Two-sample and change-point procedures based on empirical characteristic functions in higher dimension

Zdeněk Hlávka, Marie Hušková and S. Meintanis, et al

Charles University, Prague and and Kapodistrian University, Athens

Warsaw, July 2018

Outline

1 Introduction
- Goodness-of-fit tests
- Empirical characteristic function based procedures

2 Two-sample procedures based on characteristic function

3 Detection of a change (independent observations)
- Change detection in VAR models
- Computations
- Simulations and real data
- Real data analysis

4 Two-sample problem in high dimension
- Simulations and application

Introduction

Well -known from basic courses:

There is a one-to-one relationship between distribution function and characteristics function

$\mathbf{X}$ – d-dimensional random vector

$F(\mathbf{x}) = P(\mathbf{X} \leq \mathbf{x}),\ \mathbf{x} \in \mathcal{R}^d$ – distribution function

$\varphi(\mathbf{u}) = E(\exp\{i\mathbf{u}^T\mathbf{X}\}),\ \mathbf{u} \in \mathcal{R}^d$ – characteristic function

Statistical problems typically formulated in terms of distribution functions and their parameters, therefore also in terms of characteristics functions.

$$\varphi(\mathbf{u}) = E(\exp\{i\mathbf{u}^T\mathbf{X}\}) = C(\mathbf{u}) + iS(\mathbf{u}) = E\cos(\mathbf{u}^T\mathbf{X}) + i\sin(\mathbf{u}^T\mathbf{X})$$

Recently, proposed and studied a number of statistical procedures employing empirical characteristics functions for various setups

- goodness-of-fit tests,
- model specification tests
- tests for detection of changes
- with and without nuisance parameters
- mostly for univariate case, here we focus on multivariate setups

- The overview paper published by S. Meintanis (2016).

Goodness-of-fit tests shortly, simplest formulation
$X_1, \dots, X_n$ are i.i.d. random variables with d.f. F

$H_0 : F = F_0$ for a given F_0 against $H_1 : H_0$ is not true

More often:

$H_0^* : F \in \mathcal{F}$, $\mathcal{F}$ a system of distributions, typically depending on parameters-nuisance parameters

Kolmogorov-Smirnov type tests

Test procedures are based on empirical distribution functions

$$\widehat{F}_n(x) = \sum_{j=1}^{n} I\{X_i \le x\}, \quad x \in \mathbb{R}$$

Kolmogorov-Smirnov test: $\sup_{x\in\mathbb{R}} |\widehat{F}_n(x) - F_0(x)|$

Cramér-von-Mises test: $\int_{x\in\mathbb{R}} |\widehat{F}_n(x) - F_0(x)|^2 dF_0(x)$

Anderson-Darling test: $\int_{x\in\mathbb{R}} |\widehat{F}_n(x) - F_0(x)|^2 w(x) dF_0(x)$

Advantage: if F_0 is continuous the distribution of KS and CVM under H_0 does not depend on F_0 (distribution free test statistics)

Similar problem:

(i) H_0^S: distribution F is symmetric $(F(x) = 1 - F(x)\,\forall x)$

(ii)two sample tests– two independent samples, we are testing that they have the same distribution

(iii) independence tests

(iv)change-point tests

F_0 depends on nuisance parameters, particular cases

Empirical characteristic function based procedures

$X_1, \ldots, X_n$ – i.i.d. random variables

Testing problem H_0 versus H_1 can be equivalently expressed as

$H_0 : \varphi = \varphi_0$ for a given φ_0 versus $H_1 : H_0$ is not true

$\varphi(u) = E\exp\{iuX_j\},\ u \in \mathcal{R}$ – characteristic function(CF)

$\widehat{\varphi}_n(u) = \frac{1}{n}\sum_{j=1}^{n}\exp\{iuX_j\},\ u \in \mathcal{R}$ – empirical characteristic function (ECF)

Test statistic:

$$T_n(w) = \int_{\mathcal{R}} |\widehat{\varphi}_n(u) - \varphi_0(u)|^2 w(u)du$$

$w(\cdot)$- weight function (usually, nonnegative, symmetric)

Something from the history

H. Cramér (1946) – classical book, empirical characteristic function mentioned

Feuerverger and Mureika (1997), Annals of Statistics

Sandor Csörgő (1984) – Proceedings of Asymptotic Statistics, 1984, Praha

Ushakov (1999) – Selected Topics in Characteristics Functions (book)

Meintanis (2016), South African Statistical Journals – survey paper with discussions

More general setup:

Klebanov (2005)- book – N-distances and Their Applications

Procedures based on probability generating function - Hudecová

Rizzo and Székely et al (2010,...)

Two-sample procedures based on characteristic function

$\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ – independent p-dimensional random vectors
F_j – distribution function of $\mathbf{Y}_j$
Testing problem

$$H_0 : F_1 = \ldots = F_n$$
$$H_1 : F_1 = \ldots = F_m \neq F_{m+1} = \ldots = F_n \qquad \textit{for} \quad m < n,$$

F_1 and F_n are unknown, m - known.

$$T_{m,n-m}(w) = \frac{m(n-m)}{n} \int_{\mathcal{R}^p} |\widehat{\varphi}_m(\mathbf{t}) - \widehat{\varphi}^0_{n-m}(\mathbf{t})|^2 w(\mathbf{t}) d\mathbf{t},$$

$w(\cdot)$ – a nonnegative weight function
$\widehat{\varphi}_m(\mathbf{t})$ and $\widehat{\varphi}^0_{n-m}(\mathbf{t})$ – empirical characteristic functions based on $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$ and $\mathbf{Y}_{m+1}, \ldots, \mathbf{Y}_n$, respectively, i.e.

$$\widehat{\varphi}_m(\mathbf{t}) = \frac{1}{k} \sum_{j=1}^{m} \exp\{i\mathbf{t}^T\mathbf{Y}_j\}, \quad \widehat{\varphi}^0_{n-m}(\mathbf{t}) = \frac{1}{n-m} \sum_{j=m+1}^{n} \exp\{i\mathbf{t}^T\mathbf{Y}_j\}$$

Direct calculations give:

$$T_{m,n-m}(w) = \frac{m(n-m)}{n} \int_{\mathcal{R}^p} \Big(\frac{1}{m} \sum_{j=1}^{m} U_j(\mathbf{t}) - \frac{1}{n-m} \sum_{j=m+1}^{n} U_j(\mathbf{t}) \Big)^2 w(\mathbf{t}) d\mathbf{t}$$

$$= \frac{m(n-m)}{n} \Big(\frac{1}{m^2} \sum_{j,v=1}^{m} I_w(\mathbf{Y}_j - \mathbf{Y}_v) + \frac{1}{(n-m)^2} \sum_{j,v=m+1}^{n} I_w(\mathbf{Y}_j - \mathbf{Y}_v)$$

$$- \frac{2}{m(n-m)} \sum_{j=1}^{m} \sum_{v=m+1}^{n} I_w(\mathbf{Y}_j - \mathbf{Y}_v) \Big),$$

$$U_j(\mathbf{t}) = \cos(\mathbf{t}^T \mathbf{Y}_j) + \sin(\mathbf{t}^T \mathbf{Y}_j), \quad I_w(\mathbf{x}) = \int_{\mathcal{R}^p} \cos(\mathbf{u}^T \mathbf{x}) w(\mathbf{u}) d\mathbf{u}$$

possible choice of $w(\cdot)$:

$$w_a(\mathbf{u}) = \exp\{-||\mathbf{u}||^2\}, \quad a > 0$$

$$I_w(\mathbf{x}) = \left(\frac{\pi}{a}\right)^{p/2} \exp\{-a||\mathbf{u}||^2/(4a)\}$$

Simple characteristic

$$E\Big(\frac{1}{m}\sum_{j=1}^{m}U_j(\mathbf{t})-\frac{1}{n-m}\sum_{j=m+1}^{n}U_j(\mathbf{t})\Big)^2$$

$$=\frac{var(U_1(\mathbf{t}))}{m}+\frac{var(U_n(\mathbf{t}))}{n-m}+\Big(EU_1(\mathbf{t})-EU_1(\mathbf{t})\Big)^2$$

It simplifies under the null hypothesis

• for testing - null hypothesis rejected for large values of test statistic

• approximation for critical values– either simulation of the limit distribution with estimated covariance, or some bootstrap

Theorem 1 Let $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$ be a sequence of i.i.d. p-dimensional r. v. with finite second moment, let $m_n/n \to \theta_0 \in (0,1)$ and $w(\cdot)$ be a nonnegative measurable weight function defined on $\mathbb{R}^d$ such that

$$w(\mathbf{u}) = w(-\mathbf{u}), \quad \forall \mathbf{u} \in \mathbb{R}^d, \quad 0 < \int_{\mathbb{R}^p} \|\mathbf{u}\|^2 w(\mathbf{u}) d\mathbf{u} < \infty. \qquad (1)$$

Then for $(m, n-m) \to \infty$,

$$T_{m,n-m}(w) \to^d \int_{\mathcal{R}^p} V^2(\mathbf{t}) w(\mathbf{t}) d\mathbf{t},$$

$\{V(\mathbf{t});, t \in \mathcal{R}^p\}$ – Gaussian process with zero mean and covariance structure

$$cov(V(\mathbf{t}_1), V(\mathbf{t}_2)) = cov(U_j(\mathbf{t}_1), U_j(\mathbf{t}_2))$$

- consistent test

$$\frac{1}{n} T_{m,n-m}(w) \to^d (1-\theta_0)\theta_0 \int_{\mathcal{R}^p} \left|\varphi_0(\mathbf{t}) - \varphi^0(\mathbf{t})\right|^2 w(\mathbf{t}) d\mathbf{t},$$

Detection of a change for independent observations

$\{\mathbf{X}_t,\ t = 1, 2, \ldots, T\}$ – a sequence of random vectors of dimension p

$\mathbf{X}_t$ has the distribution function (DF) $F_t, 1 \le t \le T$.

Classical change-point detection problem

$$\mathcal{H}_0 : F_t \equiv F_0 \text{ for all } t = 1, \ldots, T, \quad \text{vs.}$$
$$\mathcal{H}_1 : F_t \equiv F_0,\ t \le t_0;\ F_t \equiv F^0,\ t > t_0,$$

$F_0 \neq F^0$ and t_0 are unknown

The null hypothesis equivalently formulated via characteristic functions:

$$\mathcal{H}_0 : \varphi_t \equiv \varphi_0 \text{ for all } t = 1, \ldots, T, \quad \text{vs.}$$
$$\mathcal{H}_1 : \varphi_t \equiv \varphi_0,\ t \le t_0;\ \varphi_t \equiv \varphi^0,\ t > t_0,$$

$\varphi_t(\mathbf{u}) := E(e^{i\mathbf{u}^T\mathbf{X}_t})$ the characteristic function (CF) of X_t
$\varphi_0, \varphi^0, t_0$ – unknown

The proposed test is based on

$$Q_{T,w}(\gamma) = \max_{1\le t<T} \left(\frac{t(T-t)}{T^2}\right)^{2+\gamma} T \int_{\mathcal{R}^d} \left|\widehat{\varphi}_t(\mathbf{u}) - \widehat{\varphi}^t(\mathbf{u})\right|^2 w(\mathbf{u}) d\mathbf{u},$$

$w(\cdot)$ is a suitable weight function,

$\gamma \in (-1, 1]$ is a tuning constant

$$\widehat{\varphi}_t(\mathbf{u}) = \frac{1}{t}\sum_{\tau=1}^{t} e^{i\mathbf{u}^T\mathbf{X}_\tau}, \quad \widehat{\varphi}^t(\mathbf{u}) = \frac{1}{T-t}\sum_{\tau=t+1}^{T} e^{i\mathbf{u}^T\mathbf{X}_\tau},$$

the empirical CFs computed from $\mathbf{X}_1, \dots, \mathbf{X}_t$ and $\mathbf{X}_{t+1}, \dots, \mathbf{X}_T, \; t = 1, \dots, T$, respectively.

Large values indicate that the null hypothesis is violated.

Theorem 2 Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be a sequence of i.i.d. d-dimensional r. v. with finite second moment, $\gamma \in (-1, 1]$ and $w(\cdot)$ be a nonnegative measurable weight function defined on $\mathbb{R}^d$ such that

$$w(\mathbf{u}) = w(-\mathbf{u}), \quad \forall \mathbf{u} \in \mathbb{R}^d, \quad 0 < \int_{\mathbb{R}^d} \|\mathbf{u}\|^2 w(\mathbf{u}) d\mathbf{u} < \infty. \qquad (2)$$

Then, as $T \to \infty$,

$$Q_{T,w}(\gamma) \xrightarrow{d} \sup_{s \in (0,1)} (s(1-s))^{\gamma} \int_{\mathbb{R}^d} \left(V(\mathbf{u}, s) - sV(\mathbf{u}, 1)\right)^2 w(\mathbf{u}) d\mathbf{u},$$

where $\{V(\mathbf{u}, s);\ \mathbf{u} \in \mathbb{R}^d,\ s \in (0, 1)\}$ is a Gaussian process with zero mean and covariance structure

$$\mathrm{cov}(V(\mathbf{u}_1, s_1), V(\mathbf{u}_2, s_2)) = \min(s_1, s_2) C(\mathbf{u}_1, \mathbf{u}_2),$$

$$C(\mathbf{u}_1, \mathbf{u}_2) = \mathrm{cov}\big(\cos(\mathbf{u}_1^T \mathbf{X}_1) + \sin(\mathbf{u}_1^T \mathbf{X}_1), \cos(\mathbf{u}_2^T \mathbf{X}_1) + \sin(\mathbf{u}_2^T \mathbf{X}_1)\big)$$

.

• The one-dimensional setup in Hušková and Meintanis (2006) with differently formulated the limit distribution.

• An approximation of the related limit distributions are
(i) to estimate these quantities and then simulate the limit distribution by Monte Carlo
(ii) to apply a proper version of resampling.

• The same test statistics can be used for dependent observations (e.g., α–mixing), possible further extension to testing of no change in the joint distribution of the vector $(\mathbf{X}_t, \ldots, \mathbf{X}_{t+q})'$, for given $q \geq 1$.

- Behavior of $Q_{T,w}(\gamma)$ under alternatives.

Denote the CFs before and after the change by φ_0 and φ^0, respectively, and

$$B_0(\mathbf{u}) = E\Big(\cos(\mathbf{u}'\mathbf{X}_t) + \sin(\mathbf{u}'\mathbf{X}_t)\Big), \quad 1 \le t \le t_0,$$
$$B^0(\mathbf{u}) = E\Big(\cos(\mathbf{u}'\mathbf{X}_t) + \sin(\mathbf{u}'\mathbf{X}_t)\Big), \quad t_0 + 1 \le t \le T.$$

Theorem 3 Let $\mathbf{X}_1, \dots, \mathbf{X}_T$ be independent d-dimensional random vectors and let $\mathbf{X}_1, \dots, \mathbf{X}_{t_0}$ and $\mathbf{X}_{t_0+1}, \dots, \mathbf{X}_T$ have CF φ_0 and φ^0, respectively. Let assumption (2) on the weight function $w(\cdot)$ be satisfied and assume that $t_0 = \lfloor Ts_0 \rfloor$, for some $s_0 \in (0,1)$. Then, as $T \to \infty$, for $s \in (0,1)$

$$\frac{(s(1-s))^2}{T} D_{\lfloor Ts \rfloor, w} \xrightarrow{P} (\min(s, s_0)(1 - \max(s, s_0)))^2 \int_{\mathbb{R}^d} (B_0(\mathbf{u}) - B^0(\mathbf{u}))^2 w(\mathbf{u}) d$$

Change detection in VAR models

For fixed $q > 0$, assume that we observe p-dimensional $\mathbf{X}_t, \; t = 1, \ldots, T$, coming from the VAR($q$) model

$$\mathbf{X}_t = \sum_{j=1}^{q} \mathbf{A}_j \mathbf{X}_{t-j} + \varepsilon_t, \tag{3}$$

$\{\varepsilon_t\}$ – a sequence of $(p \times 1)$ i.i.d. random vectors (innovations) with

$\mathbb{E}(\varepsilon_t) = 0$, $\mathbb{E}(\varepsilon_t \varepsilon_t') = \mathbf{\Sigma}_\varepsilon$ and $\mathbb{E}(\varepsilon_t \varepsilon_s') = 0, \; t \neq s$.

$\{\mathbf{A}_j\}_{j=1}^{q}$ – $(p \times p)$ square matrices with unknown elements fulfilling the usual stability condition

$$\det(\mathbb{I}_p - \sum_{j=1}^{q} \mathbf{A}_j z^j) \neq 0, \; |z| \leq 1$$

, with $\mathbb{I}_p$ denoting the identity matrix of dimension $(p \times p)$

Several kinds of alternatives:

- changes in the parameters $\mathbf{A}_j$,
- changes in the correlation structure of $\mathbf{\Sigma}_\varepsilon$
- a change in the shape of the conditional distribution of the innovation

Earlier work on change point detectors in the context of VAR models includes Bai et al. (1998), Bai (2000), Ng and Vogelsang (2002), Qu and Perron (2007), Dvořák and Prášková (2013), Dvořák (2015, 2016).

We consider the detection problem for model (3) where F_t denotes the distribution of $\varepsilon_t,\ t \geq 1$
our test statistic will be based on corresponding residuals

$$\widehat{\varepsilon}_t = \mathbf{X}_t - \sum_{j=1}^{q} \widehat{\mathbf{A}}_j \mathbf{X}_{t-j}, \tag{4}$$

$\widehat{\mathbf{A}}_j$, $j = 1, \ldots, q$, are $\sqrt{T}$ consistent estimators of $\mathbf{A}_j$, $j = 1, \ldots, q$,

a set of starting values $\mathbf{X}_{1-p}, \ldots, \mathbf{X}_0$, exists.

The criterion based on $\widehat{Q}_{T,w}(\gamma)$ is given by

$$\widehat{Q}_{T,w}(\gamma) = \max_{1\le t<T} \left(\frac{t(T-t)}{T^2}\right)^{2+\gamma} T\widehat{D}_{t,w}$$

$$\widehat{D}_{t,w} = \int_{\mathbb{R}^d} \left|\widehat{\phi}_t(\mathbf{u}) - \widehat{\phi}^t(\mathbf{u})\right|^2 w(\mathbf{u})d\mathbf{u},$$

$$\widehat{\phi}_t(u) = \frac{1}{t}\sum_{\tau=1}^{t} e^{i\mathbf{u}'\widehat{\varepsilon}_\tau}, \qquad \widehat{\phi}^t(u) = \frac{1}{T-t}\sum_{\tau=t+1}^{T} e^{i\mathbf{u}'\widehat{\varepsilon}_\tau}, \tag{5}$$

computed from $\widehat{\varepsilon}_1,\ldots,\widehat{\varepsilon}_t$ and $\widehat{\varepsilon}_{t+1},\ldots,\widehat{\varepsilon}_T,\ t=1,\ldots,T$, respectively.

Limit distribution under the null hypothesis does not depend the chosen estimators $\widehat{\mathbf{A}}_j, j=1,\ldots,q$. Similar limit properties as above– replacing ε_j by $\mathbf{X}_j$

Similar as in two-sample problem **Computations**
Similar as in two-sample problem. We add
Recall- Henze and Wagner (1997) proposed weight function
$w(\mathbf{u}) = e^{-a\|\mathbf{u}\|^2}, \ a > 0,$
which leads to

$$I_w(\mathbf{x}) = \left(\frac{\pi}{a}\right)^{d/2} e^{-\|\mathbf{x}\|^2/4a}, \tag{6}$$

where $\|\mathbf{z}\| = \sqrt{\sum_{m=1}^{d} z_m^2}$ denotes the Euclidian norm of an arbitrary vector $\mathbf{z}$ of dimension d.

Matteson and James (2014), motivated by Székely and Rizzo (2005), suggest $I_w(\mathbf{x})$ is given by

$$I_w(\mathbf{x}) = \int_{\mathbb{R}^d} (1 - \cos(\mathbf{u}'\mathbf{x}))w(\mathbf{u})d\mathbf{u},$$

with weight function $w(\mathbf{u}) = 1/(C\|\mathbf{u}\|^{d+a})$, where C is a fixed known constant depending on d and a.

Székely and Rizzo (2005) showed $I_w(\mathbf{x}) = \|\mathbf{x}\|^a$.
$0 < a < 2$, and Matteson and James (2014) include the extra assumption of finite moment of order $a \in (0, 2)$ for the underlying random variable.
This leads to

$$Q_{T,w}(\gamma) = \min_{1 \le t < T} \left(\frac{t(T-t)}{T^2} \right)^{2+\gamma} T\, \Psi_{t,w}, \tag{7}$$

where

$$\Psi_{t,w} = \frac{1}{t^2} \sum_{\tau,s=1}^{t} \|\mathbf{X}_{\tau,s}\|^a + \frac{1}{(T-t)^2} \sum_{\tau,s=t+1}^{T} \|\mathbf{X}_{\tau,s}\|^a - \frac{2}{t(T-t)} \sum_{\tau=1}^{t} \sum_{s=t+1}^{T} \|\mathbf{X}_{\tau,s}\|^a. \tag{8}$$

Simulations

The setup of the simulation study has been inspired by Dvořák (2015):

$$\mathbf{A}_1 = \begin{pmatrix} 0.5 & 0.2 \\ 0.2 & 0.1 \end{pmatrix}, \quad \text{and} \quad \mathbf{\Sigma}_\varepsilon = \sigma \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

with the parameter σ controlling the scale and the parameter ρ the correlation, distributions of the random error terms:

1. multivariate normal (N),
2. multivariate t_{df} with df degrees of freedom,
3. multivariate χ^2_{df} with df degrees of freedom.

All distributions are standardized, i.e., $\mathbb{E}(\varepsilon_t) = 0$ and $\mathbb{E}(\varepsilon_t \varepsilon_t') = \mathbf{\Sigma}_\varepsilon$.

The multivariate normal and t_{df} distributions were simulated using R library `mvtnorm` (Genz et al, 2014; Genz and Bretz, 2009).

The multivariate χ^2_{df} distribution was simulated according to Minhajuddin et al (2004).

Test statistic $Q_{T,w}(\gamma)$ with the weight function $w(\mathbf{u}) = e^{-a\|\mathbf{u}\|^2}, \; a > 0$.
The VAR coefficients are estimated using OLS method.

Empirical level

Table : Empirical level (in %) for five error distributions ($\gamma = -0.5$).

γ	a	α	$T = 200$					T=400				
			N	t_3	t_4	χ^2_2	χ^2_4	N	t_3	t_4	χ^2_2	χ^2_4
		0.01	0.9	0.8	0.6	1.1	0.9	1.2	0.7	1.4	1.3	1.3
	1	0.05	3.3	3.9	4.9	4.9	4.2	4.9	4.5	5.4	4.1	5.1
		0.10	7.9	9.9	10.3	10.2	9.6	9.8	9.9	11.0	8.2	11.0
		0.01	1.0	0.5	0.3	0.9	1.0	0.8	1.3	1.5	1.5	0.6
-0.5	2	0.05	4.6	3.7	5.4	5.1	5.3	4.3	4.8	5.4	4.5	4.1
		0.10	9.8	8.1	10.8	9.9	9.8	11.0	9.8	10.2	10.0	7.9
		0.01	1.0	1.3	0.6	1.1	1.0	1.0	1.1	0.5	1.1	1.4
	3	0.05	5.0	4.2	4.4	5.0	3.8	5.6	4.7	4.3	5.6	4.9
		0.10	10.4	8.2	9.2	9.8	9.3	9.1	9.6	11.2	10.9	9.9

Empirical level

Table : Empirical level (in %) for five error distributions ($\gamma = 0$).

γ	a	α	$T = 200$ N	t_3	t_4	χ^2_2	χ^2_4	T=400 N	t_3	t_4	χ^2_2	χ^2_4
		0.01	1.2	0.7	0.8	0.9	1.3	0.9	1.2	0.6	1.0	0.8
	1	0.05	4.6	4.7	3.6	3.9	6.4	5.2	4.3	3.6	5.5	3.2
		0.10	10.3	8.8	8.5	8.3	11.6	10.8	7.5	7.8	11.7	7.5
		0.01	1.6	1.1	0.5	1.0	1.1	1.2	0.9	1.6	1.0	1.2
0	2	0.05	6.4	4.4	5.0	4.7	4.9	5.2	4.5	4.9	4.4	4.5
		0.10	11.9	7.9	9.0	8.6	10.0	9.3	8.8	8.7	8.4	10.2
		0.01	1.0	0.9	1.5	1.0	1.0	1.0	0.9	0.9	1.3	0.9
	3	0.05	5.8	5.1	5.1	4.2	4.6	5.0	5.0	5.1	6.2	5.2
		0.10	10.8	8.9	10.8	8.7	10.4	9.9	11.1	9.8	10.6	11.2

Empirical level

Table : Empirical level (in %) for five error distributions ($\gamma = 0.5$).

			$T = 200$					T=400				
γ	a	α	N	t_3	t_4	χ^2_2	χ^2_4	N	t_3	t_4	χ^2_2	χ^2_4
		0.01	0.9	0.7	0.5	1.0	1.0	1.5	1.4	1.1	0.7	1.0
	1	0.05	5.2	3.8	5.5	4.3	5.2	5.2	4.8	5.3	4.8	4.9
		0.10	10.0	8.4	10.8	9.3	9.9	10.4	9.5	10.3	9.5	9.1
		0.01	0.6	0.6	1.2	1.1	0.9	0.8	0.7	1.1	1.4	0.8
0.5	2	0.05	4.7	4.4	5.5	5.3	5.2	3.7	3.6	4.0	5.3	5.0
		0.10	11.0	9.2	11.4	10.0	10.7	8.6	8.8	9.9	8.9	11.0
		0.01	0.8	0.6	0.8	1.3	0.8	0.9	0.7	1.5	1.2	1.3
	3	0.05	3.8	4.3	5.2	5.9	5.0	4.6	4.3	6.0	5.7	4.0
		0.10	8.3	8.9	10.9	10.8	9.8	8.6	9.4	10.6	12.0	8.2

Empirical power

the power of the change–point test with respect to changes in the error distribution.

the distribution before the change–point $t_0 = \tau_0 T$ is bivariate normal with the variance matrix $\boldsymbol{\Sigma}_\varepsilon$ defined in the previous section with $\sigma_1 = 1$ and $\rho_1 = 0.2$

types of change:

1. change in scale (the parameter $\sigma_1 = 1$ changes to $\sigma_2 = 2$),
2. change in correlation (the parameter $\rho_1 = 0.2$ changes to $\rho_2 = 0.6$),
3. change in distribution (normal distribution changes to t_4 or χ^2_4).

Table : Empirical power (in %) for several types of change in the error distribution with changepoint $t_0 = \tau_0 T$. The symbol $\star$ denotes 100%, $a = 2$.

			$\sigma_1 \to \sigma_2$			$\rho_1 \to \rho_2$			$N \to t_4$			$N \to \chi^2_4$		
T	τ_0	γ	−0.5	0.0	0.5	−0.5	0.0	0.5	−0.5	0.0	0.5	−0.5	0.0	0.5
200	0.1		37.0	20.6	19.4	4.7	5.9	4.5	5.7	4.4	4.7	6.8	4.8	5.1
	0.2		98.7	97.4	91.6	4.9	5.2	7.3	6.8	6.3	5.3	10.0	9.5	9.7
	0.5		$\star$	99.9	$\star$	9.2	7.9	8.7	8.9	7.2	8.6	19.0	20.0	20.5
	0.8		99.4	98.0	95.0	6.4	5.9	4.9	5.9	6.6	5.7	11.3	9.2	8.1
400	0.1		96.7	64.5	43.4	5.6	5.7	5.0	6.6	5.5	6.8	8.9	6.4	7.4
	0.2		$\star$	$\star$	$\star$	6.3	6.0	5.1	8.7	8.4	6.8	18.4	14.1	13.0
	0.5		$\star$	$\star$	$\star$	11.4	12.7	13.4	13.1	12.9	16.9	36.8	37.5	40.2
	0.8		$\star$	$\star$	$\star$	7.7	5.1	6.9	6.7	7.6	6.9	15.2	14.7	11.4
600	0.1		$\star$	97.1	74.0	6.5	5.3	6.1	5.8	4.9	5.5	11.5	8.6	6.7
	0.2		$\star$	$\star$	$\star$	8.2	8.3	6.2	11.4	10.0	7.7	24.4	21.5	19.7
	0.5		$\star$	$\star$	$\star$	15.5	19.6	20.3	18.6	21.4	21.7	50.2	57.0	56.6
	0.8		$\star$	$\star$	$\star$	8.8	7.6	7.1	7.7	7.9	7.9	21.8	20.1	18.5

The test has good power against changes in the variance of the random errors.

The empirical power against other types of alternatives is much lower.

With $T = 600$ observations, the test rejects the null hypothesis of no change with probability 20% for the change in the correlation of random errors and for the change from Normal to t_4 distribution.

The probability of detecting the change from Normal to χ_4^2 distribution with the same number of observations is approximately 50%.

Concerning the choice of the parameter γ, it seems that $\gamma = 0.5$ works somewhat better for changes occurring in the center of the time series ($\tau_0 = 0.5$) and $\gamma = -0.5$ works somewhat better especially for changes occurring earlier. In our opinion,
the value $\gamma = 0.0$ provides a reasonable compromise.

Real data analysis

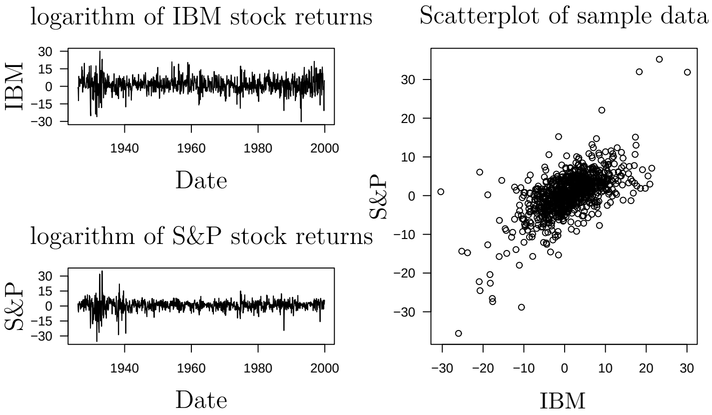
We apply the proposed test on the bivariate time series consisting of monthly log returns of IBM and S&P500 from January 1926 until December 1999 (Tsay, 2010).

This data set has been already investigated in Dvořák (2015, Section 3.6), who considered VAR(5) model and identified a change in its parameters in December 1932.

Looking at the time series ($T = 888$) and applying the proposed test with parameters $a = 2$ and $\gamma = 0$, we also reject the null hypothesis of no change (p-value= 0.0045), estimated change point only in !990.

Table : p–values for monthly IBM and S&P500 log returns for seven decades.

decade	1930s	1940s	1950s	1960s	1970s	1980s	1990s
p-value	0.2380	0.1595	0.8590	0.4740	0.2430	0.4245	0.0185


logarithm of IBM stock returns
IBM
30
15
0
−15
−30
1940
1960
1980
2000
Date
logarithm of S&P stock returns
S&P
30
15
0
−15
−30
1940
1960
1980
2000
Date
Scatterplot of sample data
S&P
30
20
10
0
−10
−20
−30
−30
−20
−10
0
10
20
30
IBM

Two-sample problem in higher dimension

Extension of on finite dimension two-sample problem to high dimension, functional data.

$X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_1}$ – two sequences of independent of random functions on $(0,1)$, $X_j = \{X_j(t);\ t \in (0,1)$, $Y_j = \{Y_j(t);\ t \in (0,1)$

$X_1, \ldots, X_{n_1}$ are i.i.d are random function on $(0,1)$,

$Y_1, \ldots, Y_{n_1}$ are i.i.d are random function on $(0,1)$

Our interest is to test that all random functions have the same distribution but we we start with subproblems:

Equality of marginal distributions

$H_0^* : \varphi_{X_j(t)}(u) =^d \varphi_{Y_j(t)}(u), \quad \forall t \in (0,1),\ u \in \mathcal{R}^1$

where $\varphi_{X_j(t)}(u)$ – characteristic function of $X_j(t)$ at u, the null hypothesis concerns only marginal distributions

Parallel to classical two-sample problem we introduce the test statistic:

$$D_w = \int_0^1 \Big(\int_{\mathcal{R}^1} \Big|\widehat{\varphi}_{X(t)}(u) - \widehat{\varphi}_{Y(t)}(u)\Big|^2 w(u)du \Big) dt,$$

$$\widehat{\varphi}_{X(t)}(u) = \frac{1}{n_1}\sum_{j=1}^{n_1} \exp\{iX_j(t)u\}, \quad \widehat{\varphi}_{Y(t)}(u) = \frac{1}{n_2}\sum_{j=1}^{n_2} \exp\{iY_j(t)u\}$$

Under the H_0^* for $min(n_1, n_2) \to \infty, n_1/(n_1+n_2) \to \theta \in (0,1)$

$$(n_1+n_2)D_w \to^d \frac{1}{\theta(1-\theta)} \int_0^1 \int_{\mathcal{R}^1} (V_\theta(t,u))^2 w(u)du$$

$\{V_\theta(t,u),\ t \in (0,1),\ u \in \mathcal{R}^1\}$ is a Gaussian process with zero mean and covariance structure

If observations obtained only in discrete time points

$t_j = j/m,\ j = 1, \dots, m$

$$\widetilde{D}_w(m) = \int_{\mathcal{R}^1} \frac{1}{m}\sum_{j=1}^{m} \Big|\widehat{\varphi}_{X(t_j)}(u) - \widehat{\varphi}_{Y(t_j)}(u)\Big|^2 w(u)du,$$

Version based on simultaneous characteristic functions at discrete points

$$\widetilde{\Delta}_w(m) = \int_{\mathcal{R}^m} \frac{1}{m} \sum_{j=1}^{m} \left| \widehat{\varphi}_{X(t_1,\dots,t_m)}(u) - \widehat{\varphi}_{Y(t_1,\dots,t_m)}(u) \right|^2 w(\mathbf{u}) d\mathbf{u},$$

$$\widehat{\varphi}_{X(t_1,\dots,t_m)}(\mathbf{u}) = \frac{1}{n_1} \sum_{j=1}^{n_1} \exp\{i \sum_{v=1}^{m} X_j(t_v) u_v\}$$

$$\widehat{\varphi}_{Y(t_1,\dots,t_m)}(\mathbf{u}) = \frac{1}{n_2} \sum_{j=1}^{n_2} \exp\{i \sum_{v=1}^{m} Y_j(t_v) u_v\}$$

$$\mathbf{X}(t_1, \dots, t_m) = (X(t_1), \dots, X(t_m))^T, \quad \mathbf{Y}(t_{1,\dots,t_m}) = (Y(t_1), \dots, Y(t_m))^T.$$

Notice that $\sum_{v=1}^{m} X_j(t_v) u_v$ is a scalar of product $\mathbf{X}(t_{1,\dots,t_m})$ and $(u_1, \dots, u_m)$

The question is the choice of weight function $w(\cdots)$????

Computational version:

$$\widetilde{\Delta}_w(m) = \frac{1}{n_1^2}\sum_{s,v=1}^{n_1} I_w(\mathbf{X}_s - \mathbf{X}_v) + \frac{1}{n_2^2}\sum_{s,v=1}^{n_2} I_w(\mathbf{Y}_s - \mathbf{Y}_v)$$

$$-\frac{2}{n_2 n_1^2}\sum_{s}^{n_1}\sum_{v}^{n_2} I_w(\mathbf{X}_s - \mathbf{Y}_v)$$

$$I_w(\mathbf{z}) = \int_{\mathcal{R}^m} \cos(\sum_{j=1}^{m} u_j z_j) w(\mathbf{u}) d\mathbf{u}$$

possible choose: $I_w(\mathbf{z}) = \exp\{-\frac{1}{2}\mathbf{z}^T \mathbf{\Sigma}_m^{-1}\mathbf{z}\}$ with $\mathbf{\Sigma}_m > 0$ -symmetric

$\sigma(t_j, t_v) = \min(t_j, t_v)$ – Wiener process behind it
or
$\sigma(t_j, t_v) = \exp\{-|t_j - t_v|/2\}$ - Ornstein-Uhlenbeck process behind it
or
matrix related to covariance matrix of $\mathbf{X}_j(t_1, \dots, t_m)$ and $\mathbf{Y}_j(t_1, \dots, t_m)$

Simulations

Test for marginal distributions

$$D_w = \int_0^1 \left(\int_{\mathcal{R}^1} \left| \widehat{\varphi}_{X(t)}(u) - \widehat{\varphi}_{Y(t)}(u) \right|^2 w(u) du \right) dt,$$

with $w(\mathbf{u}) = e^{-a\|\mathbf{u}\|^2}$, $a > 0$ (denoted by α
Test for simultaneous distributions

$$\widetilde{\Delta}_w(m) = \int_{\mathcal{R}^m} \frac{1}{m} \sum_{j=1}^{m} \left| \widehat{\varphi}_{X(t_1,\dots,t_m)}(u) - \widehat{\varphi}_{Y(t_1,\dots,t_m)}(u) \right|^2 w(\mathbf{u}) d\mathbf{u},$$

also a small comparison with tests by Horváth and Rice (2013)

Application: Australian Temperature Data
224 weather stations across Australia
monthly mean maximum temperatures in degrees Celsius
four periods: 1914 to 1933 (period 1), 1934 to 1953 (period 2), 1954 to 1973 (period 3), 1974 to 1993 (period 4)

Table B.1: Probability of rejection for the ECF test $D_{\alpha,m}$ at 5% significance based on 1000 permutations when sample noise is equidistributed between the two groups

Sample Size	Time Points	Distance Parameter: δ						
n_1=n_2	m	0	0.2	0.4	0.6	0.8	1	2
		$\alpha = 0.5$						
15	20	0.038	0.061	0.078	0.218	0.634	0.922	1
	100	0.050	0.034	0.116	0.618	0.982	1	1
25	20	0.058	0.052	0.13	0.406	0.908	1	1
	100	0.064	0.044	0.238	0.918	1	1	1
50	20	0.054	0.052	0.192	0.854	1	1	1
	100	0.040	0.062	0.522	1	1	1	1
		$\alpha = 1$						
15	20	0.058	0.048	0.082	0.242	0.652	0.966	1
	100	0.046	0.086	0.136	0.670	0.996	1	1
25	20	0.050	0.062	0.120	0.462	0.958	1	1
	100	0.036	0.076	0.252	0.946	1	1	1
50	20	0.066	0.054	0.206	0.912	1	1	1
	100	0.056	0.046	0.600	1	1	1	1
		$\alpha = 1.5$						

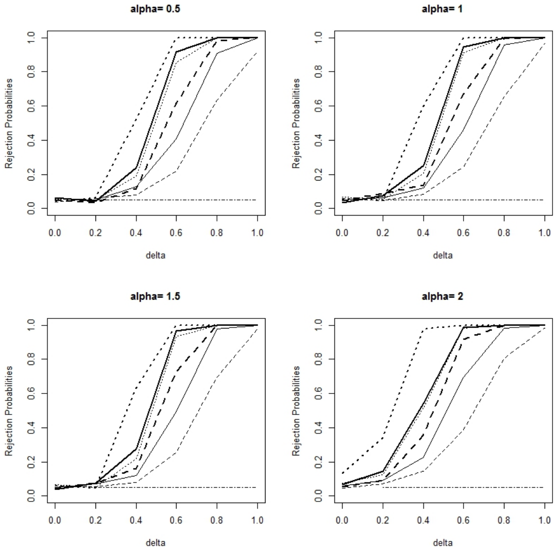


Figure B.1: Probability of rejection for the ECF test $D_{\alpha,m}$ at 5% significance based on 1000 permutations when sample noise is equidistributed

Table B.5: Probability of rejection for the ECF test $\Psi_{\alpha,m}$, $\alpha = 1$, and the HKR method at 5% significance with m time points

delta	m=10		m=30	
	ECF	HKR	ECF	HKR
0.0	0.052	0.0	0.048	0.061
0.1	0.058	0.0	0.078	0.071
0.2	0.098	0.0	0.096	0.108
0.3	0.122	0.0	0.132	0.174
0.4	0.206	0.0	0.188	0.246
0.5	0.268	0.0	0.284	0.331
0.6	0.358	0.0	0.382	0.448
0.7	0.470	0.0	0.498	0.530
0.8	0.578	0.0	0.610	0.665
0.9	0.634	0.0	0.712	0.751
1.0	0.770	0.0	0.800	0.811
1.1	0.812	0.0	0.854	0.889
1.2	0.910	0.0	0.896	0.928
1.3	0.926	0.0	0.952	0.960

7.pdf

Table B.9: Probability of rejection for ECF test $\Upsilon_{W,m}$ at 5% significance based on 1000 permutations when samples are Wiener process without noise

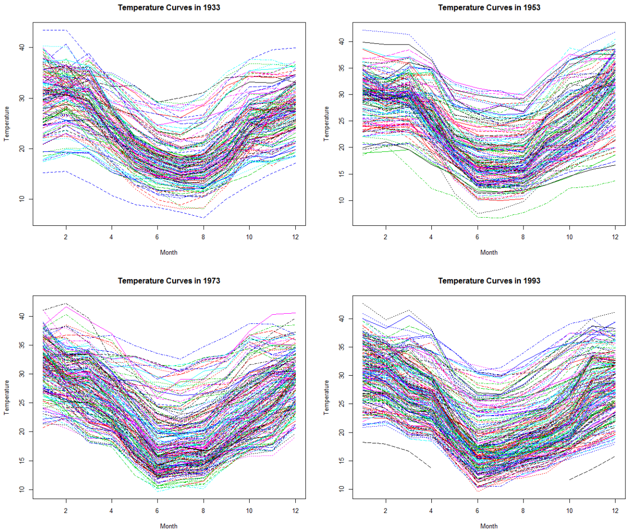
Sample Size	Time Points	Distance Parameter: δ					
$n_1 = n_2$	m	0	0.1	0.2	0.3	0.4	0.5
15	15	0.060	0.060	0.190	0.270	0.410	0.570
	25	0.050	0.090	0.150	0.420	0.520	0.640
25	15	0.040	0.140	0.180	0.410	0.650	0.700
	25	0.050	0.080	0.330	0.530	0.630	0.760
50	15	0.060	0.090	0.460	0.670	0.930	0.950
	25	0.060	0.160	0.480	0.670	0.830	0.920

10b.pdf

Table B.10: Probability of rejection for ECF test $\Upsilon_{W,m}$ at 5% significance based on 1000 permutations when samples are Ornstein-Uhlenbeck process without noise

Sample Size	Time Points	Distance Parameter: δ					
$n_1 = n_2$	m	0	0.1	0.2	0.3	0.4	0.5
15	15	0.045	0.065	0.145	0.250	0.370	0.550
	25	0.055	0.095	0.210	0.355	0.570	0.590
25	15	0.045	0.095	0.175	0.355	0.625	0.720
	25	0.060	0.115	0.285	0.490	0.815	0.840
50	15	0.060	0.140	0.325	0.670	0.890	0.930
	25	0.050	0.140	0.405	0.700	0.810	0.890

11.pdf



10.pdf Figure B.4: Temperature curves for the 224 weather stations in 1933 (Top Left), 1953 (Top Right), 1973 (Bottom Left), 1993 (Bottom Right)

Table B.8: Probability of rejection for the ECF test $D_{\alpha,m}$, $\alpha = 1$, of the pairwise tests based on the Australian weather data

Period	Period	P-value
1	2	0.0553
1	3	0.0519
1	4	0.0199
2	3	0.0521
2	4	0.0391
3	4	0.1296

9.pdf

References

• Qing Jiang, Marie Hušková, Simos G. Meintanis, Lixing Zhu: *Asymptotics, finite-sample comparisons and applications for two-sample tests with functional data*. just accepted for publication in JMVA

• L. Horváth, P. Kokoszka, *Inference for Functional Data with Applications*, Springer Series in Statistics, Springer, New York (2012

• Horváth, G. Rice, *Testing equality of means when observations are from functional time series*, J. Time Ser. Anal. 36 (2015)

• P. Hall, I. Van Keilegom,*Two-sample tests in functional data analysis starting from discrete data*, Stat. Sinica. 17 (2007)

• Z. Hlávka, M. Hušková and S. G. Meintanis: *Change point detection with multivariate observations based on characteristic functions*, Festschrift for W. Stute, Springer Verlag, 2017

THANK YOU!!!!