{J}_m and $\mathcal{J}_{m'}$:

$$H_{0,m,m'}: \begin{cases} \mathbf{C}_{\mathcal{J}_m} = \mathbf{C}_{\{1,\dots,p\} \setminus (\mathcal{J}_m \cup \mathcal{J}_{m'})} \mathbf{A} + \boldsymbol{\varepsilon}_{\mathcal{J}_m} \\ \mathbf{C}_{\mathcal{J}_{m'}} = \mathbf{C}_{\{1,\dots,p\} \setminus (\mathcal{J}_m \cup \mathcal{J}_{m'})} \mathbf{A}' + \boldsymbol{\varepsilon}_{\mathcal{J}_{m'}} \\ \boldsymbol{\varepsilon}_{\mathcal{J}_m}, \boldsymbol{\varepsilon}_{\mathcal{J}_{m'}} & \text{independent} \end{cases} \qquad H_{1,m,m'} = \boldsymbol{\Omega}_{\mathcal{J}_m \times \mathcal{J}_{m'}} \neq 0$$

• Test statistic:

$$T_{mm'} = \sum_{j \in \mathcal{J}_m, j' \in \mathcal{J}_{m'}} \left(\widehat{\Omega}_{jj'}\right)^2$$

• Xia Y., Cai T., Cai T.T. Multiple testing of submatrices of a precision matrix with applications to identification of between pathway interactions J. Am Stat. Assoc., 113(521), 328–339 (2018).



Permutation test of conditional linear independence between coefficients in blocks \mathcal{J}_m and $\mathcal{J}_{m'}$:

$$H_{0,m,m'}: \begin{cases} \mathbf{C}_{\mathcal{J}_m} = \mathbf{C}_{\{1,\dots,p\} \setminus (\mathcal{J}_m \cup \mathcal{J}_{m'})} \mathbf{A} + \boldsymbol{\varepsilon}_{\mathcal{J}_m} \\ \mathbf{C}_{\mathcal{J}_{m'}} = \mathbf{C}_{\{1,\dots,p\} \setminus (\mathcal{J}_m \cup \mathcal{J}_{m'})} \mathbf{A}' + \boldsymbol{\varepsilon}_{\mathcal{J}_{m'}} \\ \boldsymbol{\varepsilon}_{\mathcal{J}_m}, \boldsymbol{\varepsilon}_{\mathcal{J}_{m'}} & \text{independent} \end{cases} \qquad H_{1,m,m'} = \boldsymbol{\Omega}_{\mathcal{J}_m \times \mathcal{J}_{m'}} \neq 0$$

• Test statistic:

$$T_{mm'} = \sum_{j \in \mathcal{J}_m, j' \in \mathcal{J}_{m'}} \left(\widehat{\mathbf{\Omega}}_{jj'}\right)^2$$

• **Permutations**: n! permutations of the coefficients C_{ij} $j \in \mathcal{J}_m$ keeping C_{ij} $j \in \mathcal{J}_{m'}$ fixed, (or vice-versa).





Permutation test of conditional linear independence between coefficients in blocks \mathcal{J}_m and $\mathcal{J}_{m'}$:

$$H_{0,m,m'}: \begin{cases} \mathbf{C}_{\mathcal{J}_m} = \mathbf{C}_{\{1,\dots,p\} \setminus (\mathcal{J}_m \cup \mathcal{J}_{m'})} \mathbf{A} + \boldsymbol{\varepsilon}_{\mathcal{J}_m} \\ \mathbf{C}_{\mathcal{J}_{m'}} = \mathbf{C}_{\{1,\dots,p\} \setminus (\mathcal{J}_m \cup \mathcal{J}_{m'})} \mathbf{A}' + \boldsymbol{\varepsilon}_{\mathcal{J}_{m'}} \\ \boldsymbol{\varepsilon}_{\mathcal{J}_m}, \boldsymbol{\varepsilon}_{\mathcal{J}_{m'}} & \text{independent} \end{cases} \qquad H_{1,m,m'} = \boldsymbol{\Omega}_{\mathcal{J}_m \times \mathcal{J}_{m'}} \neq 0$$

• Test statistic:

$$T_{mm'} = \sum_{j \in \mathcal{J}_m, j' \in \mathcal{J}_{m'}} \left(\widehat{\mathbf{\Omega}}_{jj'}\right)^2$$

Under the null hypothesis the errors of those two linear models are exchangeable:

$$egin{aligned} \mathbf{C}_{\mathcal{J}_m} &= \mathbf{C}_{\{1,...,p\} \setminus (\mathcal{J}_m \cup \mathcal{J}_{m'})} \mathbf{A} + oldsymbol{arsigma}_{\mathcal{J}_m} \ \mathbf{C}_{\mathcal{J}_{m'}} &= \mathbf{C}_{\{1,...,p\} \setminus (\mathcal{J}_m \cup \mathcal{J}_{m'})} \mathbf{A}' + oldsymbol{arsigma}_{\mathcal{J}_{m'}} \end{aligned}$$

• **Permutations:** n! permutations of the residuals $\hat{\varepsilon}_{ij}$ $j \in \mathcal{J}_m$ keeping $\hat{\varepsilon}_{ij}$ $j \in \mathcal{J}_{m'}$ fixed, (or vice-versa).



Permutation test of conditional linear independence between coefficients in blocks \mathcal{J}_m and $\mathcal{J}_{m'}$:

$$H_{0,m,m'}: \begin{cases} \mathbf{C}_{\mathcal{J}_m} = \mathbf{C}_{\{1,\dots,p\} \setminus (\mathcal{J}_m \cup \mathcal{J}_{m'})} \mathbf{A} + \boldsymbol{\varepsilon}_{\mathcal{J}_m} \\ \mathbf{C}_{\mathcal{J}_{m'}} = \mathbf{C}_{\{1,\dots,p\} \setminus (\mathcal{J}_m \cup \mathcal{J}_{m'})} \mathbf{A}' + \boldsymbol{\varepsilon}_{\mathcal{J}_{m'}} \\ \boldsymbol{\varepsilon}_{\mathcal{J}_m}, \boldsymbol{\varepsilon}_{\mathcal{J}_{m'}} & \text{independent} \end{cases} \qquad H_{1,m,m'} = \boldsymbol{\Omega}_{\mathcal{J}_m \times \mathcal{J}_{m'}} \neq 0$$

• Test statistic:

$$T_{mm'} = \sum_{j \in \mathcal{J}_m, j' \in \mathcal{J}_{m'}} \left(\widehat{\Omega}_{jj'}\right)^2$$

- **Permutations:** n! permutations of the residuals $\widehat{\boldsymbol{\varepsilon}}_{ij}$ $j \in \mathcal{J}_m$ keeping $\widehat{\boldsymbol{\varepsilon}}_{ij}$ $j \in \mathcal{J}_{m'}$ fixed, (or vice-versa).
- p-value:

$$p = \frac{\#\{T_{mm'}^* \ge T_{mm'}\}}{n!}$$



THEORETICAL PROPERTIES: TEST ON PRECISION MATRIX

• The test of conditional linear independence is **asymptotically exact** and **consistent**. However, only linear conditional independence is tested (which can not be generalized to conditional independence).

$$\lim_{n \to \infty} \mathbb{P}_{H_{0,m,m'}}[p_{m,m'} \le \alpha] = \alpha$$
$$\lim_{n \to \infty} \mathbb{P}_{H_{1,m,m'}}[p_{m,m'} \le \alpha] = 1$$

• For Gaussian data: the test of conditional linear independence is a test for conditional independence. It is **asymptotically exact** and **consistent** under every alternative.



The total number of tests is **M(M-1)/2**. So, we need to adjust p-values.

Available methods include Bonferroni adjustment (Holm, 1976), Closed Testing Procedure (Marcus et al, 1976), or methods specifically designed for the precision matrix (Xia et al. 2018):



generally very conservative;

do not exploit the proximity structure between blocks.

Instead, we generalize the interval-wise testing procedure (Pini and Vantini, 2016, 2017), a method to adjust for multiplicity controlling the FWER over closed sub-intervals of the domain.

[•] Holm, S.: A simple sequentially rejective multiple test procedure Scand. J. Stat., 6(2), 65–70 (1979)

[•] Marcus, R., Eric, P., & Gabriel, K. R. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, *63*(3), 655-660 (1976).

[•] Pini, A. and Vantini, S. The interval testing procedure: a general framework for inference in functional data analysis Biometrics 72(3) 835–845 (2016).

[•] Pini, A., Vantini, S. Interval-wise testing for functional data. J. Nonparametr. Stat. 29(2), 407–424 (2017).

[•] Xia Y., Cai T., Cai T.T. Multiple testing of submatrices of a precision matrix with applications to identification of between pathway interactions J. Am Stat. Assoc., 113(521), 328–339 (2018).



- 1. Perform a test on each couple of blocks.
- 2. Perform a tests on each couple of non-overlapping intervals of blocks.
- 3. For each couple of blocks \mathcal{J}_m and $\mathcal{J}_{m'}$, compute the adjusted *p*-value as the maximum between all *p*-values of tests of intervals of blocks including \mathcal{J}_m and $\mathcal{J}_{m'}$ (including blocks \mathcal{J}_m and $\mathcal{J}_{m'}$ themselves).

$$\tilde{p}_{m,m'} = \max_{\substack{\mathcal{I}:m\in\mathcal{I}\\\mathcal{I}':m'\in\mathcal{I}'}} p_{\mathcal{I},\mathcal{I}'}$$



Example: three RoI

• Performed tests:





Example: three RoI

- Performed tests
- Adjusted p-value for couple (1,2):





THEORETICAL PROPERTIES OF THE PROCEDURE

Theorem 1 The adjusted p-value $\tilde{p}_{m,m'}$ is provided with the following error control: for all non-overlapping intervals of blocks $\mathcal{J}_{\mathcal{I}} = \bigcup_{i=m}^{m+h} \{\mathcal{J}_i\}$ and $\mathcal{J}_{\mathcal{I}'} = \bigcup_{i=m'}^{m'+h'} \{\mathcal{J}_i\}$, with $1 \leq m \leq m+h < m' \leq m'+h' \leq M$, $\forall \alpha \in (0,1)$

if
$$H_{0,m,m'}$$
 is true $\forall (m,m') \in \mathcal{I} \times \mathcal{I}',$
$$\limsup_{n \to \infty} \mathbb{P} \left[\exists (m,m') \in \mathcal{I} \times \mathcal{I}' : \tilde{p}_{m,m'} \leq \alpha \right] \leq \alpha.$$

Theorem 2 The adjusted p-value $\tilde{p}_{m,m'}$ is consistent, that is: for all nonoverlapping intervals of blocks $\mathcal{J}_{\mathcal{I}} = \bigcup_{i=m}^{m+h} \{\mathcal{J}_i\}$ and $\mathcal{J}_{\mathcal{I}'} = \bigcup_{i=m'}^{m'+h'} \{\mathcal{J}_i\}$, with $1 \leq m \leq m+h < m' \leq m'+h' \leq M, \forall \alpha \in (0,1)$

if
$$H_{0,m,m'}$$
 is false $\forall (m,m') \in \mathcal{I} \times \mathcal{I}',$
$$\lim_{n \to \infty} \mathbb{P} \left[\forall (m,m') \in \mathcal{I} \times \mathcal{I}' : \tilde{p}_{m,m'} \leq \alpha \right] = 1.$$



FRUITS PUREES DATASET INDEPENDENCE



Adjusted p-value – covariance matrix



FRUITS PUREES DATASET CONDITIONAL INDEPENDENCE



Adjusted p-value - precision matrix



DISCUSSION

- Two methods for performing inference on the covariance structure of functional data:
 - A test of linear independence.
 - test of conditional linear independence.
- A novel way to adjust for multiplicity when testing blocks of the covariance/precision matrix.
- The adjustment method can be plugged-in with every available testing procedure.



Future works

- How to define blocks when they are not available?
- How to extend the procedure without relying on a basis expansion?



SELECTED REFERENCES

- Briandet, R., Kemsley, E. K., & Wilson, R. H. Approaches to adulteration detection in instant coffees using infrared spectroscopy and chemometrics. *Journal of the Science of Food and Agriculture*, *71*(3), 359-366 (1996).
- Freedman, D., Lane, D. A nonstochastic interpretation of reported significance levels Journal of Business & Economic Statistics, 1(4), 292–298 (1983).
- Holm, S.: A simple sequentially rejective multiple test procedure Scand. J. Stat., 6(2), 65–70 (1979)
- Marcus, R., Eric, P., & Gabriel, K. R. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, *63*(3), 655-660 (1976).
- Pini, A. and Vantini, S. The interval testing procedure: a general framework for inference in functional data analysis Biometrics 72(3) 835–845 (2016).
- Pini, A., Vantini, S. Interval-wise testing for functional data. J. Nonparametr. Stat. 29(2), 407–424 (2017).
- Xia Y., Cai T., Cai T.T. Multiple testing of submatrices of a precision matrix with applications to identification of between pathway interactions J. Am Stat. Assoc., 113(521), 328–339 (2018).