Large sample distribution for fully functional periodicity tests

Siegfried Hörmann

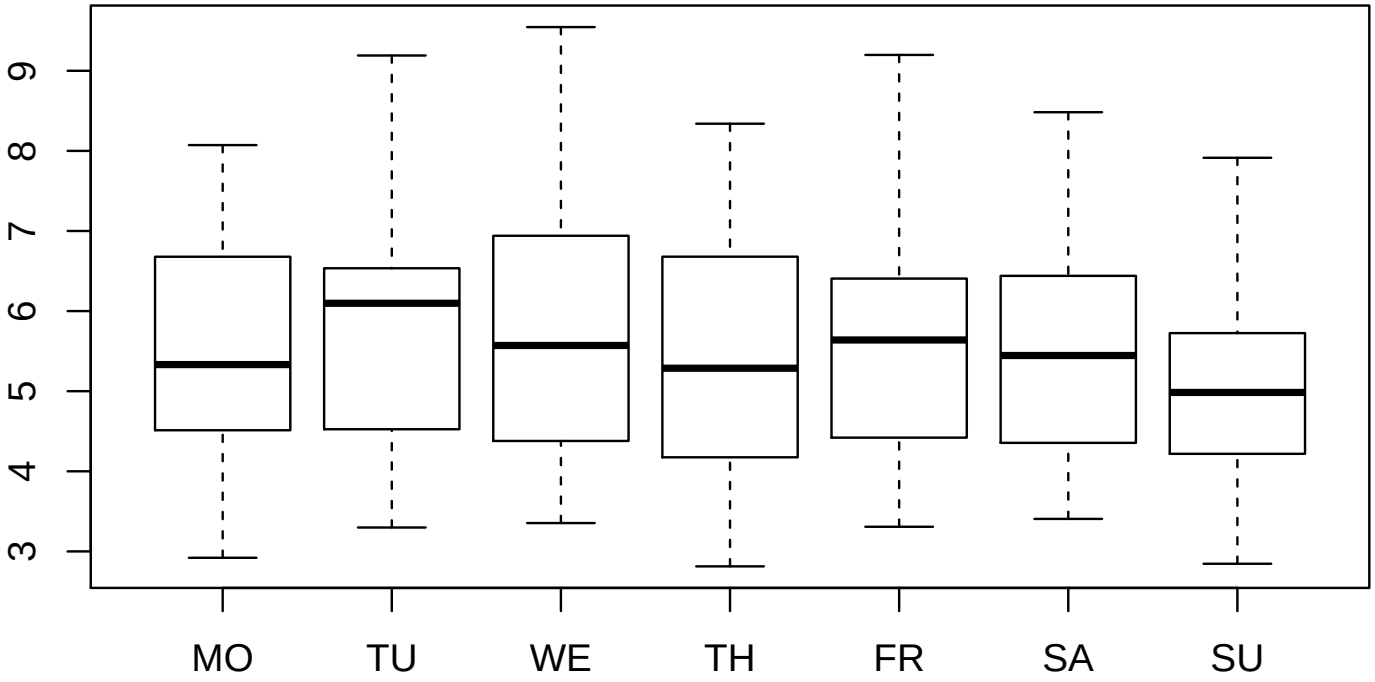
Institute for Statistics
Graz University of Technology

Based on joint work with
Piotr Kokoszka (Colorado State) and Gilles Nisol (ULB)
Clément Cerovecki (ULB) and Vaidotas Characiejus (ULB)

Sept 21, 2018

Pollution data

$N = 167$ observations of `PM10` measured in Graz (Austria) from 10/1/15 to 03/15/16.



Is there a significant difference in the daily average w.r.t. day of week?

Pollution data

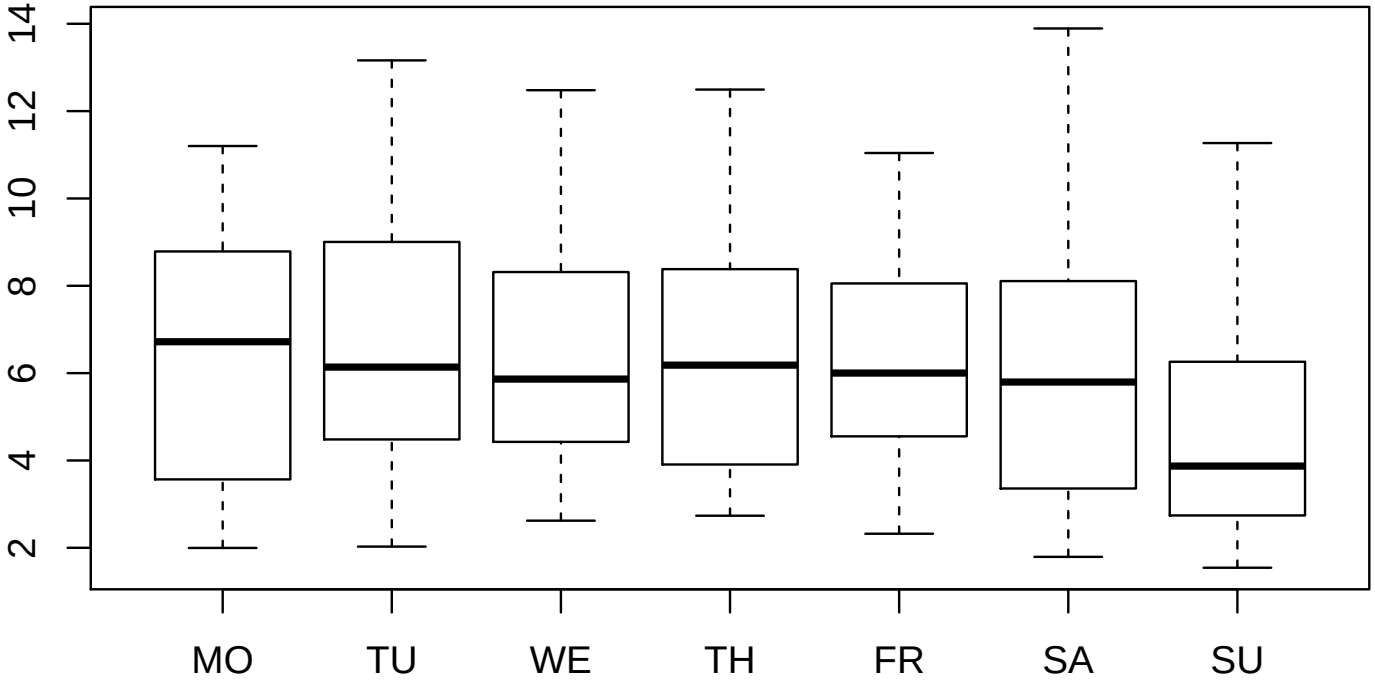
We use ANOVA:

$$\mathcal{H}_0 \colon \mu_{\mathrm{MO}} = \cdots = \mu_{\mathrm{SU}} \quad \text{v.s.} \quad \mathcal{H}_A \colon \mathcal{H}_0 \text{ doesn't hold.}$$

We obtain p-value of 0.72 $\Longrightarrow$ do not reject $\mathcal{H}_0$.

Pollution data

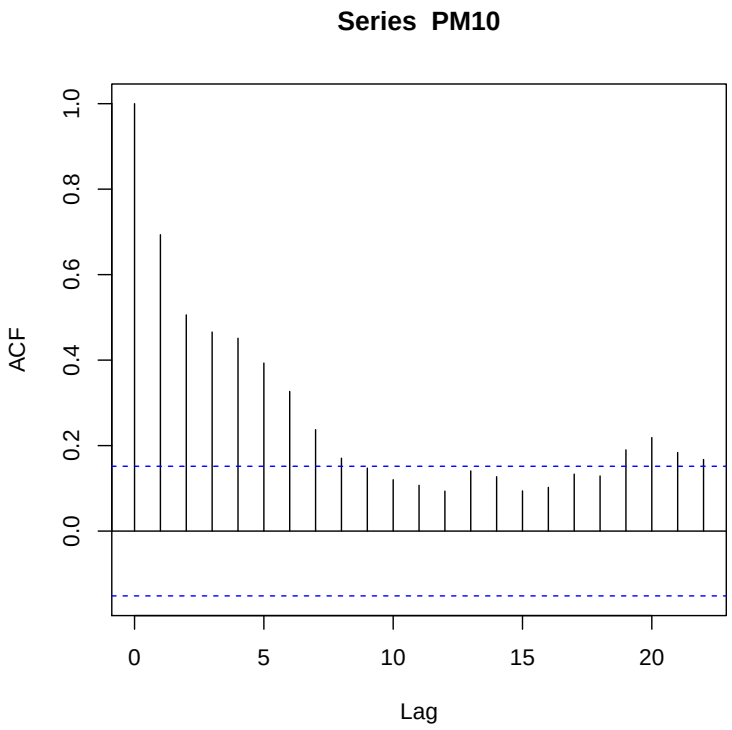
Same time period with `NO` (nitrogen monoxide).

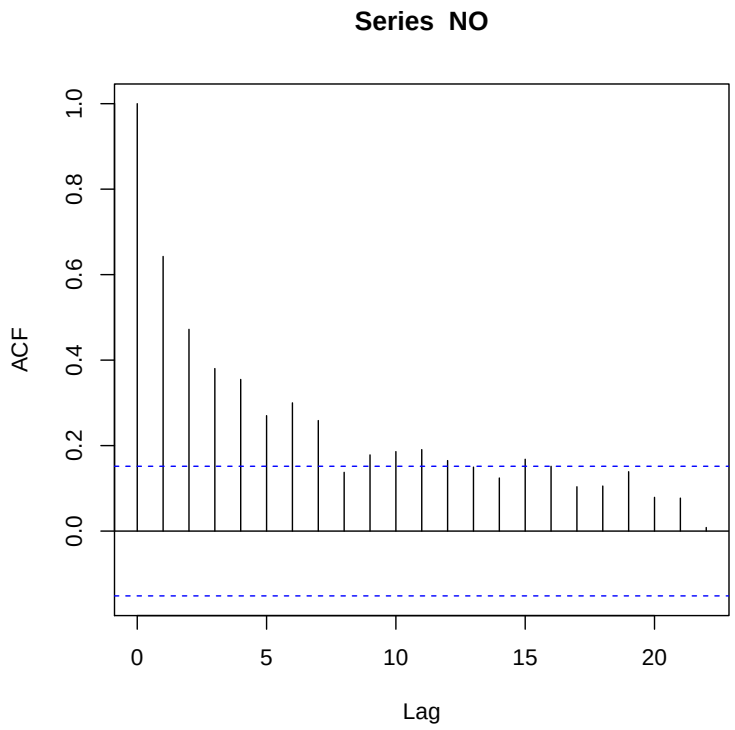


We obtain p-value of 0.25 $\Longrightarrow$ do not reject $\mathcal{H}_0$.

Dependence

Autocorrelation functions of daily averages of `PM10` and `NO`.



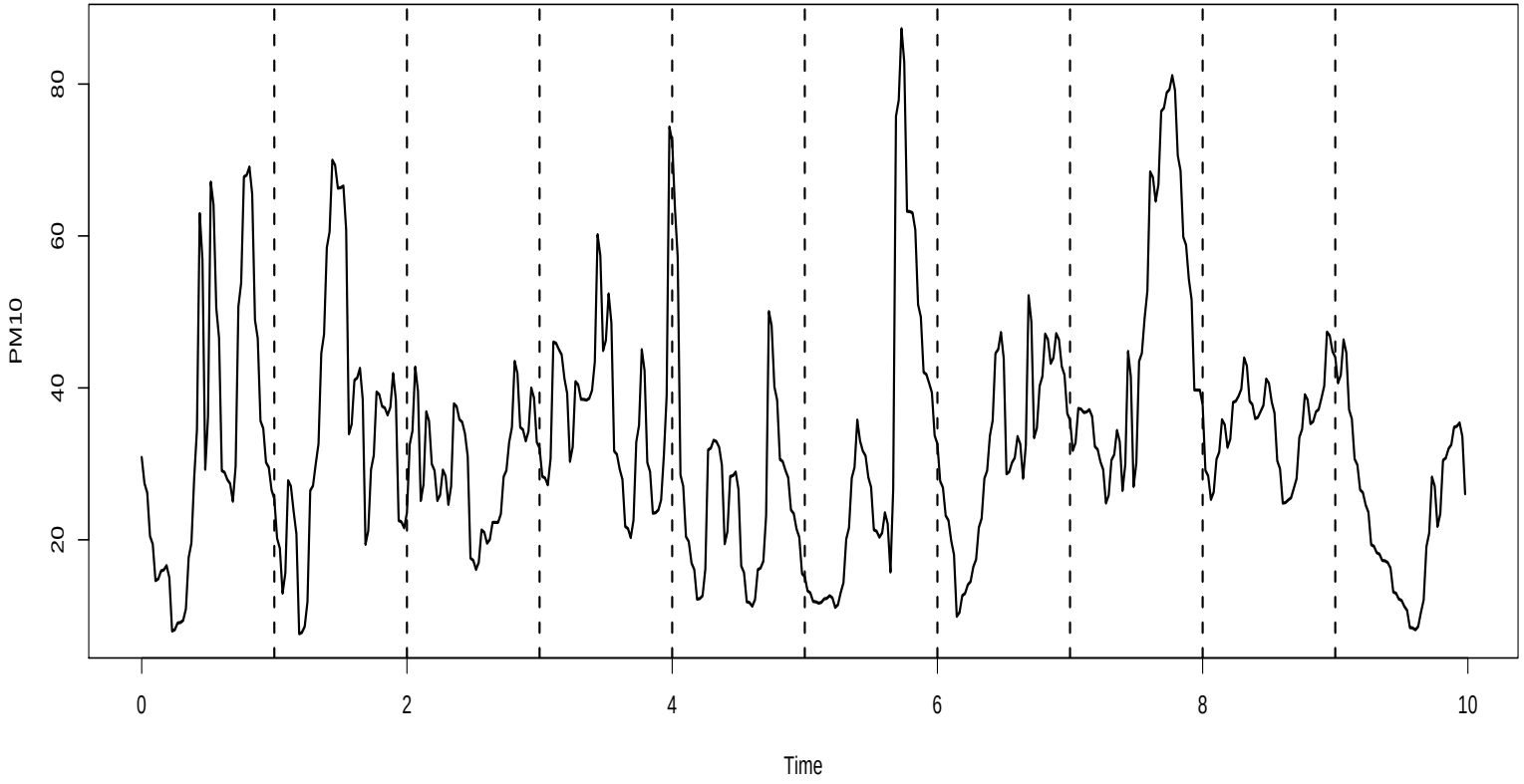


Functional data

PM10–data in Graz are available in 30 minutes resolution.

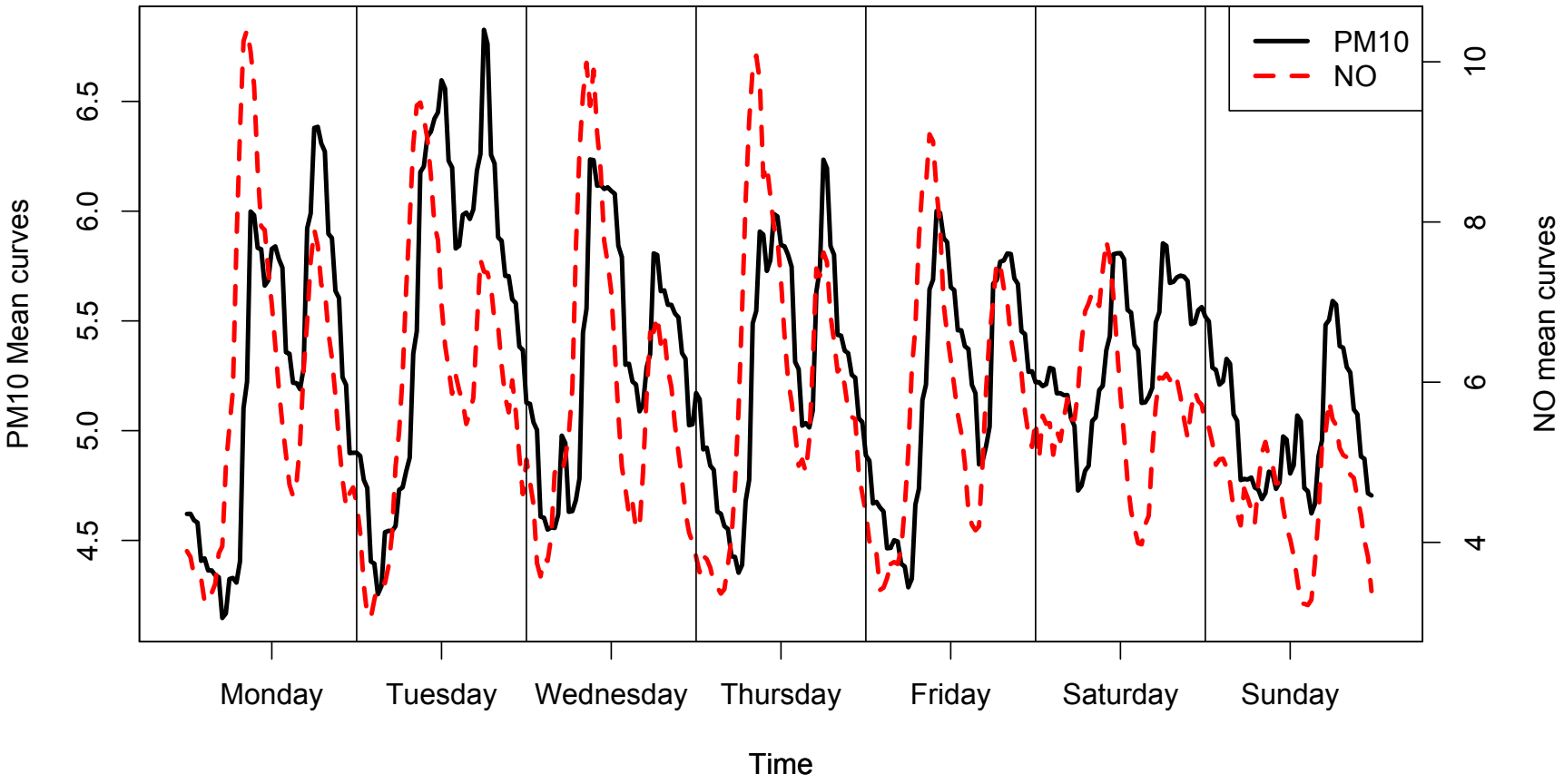
$\Longrightarrow$ Functional time series

Goal: check whether there is underlying periodic signal



Weekday averages

Intraday mean curves from half-hourly observations.



Three models for periodic functional data

$$Y_t(u) = \mu(u) + [\alpha \cos(t\theta) + \beta \sin(t\theta)] \times w_0(u) \quad + Z_t(u), \quad (1)$$

$$Y_t(u) = \mu(u) + s_t \times w_0(u) \quad + Z_t(u), \quad (2)$$

$$Y_t(u) = \mu(u) + w_t(u) \quad + Z_t(u), \quad (3)$$

with

$$s_t = s_{t+d} \quad \text{and} \quad w_t(u) = w_{t+d}(u)$$

and

$$\sum_{t=1}^{d} s_t = 0 \quad \text{and} \quad \sum_{t=1}^{d} w_t(u) = 0.$$

Settings

Part 1: θ is known.

Part 2: θ is unknown.

Noise: We start with

$$Z_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Gamma).$$

Later we generalize to very general stationary processes.

Projections

Choose a set of functions $v_1(u), \ldots, v_p(u)$ and compute

$$Y_{tk} = \int Y_t(u) v_k(u) du,$$

which gives rise to the vector models

$$\boldsymbol{Y}_t = \boldsymbol{\mu} + [\alpha \cos(t\theta) + \beta \sin(t\theta)] \times \boldsymbol{w}_0 + \boldsymbol{Z}_t;$$

$$\boldsymbol{Y}_t = \boldsymbol{\mu} + s_t \times \boldsymbol{w}_0 + \boldsymbol{Z}_t;$$

$$\boldsymbol{Y}_t = \boldsymbol{\mu} + \boldsymbol{w}_t + \boldsymbol{Z}_t.$$

LR tests for projections

Let us assume that d is odd and set $q = (d-1)/2$, $\vartheta_k = \frac{2\pi k}{d}$.

Theorem

Assume $\boldsymbol{\Sigma} = \mathrm{Var}(\boldsymbol{Z}_1) > 0$ *known. LR-test statistics are:*

$$T^{\mathtt{M}_1} := \left\| \boldsymbol{A}(\vartheta_1)\boldsymbol{\Sigma}^{-1}\boldsymbol{A}'(\vartheta_1) \right\|$$

$$T^{\mathtt{M}_2} := \left\| \boldsymbol{A}(\vartheta_1,\ldots,\vartheta_q)\boldsymbol{\Sigma}^{-1}\boldsymbol{A}'(\vartheta_1,\ldots,\vartheta_q) \right\|$$

$$T^{\mathtt{M}_3} := \left\| \boldsymbol{A}(\vartheta_1,\ldots,\vartheta_q)\boldsymbol{\Sigma}^{-1}\boldsymbol{A}'(\vartheta_1,\ldots,\vartheta_q) \right\|_{\mathbf{tr}}$$

Here $\boldsymbol{A}(\theta_{i_1},\ldots,\theta_{i_k}) = \left[\boldsymbol{R}(\theta_{i_1}),\ldots,\boldsymbol{R}(\theta_{i_k}),\boldsymbol{C}(\theta_{i_1}),\ldots,\boldsymbol{C}(\theta_{i_k})\right]'$ where

$$\boldsymbol{R}(\theta) + \mathrm{i}\boldsymbol{C}(\theta) = \frac{1}{\sqrt{N}}\sum_{k=1}^{N}\boldsymbol{Y}_k e^{-\mathrm{i}k\theta}.$$

Model 3: relation to MANOVA

Note that

$$T^{\mathtt{M}_3} = \left\| \boldsymbol{A}(\vartheta_1, \ldots, \vartheta_q) \boldsymbol{\Sigma}^{-1} \boldsymbol{A}'(\vartheta_1, \ldots, \vartheta_q) \right\|_{\mathbf{tr}} \sim T^{\mathtt{MAV}}$$

where

$$T^{\mathtt{MAV}} = \sum_{k=1}^{d} \frac{N}{d} (\overline{\boldsymbol{Y}}_k - \overline{\boldsymbol{Y}})' \boldsymbol{\Sigma}^{-1} (\overline{\boldsymbol{Y}}_k - \overline{\boldsymbol{Y}}),$$

where $\overline{\boldsymbol{Y}}_k = \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{Y}_{(t-1)d+k}, 1 \leq k \leq d$. (Balanced design).

Remarks

- Approach is easy and purely multivariate.
- In practice we replace $\mathbf{\Sigma}$ by estimator.

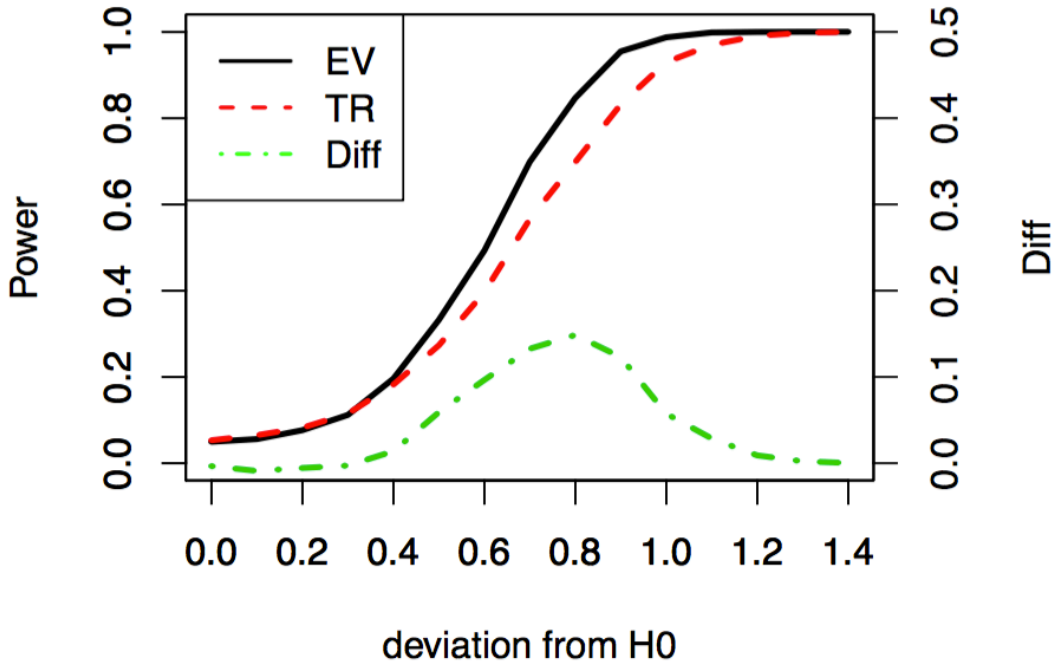
  NOTE: If we plug in $\widehat{\mathbf{\Sigma}}_A$ this test is asymptotically equivalent to Wilks' Lambda.
- For all three tests explicit null-distributions are available.
- The downside of this method: we need to have a good basis which picks up the signal.

Remarks

The standard multivariate test of MacNeill for Model 1 is

$$S^{\mathtt{M}_1} := \|\widetilde{D}_N(\vartheta_1)\|^2 = \left\|\boldsymbol{A}(\vartheta_1)\boldsymbol{\Sigma}^{-1}\boldsymbol{A}'(\vartheta_1)\right\|_{\mathbf{tr}}.$$

(Here $\widetilde{D}_N$ is the DFT of standardized data.)



Rescaling

The tests are computed from the standardized data

$$\boldsymbol{\Sigma}^{-1/2}\boldsymbol{Y}_1, \ldots, \boldsymbol{\Sigma}^{-1/2}\boldsymbol{Y}_N.$$

$\Longrightarrow$ asymptotically pivotal tests;

$\Longrightarrow$ component processes with larger variation are not concealing potential periodic patterns in components with little variance.

While this is clearly a very desirable property for multivariate data samples, one may favor a different perspective for functional data.

Fully functional approach

Let $A(\theta_{i_1}, \ldots, \theta_{i_k})$ be a $2k$-vector of functions computed from functional DFT

$$\frac{1}{\sqrt{N}} \sum_{t=1}^{N} Y_t(u) e^{-\mathrm{i}t\theta}.$$

Fully functional test for Model (3) $\Longleftrightarrow$ ANOVA model:

$$T^{\mathrm{F}_3} := \|A(\vartheta_1, \ldots, \vartheta_q) A'(\vartheta_1, \ldots, \vartheta_q)\|_{\mathrm{tr}}.$$

Again

$$T^{\mathrm{F}_3} \sim T^{\mathtt{FAV}} = \sum_{k=1}^{d} \frac{N}{d} \|\overline{Y}_k - \overline{Y}\|^2.$$

Asymptotic distribution of T^{F_3} (functional of DFT's) can be derived and then carried over to $T^{\mathtt{FAV}}$.

Theorem

Under L^2-m-approximability the functional ANOVA test statistic satisfies

$$\sum_{k=1}^{d} \frac{N}{d} \|\overline{Y}_k - \overline{Y}\|^2 \xrightarrow{d} 2 \sum_{k=1}^{(d-1)/2} \Xi_k,$$

where $\Xi_k \overset{ind.}{\sim} \mathrm{HExp}(\lambda_1(\vartheta_k), \lambda_2(\vartheta_k), \ldots)$, *and* $\lambda_\ell(\vartheta_k)$ *are the eigenvalues of* $\mathcal{F}_{\vartheta_k}$.

Remarks:

1. non-pivotal test;
2. consistent estimation of spectral density operator and its eigenvalues is possible;
3. good empirical approximation.

Empirical levels

- Synthetic but realistic data: PM-residuals resampled.
- Used to generate a functional MA(5) process.
- 1.000 repetitions.

	$\alpha = 5\%$				$\alpha = 10\%$			
	$T^{\mathtt{M}1}$	$T^{\mathtt{M}2}$	$T^{\mathtt{M}3}$	$T^{\mathtt{F}3}$	$T^{\mathtt{M}1}$	$T^{\mathtt{M}2}$	$T^{\mathtt{M}3}$	$T^{\mathtt{F}3}$
FF				5.0				8.3
				4.7				9.9
$p = 1$	5.0	3.9	3.9		9.6	8.7	8.7	
	5.9	4.4	4.4		9.8	9.9	9.9	
$p = 2$	6.4	4.3	3.9		11.2	8.7	8.0	
	6.0	5.4	4.1		10.6	9.8	9.1	
$p = 3$	6.8	3.8	3.9		12.2	7.9	6.9	
	5.8	5.5	4.2		10.6	9.6	7.9	
$p = 5$	7.4	6.4	6.2		15.5	11.6	11.4	
	6.7	6.4	5.7		12.5	12.2	10.7	

Table 1: Empirical size (in %) at the nominal level α of 5% and 10% for dependent time series with sample sizes $N = 210$ (top rows) and $N = 420$ (bottom rows).

Power

To study a realistic alternative we add back the weekday means to previous time series.

The signal-to-noise ratio in this example is approx 1/30.

	$N = 210$				$N = 420$			
	$T^{\mathtt{M}_1}$	$T^{\mathtt{M}_2}$	$T^{\mathtt{M}_3}$	$T^{\mathtt{F}_3}$	$T^{\mathtt{M}_1}$	$T^{\mathtt{M}_2}$	$T^{\mathtt{M}_3}$	$T^{\mathtt{F}_3}$
FF				72.9				99.9
$p = 1$	14.3	26.5	26.5		21.7	56.6	56.6	
$p = 2$	50.2	89.4	89.9		76.9	99.7	99.8	
$p = 3$	73.4	96.4	98.1		92.2	100	100	
$p = 5$	99.22	100	100		100	100	100	

Table 2: Empirical rates (in %) when testing at nominal level α of 5%. Results are based on 1,000 Monte Carlo simulation runs.

Real data tests for NO data

	$T^{\mathtt{M}_1}$	$T^{\mathtt{M}_2}$	$T^{\mathtt{M}_3}$	$T^{\mathtt{F}_3}$
FF (100%)				0.006
$v(u) \equiv 1$	0.305	0.099	0.099	
$p = 1$ (68%)	0.247	0.076	0.076	
$p = 2$ (81%)	0.495	0.204	0.172	
$p = 3$ (87%)	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$	
$p = 5$ (97%)	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$	

Table 3: The p-values for NO data. In parentheses, the percentage of variance explained by the first p principal components on which the curves are projected.

Uniform Test

Assume now that we would like to test for a periodic pattern when the frequency θ is unknown.

We consider a test based on

$$M_N = \max_{j=1,\ldots,q} \|D_N(\theta_j)\|^2 = \max_{j=1,\ldots,q} \|A(\theta_j)A'(\theta_j)\|_{\mathrm{tr}}.$$

Scalar setting

In the scalar setting, we know that

$$M_N - \log q \xrightarrow{d} \mathcal{G}, \quad q \sim N/2 \tag{4}$$

where $\mathcal{G}$ is the standard *Gumbel* distribution.

Simple explanation: If data are iid Gaussian, then

$$|D_N(\theta_j)|^2 \overset{iid}{\sim} \text{Exp}(1).$$

Without Gaussianity: we typically only have that

$$\left(|D_N(\omega_1)|^2, \ldots, |D_N(\omega_q)|^2\right) \xrightarrow{d} (E_1, \ldots, E_q), \quad E_j \overset{iid}{\sim} \mathrm{Exp}(1)$$

if q is fixed. Nevertheless, Davis and Mikosch (1999) obtained (4) for i.i.d. sequences and linear processes, and later to a more general notion of dependence in Lin and Liu (2009).

Main results

For $p \geq 1$, we define $M_N^p = \max_{j=1,\dots,q} \|D_N^p(\omega_j)\|^2$, where

$$D_N^p(\omega) \overset{\text{K.L.}}{=} N^{-1/2} \sum_{t=1}^{N} \sum_{k=1}^{p} \langle X_t, v_k \rangle v_k e^{-\mathbb{i} t\omega}$$

and the v_k's are the **eigenfunctions** of the operator C_0, associated to eigenvalues $\lambda_1 > \cdots > \lambda_p$.

Theorem 1. $\frac{M_N^p - b_N^p}{\lambda_1} \xrightarrow{d} \mathcal{G}$, for a fixed $p \geq 1$

Theorem 2. $\frac{M_N^{p_N} - b_N^{p_N}}{\lambda_1} \xrightarrow{d} \mathcal{G}$, for some $p = p_N \nearrow +\infty$

Theorem 3. $\frac{M_N - b_N}{\lambda_1} \xrightarrow{d} \mathcal{G}$

$$\text{where} \quad \begin{array}{rcl} b_N^p & = & \lambda_1 \log q - \lambda_1 \sum_{j=2}^{p} \log(1 - \lambda_j/\lambda_1) \\ b_N & = & \lim\limits_{p \to +\infty} b_N^p \quad \geq \lambda_1 \log q + E\|X_0\|^2 \end{array}$$

Idea of proof (with $\lambda_1 = 1$)

Write

$$M_N - b_N = \underbrace{M_N - M_N^p}_{A_1} + \underbrace{M_N^p - b_N^p}_{A_2} + \underbrace{b_N^p - b_N}_{A_3}.$$

- Obviously $A_3 \to 0$ when $p \to \infty$. ✓
- We can show that $A_1 \to 0$ if $p \to \infty$ fast enough. ✓
- The challenge is A_2! We will need that p grows slowly enough and hope that the intersection of sequences $p = p_N$ in A_2 and A_1 is not empty.

Idea of proof

Let $\widetilde{M}_N^p$ be computed as M_N^p, with a Gaussians sequence $Y_1, \ldots, Y_N$ such that $\mathrm{Var}(Y_1) = \mathrm{Var}(X_1)$, and we define

$$\rho_{N,p} = \sup_{x \in \mathbb{R}} |P(M_N^p \leq x) - P(\widetilde{M}_N^p \leq x)|$$

Then

$$\left| P\left(M_N^p - b_N^p \leq x\right) - e^{-e^{-x}} \right| \leq \rho_{N,p} + \underbrace{\left| P(\widetilde{M}_N^p - b_N^p \leq x) - e^{-e^{-x}} \right|}_{\checkmark}.$$

Gaussian approximation

We have that

$$\{M_N^p \leq x\} = \{S_N^{\mathbf{X}} \in A\},$$

where

$$\mathbf{X}_t = \begin{pmatrix} \langle X_t, v_1 \rangle \cos(\omega_1) \\ \langle X_t, v_1 \rangle \sin(\omega_1) \\ \vdots \\ \langle X_t, v_p \rangle \sin(\omega_q) \end{pmatrix} \in \mathbb{R}^{2pq} \quad (2pq \approx Np)$$

and A is $2p$–sparsely convex, i.e. finite intersections of sets whose indicator function depends on at most $2p$ components.

Gaussian approximation

We use a result of Chernozhukov, Chetverikov, and Kato (2017)

$$\sup_{A \in \mathcal{A}^{\mathrm{sp}}} |P(S_N^{\mathbf{X}} \in A) - P(S_N^{\mathbf{Y}} \in A)|,$$

where $\mathcal{A}^{\mathrm{sp}}$ is the class of s–sparsely convex subsets of $\mathbb{R}^d$.

We had to work a bit around since in this bound there was one constant left, which depends (translated into our setup) in an unspecified way on λ_p. Doesn't matter if p is fixed!

In the end we deduce that

$$\rho_{N,p} \lesssim \frac{p^3 \log(N)}{\lambda_p^{1/2} N^{1/6}}.$$

Conclusion of the proof

In particular Th. 3 holds if:

- $\lambda_k \sim \rho^k$, $\rho < 1$ and $E\|X_1\|^s < \infty$, with $s > 10$.
  $(p_N \sim \log(n))$
- $\lambda_k \sim \frac{1}{k^\nu}$, $\nu > 1$ and $E\|X_1\|^s < \infty$, with s *very big*.
  $(p_N \sim N^\gamma)$

Comments

- All the results are still valid when b_N is replaced by $\widehat{b}_n$.
- We work under $E\|X_1\|^4 < \infty$, but when p is *fixed*, we can use a **truncation** argument to relax it to $E\|X_1\|^{2+\epsilon} < \infty$. Hence, we have extended the result of Davis and Mikosch (1999).
- If p is *fixed*, we have considered the **dependent** case, i.e. of multivariate linear processes $X_t = \sum_{\ell=1}^{\infty} \Psi_\ell(Z_{t-\ell})$, where the Ψ_ℓ's are matrices and $(Z_t)_{t\in\mathbb{Z}}$ is an i.i.d. sequence in H. In that case, we rather work with

$$L_N = \max_{j=1,\dots,q} \|\widehat{\mathcal{F}}_{\omega_j}^{-\frac{1}{2}} D_N(\omega_j)\|^2$$

Uniform test

Consider a functional time series

$$X_t \;=\; s_t \;+ \varepsilon_t$$

where $t \mapsto s_t$ is **periodic** of **unknown** frequency ω and $(\varepsilon_t)_{t\in\mathbb{Z}}$ is some iid noise in H. We wish to test whether

$$\begin{cases} \mathcal{H}_0 & : \; \|s_t\| = 0 \quad \forall t, \\ \mathcal{H}_1 & : \; \exists t \text{ such that } \|s_t\| > 0 \end{cases}$$

From the previous result, we deduce that *under* $\mathcal{H}_0$,

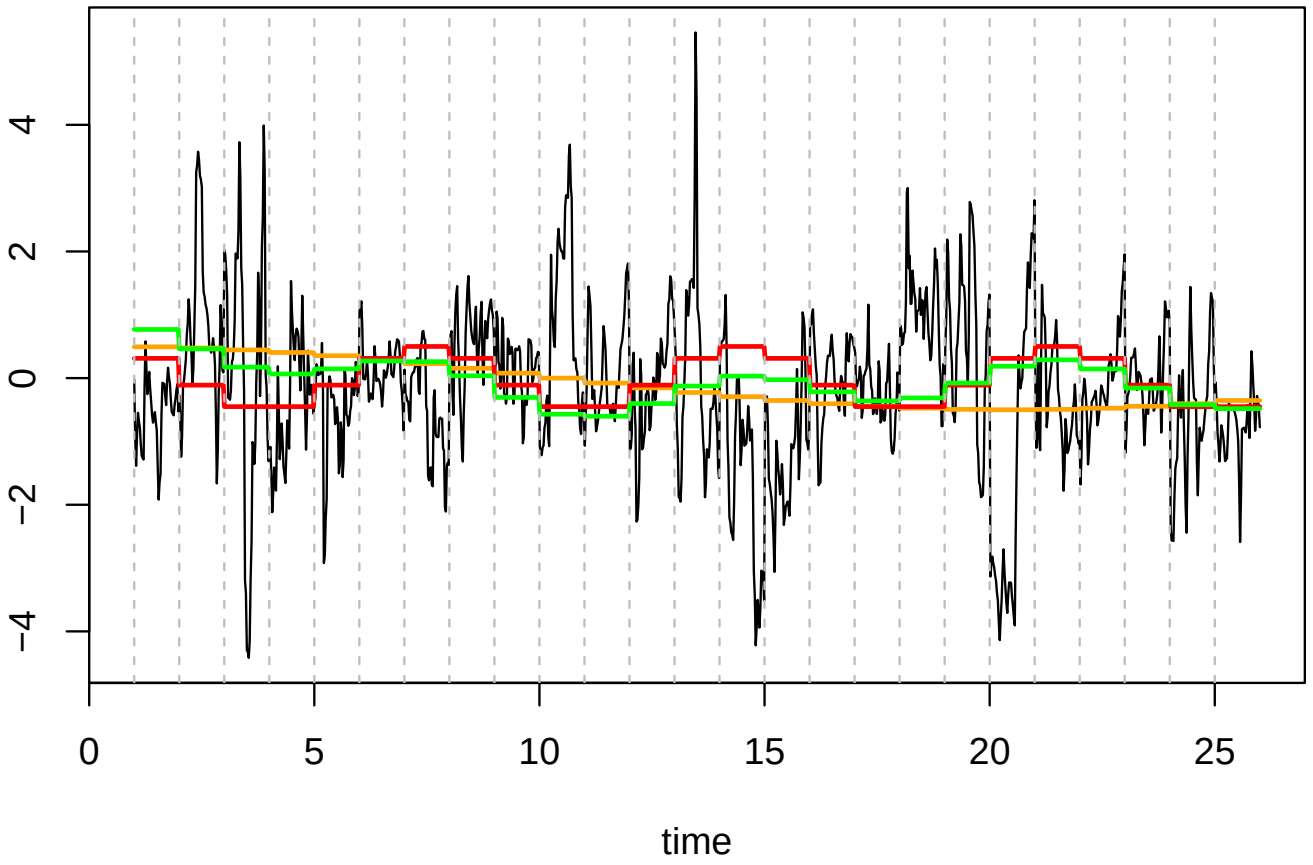
$$\mathbf{T} \;=\; \frac{M_N - b_N}{\lambda_1} \;\xrightarrow{d}\; \mathcal{G}.$$

On the other hand, we have that $\mathbf{T} \xrightarrow{P} +\infty$ *under* $\mathcal{H}_1$.

Simulation

We consider 3 different *alternative hypothesis*:

- **low frequency**, $s_t^{(1)} = .5\cos(t \cdot \frac{2\pi}{40})$
- **high frequency**, $s_t^{(2)} = .5\cos(t \cdot \frac{2\pi}{7})$
- **mixture**, $s_t^{(3)} = .3\cos(t \cdot \frac{2\pi}{40}) + .3\cos(t \cdot \frac{2\pi}{20}) + .3\cos(t \cdot \frac{2\pi}{7})$



Simulations

Rejection rates under the *null* and 3 different *alternatives* at level $\alpha = 0.05$
with increasing sample size, computed with 1000 replications.

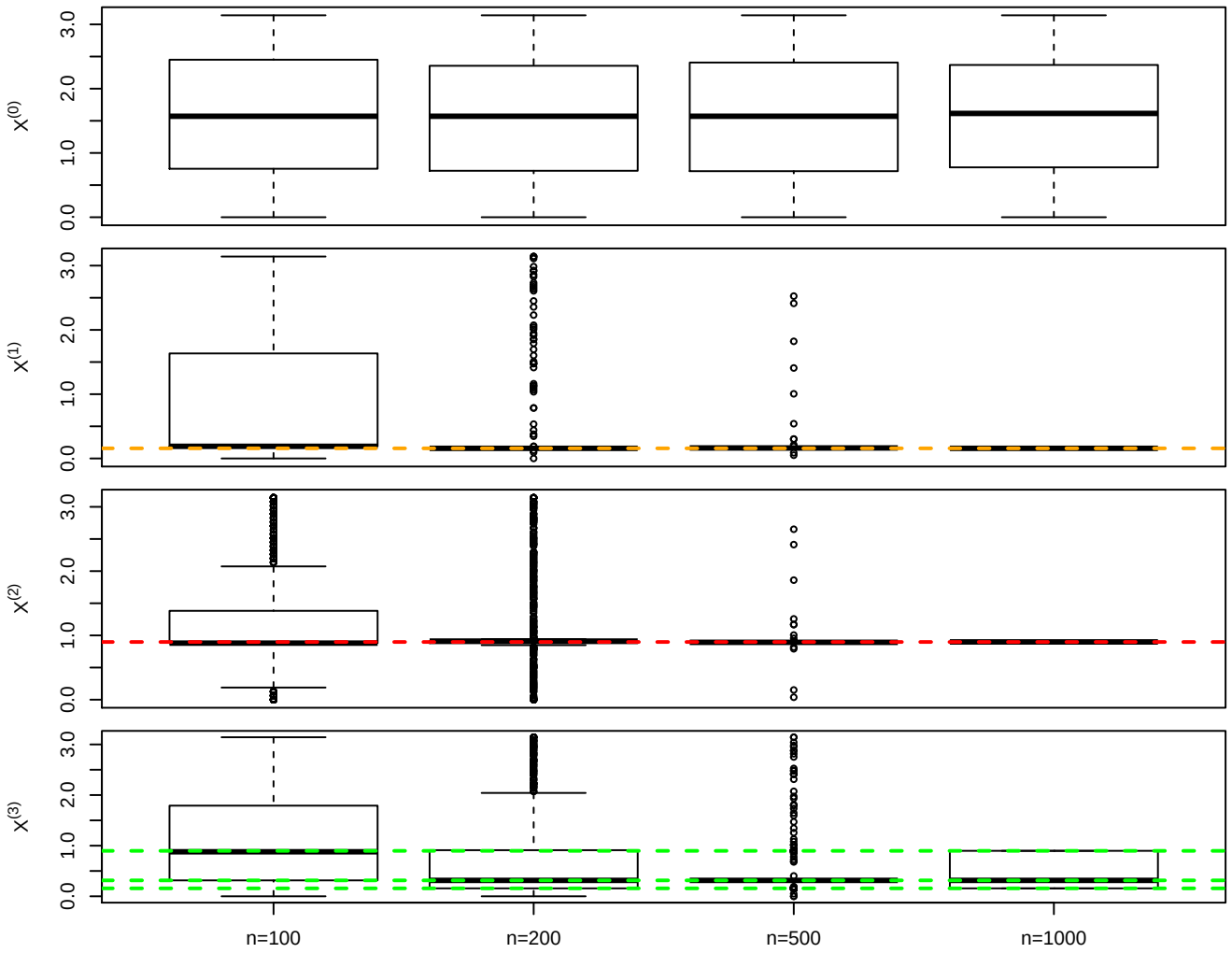
	$N = 100$	$N = 200$	$N = 500$	$N = 1000$
$X^{(0)}$	0.06	0.05	0.06	0.06
$X^{(1)}$	0.18	0.87	0.97	1.00
$X^{(2)}$	0.31	0.49	0.97	1.00
$X^{(3)}$	0.14	0.39	0.87	1.00

$$X_t^{(i)} = s_t^{(i)} + \varepsilon_t, \qquad t = 1, \ldots, n$$

with $s_t^{(0)} := 0$, and the curves ε_t are bootstrapped from the PM_{10} data set.

Simulations

We set $\omega^* = \underset{j=1,\ldots,q}{\operatorname{argmax}} \|D_N(\omega_j)\|^2$



Summary

- We have proposed several tests for periodicity of functional data.
- We distinguish between projection based approach and fully functional approach.
- Fully functional approach theoretically and practically appealing.
- Projection based approach very powerful, but needs tuning.
- We developed our tests from LR approach which improves also multivariate theory.
- Extension to dependent data.
- Extension to unknown length of period.