CrossMark

Detecting a structural change in functional time series using local Wilcoxon statistic

Daniel Kosiorowski[1] · **Jerzy P. Rydlewski**[2] ·
Małgorzata Snarska[3]

Abstract Functional data analysis is a part of modern multivariate statistics that analyzes data that provide information regarding curves, surfaces, or anything that varies over a certain continuum. In economics and empirical finance, we often have to deal with time series of functional data, where decision cannot be made easily, for example whether they are to be considered as homogeneous or heterogeneous. A discussion on adequate tests of homogenity for functional data is carried out in literature nowadays. We propose a novel statistic for detecting a structural change in functional time series based on a local Wilcoxon statistic induced by a local depth function proposed by Paindaveine and Van Bever, and where a point of the hypothesized structural change is assumed to be known.

Keywords Functional data analysis · Local depth · Functional depth · Detecting structural change · Heterogenity · Wilcoxon test

Mathematics Subject Classification 62G30 · 62-07 · 62G35 · 62P20

✉ Jerzy P. Rydlewski
ry@agh.edu.pl

Daniel Kosiorowski
daniel.kosiorowski@uek.krakow.pl

Małgorzata Snarska
malgorzata.snarska@uek.krakow.pl

[1] Department of Statistics, Cracow University of Economics, Kraków, Poland

[2] AGH University of Science and Technology, Faculty of Applied Mathematics,
al. A. Mickiewicza 30, 30-059 Kraków, Poland

[3] Department of Financial Markets, Cracow University of Economics, Kraków, Poland

1 Introduction

Many objects in economics takes the form of a function of certain continuum, for example, utility curves, yield curves, electricity demand trajectories during day and night, time series of concentration of dangerous particulates in the atmosphere, the Internet traffic intensity within a day (see Figs. 1, 2). Economic phenomena are often observed through a certain number of non-homogenous components, that is, they exhibit multimodality. The phenomena seem to be similar globally, but locally they differ significantly. Global methods of population comparison (visual, inferential, and descriptive) using popular centrality measures, that is, mean or median, may be misleading. Further problems occur when functional outliers (defined with respect to functional boxplot) are present in the data set. Because of the lack of reliable economic theory on data generating processes, which may be used for describing economic phenomena, functional generalizations of well-known statistical procedures (e.g., ANOVA) are inefficient (Ramsay and Silverman 2005; Ramsay et al. 2009; Horváth and Kokoszka 2012). If an economic system in each period of time is described by a certain number of functions (e.g., individual demand and supply curves or investment strategies), then the characteristics of this dynamic system are observed as a multi-regime functional time series where heterogenity is related to change in probability distribution over the considered space of functions. The aim of this study is to detect a structural change related to local differences between populations or between population characteristics in two or more periods of study, and a point of the hypothesized structural change is assumed to be known (e.g., before and after financial crash).

The approach that is proposed will detect changes in the behavior of Internet service users behaviour related to a certain event, for example, introducing a new marketing strategy, modification of an online banking service, or some political event. This procedure can be used to detect changes in the distribution of development paths of a small company in the context of introducing a new taxation system. It can also detect
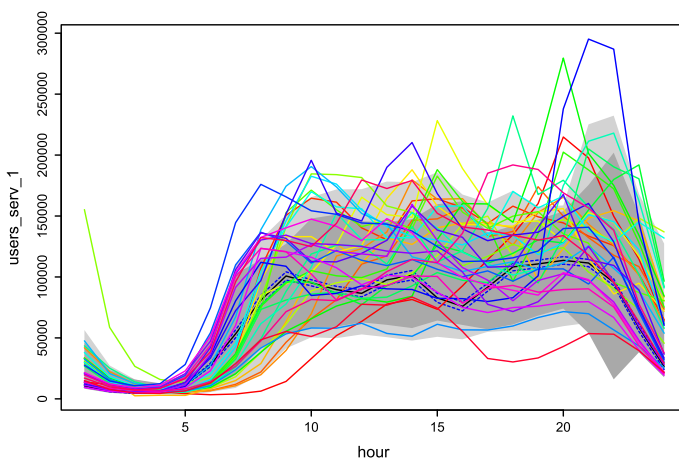

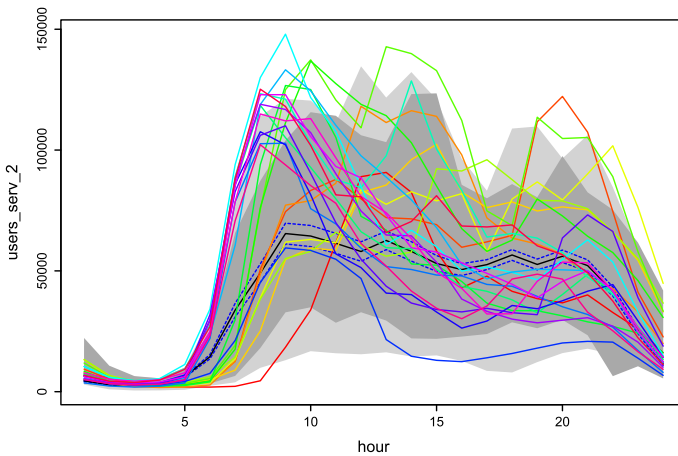**Fig. 1** Functional boxplot for number of users in service 1 during day and night

Fig. 2 Functional boxplot for number of users in service 2 during day and night

shifts in supply or demand for certain goods, for example, electricity, within day and night. Finally, this method can be used to monitor daily concentration of dangerous particles in the atmosphere.

In one dimensional case, Wilcoxon rank sum test correctly detects the differences in the location of a rich class of populations (Wilcox 2014). For many economic phenomena described by means of certain curves (e.g., yield curves, utility curves, and dangerous particles in atmosphere concentration curves), available structural change tests assume parametric form for each curve and rely on performing independent tests of parameter equality of the curves. This study proposes a novel nonparametric and robust test for the detection of a structural change in economic system, namely local extension of Wilcoxon test for two functional samples. The test can be effectively used for detecting a structural change in functional time series. The underlying idea is to compare populations at different locality level, which may be interpreted as data resolution. In our proposal, the local Wilcoxon test statistic is induced by the corrected modified band depth (López–Pintado and Romo 2007) with the concept of locality proposed by Paindaveine and Van Bever (2013).

The rest of the paper is organized as follows. Section 2 sketches the basic concepts of a two-sample test for homogenity in the context of functional time series. Section 3 introduces two-sample local Wilcoxon test statistic for detecting a structural change in functional time series. Section 4 discusses the properties of the procedure through numerical simulations and tests the applicability of the proposed methodology on empirical examples (i.e., activities of Internet users and monitoring yield curves). Section 5 conducts a short sensitivity analysis. Section 6 comprises a brief summary.

2 A concept of homogenous functional data

Nowadays, an intensive debate on adequate tests of homogeneity for functional data is carried out in literature (Flores et al. 2015, see the paper and references therein).

In Flores et al. (2015), selected two-sample homogenity tests were discussed on the basis of the comparison of maximal depth elements.

Functional time series are usually defined in terms of functional stochastic processes with values in Banach or Hilbert spaces (Bosq 2000; Horváth and Kokoszka 2012). Bosq (2000) explained that probability distribution $\mathbf{F}$ of functional random variable does exist. We look at the random curve $X = \{X(t), t \in [0, T]\}$ as a random element of the space $L^2 = L^2([0, T])$ equipped with the Borel $\sigma-$algebra. L^2 is a separable Hilbert space with the inner product $\langle x, y \rangle = \int x(t)y(t)dt$. We consider a sample of curves, but each curve is observed at discrete and finite grid of points in practice. Discrete data are transformed into curves using various techniques including nonparametric smoothing (Ramsay et al. 2009).

Horváth and Kokoszka (2012) prove various properties of the functional estimators and show that, under some regularity conditions, mean value and variance are unbiased and mean square error consistent estimators. Horváth et al. (2014) formalize the assumption of stationarity in the context of functional time series and propose several procedures to test the null hypothesis of stationarity, which in turn may be used to detect a structural change in functional time series (FTS) setup. Furthermore, Horváth et al. (2014) have noted that spectral analysis of nonstationary functional time series has not been developed to a point where usable extensions could be readily derived; hence they developed a general methodology for testing the assumption that a modeled functional time series is indeed stationary and analyzed the behavior of the tests under several alternatives, that is, change point alternative. The tests developed by the authors are consistent against any other sufficiently large departures from stationarity and weak dependence. They warn that in the functional setting, there is a fundamentally new aspect, that is, convergence of a scalar estimator of the long-run variance must be replaced by the convergence of the eigenvalues and the eigenfunctions of the long-run covariance function. They note that their method is extremely computationally intensive. Obtaining an obvious increasing/decreasing trend in functional data is uncertain. Fraiman et al. (2014) considered functional time series, where a trend is expected. They defined different kinds of trend and then performed various tests that could detect the trends. The authors developed the nonparametric tests for the proposed increasing trends for a sequence of functional data and established their results for a multiple time series of functional data.

Let us pose our hypotheses. If $\mathbf{F}$ and $\mathbf{G}$ denote a probability distribution of the first and second population, respectively, we can formulate null and alternative hypothesis:

$$H_0 : \mathbf{F} = \mathbf{G} \ \textit{vs.} \ H_1 : \mathbf{F} \neq \mathbf{G}. \tag{1}$$

Our first aim is to test the null hypothesis against its alternative having two samples in a disposal. In this situation, our first set of hypotheses states that two samples are drawn from the same distribution, whereas the alternative states the opposite. We use a local Wilcoxon statistic to deal with the problem.

Our second aim is to use the local Wilcoxon statistic to detect a structural change in functional time series. In other words, we intend to test a set of the following hypotheses:

$$H_0 : \mathbf{F}_{X_1} = \mathbf{F}_{X_2} = \cdots = \mathbf{F}_{X_N} \; vs. \; H_1 : \mathbf{F}_{X_1} = \cdots = \mathbf{F}_{X_k} \neq \mathbf{F}_{X_{k+1}} = \cdots = \mathbf{F}_{X_N}$$
(2)

for some $k \in \{1, 2, \ldots, N\}$, where $\mathbf{F}_{X_i}$ is a probability distribution of a functional random variable X_i.

We use a moving local Wilcoxon statistic for the purpose.

3 Our proposals

This section introduces a two-sample local Wilcoxon test for homogenity, that is, for veryfying (1) set of hypotheses.

3.1 Ranks induced by depth functions

Consider an functional data analysis (FDA) setup in which each observation is a real function defined on a common interval in $\mathbb{R}$. In order to introduce rank-based statistic for comparing samples of functional data, we focus our attention on statistical depth functions for functional objects. It enables us for ordering these objects in terms of departure of an object from a center—the functional median. *The data depth concept* was originally introduced to generalize the concept of order statistics to a multivariate case (Mosler 2013), but is presently treated as a very powerful data analytic tool that is able to express various features of the underlying distribution. The depth function yields information about spread, shape, and asymmetry of a distribution, through depth regions (Liu et al. 1999; Mosler 2013 and references therein).

Within the depth concept, it is possible to propose effective methods of location and scale differences testing (Li and Liu 2004).

Classical depth functions associate a maximal depth value with any center of symmetry. Together with the fact that depth decreases along any half-line originating from any deepest point, this leads to nested star-shaped (in most cases convex) depth regions, although the underlying distribution may be nonconvex (Zuo and Serfling 2000). However, distributions that are multimodal or have nonconvex support are present in many economic applications (mixture models, multi-regime time series, or issues solved by means of clustering and classification procedures, see, for example, Lange et al. 2015). These facts motivated several authors to extend the concept of depth to make it flexible enough to deal with such distributions. Such extensions are available in the literature, under the name *local depths*. This study uses the concept of local depth proposed by Paindaveine and Van Bever (2013) and implemented among others in Kosiorowski and Zawadzki (2014).

Thorough presentation of the depth concept may be found in Liu (1990), Zuo and Serfling (2000), Serfling (2006), Mosler (2013), and Nieto-Reyes and Battey (2016).

In recent years, some definitions of depth for functional data have been proposed as well. Fraiman and Muniz (2001) considered a concept of depth based on the integral of univariate depths, and López-Pintado and Jörnsten (2007) introduced functional depths taking into consideration the shape of the selected curve. Cuesta-Albertos and Nieto-Reyes (2008) introduced a random Tukey depth. Cuevas et al. (2007) proposed classification procedures that are implemented within fda.usc R package

(Febrero-Bande and de la Fuente 2012). A very useful theoretical considerations related to the definition of the functional depth and comparative study of several functional depths may be found in Nieto-Reyes and Battey (2016).

In our opinion, statistics induced by the functional depths may effectively be used for non-parametric and robust monitoring of certain properties of functional time series. In this context, we propose to use novel tools offered by robust functional analysis to test a reasonable hypothesis of equality distributions of the two given sets of functional sequences. Consider a situation in which we would like to compare two functional sequences $\{X_i\}_{i=1}^{n-m}$ and $\{Y_i\}_{i=1}^{m}$.

We suggest to proceed in the following manner. Using a concept of corrected generalized band depth (López-Pintado and Jörnsten 2007), we rank the original observations from one that is the closest to the functional median up to one that is the furthest one. Then, the Wilcoxon test or another rank test is conducted (see Hájek and Šidák 1967 for alternative rank tests).

Let us examine our procedure in detail. First, we combine both the samples $\{X_i\}_{i=1}^{n-m}$ and $\{Y_i\}_{i=1}^{m}$. Let now $\mathbf{X} = \{x_1, \ldots, x_n\}$ denote a combined sample of continuous curves defined on the compact interval T. Let λ denote the Lebesgue measure and let $a(i_1, i_2) = \{t \in T : x_{i_2} - x_{i_1} \geq 0\}$, where x_{i_1} and x_{i_2} are band delimiting objects. Let $L_{i_1,i_2} = \frac{\lambda(a(i_1,i_2))}{\lambda(T)}$. *A corrected generalized band depth* of a curve x with respect to the sample $\mathbf{X}$ is (López-Pintado and Jörnsten 2007; López–Pintado and Romo 2007)

$$cGBD(x|\mathbf{X}) = \frac{2}{n(n-1)} \sum_{1 \leq i_1 < i_2 \leq n} \frac{\lambda(A^c(x; x_{i_1}, x_{i_2}))}{\lambda(T)} \tag{3}$$

where

$$A^c(x; x_{i_1}, x_{i_2}) = \{t \in a(i_1, i_2) : x_{i_1}(t) \leq x(t) \leq x_{i_2}(t)\}, \quad \text{if } L_{i_1,i_2} \geq \frac{1}{2}$$

and

$$A^c(x; x_{i_1}, x_{i_2}) = \{t \in a(i_2, i_1) : x_{i_2}(t) \leq x(t) \leq x_{i_1}(t)\}, \quad \text{if } L_{i_2,i_1} > \frac{1}{2}.$$

Band depth is thus modified so as to consider only the proportion of the domain where the delimiting curves define a contiguous region that has non-zero width. To conduct the construction, we evaluate the depth regions of order α for cGBD, that is,

$$R_\alpha(P) = \{x : cGBD(x, P) \geq \alpha\}.$$

For any depth function $D(x, P)$, the depth regions, $R_\alpha(P) = \{x \in L^2([0, T]) : D(x, P) \geq \alpha\}$ are of paramount importance as they reveal very diverse characteristic of probability distribution P: location, scatter, and dependency structure (clearly these regions are nested and inner regions contain larger depth). While defining local depth, following the concept of Paindaveine and Van Bever (2013), it will be more appropriate to index the family $\{R_\alpha(P)\}$ by means of probability contents. Consequently, for any $\beta \in (0, 1]$, we define the smallest depth region with P-probability equal or larger than β as

$$R^\beta(P) = \bigcap_{\alpha \in A(\beta)} R_\alpha(P),$$

where $A(\beta) = \{\alpha \geq 0 : P(R_\alpha(P)) \geq \beta\}$. The depth regions $R_\alpha(P)$ or $R^\beta(P)$ provide neighborhood of the deepest point only. However, we can replace P by its symmetrized version $P_x = \frac{1}{2}P^{\mathbf{X}} + \frac{1}{2}P^{2x-\mathbf{X}}$. Let $D(\cdot, P)$ be a depth function. The corresponding *sample local depth function at the locality level* $\beta \in (0, 1]$ is $LD^\beta(x, P^{(n)}) = D(x, P_x^{\beta(n)})$, where $P_x^{\beta(n)}$ denotes the empirical measure of those data points that belong to $R_x^\beta(P^{(n)})$. $R_x^\beta(P^{(n)})$ is the smallest sample depth region that contains at least a proportion β of the $2n$ random functions $x_1, \ldots, x_n, 2x - x_1, \ldots, 2x - x_n$. Depth is always well defined—it is an affine invariant from original depth. For $\beta = 1$, we obtain global depth, whereas for $\beta \simeq 0$ we obtain extreme localization. As in the population case, our sample local depth will require considering, for any $x \in \mathbb{L}^2$, the symmetrized distribution P_x^n, which is an empirical distribution associated with $x_1, \ldots, x_n, 2x - x_1, \ldots, 2x - x_n$. Sample properties of the local versions of depths result from the general findings presented in Zuo and Serfling (2000).

Implementations of local versions of several depths including projection depth, Student, simplicial, L^p depth, regression depth, and modified band depth can be found in free R package *DepthProc* (Kosiorowski and Zawadzki 2014). For choosing the locality parameter β, we recommend using cross-validation related to an optimization of a certain merit criterion (the resolution being appropriate for comparing the phenomena in terms of their aggregated local shape differences, which relies on our knowledge on the considered phenomena).

3.2 Local Wilcoxon test for testing homogeneity

Let us consider two samples and $\mathbf{X} = \{x_1, x_2, \ldots, x_n\} = \{X_i\}_{i=1}^{n-m} \cup \{Y_i\}_{i=1}^m$. The ranks induced by a local corrected generalized band depth with prefixed locality parameter $\beta \in (0, 1]$ are

$$R_l = \# \left\{ x_j \in \mathbf{X} : cGDB(\beta)(x_j, \mathbf{X}) \leq cGDB(\beta)(x_l, \mathbf{X}) \right\}, \tag{4}$$

$l = 1, \ldots, n$. Ranking the original observations according to the $cGBD$ is done subsequently. Let the unified ranks in the combined sample of all observations be R_l, $l = 1, 2, \ldots, n$ or $R_{x_1}, \ldots, R_{x_{n-m}}$ ranks of X_i's and $R_{y_1}, \ldots, R_{y_m}$ ranks of Y_i's.

Proposal 1 We propose to conduct a proper Wilcoxon test to test the hypothesis of equality of the two distributions generating two given sets of functional sequences. The β-local (two independent samples) Wilcoxon rank sum statistic for functional data takes the following form:

$$S^\beta = \sum_{i=1}^{n-m} R_{x_i}, \tag{5}$$

where ranks are induced by local cGBD with locality parameter β. Following Li and Liu (2004), it is worth noticing that if two samples X and Y are found in a disposal

and any depth function, one can calculate depth in the combined sample $X \cup Y$, assuming empirical distribution calculated using all observations, or calculating this distribution assuming only one of the samples X or Y. If we observe that depths for $X'_l s$ indicate the center of the combined sample and depths for $Y'_l s$ indicate peripheries, we conclude Y was taken from distribution with bigger scatter. In general, differences in allocations of ranks between samples indicate various differences in shapes of underlying distributions and hence a departure from their equality (Liu et al. 1999; Kosiorowska et al. 2014). In the functional setting, a difference in scale means that an α-central region drawn from the population X consists of smaller amount of probability mass than taken from the population Y, and hence, Y is more scattered than X. The locality parameter β indicates a resolution in which we compare the populations. It ranges from a very misty comparison (parameter close to 1) to a very sharp comparison (the parameter close to 0). From other point of view, we can treat the statistic (5) as an aggregate representing local asymmetry in data set (Paindaveine and Van Bever 2013). Differences in value of (5) for two samples indicate differences in aggregated local asymmetry but simultaneously in local location and scale. Notice that for $\beta = 1$, we perform classical Wilcoxon rank sum test, and hence, we can use tables for this test to obtain critical values and use well-known tie-breaking schemes in case of ties (Jurečková and Kalina 2012). Big or small values of test statistics indicate differences in distributions between the samples then. For other β values, for each point we calculate depth with respect to empirical distribution symmetrized in this point. It may happen that two points have the same depth value and hence the same rank. However, we expect significant differences in sums of ranks for samples drawn from different continuous distributions (different distributions should be characterized by different kinds of local asymmetry). The differences we underline are related to the parameter of resolution β in which we conduct the comparison. On the other hand, the β parameter may be treated as parameter of data peeling of the combined sample—a parameter of desired sensitivity to contamination of our procedure. "A power" of the test depends not only on the differences between location and scale of the underlying distributions but also on the differences in "shape" of the underlying distributions in appropriate functional space. For practical purposes, we recommend Monte-Carlo evaluation of the "power" in case of selected alternatives being especially important for a decision–maker from a merit point of view. Merit properties of the proposal depends on the properties of functional depth that is being used. Sample properties and other asymptotic properties of the proposed statistic result from Nieto-Reyes and Battey (2016) and Paindaveine and Van Bever (2013). However, notice that Nieto-Reyes and Battey (2016) did not consider local functional depth but only global versions.

Flores et al. (2015) constructed four different statistics to measure the distance between two samples based on the comparison of maximal depth elements. They proposed two-sample tests for homogenity in the context of FDA. They did not use a concept of local depth in their considerations. Our approach enables to use a locality parameter β, which indicates a resolution in which we compare the populations. It ranges from a very misty comparison (parameter close to 1) to a very sharp comparison (the parameter close to 0). The researcher may adjust the locality parameter on the grounds of the matter being considered and her/his experience. Our proposal (1) outperforms their proposals in cases of multimodal distributions.

Proposal 2: detecting a structural change For a given functional reference sample $X_1, \ldots, X_M$, we would like to compare the stream of functional data $Y_1, \ldots, Y_N$, where $N >> M$ with our reference sample, that is, to detect a structural change in a functional data stream.

We construct a moving window of length L and sequentially test a homogeneity of $X_1, \ldots, X_N$ and $Y_k, \ldots, Y_{k+L-1}$ for $k \in \{1, \ldots, N-L+1\}$ using the statistic (5). Our procedure detects a structural change in a functional data stream. In order to obtain sample distribution of the test statistic and in consequence the necessary p-values, we propose to use a maximal entropy bootstrap methodology proposed by Vinod and de Lacalle (2009) and implemented in *meboot* R package. Note that in the time series setting, because of the temporal dependence between observations, resampling and especially bootstrap seem to be the only solution to conduct statistical inference (Shang (2016)). Having empirical time series under our study, we generate bootstrap samples using *meboot* R package, and then we calculate our sample Wilcoxon statistic distribution to obtain appropriate p-values.

4 Properties of the proposals: simulation studies

In order to determine the finite sample properties of our proposal, we conducted simulation studies. In order to find the finite sample properties of our "static" proposal (1), we conducted extensive simulation studies involving various differences in location and scale and shape of distributions generating samples. We generated two samples 100 times from the same distribution (situation representing null hypothesis) and from two distributions of the same kind but differing with respect to location and scale. Similarly to Horváth et al. (2014) and Didericksen et al. (2012), we considered samples with functional errors being generated by Wiener process and Brownian bridge divided into 1440 and 120 time points (24 hours divided into 1- and 12-min time segments). We considered samples of equal and different sizes. In general, in case of simple differences in location and scale, our proposal performed comparable to proposals introduced by Flores et al. (2015) based on maximal depth elements in two sample comparison but significantly outperformed them in cases of existence of multimodality—local differences between samples. Figure 3 presents a sample of 50 curves generated from Wiener process observed at 120 points. Figure 4 presents a sample of 50 curves generated from 5% mixture of two Wiener processes differing with respect to location. Figure 5 presents estimated density of the statistic (5) under the hypothesis that both samples are generated from the population related to Fig. 3, and Fig. 6 presents the estimated density of statistic (5) under an alternative in which the first sample is generated from population related to Fig. 3, and the second population is generated from a population related to Fig. 4. It is easy to notice that the estimated densities differ with respect to the location and hence may be used to discriminate between populations. Further results and R codes are available upon request (we performed sensitivity analysis similar to given in Flores et al. 2015).

For checking the proposed structural change detection procedure, we generated time series from the following models having economic justification in the context of cyclical properties modeling. We used functional autoregression model FAR(1), that
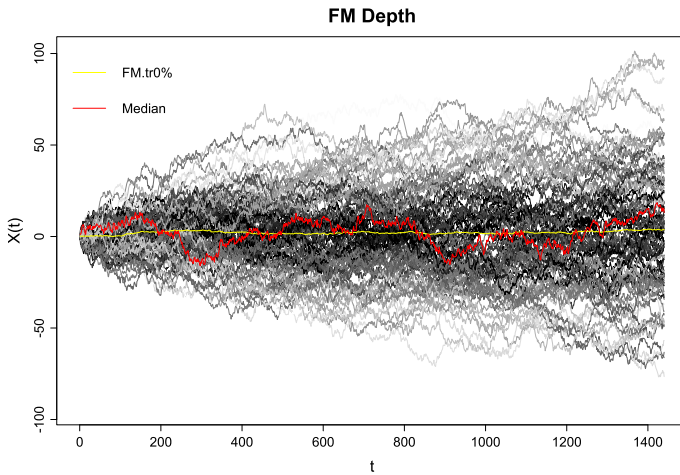
FM Depth

Fig. 3 Sample of 50 curves generated from Wiener process observed at 120 points
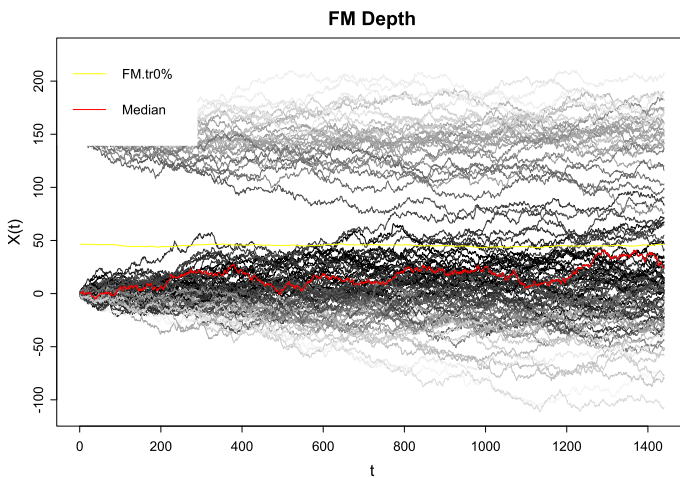
FM Depth

Fig. 4 Sample of 50 curves generated from 5% mixture of two Wiener processes differing with respect to location

is, $X_{n+1} = \Psi(X_n) + \epsilon_{n+1}$, in which the errors ϵ_n and the observations X_n are curves, and Ψ is a linear operator transforming a curve into another curve. The operator Ψ is defined as $\Psi(X)(t) = \int \psi(t, s)X(s)ds$, where $\psi(t, s)$ is a bivariate kernel assumed to satisfy $||\psi|| < 1$, where $||\Psi||^2 = \int \psi^2(t, s)dtds$. The condition $||\Psi|| < 1$ ensures the existence of a stationary causal solution to FAR(1) equations.

The FAR(1) data generating process series are thus generated according to model $X_{n+1}(t) = \int_0^1 \psi(t, s)X_n(s)ds + \epsilon_{n+1}(t)$, where n = 1, 2, ..., N. We used the following designs of a simulation study (Didericksen et al. 2012).

1. In experiment 1, we generated 100 curves using a Gaussian kernel $\psi(t, s) = C \exp\{-\frac{1}{2}(t^2 + s^2)\}$, and errors of type (8) from Didericksen et al. (2012) and
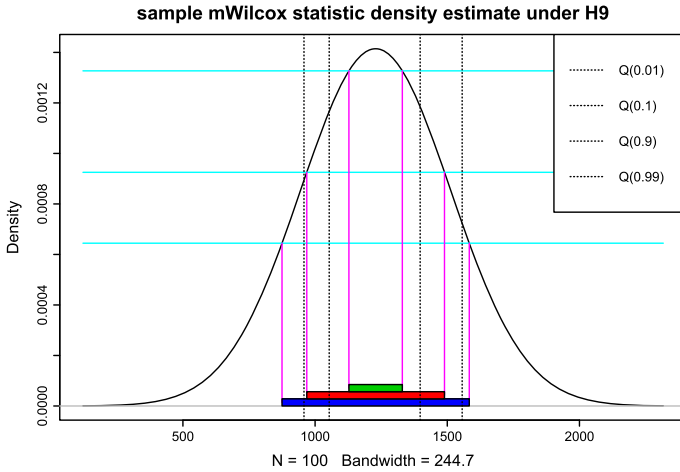
sample mWilcox statistic density estimate under H9

Fig. 5 Estimated density of the statistic (5) under the hypothesis that both samples are generated from the population related to Fig. 3

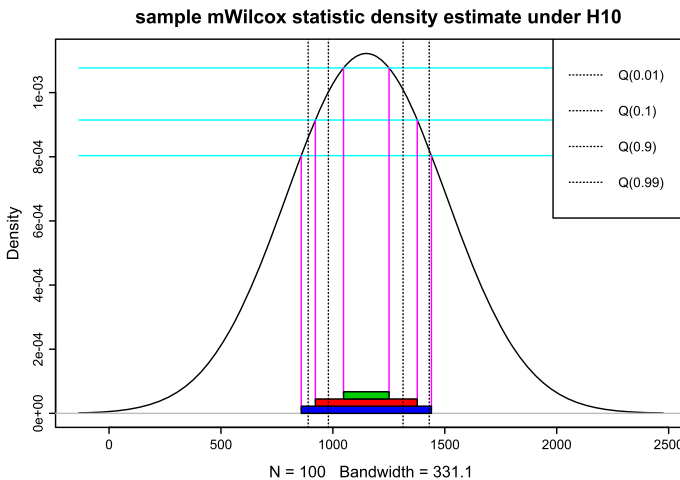**sample mWilcox statistic density estimate under H10**

Fig. 6 Estimated density of statistic (5) under an alternative in which the first sample is generated from population related to Fig. 3 and the second is generated from a population related to Fig. 4

then 100 curves using a kernel $\psi(t, s) = C$. Figure 7 presents an illustration for experiment 1.

We repeated the whole experiment 100 times. Figure 8 present the results of the simulations for the experiment 1 using functional boxplot and FM depth correspondingly.

2. In experiment 2, we generated 100 curves using a Gaussian kernel with appropriate constant C and 100 curves from a mixture of two processes considered in the experiment 1 but differing with respect to parameters of the error term (8) taken from Didericksen et al. (2012). Figure 9 presents an illustration for experiment 2,
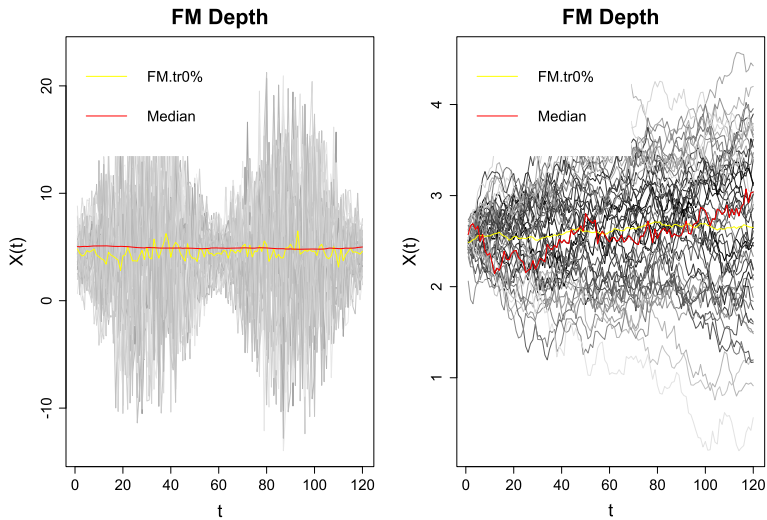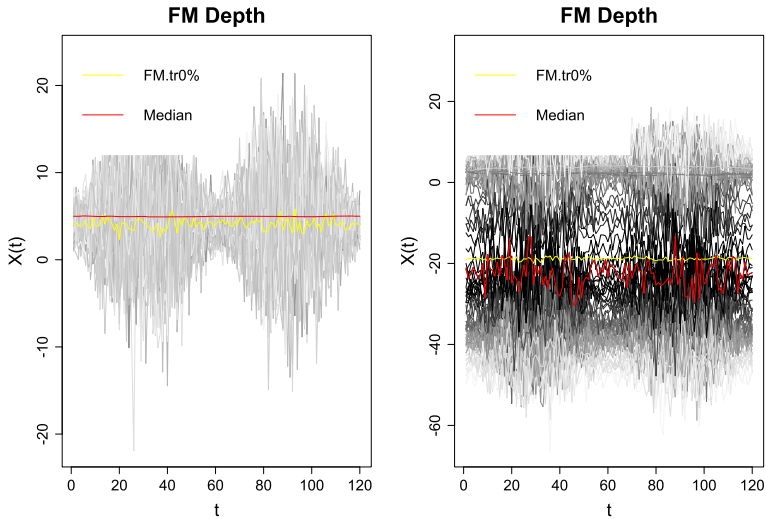
Fig. 7 Structural change detection in Scheme 1
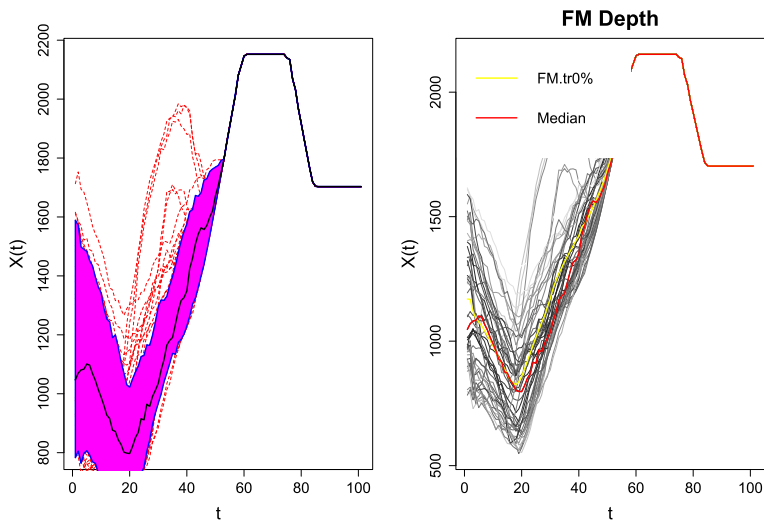
Fig. 8 100 detections of structural change in Scheme 1 using moving Wilcoxon statistic

and Fig. 10 presents the results of the simulations for experiment 2 using functional boxplot and FM depth correspondingly. Left panel of Fig. 11 shows sample density estimate (H_{d0}, see Hyndman et al. 2013) for a situation in which samples are generated from process presented on the left panel of Fig. 12, which is in turn a mixture of two processes. Right panel of Fig. 11 presents the sample density estimate (H_{d1}) where the first sample is generated from the mixture of processes presented on the left panel of Fig. 12, and the second sample is generated from the mixture of processes presented on the right panel of Fig. 12. The estimated

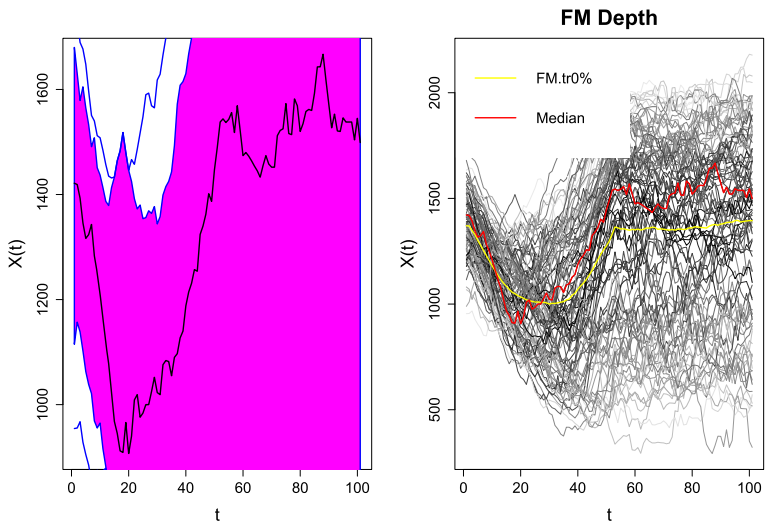**Fig. 9** Structural change detection in Scheme 2

Fig. 10 100 detections of structural change in Scheme 2 using moving Wilcoxon statistic

densities differ significantly with respect to location and hence our procedure correctly detects the change of type of mixture—this is a situation in which our procedure performs much better than the proposals of Flores et al. (2015), taking into account the deepest elements in both the samples.

It is easy to notice that our procedure correctly detects the structural change appearing after the 100th observation.

Figure 12 presents a very interesting example of structural change in which our proposal (1) outperforms the proposals based on statistics introduced in Flores et al.
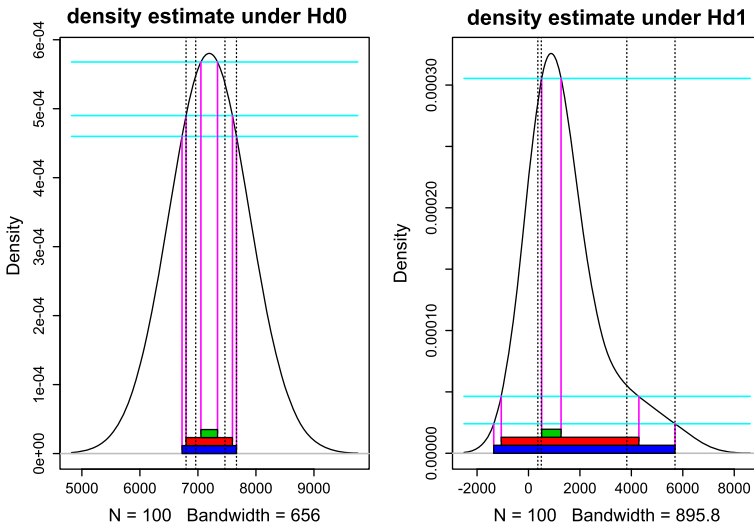
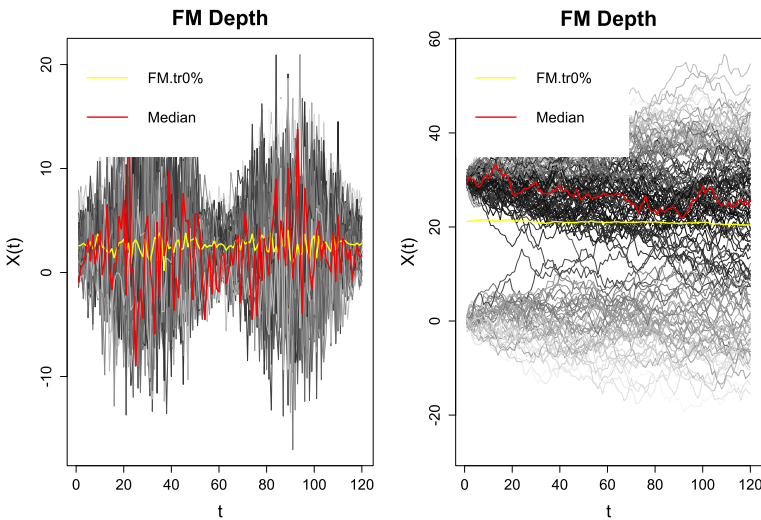**Fig. 11** Sample statistic density estimate under H_{d0} and under H_{d1}

Fig. 12 Two processes that are mixtures of two different processes

(2015). Structural change relates to change of type of the mixture of processes generating curves.

4.1 Properties of the proposal: empirical example 1

In order to verify the empirical usefulness of the proposal, we considered two Internet services with respect to number of users and numbers of page views based on the real data that were made available for us by the owners of the services. Figure 1 presents
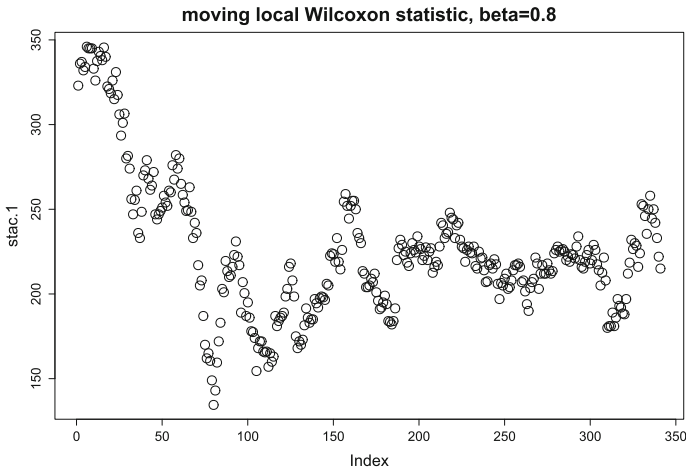
Fig. 13 Moving Wilcoxon statistic for numbers of users in service 1, $\beta = 0.8$
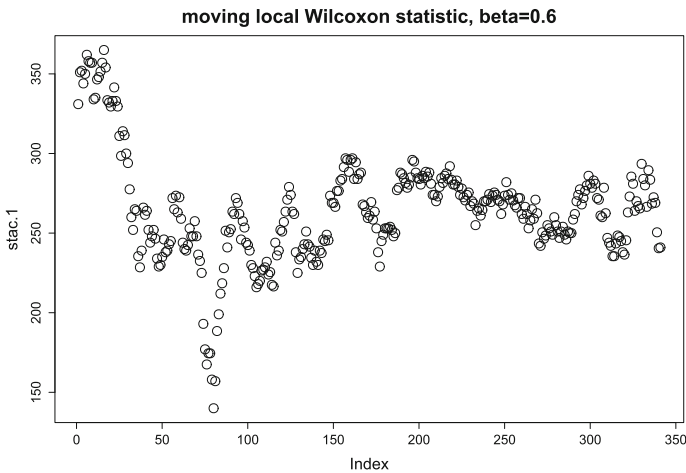

Fig. 14 Moving Wilcoxon statistic for numbers of users in service 1, $\beta = 0.6$

the functional boxplot for hourly numbers of users of the service 1 in 2013, Fig. 2 presents the functional boxplot for hourly numbers of users of the service 2 in 2013.

Figures 13, 14, 15 and 16 show the behavior of our proposal calculated from moving window for selected values of locality parameter β, and the reference sample consisted of the first 100 observations. One can notice a general tendency to stabilization of values of the statistic. The considered process seems to tend toward stationarity.

4.2 Empirical example 2: yield curves

Our second empirical example introduces FDA into modeling and predicting yield curves. Yield curves originate from the concept of risk-free interest rate, that is,
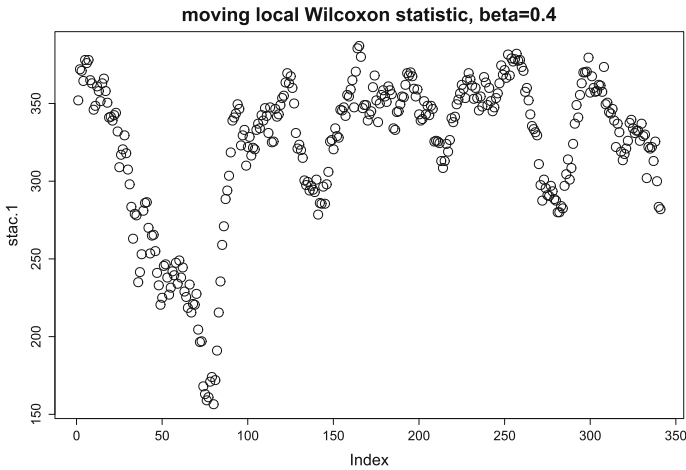
moving local Wilcoxon statistic, beta=0.4

Fig. 15 Moving Wilcoxon statistic for numbers of users in service 1, $\beta = 0.4$

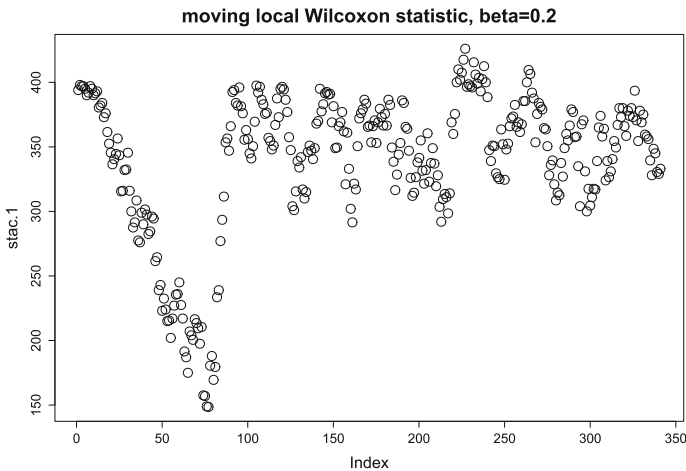**moving local Wilcoxon statistic, beta=0.2**

Fig. 16 Moving Wilcoxon statistic for numbers of users in service 1, $\beta = 0.2$

theoretical price which is paid for investments in safe assets. In practice, however, risk-less instruments do not exist, the risk-free rate is not directly observable and must be approximated by products traded on the market, like treasury bills, treasury and corporate bonds, inter-bank lending rates, forward rate agreements or swaps, and so on. From our point of view, yield curves are functions of time to maturity τ. A change in yield curve shape is considered to be the sign of change of expectations and the sign for change in real business cycle phase. Unfortunately, one cannot observe full functions' shape, since bonds and other interest rate derivatives have fixed dates of expiration. The detailed theory of shape of yield curve and factors affecting it are not fully developed. Estimation of yield curve is usually done in two ways: in
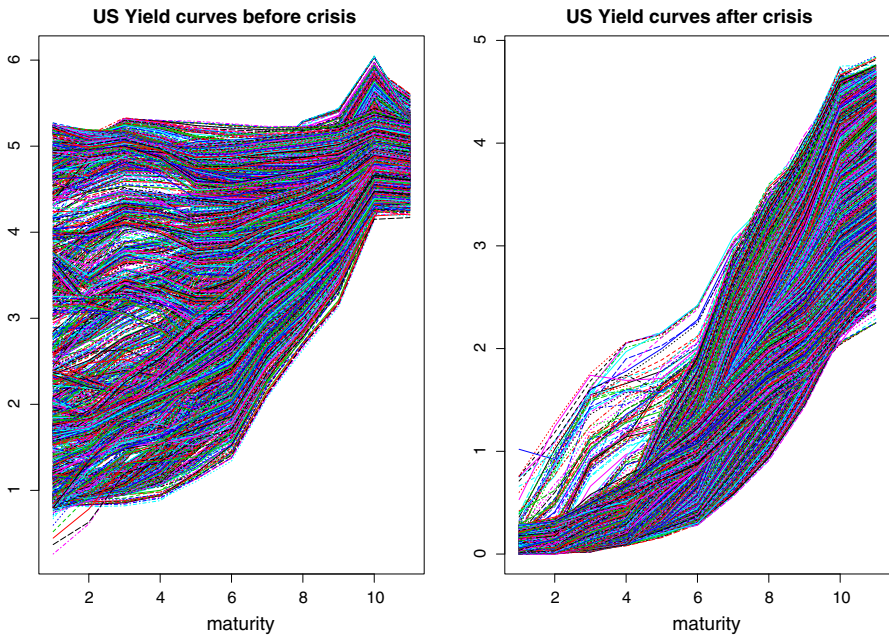
Fig. 17 Bundles of yield curves before and after crisis represented as functional object in Fourier basis

a non-parametric setting through linear or spline approximation and using bootstrap techniques or using parametric approach (Diebold and Li 2006).

Since US economy is a precursor of changes in global economy, we focus our attention on US yield curve. For our study, we use daily observed US yields from the period of January 3, 2000 to March 30, 2016 with maturities between 1 month and 30 years. In order to check whether our test is able to detect financial crisis, this subset is divided into two parts:

1. X—before Lehman Brothers bankruptcy (January 3, 2000 to September 14, 2008),
2. Y—since the beginning of sub-prime crisis in September 15, 2008 till present.

Both the subsets are converted into FDA objects and described in the Fourier basis (Fig. 17).

Figure 17 clearly shows that the resultant shape (i.e., slopes and curvatures) for both the bundles—before and after crisis—are different, as they should, because of the change of business cycle phase in global economy. The next step of the procedure requires the estimation of cGBD for both the subsets. The functional boxplot in Fig. 18 displays the median curve (the deepest location), along with the selected α central regions (see Hyndman and Shang 2010). Any point beyond the highest value of α may be considered as an outlier. As shown in Fig. 18, the central tendency of the shape of yield curve before and after crisis is the same, whereas the shape and nature of outliers differ significantly.

In order to check whether our procedure is able to detect changes in the structure of functional time series, the final step of this simulation involved calculation of local
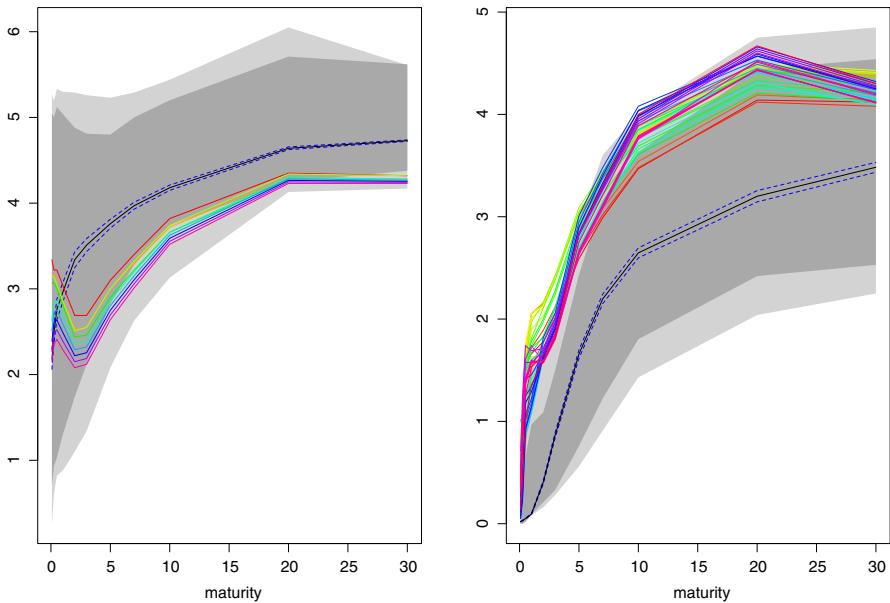
Fig. 18 Functional boxplot for both the subsets of yield curves

Wilcoxon statistics for functional yield curve in the rolling window scheme (proposal 2). In this scheme, we assume two windows of specified length—the reference window (or ref for short) has a fixed length and initially includes monthly functional yield curves data starting from January 2000 and the second window of the same size is shifted by a fixed number of observations. The shift size and sample length are kept fixed. Results for windows of length 10, 50, and 40 observations shifted by 10, 20, or 30 points in time are presented in Fig. 19. The obtained results clearly depends on the window size. For relatively large windows and partially overlapping samples, changes in the local Wilcoxon statistics are less volatile and can be related to the phases in business cycle, for example, for windows of length 40 and 50 observations one can observe the regime change at the end of 2007 (or the begin of 2008) and changes between the end of 2009 and the begin of 2010 and in 2014. As yield curve is a predictor of the phase of business cycle, one can be related relate this sharp peaks to two crisis waves—sub-prime crisis and euro-zone debt crisis. Final peak can easily be related to the period of time when US economy entered a growth phase again.

5 Sensitivity analysis

Classical one-dimensional Wilcoxon rank sum test effectively detects difference in location for family of logistic distributions. Multivariate tests induced by depths were proposed in Liu and Singh (1995) and Li and Liu (2004). Theoretical properties of multivariate Wilcoxon test (unbiasedness as well as its consistency were critically discussed in Jurečková and Kalina 2012). In our proposal, ranks are induced by outly-
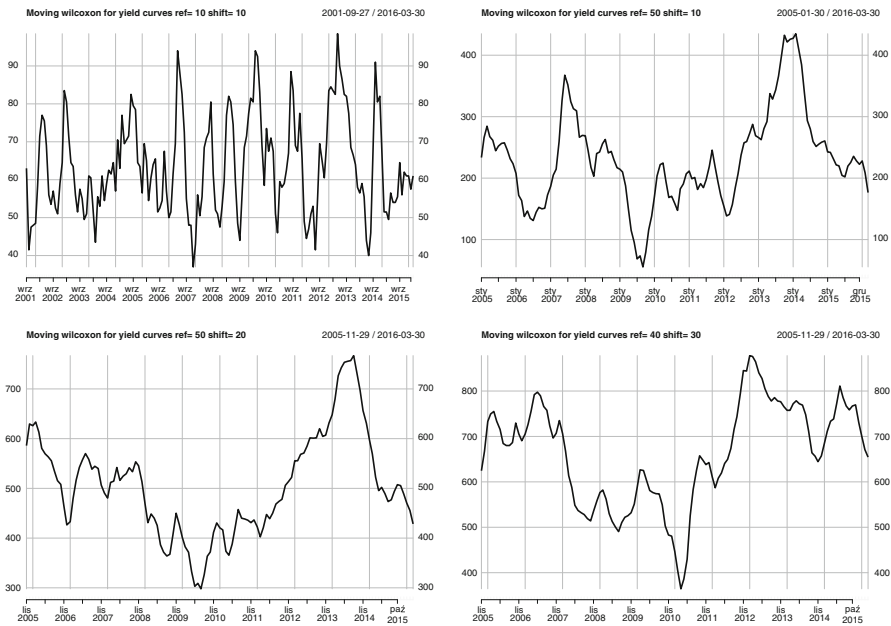
Fig. 19 Local Wilcoxon statistics for yield curves

ingness relative to the local centrality characteristic. Observations are ranked from the closest to the local median to the furthest to the local median. Relatively big or small values of the proposed statistic indicate differences in structure of outlyingness (considered on a locality level β) and should lead us to rejecting a hypothesis of equality of distributions.

In the functional case, significantly differing curves may have the same depth and hence the same rank. On the other hand, different empirical depths indicate differences in the underlying distributions, because under very mild conditions depths characterize multivariate distributions (and if distribution in a functional space as well it is still an opened question) (Kong and Zuo 2010). However, the simulation studies lead to a hypothesis that in a functional case we can expect a similar result: *the corrected generalized band depth characterizes a distribution in a functional space or at least effectively describes its merit important features*. Considering a reference sample and a moving window from a process, we can use our proposal for detecting not only a structural change but also a departure from stationarity (represented by the reference sample). Results of the simulations lead to a conclusion that our proposal is at least qualitative robust in the Hampel sense (Wilcox 2014). Small changes in null and alternative hypotheses do not significantly change the size and power of our proposal. We considered distance in the input space in terms of a median of all distances for pairs of functions, where one function belong to assumed model and the second to a model representing a departure from assumptions. In the output space, we considered euclidean distance between the values of our test statistic. In these terms, small changes of input data lead to small changes of a decision process based on monitoring stationarity of the

functional time series (Hall et al. 2003). Understanding various possible outlyingness in functional time series setting is worth to notice. It is possible to consider the outliers in space of functions or outlyingness related to vertical point-wise contamination. Contamination may affect the reference sample or the working window. We considered functional outliers with respect to functional boxplot induced by the corrected generalized band depth. Notice that our proposal is robust but not very robust (it copes with up to 10% of contamination). It is robust to a moderate fraction of outliers or inliers (they lead to small change of ranking induced by depth) but sensitive to a time series regime change. Therefore, the procedure may be used for monitoring the data streams (Kosiorowski 2016). In our opinion, alternative procedures for monitoring a homogeneity in functional time series are less robust to functional outliers than our proposal.

We can evaluate the "size" and the "power" of our procedure in a similar manner as in Li and Liu (2004) and Jureĉková and Kalina (2012). A central issue in the analysis of functional data is to take into account the temporal dependencies between the functional observations. Because of this temporal dependence, even the most elementary statistics became inaccurate. In this context, resampling methodology, especially bootstrapping, proves to be the only alternative. In order to obtain bootstrap p-values for our test, we propose to use a maximum entropy methodology proposed by Vinod and de Lacalle (2009) and used among others by Shang (2016). The meboot R package together with DepthProc R package gives the appropriate computational support.

6 Summary

The proposed procedure based on moving local Wilcoxon statistic may effectively be used for detecting heterogenity in functional time series. Simulation studies indicate that properties of our proposal depends on the Kolmogorov distance between functional medians in the distributions generating samples, one representing null hypothesis of stationarity and the second alternative representing a fixed departure from the stationarity. The locality parameter β may be interpreted as a resolution or a sensitivity to details (e.g., local asymmetry) at which we monitor a process.

Merit properties of the proposed procedure strongly depend on the functional depth being used (on which conditions we choose a center in a sample of functions (Sguera et al. 2016; Nagy et al. 2016). The conducted simulation studies as well as the studied empirical examples show a big potential of our proposal in the context of discrimination between the alternatives and in a consequence in detecting a structural change. Implementations of the local Wilcoxon test and our proposal may be found in *DepthProc* R package, which is available through CRAN servers. Note that for detecting special kinds of nonstationarity, it is possible to replace the local Wilcoxon statistic by means of local Kamat or Haga statistics (or other rank statistic). Further theoretical properties of our proposal are still under our consideration and constitute a part of our future work.

References

Bosq D (2000) Linear processes in function spaces. Springer, New York

Cuesta-Albertos J, Nieto-Reyes A (2008) The random Tukey depth. Comput Stat Data Anal 52:4979–4988

Cuevas A, Febrero-Bande M, Fraiman R (2007) Robust estimation and classification for functional data via projection-based depth notions. Comput Stat 22(3):481–496

Diebold F, Li C (2006) Forecasting the term structure of government bond yields. J Econom 130(2):337–364

Didericksen D, Kokoszka P, Zhang X (2012) Empirical properties of forecasts with the functional autoregressive model. Comput Stat 27(2):285–298

Febrero-Bande M, de la Fuente M (2012) Statistical computing in functional data analysis: the R package fda.usc. J Stat Softw 51(4):1–28

Flores R, Lillo R, Romo J (2015) Homogenity test for functional data. arXiv:1507.01835v1

Fraiman R, Muniz G (2001) Trimmed means for functional data. Test 10(2):419–440

Fraiman R, Justel A, Liu R, Llop P (2014) Detecting trends in time series of functional data: a study of Antarctic climate change. Can J Stat 42(4):597–609

Hájek J, Sidák Z (1967) Theory of rank tests. Academic Press, New York

Hall P, Rodney CL, Yao Q (2003) Comprehensive definitions of breakdown points for independent and dependent observations. J R Stat Soc B 65:81–84

Horváth L, Kokoszka P (2012) Inference for functional data with applications. Springer, New York

Horváth L, Kokoszka P, Rice G (2014) Testing stationarity of functional time series. J Econom 179:66–82

Hyndman RJ, Shang HL (2010) Rainbow plots, bagplots, and boxplots for functional data. J Comput Graph Stat 19(1):29–45

Hyndman R, Einbeck J, Wand M (2013) The R package hdrcde

Jureĉková J, Kalina J (2012) Nonparametric multivariate rank tests and their unbiasedness. Bernoulli 18(1):229–251

Kong L, Zuo Y (2010) Smooth depth contours characterize the underlying distribution. J Multivar Anal 101:2222–2226

Kosiorowski D (2016) Dilemmas of robust analysis of economic data streams. J Math Sci 1(2):59–72

Kosiorowski D, Zawadzki Z (2014) DepthProc: an R package for robust exploration of multidimensional economic phenomena. http://arxiv.org/pdf/1408.4542. Accessed 5 April 2016

Kosiorowska E, Kosiorowski D, Zawadzki Z (2014) Evaluation of the fourth millenium development goal realisation using multivariate nonparametric depth tools offered by DepthProc R package. Folia Oecon Stetin 15(1):34–52

Li J, Liu R (2004) New nonparametric tests of multivariate locations and scales using data depth. Stat Sci 19(4):686–696

Liu R (1990) On a notion of data depth based on random simplices. Ann Stat 18:405–414

Liu R, Singh K (1995) A quality index based on data depth and multivariate rank tests. J Am Stat Assoc 88:252–260

Liu R, Parelius JM, Singh K (1999) Multivariate analysis by data depth: descriptive statistics, graphics and inference (with discussion). Ann Stat 27:783–858

López-Pintado S, Jörnsten R (2007) Functional analysis via extensions of the band depth. IMS lecture notes–monograph series complex datasets and inverse problems: tomography, networks and beyond, vol 54. Institute of Mathematical Statistics, Hayward, pp 103–120

López-Pintado S, Romo J (2007) Depth-based inference for functional data. Comput Stat Data Anal 51:4957–4968

Mosler K (2013) Depth statistics. Robustness and complex data structures. Springer, Heidelberg, pp 17–34

Lange T, Mosler K, Mozharovsky P (2015) Fast nonparametric classification based on data depth. Stat Pap 55(1):49–69

Nagy S, Hlubinka D, Gijbels I (2016) Integrated depth for functional data: statistical properties and consistency. ESIAM Probab Stat 20:95–130

Nieto-Reyes A, Battey H (2016) A topologically valid definition of depth for functional data. Stat Sci 31(1):61–79

Paindaveine D, Van Bever G (2013) From depth to local depth: a focus on centrality. J Am Stat Assoc 108(503):1105–1119

Ramsay J, Silverman B (2005) Functional data analysis. Springer, New York

Ramsay J, Hooker G, Graves S (2009) Functional data analysis with R and Matlab. Springer, New York

Serfling R (2006) Depth functions in nonparametric multivariate inference. In: Liu R, Serfling R, Souvaine D (eds) Series in discrete mathematics and theoretical computer science, vol 72. AMS, Providence, pp 1–15

Sguera C, Galeano P, Lillo RE (2016) Global and local functional depths. arXiv: 1607.05042v1

Shang HL (2016) Bootstrap methods for stationary functional time series. Stat Comput (to appear)

Vinod HD, de Lacalle JL (2009) Maximum entropy bootstrap for time series: the meboot R package. J Stat Softw 29:5

Wilcox R (2014) Introduction to robust estimation and hypothesis testing. Academic Press, San Diego

Zuo Y, Serfling R (2000) Structural properties and convergence results for contours of sample statistical depth functions. Ann Stat 28:483–499