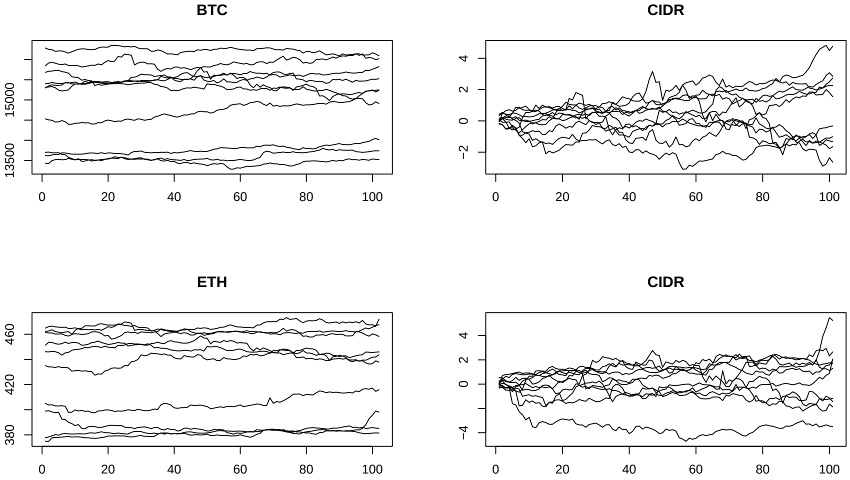
Robust tests of mutual independence between functional time series

CHARLES UNIVERSITY
Faculty of mathematics
and physics

Z. Hlávka, M. Hušková & S.G. Meintanis

Statistical Computing and Robust Inference for High Dimensional Data
SCRI 2023, Taipei

Functional data



Functional data

Usual mathematical framework:

- square integrable functions on $[0,1]$,
- generalization of random vectors.

In practice, 'functional data' may describe many different setups:

- growth functions (observed discretely or smoothed),
- random processes (continuous, discrete),
- images (in more dimensions).

Statistical methods for FDA (starting from a functional random sample) are typically generalizations of multivariate methods (often using suitable basis expansions, e.g., principal components).

Mean and covariance function

Random sample (or time series): $\{X_j, j = 1, \dots\}$ iid in L^2.

The **mean function** $\mu(t) = EX(t)$ can be estimated by $\widehat{\mu}(t) = \sum X_j(t)/n$.

The **covariance operator** $C(.) = E[\langle (X-\mu), . \rangle (X-\mu)]$ can be estimated by $\widehat{C}(x) = \sum \langle X_j - \widehat{\mu}, x \rangle (X_j - \widehat{\mu})/n$, $x \in L^2$.

In L^2, $C(t)(t) = \int c(t,s) y(s) ds$ and the **covariance function** $c(t,s) = \mathrm{Cov}(X(t), X(s))$ is estimated by $\sum (X_j(t) - \widehat{\mu}(t))(X_j(s) - \widehat{\mu}(s))/n$.

Statistical inference is usually based on suitable lower dimensional characteristics (e.g. norms or integrals) or projections (e.g. functional principal components).

Characterization of functional distribution

Generalizations of distribution function
$F_X(x) := \Pr[X(t) \leq x(t),\ \forall t \in [0,1]],\ x \in \mathbb{H}$ and density are problematic.

More promising: generalizations of the **characteristic function**
$\phi(t) = E\exp(it^\top X)$ characterizing distribution of random vector X.

The **characteristic functional** (CFL):

$$\phi(x) = E\exp\left\{i\int_0^1 x(t)X(t)dt\right\}$$

can be estimated by the empirical CFL (ECFL):

$$\widehat{\phi}(x) = \frac{1}{n}\sum\exp\left\{i\int_0^1 x(t)X_j(t)dt\right\}.$$

Tests based on characteristic functions

CF-based tests have some attractive properties for high-dimensional data, see the review paper Meintanis (2016).

CFL-based tests for functional data were also already considered: but in practice, we mostly work with discretized $\boldsymbol{X} = (X(t_1), \ldots, X(t_p))^\top$ and test statistics based on the empirical CF

$$\widehat{\varphi}_n(u_1, \ldots, u_p) = n^{-1} \sum_{j=1}^{n} e^{\mathrm{i} \sum_{\ell=1}^{p} u_\ell X_j(t_\ell)}.$$

In the following, we use CFs to characterize independence because covariance works only for linear dependencies and densities and distribution functions don't work well for functional data.

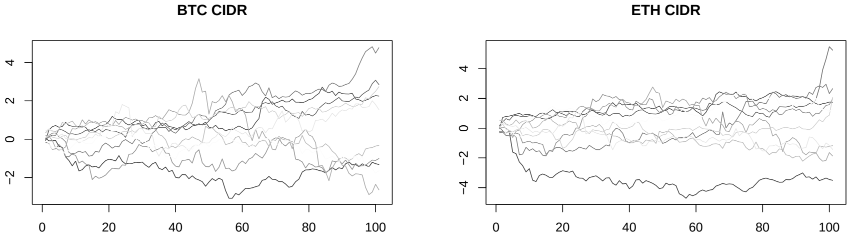
Functional time series (FTS)

FTS: functional data $\{X_j, j = 1, \dots\}$ observed sequentially over time.

This talk concerns tests of dependencies in FTS:

serial independence: independence of $\{X_j, j = 1, \dots\}$,

mutual independence: independence between two simultaneously observed FTS $\{X_j, j = 1, \dots\}$ and $\{Y_j, j = 1, \dots\}$.



Tests of mutual independence

Testing independence of random vectors in finite dimension: Herwartz, Maxand, Nonparametric tests for independence: a review and comparative simulation study with an application to malnutrition data in India, Statistical Papers 61 (2020) 2175–2201.

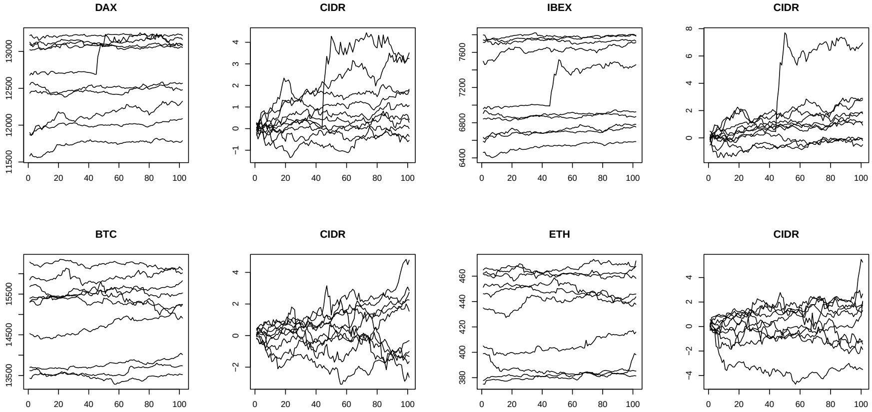
Functional data: Horváth, Rice, Testing for independence between functional time series, Journal of Econometrics 189 (2015) 371–382.

The setup: Observing a pair of stationary random curves $\{X_1(t), Y_1(t)\}, \dots, \{X_n(t), Y_n(t)\}, \dots,\ t \in [0,1]$, we want to test the null hypothesis

$$\Upsilon_0 : \{X_j\}_{j=1}^{\infty} \text{ and } \{Y_j\}_{j=1}^{\infty} \text{ are independent.}$$

Cryptocurrencies and market indices

DAX, IBEX, BTC, and ETH from November 2nd, 2020 to November 13th, 2020, intraday values (5 minutes) and corresponding cumulative intraday returns (CIDR) from 9:00 to 17:25, source `https://stooq.com/db/h/`



Fully functional approach (not used here)

Characteristic functional (CFL)

$$\varphi_X(u) = \mathbb{E}(e^{i\langle u, X\rangle}),$$

where $\langle u, X\rangle = \int_0^1 u(t)X(t)\mathrm{d}t$, $u \in \mathbb{H}$

Dependence between a pair $Z_h(t) = (X(t), Y_h(t))$ of $\mathbb{H}$–valued random elements could be measured by

$$\Delta_h^{(Q)} = \int_{\mathbb{H}\times\mathbb{H}} \left|\varphi_{Z_h}(U) - \varphi_X(u_1)\varphi_{Y_h}(u_2)\right|^2 \mathrm{d}Q(U), \tag{1}$$

where $U = (u_1, u_2)$, and Q is a Borel probability measure on $\mathbb{H} \times \mathbb{H}$.

Problem: choice of the measure Q (chosen as a Gaussian measure in [Hlávka, Hlubinka, Koňasová, K. (2022). Functional ANOVA based on empirical characteristic functionals. JMVA, 189, 104878.])

The notation and the alternative

$\varphi_{X,Y,h}(t_1, t_2; u_1, u_2)$ is the joint CF of $X_1(t_1)$ and $Y_{1+h}(t_2)$

$\varphi_X(t_1; u_1)$ and $\varphi_{Y,h}(t_2; u_2)$: marginal CFs

Denoting

$$D_h(t_1, t_2; u_1, u_2) = \varphi_{X,Y,h}(t_1, t_2; u_1, u_2) - \varphi_X(t_1; u_1)\varphi_{Y,h}(t_2; u_2),$$

we consider the alternative

$$\Upsilon_H: \text{ there exists an integer } |h_0| \leq H, \text{ such that}$$
$$\int_0^1 \int_0^1 \int_{\mathbb{R}} \int_{\mathbb{R}} \left| D_{h_0}(t_1, t_2; u_1, u_2) \right|^2 w(u_1) w(u_2) \mathrm{d}u_1 \mathrm{d}u_2 \mathrm{d}t_1 \mathrm{d}t_2 > 0,$$

where H is an a priori chosen fixed integer $0 \leq H < \infty$

ECF-based test statistic

Defining the marginal and joint ECFs, e.g.,

$$\widehat{\phi}_{X,Y,n,h}(t,s;u,v) = \frac{1}{n-|h|} \sum_{j=\max(1,1-h)}^{\min(n-h,n)} e^{\mathrm{i}\left(uX_j(t)+vY_{j+h}(s)\right)},$$

we propose the test statistic

$$T_{n,H}^{(w)} = \sum_{h=-H}^{H} (n-|h|) \int_0^1 \int_0^1 \Delta_{n,h}^{(w)}(t,s)\mathrm{d}s\mathrm{d}t,$$

where

$$\Delta_{n,h}^{(w)}(t,s) = \int_{\mathbb{R}} \int_{\mathbb{R}} \left|\mathbb{D}_{n,h}(t,s;u,v)\right|^2 w(u)w(v)\mathrm{d}u\mathrm{d}v$$

and

$$\mathbb{D}_{n,h}(t,s;u,v) = \widehat{\phi}_{X,Y,n,h}(t,s;u,v) - \widehat{\phi}_{X,n}(t;u)\widehat{\phi}_{Y,n,h}(s;v).$$

Some remarks

$$T_{n,H}^{(w)} = \sum_{h=-H}^{H} (n - |h|) \int_0^1 \int_0^1 \int_{\mathbb{R}} \int_{\mathbb{R}} \left| \widehat{\phi}_{X,Y,n,h}(t,s;u,v) - \widehat{\phi}_{X,n}(t;u) \widehat{\phi}_{Y,n,h}(s;v \right|^2 \mathrm{d}s\mathrm{d}t$$

- Advantages of using CFs: less moment assumptions, true dependence instead of lack of covariance.
- Complicated limit null distribution, need for resampling (in time series).
- In practice: effects of discretization.
- In practice, adequate value of H is unknown (we consider H fixed and rather small).
- For asymptotics, the weight function is general but using certain functional forms of $w(\cdot)$ improve computational expediency.

Assumptions

(A.1) The sequence $\{(X_j, Y_j)\}_{j=0}^{\infty}$ is a 2-dimensional strictly stationary α-mixing with coefficient $\alpha(k)$ such that $\sum_{k=0}^{\infty}(k+1)\alpha(k) \leq C$ for some positive constant C.

(A.2) The random functions satisfy

$$\mathbb{E}\int_0^1 |X_j(t)|^2 + |Y_j(t)|^2 \mathrm{d}t < \infty,$$

$$\mathbb{E}|X_j(t_1) - X_j(t_2)|^2 \leq C|t_1 - t_2|^{\kappa},\ \mathbb{E}|Y_j(t_1) - Y_j(t_2)|^2 \leq C|t_1 - t_2|^{\kappa},$$

$\forall (t_1, t_2) \in [0,1]^2$, and for some positive constants C and κ.

(A.3) The weight function $w : \mathbb{R} \to \mathbb{R}$ is nonnegative, measurable, and such that $w(-u) = w(u)$, $u \in \mathbb{R}$, $0 < \int_{\mathbb{R}} u^2 w(u) \mathrm{d}u < \infty$.

Null distribution

Under Υ_0,as $n \to \infty$, $T^{(w)}_{n,H}$ is distributed as

$$\int_0^1 \int_0^1 \int_{\mathbb{R}} \int_{\mathbb{R}} \sum_{h=-H}^{H} \left\{ V_h(t,s;u,v) \right\}^2 w(u)w(v)\mathrm{d}u\mathrm{d}v\mathrm{d}t\mathrm{d}s$$

where $\{V_h(t,s;u,v)\}$ are centered Gaussian processes with

$$\begin{aligned}
\mathrm{cov}(\mathrm{V_{h_1}(t_1,s_1;u_1,v_1)}, V_{h_2}(t_2,s_2;u_2,v_2)) = \\
&\mathbb{E}\left[\widetilde{g}\{u_1X_0(t_1)\}\widetilde{g}\{u_2X_0(t_2)\}\right]\mathbb{E}\left[\widetilde{g}\{v_1Y_{h_1-h_2}(s_1)\}\widetilde{g}\{v_2Y_0(s_2)\}\right] \\
&+\sum_{q\geq 1}\mathbb{E}\left[\widetilde{g}\{u_1X_q(t_1)\}\widetilde{g}\{u_2X_0(t_2)\}\right]\mathbb{E}\left[\widetilde{g}\{v_1Y_{q+h_1-h_2}(s_1)\}\widetilde{g}\{v_2Y_0(s_2)\}\right] \\
&+\sum_{q=1}^{h_1-h_2}\mathbb{E}\left[\widetilde{g}\{u_1X_0(t_1)\}\widetilde{g}\{u_2X_q(t_2)\}\right]\mathbb{E}\left[\widetilde{g}\{v_1Y_{-q+h_1-h_2}(s_1)\}\widetilde{g}\{v_2Y_0(s_2)\}\right] \\
&+\sum_{q=h_1-h_2+1}^{\infty}\mathbb{E}\left[\widetilde{g}\{u_1X_0(t_1)\}\widetilde{g}\{u_2X_q(t_2)\}\right]\mathbb{E}\left[\widetilde{g}\{v_1Y_0(s_1)\}\widetilde{g}\{v_2Y_{q+h_2-h_1}(s_2)\}\right]
\end{aligned}$$

where $g(x) = \sin x + \cos x$, $x \in \mathbb{R}$ and $\widetilde{g}(Z) = g(Z) - \mathbb{E}g(Z)$.

Fixed and local alternatives

The test is consistent under fixed alternatives.

In case of local alternatives satisfying

$$0 < \lim_{n\to\infty} n \sum_{n=H}^{H} \int_0^1 \int_0^1 \int_{\mathbb{R}} \int_{\mathbb{R}} |\phi_{X_n,Y_n;h}(u_1, u_2; t_1, t_2) - \phi_{X_n}(u_1; t_1)\phi_{Y_{j+h,n}}(u_2; t_2)|^2 w(u_1)w(u_2)\mathrm{d}t_1\mathrm{d}t_2 < \infty,$$

the limit distribution is of the same type but the limiting Gaussian process has nonzero expectation depending on

$$\phi_{X_n,Y_n;h}(u_1, u_2; t_1, t_2) - \phi_{X_n}(u_1; t_1)\phi_{Y_{j+h,n}}(u_2; t_2).$$

Computations

$$T_{n,H}^{(w,p)} = \frac{1}{p^2} \sum_{h=-H}^{H} n_h \sum_{i=1}^{p} \sum_{j=1}^{p} \Delta_{n,h}^{(w)}(t_i, s_j),$$

where

$$\Delta_{n,h}^{(w)}(t,s) = \iint \left|\mathbb{D}_{n,h}(t,s;u,v)\right|^2 w(u)w(v)\mathrm{d}u\mathrm{d}v$$

$$= \frac{1}{n_h^2} \sum_{j,k} I_w\{X_{j,k}(t)\} I_w\{Y_{j+h,k+h}(s)\} - \frac{2}{n_h^3} \sum_{j,k,l} I_w\{X_{j,k}(t)\} I_w\{Y_{j+h,l+h}(s)$$

$$+ \frac{1}{n_h^4} \sum_{j,k} I_w\{X_{j,k}(t)\} \sum_{l,m} I_w\{Y_{l+h,m+h}(s)\},$$

with $I_w(x) = \int_{\mathbb{R}} \cos(ux) w(u) \mathrm{d}u$, $X_{j,k}(t) = X_j(t) - X_k(t)$, $n_h = n - |h|$.

We choose $w(u) = \{a/(2\pi)\}^{1/2} \exp\{-(a/2)u^2\}$, $a > 0$, so that $I_w(x) = \exp\{-x^2/(2a)\}$.

Simplification

Under independence:

$$|\varphi_{X,Y,h}(t_1, t_2; u_1, u_2)|^2 = |\varphi_X(t_1; u_1)|^2 |\varphi_{Y,h}(t_2; u_2)|^2.$$

Simplified test statistic

$$\widetilde{T}_{n,H}^{(w,p)} = \frac{1}{p^2} \sum_{h=-H}^{H} \sqrt{n_h} \sum_{i=1}^{p} \sum_{j=1}^{p} \widetilde{\Delta}_{n,h}^{(w)}(t_i, s_j),$$

with

$$\widetilde{\Delta}_{n,h}^{(w)}(t,s) = \iint \left(|\widehat{\phi}_{X,Y,n,h}(t,s;u,v)|^2 - |\widehat{\phi}_{X,n}(t;u)|^2 |\widehat{\phi}_{Y,n,h}(s;v)|^2 \right) w(u) w(v) \mathrm{d}u \mathrm{d}v.$$

Distance covariance

Setting $w(u) = (\pi u^2)^{-1}$ leads

$$T_{n,H}^{DCov} = \frac{1}{p^2} \sum_{h=-H}^{H} n_h \sum_{i=1}^{p} \sum_{j=1}^{p} \Delta_{n,h}^{(DCov)}(t_i, s_j),$$

where

$$\begin{aligned}
\Delta_{n,h}^{(DCov)}(t,s) = \\
&\frac{1}{n_h^2} \sum_{j,k} |X_{j,k}(t)|\,|Y_{j+h,k+h}(s)| - \frac{2}{n_h^3} \sum_{j,k,l} |X_{j,k}(t)|\,|Y_{j+h,l+h}(s)| \\
&+ \frac{1}{n_h^4} \sum_{j,k} |X_{j,k}(t)| \sum_{l,m} |Y_{l+h,m+h}(s)|.
\end{aligned}$$

Empirical level

We simulate:

IID two independent samples of IID Brownian motions, say $W_{X,j}(t)$ and $W_{Y,j}(t)$, for $j = 1, \ldots, n$,

$\text{FAR}_q(1)$ two independent functional autoregressive models of order one with autoregressive parameter $\psi_q(t,u) = q \min(t,u)$, i.e.,

$$
\begin{aligned}
X_j(t) &= \int \psi_q(t,u) X_{j-1}(u) \mathrm{d}y + W_{X,j}(t), \\
Y_j(t) &= \int \psi_q(t,u) Y_{j-1}(u) \mathrm{d}y + W_{Y,j}(t).
\end{aligned}
$$

Critical values: block bootstrap vs. dependent wild bootstrap.

Empirical level, $a = 0.5$

			HR				CF $a = 0.5$			
	b	$n \setminus H$	0	1	2	5	0	1	2	5
IID	3	10	5.0	3.8	3.2	4.0	3.8	3.8	5.5	3.5
		20	4.0	5.8	3.8	5.0	6.5	4.0	4.8	4.2
$FAR_{0.75}(1)$	10	40	4.8	5.8	4.5	6.0	5.8	4.5	5.0	6.0
		80	7.8	6.5	5.8	8.0	3.8	4.8	5.2	6.2
$FAR_{1.50}(1)$	10	40	6.8	6.8	7.8	10.2	8.0	7.8	7.5	7.8
		80	6.0	5.5	7.8	8.0	7.8	6.8	9.2	10.0
	15	40	6.8	7.0	5.8	6.8	6.5	8.2	8.5	6.80
		80	7.2	7.2	6.5	8.8	6.5	6.5	9.2	8.8
$FAR_{2.25}(1)$	10	40	18.8	18.2	17.2	24.5	17.8	18.5	22.8	27.0
		80	23.2	25.0	25.0	31.2	21.5	26.0	27.8	41.5
	55	120	9.0	7.0	7.2	10.0	8.5	7.5	7.8	8.5
		200	5.2	8.8	6.2	11.5	8.5	6.5	7.0	9.2

Nonlinear dependency

First sample: IID Brownian motion $X_j(t) = W_{X,j}(t)$

Second sample: $Y_j(t) = X_j^2(t)/2$, for $j = 1, \ldots, n$.

Simulation setup: $n = 10$, $H = 0$, block-length $b = 3$:

HR empirical power 36.25%,

CF empirical power 85.5%, 70%, and 59% (for $a \in \{0.5, 1, 2\}$)

Level–robustness and power

'Heavy-tailed' functional observations

$\Upsilon_0(T_d)$ two independent processes T_X and T_Y defined as a ratio of mutually independent Wiener processes (W_X and W_Y) and d-dimensional Bessel processes (B_X and B_Y), i.e., $T_{X,j} = d^{1/2} W_{X,j}/B_{X,j}$ and $T_{Y,j} = d^{1/2} W_{Y,j}/B_{Y,j}$.

The empirical power will be investigated for samples from the following sequences of functional observations:

$\Upsilon_1(W)$: two processes U_x and V_x such that $U_{X,j} = (W_{X,j} + Z_j)/\sqrt{2}$ and $V_{X,j} = (W_{Y,j} + Z_j)/\sqrt{2}$, where W_X, W_Y, and Z are IID samples of mutually independent Wiener processes,

$\Upsilon_1(T_d)$: two processes $V_{X,j} = (d/2)^{1/2}(W_{X,j} + Z_j)/B_{X,d,j}$ and $V_{Y,j} = (d/2)^{1/2}(W_{Y,j} + Z_j)/B_{X,d,j}$.

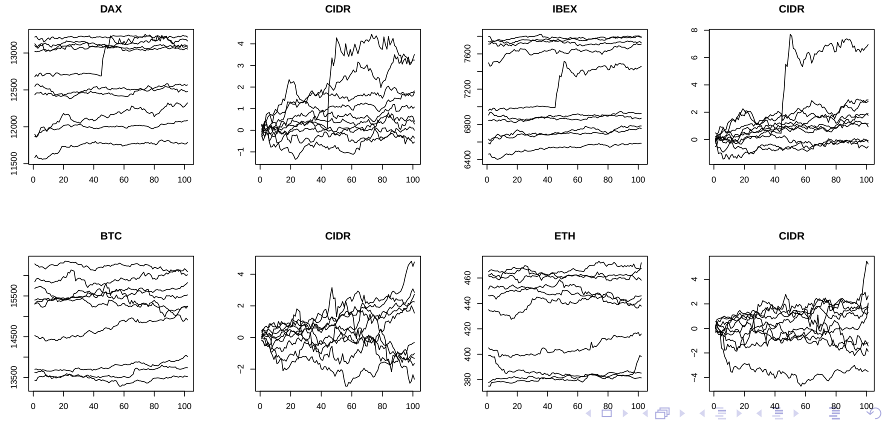
Level–robustness and power, $b = 5$

		HR		DCov		CF					
						$a = 0.5$		$a = 1$		$a = 2$	
	$n \setminus H$	0	1	0	1	0	1	0	1	0	1
$\Upsilon_0(T_1)$	20	4.8	5.2	5.7	6.0	6.0	5.0	4.2	6.0	4.2	5.0
	40	6.5	5.2	4.7	5.0	4.2	3.5	7.0	6.8	4.8	5.0
$\Upsilon_0(T_3)$	20	4.8	2.8	4.0	6.2	5.2	6.2	4.8	5.0	6.0	4.2
	40	4.0	5.8	3.2	4.2	7.2	5.5	6.0	4.0	6.5	3.5
$\Upsilon_1(T_1)$	20	5.5	5.5	8.5	5.2	24.2	13.5	27.5	14.2	**32.0**	17.8
	40	7.2	5.8	7.2	7.0	48.0	28.0	59.5	35.5	**62.7**	41.2
$\Upsilon_1(T_3)$	20	41.2	25.2	**51.2**	34.7	36.8	21.5	46.0	24.0	50.5	26.8
	40	73.8	50.5	**90.0**	68.7	68.2	44.8	77.5	56.5	85.5	59.2
$\Upsilon_1(W)$	20	**63.5**	41.8	**63.5**	34.7	48.8	32.0	56.2	31.8	60.2	38.8
	40	**95.0**	80.0	90.2	79.5	84.5	64.0	87.5	76.0	92.5	77.5

Application

DAX, IBEX, BTC, and ETH: outlier: November 9th, 2020.

Test of stationarity (Horváth et al, 2014, R library `ftsa`):
p-values DAX: 0.549, IBEX: 0.387, BTC: 0.192, and ETH: 0.168.



Application, B=400, $b = 3$, $n = 15$

Test of pairwise independence, October 26th to November 13th ($n = 15$).

H		DAX×IBEX	IBEX×BTC	IBEX ×ETH	DAX×BTC	DAX×ETH	BTC×ETH
0	HR	0.0025	0.4125	0.6525	0.2850	0.7250	0.0150
	DCov	0.0025	0.6350	0.7025	0.3050	0.4500	0.0300
	CF	0.0000	0.8825	0.7300	0.4700	0.2550	0.0475
	CF-S	0.0000	0.6350	0.6100	0.1050	0.1625	0.0000
1	HR	0.0100	0.6425	0.7950	0.5400	0.9850	0.1150
	DCov	0.0100	0.6350	0.6850	0.7100	0.7250	0.2350
	CF	0.0125	0.6975	0.3950	0.8350	0.1075	0.3200
	CF-S	0.0175	0.4450	0.7175	0.2600	0.1400	0.1425

Mutual independence ($H = 0$)

$X_{m;j}(t)$, for $m = 1, \ldots, 4$ and $j = 1, \ldots, n$, for simplicity, extension of the computationally simpler CF-S test.

For $H = 0$, the test statistic can be defined as:

$$\widetilde{T}_{n,0}^{(w,p)} = \frac{1}{p^4}\sqrt{n}\sum_{j_1=1}^{p}\sum_{j_2=1}^{p}\sum_{j_3=1}^{p}\sum_{j_4=1}^{p}\widetilde{\Delta}_n^{(w,4)}(t_{j_1}, t_{j_2}, t_{j_3}, t_{j_4})$$

with

$$\widetilde{\Delta}_n^{(w)}(t_1, t_2, t_3, t_4) = \iiiint \Big(|\widehat{\phi}_{X_1,X_2,X_3,X_4,n}(t_1, t_2, t_3, t_4; u_1, u_2, u_3, u_4)|^2 - |\widehat{\phi}_{X_1,n}(t_1; u_1)|^2|\widehat{\phi}_{X_2,n}(t_2; u_2)|^2|\widehat{\phi}_{X_3,n}(t_3; u_3)|^2|\widehat{\phi}_{X_4,n}(t_4; u_4)|^2\Big) w(u_1)w(u_2)w(u_3)w(u_4)\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}u_3\mathrm{d}u_4.$$

Define the set $\mathcal{H}_{H=1}^{(4)} := \{\boldsymbol{h} = (h_1, h_2, h_3, h_4)^\top, h_m \in \{0, 1\}; \min_m h_m = 0\}$ and the test statistic:

$$\widetilde{T}_{n,1}^{(w,p)} = \frac{1}{p^4} \sum_{\boldsymbol{h} \in \mathcal{H}_1^{(4)}} \sqrt{n_{\boldsymbol{h}}} \sum_{j_1=1}^{p} \sum_{j_2=1}^{p} \sum_{j_3=1}^{p} \sum_{j_4=1}^{p} \widetilde{\Delta}_{n,\boldsymbol{h}}^{(w,4)}(t_{j_1}, t_{j_2}, t_{j_3}, t_{j_4}),$$

where $n_{\boldsymbol{h}} = n - \max_m h_m$ (number of summands is 15) and

$$\widetilde{\Delta}_{n,\boldsymbol{h}}^{(w)}(t_1, t_2, t_3, t_4) = \iiiint \Big(|\widehat{\phi}_{X_1,X_2,X_3,X_4,n,\boldsymbol{h}}(t_1, t_2, t_3, t_4; u_1, u_2, u_3, u_4)|^2 - |\widehat{\phi}_{X_1,n,h_1}(t_1; u_1)|^2 |\widehat{\phi}_{X_2,n,h_2}(t_2; u_2)|^2 |\widehat{\phi}_{X_3,n,h_3}(t_3; u_3)|^2 |\widehat{\phi}_{X_4,n,h_4}(t_4; u_4)|^2 \Big) w(u_1)w(u_2)w(u_3)w(u_4)\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}u_3\mathrm{d}u_4.$$

Deviations from the null hypothesis are indicated both by large negative and large positive values and, therefore, this test is two-sided.

Application

A computationally feasible expression:

$$\widetilde{\Delta}^{(w)}_{n,\boldsymbol{h}}(t_1, t_2, t_3, t_4) =$$

$$\frac{1}{n_{\boldsymbol{h}}^2} \sum_{j,k} \prod_{l=1}^{4} I_w\{X_{l;j+h_l,k+h_l}(t_l)\} - \prod_{l=1}^{4} \frac{1}{n_{\boldsymbol{h}}^2} \sum_{j,k} I_w\{X_{l;j+h_l,k+h_l}(t_l)\},$$

where $X_{l;j,k}(t) = X_{l;j}(t) - X_{l;k}(t)$.

For the DAX, IBEX, BTC, ETH dataset, we reject the mutual independence both for $H = 0$ and $H = 1$ (p-value=0.0000).

For $n = 10$, the p-values are 0.0001 (for $H = 0$) and 0.0010 (for $H = 1$).

In practice, pairwise dependencies could be investigated after rejecting the mutual independence.

Generalizations: vectorial versions

More general approach is to use a test based on

$$\int_0^1 \int_0^1 \int_{\mathbb{R}^{2h}} \int_{\mathbb{R}^{2h}} |\widetilde{D}_h(t_1, t_2; \boldsymbol{u}_1, \boldsymbol{u}_2)|^2 w(\boldsymbol{u}_1) w(\boldsymbol{u}_2) \mathrm{d}\boldsymbol{u}_1 \mathrm{d}\boldsymbol{u}_2 \mathrm{d}t_1 \mathrm{d}t_2,$$

where $\boldsymbol{u}_1 = (u_{1,-h}, \dots, u_{1,h})^\top$, $\boldsymbol{u}_2 = (u_{2,-h}, \dots, u_{2,h})^\top$ and

$$\widetilde{D}_h(t_1, t_2; \boldsymbol{u}_1, \boldsymbol{u}_2) = \\ \mathbb{E} \exp\left(\mathrm{i} \left[\sum_{v_1=-h}^{+h} \{u_{1,v_1} X_{j+v_1}(t)\} + \sum_{v_2=-h}^{+h} \{u_{2,v_2} Y_{j+v_2}(s)\} \right] \right) \\ - \mathbb{E} \exp\left\{ \mathrm{i} \sum_{v_1=-h}^{+h} u_{1,v_1} X_{j+v_1}(t) \right\} \times \mathbb{E} \exp\left\{ \mathrm{i} \sum_{v_2=-h}^{+h} u_{2,v_2} Y_{j+v_2}(s) \right\},$$

i.e., independence of $(X_{j-h}(t), \dots, X_{j+h}(t))$ and $(Y_{j-h}(s), \dots, Y_{j+h}(s))$.

Mutual independence in general

M stationary random curves
$\{X_{1,j}(t), \ldots, X_{M,j}(t), j = 1, \ldots, n, \ldots, \; t \in [0,1]\}$, the null hypothesis

$$\Upsilon_0 : \{X_{m,j}\}_{j=1}^{\infty}, \text{ are mutually independent}, \; m = 1, ..., M,$$

against the alternative hypothesis

$$\Upsilon_H : \text{there exists } \boldsymbol{h}_0 = (h_{0,1}, \ldots, h_{0,M})^\top, \text{ with } 0 \le h_{0,m} \le H,$$
$$\text{such that } \int_{[0,1]^M} \int_{\mathbb{R}^M} \left| D_{\boldsymbol{h}_0}(\boldsymbol{t}; \boldsymbol{u}) \right|^2 W(\boldsymbol{u}) \mathrm{d}\boldsymbol{u} \mathrm{d}\boldsymbol{t} > 0,$$

where $0 \le H < \infty$ is an apriori chosen integer, and
$D_{\boldsymbol{h}}(\boldsymbol{t}; \boldsymbol{u}) = \varphi_{\boldsymbol{h}}(\boldsymbol{t}; \boldsymbol{u}) - \prod_{m=1}^{M} \varphi_{m,h_m}(t_m; u_m), \; \boldsymbol{h} := (h_1, h_2, ..., h_M)^\top$.

Mutual independence

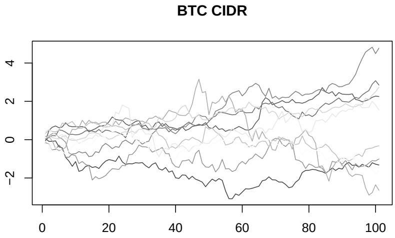
Consider set of indices
$\mathcal{H}_H^{(M)} := \{\boldsymbol{h} = (h_1, h_2, ..., h_M)^\top, h_m \in \{0, \ldots, H\};\ \min_m h_m = 0\}$, with cardinality $(H+1)^M - H^M$, test statistics may be defined similarly as before.

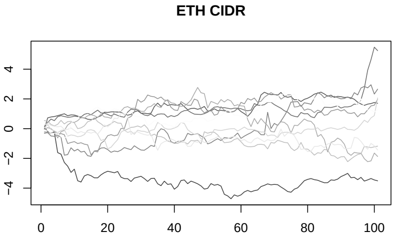
After rejecting mutual indendence, we recommend to:

1. Investigating pairwise dependencies or dependencies between other possibly interesting subsets of the M time series,
2. Investigating the effect that H has on the test decision by varying the value of this parameter,
3. Investigating the effect of $\mathcal{H}_H^{(M)}$ by considering smaller subsets suggested by prior knowledge, such as, e.g., a causal relationship that might effect a monotonicity condition $h_1 \leq \cdots \leq h_M$ on $\mathcal{H}_H^{(M)}$.

Conclusion

ECF-based tests of independence can be used also for more complicated functional objects (and also for testing specific types of dependence).





In general, investigating dependencies between functional time series is quite complicated.

References and acknowledgement

More details in:

Hlávka, Hušková, Meintanis, (2021). Testing serial independence with functional data. Test, 30(3), 603-629.

Meintanis, Hušková, Hlávka, Z. (2022). Fourier-type tests of mutual independence between functional time series. Journal of Multivariate Analysis, 189, 104873

The research has been supported by grant number GAČR 18-08888S provided by the Czech Science Foundation.