Daniel Kosiorowski iD

Cracow University of Economics, Department of Statistics, daniel.kosiorowski@uek.krakow.pl

Dominik Mielczarek iD

AGH University of Science and Technology, Faculty of Applied Mathematics
Department of Mathematical Analysis, Computational Mathematics and Probability Methods
dmielcza@wms.mat.agh.edu.pl

Jerzy Piotr Rydlewski iD

AGH University of Science and Technology, Faculty of Applied Mathematics
Department of Financial Mathematics, ry@agh.edu.pl

A Critical Study of Usefulness of Selected Functional Classifiers in Economics

Abstract: In this paper we conduct a critical analysis of the most popular functional classifiers. Moreover, we propose a new classifier for functional data. Some robustness properties of the functional classifiers are discussed as well. We can use an approach worked out in this paper to predict the expected state of the economy from aggregated Consumer Confidence Index (CCI, measuring consumers optimism) and Industrial Price Index (IPI, reflecting a degree of optimism in industry sector) exploiting not only scalar values of the indices but also the trajectories/shapes of functions describing the indices. Thus our considerations may be helpful in constructing a better economic barometer. As far as we know, this is the first comparison of functional classifiers with respect to a criterion of their usefulness in economic applications. The main result of the paper is a presentation of how a small fraction of outliers in a training sample, which are linearly independent from the training sample, consisting of almost linearly dependent functions, corrupt all analysed classifiers.

Keywords: functional classifier, functional data analysis, robust methods, economic optimism barometer

JEL: C38, C63, C32

1. Introduction

Our perception of an economic phenomenon often relates to an evaluation of properties of a function of a certain continuum. One may consider probability density function of random variable describing an income of a household, one may consider GDP per capita trajectory of a country during a decade, day and night number of visits of an Internet user in an Internet service or a behaviour of an investor's optimism indicator within a month. Reducing the whole function to a certain set of scalars (e.g., mean, variance) very often denotes a significant loss of valuable information on the phenomenon and in a consequence may lead to inappropriate perception of the phenomenon. A "shape" of the consumer price index (CPI) during a month may better express investor optimism during the considered period, as a specific sequence of "peaks" and "valleys" in a CPI trajectory and may denote sequence of activity bursts and consumer hesitations, and hence "a spectrum of moods" called optimism.

In the recent decades a very useful statistical methodology has been proposed in this context and is now being intensively developed. The methodology named functional data analysis (FDA) enables functional generalizations of the well-known uni- and multivariate statistical techniques like analysis of variance, kernel regression or classification techniques (see Ramsay, Silverman, 2005; Ferraty, Vieu, 2006; Ramsay, Hooker, Graves, 2009; Horváth, Kokoszka, 2012; Kosiorowski, Rydlewski, Snarska, 2019).

The FDA offers novel methods for decomposition of income densities or yield curves, analyzing huge, sparse economic data sets. The FDA enables effective statistical analysis when number of variables exceeds number of observations. FDA enables effective analysis of economic data streams, e.g., analysis of non-equally spaced observed time series, prediction of a whole future trajectory rather than single observations (Kosiorowski, 2016).

There are many important economic issues, which may be translated into language of statistical classification analysis. Economic agents choose their investment, cooperation or production strategies taking into account an actual situation and knowledge of the issue preserved in historical data. In a credit scoring, one may classify a client as potentially credible or not. An evaluation of a candidate for a certain position with regard to a category of her usefulness or a diagnosis of a team as to its collaboration performance or a company as to its bankruptcy closeness are direct and popular examples in this context.

Focusing our attention on certain more recent economic phenomena, one may indicate, for example, a problem of choosing a time dependent strategy for an investment, e.g., "bid/ask" trajectory in an algorithmic trading, a "real time" choosing contents of SMS alerts in a process of air quality monitoring in a city or choosing a type of administrator answer in a process of Internet service monitoring for possible intrusions. More precisely: having at a disposal a so called *training sample*

$$Z_1 = (X_1, Y_1), \ldots, Z_n = (X_n, Y_n),$$

where X_i denotes a functional observation and Y_i denotes its label, our aim is to predict the label for a new observation basing on functional observation.

In other words a classification rule (a classifier) is a function

$$d : X \rightarrow Y,$$

which assigns to a new functional observation X a prognosis of its label $d(X)$. The main aim of the classification analysis is to find a precise classifier in a certain sense (see Steinwart, Christmann, 2008).

The real classification error is defined as

$$L(d) = P(\{d(x) \neq Y\}).$$

For known joint distribution of (X, Y) the best classification rule is called the Bayes classifier (see Devroye, Giörfe, Lugosi, 1996). The Bayes classifier is a reference classifier for other classifiers, which at least partly are estimated from the training sample. Classifiers' performance generally depends on the underlying distribution. There are exceptions however (see Devroye, Giörfe, Lugosi, 1996). In fact, we seldom know joint distribution of (X, Y), so the Bayes classifier cannot be directly used to obtain the optimal classifier. In practice, the information provided by the training sample is used to construct classifier, whose conditional error is as close as possible to the Bayes error (see Vencálek, Pokotylo, 2018).

Although there is no agreement on how to understand robustness of a classification rule, we may apply a general idea of robustness stating that small changes of an input of a statistical procedure should lead to small changes in an output of the procedure (see Cuevas, 1988; Christmann, Salibian-Barrera, Van Aelst, 2013). By the output we can understand certain loss function related to the classification procedure or a quality measure of the procedure in the real classification error style, for example, an empirical risk of the classifier.

Robust classification rule denotes a rule, which focuses on an influential majority of data and which copes with certain amount of problems with data. In multivariate case the concept of robustness in a context of classifiers was studied, among others, by Hubert, Rousseeuw, Segaert (2016), and Christmann, Salibian-Barrera, Van Aelst (2013). Hubert, Van Driessen (2004) considered an overall robustness of a classifier in terms of breakdown point for the worst class performance. Their proposals rely on "robustifying" classical approaches using for example M-estimators or trimming.

In this paper we focus our attention on an issue of robust classification of functional objects and its effective applications in current macro-economic issues.

The performance of a country's economy strongly depends on expectations as to its future behaviour. These expectations are very often operationalized in a form of various ratings cyclically published by leading banking or consulting groups. On a technical level one may express a rating as certain function of classifiers. A better FDA classifier enables a better forecasting of the state of the economy, for example, aggregated CCI (measuring consumers optimism) and IPI (measuring industry optimism), exploiting not only scalar values of the indices but also the trajectories/shapes of functions describing the indices, allow for a construction of a better economic barometer or rating. The classifiers comparison should take into account the problem of outlying observations, wrong labelling and missing data problem. That is why the classifiers' robustness should be compared. In our opinion, a comparison based on the misclassification rate and computational complexity has strong justification in the area of modern e-economy and empirical finance (Kosiorowski, Mielczarek, Rydlewski, 2017; 2018).

2. Review of functional classifiers

In the recent years several algorithms for classification of functional data have been proposed. Generally speaking, the proposed classifiers are not uniformly robust i.e., their performance may strongly depend on a very small fraction of especially "bad" outlying (in a functional sense) observations. It should be stressed that commonly acceptable definition of robustness of a classification procedure does not currently exist. We suppose that robustness in this case should take into account a local nature of classification procedure – perhaps robustness should be defined with respect to specified class rather than regarding the whole data set.

Classification methods for functional data include:
1) *k*-nearest neighbours (kNN) methods (for example see Cuevas, Febrero-Bande, Fraiman, 2007);
2) reproducing kernel of a Hilbert space (RKHS) methods (for example see Schölkopf, Smola, 2002; Berlinet, Thomas-Agnan, 2004; Preda, 2007);
3) methods based on depth measures (for example see Cuevas, Fraiman, 2009);
4) methods based on depth-depth plot (see Li, Cuesta-Albertos, Liu, 2012);
5) neural networks and other methods (see Haykin, 2009).

1. In the *k*-nearest neighbours methods we fix $k \in N$ and a dissimilarity measure. The classified function is then assigned to a class, which is most common among its k nearest neighbours. Note, that different dissimilarity measures give different neighbourhoods. The choice of the number k and the dissimilarity measure defining neighbourhood is still an open problem (Ferraty, Vieu, 2006). Some variant of the method is the nearest centroid method, where the functional observation is assigned the label of the class of training samples whose centroid is clos-

est to the considered observation. Centroid is a functional mean, or a functional median induced by functional depth. Some modifications of the *k*-nearest neighbours methods were proposed (for example see Vencálek, 2013).

2. For a second family of methods, let *X* be a nonempty set and let *H* be a Hilbert space of functions $f: X \to R$ equipped with an inner product $< , >$. The space *H* is called reproducing kernel Hilbert space (RKHS), if there exists a non-negative and symmetric function $\textbf{\textit{K}}: X \times X \to R$, which possess the following properties:

$$\langle f, \textbf{\textit{K}}(X, \cdot) \rangle = f(X) \text{ for any } f \in H \tag{1}$$

(specifically, a formula $\langle \textbf{\textit{K}}(X, \cdot), \textbf{\textit{K}}(Z, \cdot) \rangle = \textbf{\textit{K}}(X, Z)$ holds true)

$$H = \overline{\text{span}\left\{ \textbf{\textit{K}}(Z, \cdot) : Z \in X \right\}} \tag{2}$$

Exploiting Mercer's Theorem, a mapping $\Phi: X \to R$ can be constructed such, that $<\Phi(Z), \Phi(Y)> = \textbf{\textit{K}}(Z, Y)$ for almost all $Y, Z, \in X$, where learning sample is $(X, Y) = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, where X_i is a function, and Y_i is a label. Then the following formulas are calculated

$$\textbf{\textit{K}}(X_i, X_j) = < \Phi(X_i), \Phi(X_j) >, \tag{3}$$

$$c(X) = \text{sgn}\left(\sum\nolimits_{k=1}^{n} c_k \textbf{\textit{K}}(X_k, X) \right), \tag{4}$$

where coefficients c_k are chosen so that congruency condition holds true, namely $c(X_i) = Y_i$.

The coefficients c_k can be chosen if the matrix of elements $\textbf{\textit{K}}(X_i, X_j)$ is nonsingular (invertible). So it suffices that the functional data are linearly independent. The formula (2) enables conducting of a classification. Note, however, that most packages, i.e. *fda.usc* do not explain how to deal with the problem of linearly dependent functional data. It is an important problem, because the coefficients in the sum may not be unique in such a case.

It is worth noticing, that Górecki, Krzyśko, Wołyński (2018) constructed an independence measure and independence test between kernels related to multivariate functional data, which also may be incorporated into a construction of a new barometer of economic optimism.

Note that in practice, at the beginning, a kernel is chosen. A feature space *H* is then constructed so that the chosen kernel produces an inner product in that space. Observations are transformed into a Hilbert space. It turns out that, if some conditions are fulfilled, it suffices to know the inner product only.

We consider a space of all functions mapping a space $L^2(\Omega)$ into *R*, which is denoted with *X*, namely

$$X = R^{L^2(\Omega)} \tag{5}$$

and any element of that space can be identified with a classifier.

Let $\mathbf{K} : L^2(\Omega) \times L^2(\Omega) \to R$ be any kernel, i.e.:

$$\mathbf{K}(f,g) = \mathbf{K}(g,f), \tag{6}$$

$$\mathbf{K}(f,f) \geq 0, \tag{7}$$

$$\mathbf{K}(f,f) = 0 \Leftrightarrow f = 0 \tag{8}$$

for any functions $f, g \in L^2(\Omega)$, and H be a reproducing kernel Hilbert space. From Mercer's Theorem there exists a mapping $\Phi : L^2(\Omega) \to H$, such that

$$< \Phi(f), \Phi(g) >_H = \mathbf{K}(f,g) \tag{9}$$

for any functions $f, g \in L^2(\Omega)$. In Hilbert space H, any function $f \in X$ has the following form

$$f(X) = < f, \mathbf{K}(X, \cdot) >_H, \tag{10}$$

where X is an element of $L^2(\Omega)$ space.

Specifically, there exists a classifier fulfilling a congruency condition $f(X_i) = Y_i$.

Note that the reproducing kernel imposes a distinctive form of the classifier, i.e.

$$f(\cdot) = \sum_{x \in X} \alpha_x \mathbf{K}(Z_x, \cdot), \tag{11}$$

where the family is summable with respect to the norm induced by the inner product in the Hilbert space H. The above formula is difficult to implement numerically as family of sets X is uncountable. If the training sample is linearly dependent, then the above sum cannot be reduced to a finite sum. In other words, if the training sample is linearly dependent, it is not clear how to approximate the sum in formula (9). Moreover, *fda.usc* package description does not explain, how to cut off the rest of the infinite sum in the formula (9). If the training sample is linearly independent, the determinant of the matrix $\mathbf{K}(X_i, X_j)_{i,j=1,\ldots,n}$ is nonzero and the following formula holds true

$$f(X) = \sum_{i=1}^{n} \alpha_i \mathbf{K}(Z_i, \cdot). \tag{12}$$

3. Cuevas, Febrero-Bande, Fraiman (2007) considered the random projection depth. It measures the depth of the functional data under projections and takes additional information of their derivatives. Each function and its first derivative are projected along a random direction. Thus a point in R^2 is defined and a two-dimensional depth enables ordering of the projected points. Cuevas, Febrero-Bande, Fraiman (2007) showed that if we use a lot of random projections, the average of the depths of the projected two-dimensional points defines the depth for functional data. Our computations, conducted by means of *fda.usc* R package are based on this approach, where a Fraiman-Muniz depth is considered.

4. The DD-plot classifier was proposed by Li, Cuesta-Albertos, Liu (2012). First, it transforms the data into depth versus depth space (DD-space). Next, the data points are separated by suitable curve from a given family of functions, so that the number of errors when classifying points from the training sample is minimized. The authors showed that their DD-classifier is asymptotically equivalent to the Bayes rule under some conditions. DD-classifier can be extended to multiclass problem by using majority vote method, i.e., DD-classifier is applied to each of the possible pairs from all the considered classes and then the majority vote method determines the final memberships of the functional observations. Other methods based on the concept of DD-plot can be proposed as well (i.e., see Kosiorowski, Mielczarek, Rydlewski, 2017).

3. Our proposal

This section describes numerically stable and effective algorithm of affine classifier for functional data, basing on properties of Gram-Schmidt matrix. Outline of our method for two-class classifier has been recently presented in Kosiorowski, Mielczarek, Rydlewski (2018). Let us come to our proposition's full description.

Let $X_1, X_2, …, X_m$ be any functional data from Hilbert space $L^2(\Omega)$ and numbers $Y_1, Y_2, …, Y_m$ be labels i.e. $Y_i \in \{-1,1\}, i \in \{1,…,m\}$. Patterns X_i are functions mapping set Ω into real numbers and the following inequality is true

$$\int_{\Omega} |X_i(\omega)|^2 \, d\omega < \infty \tag{13}$$

for any $i \in \{1, 2, …, m\}$.

We assume in the whole paper, that set Ω is bounded, and then the space $L^2(\Omega)$ is separable. Hence, there exists an orthonormal basis $\{Z_n\}_{n=1}^{\infty}$ and every function X from the space $L^2(\Omega)$ can be described as the following series

$$X = \sum_{n=1}^{\infty} <X, Z_n> \cdot Z_n , \tag{14}$$

where the series convergence is a convergence in the sense of norm of the space $L^2(\Omega)$. In practice, we fix a natural number K and we determine factors $\{c_n\}_{n=1}^{K}$ such, that

$$\hat{X} = \sum_{n=1}^{K} c_n \cdot Z_n, \tag{15}$$

so that they minimize a function $\phi : R^K \to R$ given by the following formula

$$\phi(\boldsymbol{c}) = (\boldsymbol{X} - \phi \cdot \boldsymbol{c})^T \cdot (\boldsymbol{X} - \phi \cdot \boldsymbol{c}), \tag{16}$$

where $\boldsymbol{X}^T = (X(t_1), \ldots, X(t_M))$ and ϕ is a matrix of the form $\left[Z_j(t_i) \right]_{i=1,\ldots,M}^{j=1,\ldots,K}$.

We propose a classifier for functional data in the form

$$f(X) = \int_{\Omega} X(\omega) W(\omega) d\omega + b, \tag{17}$$

where b is any real number and weight function W is essentially bounded, i.e. $W \in L_{\infty}(\Omega)$, and chosen so that affine functional f be data-consistent (congruent), i.e.

$$Y_i f(X_i) = 1, \tag{18}$$

for any $i \in \{1, 2, \ldots, m\}$.

In other words, we are given empirical data $(X_1, Y_1), (X_2, Y_2), \ldots, (X_m, Y_m)$,. Basing on the data, we classify a new functional observation X into one of the groups looking only at $sgn(f(X))$. The classifier doesn't work, if $f(X) = 0$.

Existence of the weight function W, as we show in the paper, is guaranteed with linear independence of the random functions $X_1, X_2, \ldots, X_m$.

We show, that assuming linear independence, an operator $\boldsymbol{A} : L^2(\Omega) \to R^m$ given for any function $W \in L^2(\Omega)$ with formula

$$\boldsymbol{A}(W) = \left(\int_{\Omega} X_1(\omega) W(\omega) d\omega, \int_{\Omega} X_2(\omega) W(\omega) d\omega, \ldots, \int_{\Omega} X_m(\omega) W(\omega) d\omega \right) \tag{19}$$

is a surjection. Particularly, there exists a weight function W such, that

$$\boldsymbol{A}(W) + \overbrace{(b, b, \ldots, b)}^{m} = (Y_1, Y_2, \ldots, Y_m), \tag{20}$$

so for any $i \in \{1, \ldots, m\}$ we have

$$f(X_i) = \int_{\Omega} X_i(\omega) W(\omega) d\omega + b = Y_i. \tag{21}$$

Hence

$$Y_i f(X_i) = Y_i \cdot Y_i = 1, \tag{22}$$

where $\cdot$ is a standard multiplication.

For any subset $\{i_1, i_2, \ldots, i_{m-1}\} \in \{1, 2, \ldots, m\}$ let

$$V_{i_1, i_2, \ldots, i_{m-1}} = span\{X_{i_1}, X_{i_2}, \ldots, X_{i_{m-1}}\} \tag{23}$$

we then get $X_{i_m} \notin V_{i_1, i_2, \ldots, i_{m-1}}$, where $i_m \in \{1, 2, \ldots, m\} \setminus \{i_1, i_2, \ldots, i_{m-1}\}$.

Hahn-Banach Theorem states, that there exists a bounded functional $g_{i_1, i_2, \ldots, i_{m-1}} : L^2(\Omega) \to R$ such, that

$$g_{i_1, i_2, \ldots, i_{m-1}}(X_{i_j}) = 0, \tag{24}$$

for $j \in \{1, \ldots, m-1\}$ and

$$g_{i_1, i_2, \ldots, i_{m-1}}(X_{i_m}) = 1. \tag{25}$$

The functional $g_{i_1, i_2, \ldots, i_{m-1}}$ must be of the form

$$g_{i_1, i_2, \ldots, i_{m-1}}(X) = \int_\Omega X(\omega) W_{i_1, i_2, \ldots, i_{m-1}}(\omega) d\omega, \tag{26}$$

where $W_{i_1, i_2, \ldots, i_{m-1}}$ is a function from the $L^2(\Omega)$ space.

Particularly

$$\mathbb{R}^m = span\{A(W_{i_1, i_2, \ldots, i_{m-1}}) : \{i_1, i_2, \ldots, i_{m-1}\} \subset \{1, 2, \ldots, m\}\}. \tag{27}$$

We have just shown the following

Theorem:

For any real number b there exists a function $W \in L^2(\Omega)$ such, that a set of equalities

$$Y_i\left(\int_\Omega X_i(\omega) W(\omega) d\omega + b\right) = 1, \tag{28}$$

for $i \in \{1, 2, \ldots, m\}$ has a solution. Particularly, the weight function W satisfies the following set of equalities

$$\int_\Omega X_i(\omega) W(\omega) d\omega + b = Y_i, \, for \, i \in \{1, \ldots, m\}. \tag{29}$$

It is now obvious, that in order to solve (15) it suffices to solve the following set of equations, that is, it suffices to determine a weight function W, or equivalently to find a functional g such, that

$$g(X_i) = \int_{\Omega} X_i(\omega) W(\omega) d\omega = Y_i(1 - Y_i b), \tag{30}$$

for $i \in \{1, 2, \ldots, m\}$.

When we determine the functional g, then we obtain

$$Y_i g(X_i) = Y_i \cdot Y_i (1 - Y_i b) = 1 - Y_i \cdot b. \tag{31}$$

In consequence

$$Y_i g(X_i) = Y_i \cdot Y_i (1 - Y_i b) = 1 - Y_i \cdot b. \tag{32}$$

for $i \in \{1, 2, \ldots, m\}$.

Hence, the hyperplane separating for functional data can be determined.

Our proposition of classifier can be generalized to the multiclass case. Let $X_1, X_2, \ldots, X_m$ be any functional data from the Hilbert space $L^2(\Omega)$, and let $Y_1, Y_2, \ldots, Y_m$ be a sequence of k-element classes, i.e., $Y_i \in \{1, \ldots, k\}, i \in \{1, \ldots, m\}$, where $k \ll m$.

First, all possible two class classifications are performed. Subsequently, the majority vote method is applied in order to obtain a final classification. However, it is computationally very demanding, so it is available for smaller training sets. For larger data sets we recommend the following procedure. The training data $\{(X_i(t), Y_i)\}_{i=1}^M$ is divided into two classes. Subsequently, the separating hyperplane is determined. We repeat the process of the training data division into two classes until all classes are separated. The training data division order is established empirically. We recommend to make outliergrams (see Arribas-Gil, Romo, 2013) or functional boxplots (see Kosiorowski, Rydlewski, Zawadzki, 2018; Kosiorowski, Zawadzki, 2019, and references therein) in order to divide the similar classes of functions at the most distant step of the training sample division procedure.

4. Robustness of a classification rule for functional data

Generally speaking, by a robust statistical procedure we mean a procedure which correctly expresses a tendency represented by an influential majority of probability mass, or a fraction of data (Hubert, Rousseeuw, Segaert, 2016). In the context

of a classifier, we usually consider its robustness with respect to a contamination of a training sample. We evaluate it in terms of an error of classification. It is worth underlining, that in general, robustness of the procedure depends on an underlying model of the training sample. Robustness issues in functional setup are especially difficult and in a great part are still open. Let us only consider, that in the functional setup there exist various types of outlyingness that are not present elsewhere. One may indicate shape outliers, amplitude outliers and outliers with respect to the co-variance structure. For assessing the robustness of a procedure, one can propose a useful variant of qualitative robustness (see Cuevas, 1988; Christmann, Salibi-an-Barrera, Van Aelst, 2013): small changes of input should lead to small changes of output or a measure of quality of output.

The robustness of the classifying rule toward outliers depends on the functional outliers' type. It should be different for the functional shape outliers, functional amplitude outliers and for functional outliers with respect to the co-variance structure.

That is why it is not easy to approximate breakdown point or influence function of the procedure. It should be stressed, that there is no agreement as to the breakdown point or influence function concepts even in the multivariate classification case, however some important results on influence functions were obtained by Christmann, Van Messem (2008) (see also Steinwart, Christmann, 2008). Some attempts to tackle the robustness issue in functional classification case have been made (for example, see Hubert, Rousseeuw, Segaert, 2016). We follow the qualitative robustness concept and adapt it to the functional classification case.

Definition 1 (Cuevas, 1988): We say that the sequence of functionals is qualitatively robust at $P \in P$, if for any $\varepsilon > 0$, there exists a $\delta > 0$ and a positive integer n_0 such that, for all $Q \in P$ and $n > n_0$

$$d_P(P,Q) < \delta \Rightarrow d_T\left(L_P(T_n), L_Q(T_n)\right) < \varepsilon, \tag{33}$$

where P, Q denote two mixtures of distributions in L^2 Hilbert space of functions and L_P, L_Q denote estimated characteristics of P, Q (i.e., e.g. their functional medians).

In a sample case we replace P, Q by means of empirical measures P_n, Q_m estimated from two samples X^n and Y^m, L_{P_n}, L_{Q_m}, may denote values of quality measures of classification outputs i.e., e. g., classification error.

The qualitative robustness concept has been used by Christmann, Salibi-an-Barrera, Van Aelst (2013), who show that the bootstrap distribution estimates of estimators defined by a functional, which is continuous uniformly over neighbourhoods of distributions, are qualitatively robust. The equicontinuity of relevant functionals seems to be equivalent to the qualitative robustness. Note that in the functional classification case at least one obvious problem arises. We do not know, how to operationize the distance (i.e., d_P in formula (19)) between probabil-

istic measures defining distribution of functional random variables. The distributions are theoretically known (see Bosq, 2000), but it is still an open question how to obtain their characteristics, e.g., cumulative distribution function, probability density function, or d_P. The first possible solution is to make PCA projections of the functional data, thus reducing the problem to the multidimensional case (see Górecki et al., 2018). Finally, the qualitative robustness is analysed with tools designed for a multidimensional case. The second possibility is to apply a data-analytic approach, where we evaluate empirical classification error within simulation studies. We follow this approach in our paper. Another possibility is to bypass the problem of calculating the distance between probabilistic measures defining distribution of functional random variables and to focus on estimating the functional distributions characteristics, we are able to obtain, namely, expected value, or other selected moments of functional random variable. Hence, for example, we can estimate substitute the condition $d_P(P, Q) < \varepsilon$ in formula (19) with $\|EP - EQ\| < \delta$. It is, no doubt, a simplification of the problem, but it allows for rough evaluation of the qualitative robustness.

5. Properties of the proposal

A performance of a classifier is commonly evaluated in terms of the classification error. It seems that it is often the most reasonable approach, as we try to justify in Section 4. Let g denote our classification rule. The distribution of (X, Y) is unknown, so we estimate the empirical risk

$$\hat{L}(g) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{g(\mathbf{X_i}) \neq \mathbf{Y_i}\}},\tag{34}$$

where $\mathbf{1}_S$ denotes the indicator function of the set S.

We implemented our method ourselves, but other classification methods were calculated with R packages *fda.usc* (see Febrero-Bande, de la Fuente, 2012) and *roahd* (see Tarabelloni, 2017).

5.1. Simulation Studies

In order to evaluate properties of the classifiers we conducted rich simulation studies. We used, among others, the following scheme. We generated 500 observations from four Gaussian processes centered in 5, 10, 15, 20, respectively, and with constant covariance function equal to 7.5. We gave four relevant labels to each function, and grouped all the functional observations. Subsequently, we estimated the functional classifiers' quality with cross-validation method. In Table 1 the empir-

ical risk comparison of selected functional classifiers is presented. Fraction of outliers denotes functional amplitude outliers, which represent 5%, 10% and 15% of the training set. Random projection depth, where Fraiman-Muniz depth (FM) is used, and DD-plot classifier, appeared to be the best in our simulation studies, where there were no outliers in the training set, and when we exchanged 5%, 10% and 15% of the training set with functional amplitude outliers, which have been generated from the process

$$F(t) = 60 \sin \frac{t}{2\pi} \cdot \psi_1 + 8\sqrt{2} \cos \frac{t}{2\pi} \cdot \psi_2, \tag{35}$$

where ψ_1 and ψ_2 are independent standard Gaussian random variables.

In Table 2 the empirical risk comparison of selected functional classifiers is presented. Fraction of outliers denotes functional shape outliers, which represent 5%, 10% and 15% of the training set. Note, that even in the case of 5% shape outliers in the training set, all classifiers give rather useless results. Moreover, for some classifiers the increase in shape outliers' number may decrease the empirical risk. It seems counterintuitive, but we chose the special shape outliers in order to obtain the results, namely the shape outliers have been generated from the process

$$G(t) = \psi_1 e^{-\sin(2\pi t)} + \psi_2 e^{\cos(2\pi t)}, \tag{36}$$

where ψ_1 and ψ_2 are independent standard Gaussian random variables.

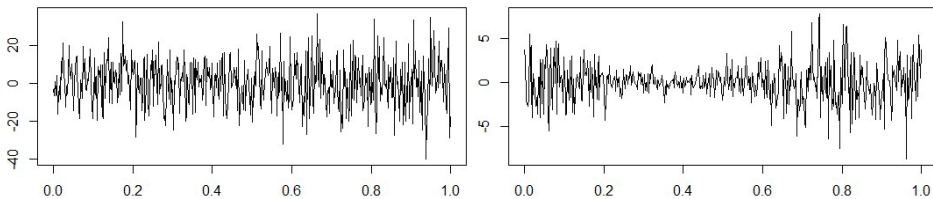In Figure 1 the example sample trajectories from $F(t)$ and $G(t)$ are presented.

Figure 1. Sample trajectories of $F(t)$ (left) and $G(t)$ (right)

Source: own calculations

The form of the trajectories of the process means that the shape outliers are virtually the shape outliers, which was checked with outliergrams, and furthermore they are linearly independent with the training set. Notwithstanding, the clean training set consists of almost linearly dependent functions. The latter fact causes that for the uncontaminated training set, the determinant of matrix $\mathbf{K}(X_i, X_j)_{i,j=1,\dots,n}$ is close to zero, which explains why RKHS method and our method do not work well. It also explains, why depth-based methods do not give satisfactory results. That is why the *knn* method appeared to be relatively the best one. Almost linear

independence of the outlying functional outliers to the original data caused that empirical risk decreased with the number of outliers. The fact is even more visible in an empirical example of CCI.

Table 1. Empirical risk comparison of selected functional classifiers, where functional amplitude outliers represent some fraction of the training set

Classifier	Fraction of outliers			
	0%	5%	10%	15%
Our method	37%	37%	38%	43%
Gaussian kernel	32%	39%	41%	47%
Epanechnikov kernel	29%	29%	35%	38%
Polynomial kernel	34%	49%	50%	63%
knn classifier	32%	46%	48%	53%
random projection depth	22%	22%	34%	35%
DD-plot classifier	18%	20%	22%	28%

Source: own calculations

Table 2. Empirical risk comparison of selected functional classifiers, where functional shape outliers represent some fraction of the training set

Classifier	Fraction of outliers			
	0%	5%	10%	15%
Our method	37%	63%	59%	67%
Gaussian kernel	32%	69%	71%	87%
Epanechnikov kernel	29%	85%	83%	89%
Polynomial kernel	34%	79%	84%	85%
knn classifier	32%	54%	65%	65%
random projection depth	22%	78%	76%	75%
DD-plot classifier	18%	80%	88%	88%

Source: own calculations

6. Empirical analysis of CCI

We considered Consumer Confidence Index (CCI) for USA from January, 1960 to December, 2017 (see OECD, 2018). For discussion of economical indices see Białek (2012). Basing on monthly CCI we constructed a CCI function for every year. In other words, we had only 58 pieces of functional data in the clean training set. Every function was then labelled, in order to describe the state of the USA economy. The labelling scheme consisted of checking whether CCI increased, or decreased in the considered year. Subsequently, we evaluated, whether the CCI was monthly more often above, or below the base level of a 100. Thus four different

labels have been given. In Figure 2 the four groups of the considered CCI functions are presented, as well as the empirical functional mean functions.
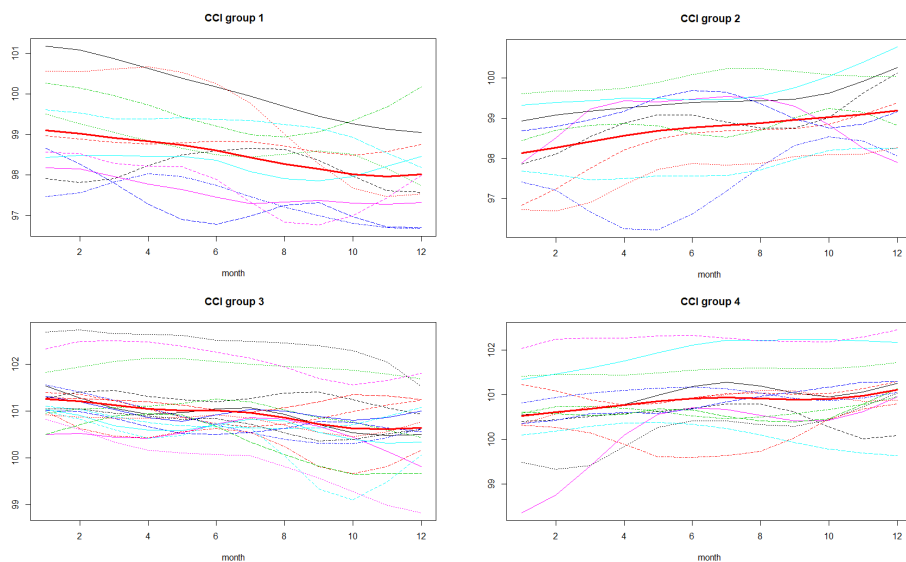

Figure 2. Four groups of CCI functions, functional means are marked in red

Source: own calculations

Subsequently, we estimated the functional classifiers' quality with the cross-validation method. In Table 3 the empirical risk comparison of selected functional classifiers is presented. Fraction of outliers denotes functional amplitude outliers, which represent 5%, 10% and 15% of the training set. The amplitude outliers have been generated from (35), as in the preceding empirical example.

In Table 4 the empirical risk comparison of selected functional classifiers is presented. Fraction of outliers denotes functional shape outliers, which represent 5%, 10% and 15% of the training set. The shape outliers have been generated from (36), as in the preceding empirical example as well. Random projection depth classifier, where Fraiman-Muniz depth (FM) is used, appeared to be the best for training sets uncontaminated and contaminated by outliers. As mentioned earlier, the linear independence of both the outlying shape and amplitude functions from the original data caused that empirical risk decreased with the number of outliers. Moreover, the increase in shape outliers' number may decrease the empirical risk. It is not only the result of the special shape outliers, but it is also caused by the small training set of 58 functions. Exchanging even a small fraction of the training set with outliers easily alters the result.

Table 3. Empirical risk comparison of selected functional classifiers of CCI function, where functional amplitude outliers represent some fraction of the training set

Classifier	Fraction of outliers			
	0%	5%	10%	15%
Our method	65%	69%	69%	69%
Gaussian kernel	62%	70%	70%	73%
Epanechnikov kernel	58%	64%	67%	71%
Polynomial kernel	49%	64%	66%	69%
knn classifier	69%	75%	73%	77%
random projection depth	28%	27%	31%	34%
DD-plot classifier	30%	42%	49%	66%

Source: own calculations

Table 4. Empirical risk comparison of selected functional classifiers of CCI function, where functional shape outliers represent some fraction of the training set

Classifier	Fraction of outliers			
	0%	5%	10%	15%
Our method	65%	64%	62%	55%
Gaussian kernel	62%	51%	51%	36%
Epanechnikov kernel	58%	64%	39%	31%
Polynomial kernel	49%	45%	42%	39%
knn classifier	69%	35%	31%	72%
random projection depth	28%	27%	32%	44%
DD-plot classifier	30%	32%	59%	51%

Source: own calculations

7. Conclusions and recommendations

It is quite obvious that there is no uniformly best classification method. If the data are linearly dependent or almost linearly dependent, than all tested functional classifiers fail. The main result of the paper is that our study shows that even a small number of outliers linearly independent from the training sample, which is in turn almost linearly dependent itself, corrupt all analysed classifiers, even if they are 5% shape or amplitude outliers in the training sample consisting of functional elements. It is the result of the fact that relevant matrices equal zero or are close to zero.

In Table 5 average computation times (in seconds) for presented methods, where functional shape outliers represent some fraction of the training set, are presented. The average computation times for presented methods, where functional amplitude outliers represent some fraction of the training set, are comparable.

Table 5. Average computation times (in seconds) for presented methods, where functional shape outliers represent some fraction of the training set

Classifier	Fraction of outliers			
	0%	5%	10%	15%
Our method	247 s	290 s	312 s	301 s
Gaussian kernel	610 s	643 s	623 s	618 s
Epanechnikov kernel	646 s	645 s	655 s	662 s
Polynomial kernel	607 s	631 s	598 s	608 s
knn classifier	203 s	236 s	225 s	220 s
random projection depth	827 s	820 s	851 s	817 s
DD-plot classifier	856 s	843 s	862 s	849 s

Source: own calculations

Random projection depth classifier, where Fraiman-Muniz depth is used, appeared to be the best for training sets uncontaminated and contaminated by a small number of shape outliers. Depth-based methods have at least one important disadvantage, namely they require a large memory pool, due to the necessity of functional depth computations. That is why they are inadequate when analysed data set is large. If the training set is contaminated with greater fraction of shape outliers, then *knn*-method works relatively well. The method we proposed works well, if we consider two class classification, and note that our method is computationally less intensive – it requires less memory pool (see Kosiorowski, Mielczarek, Rydlewski, 2018), so it is worth recommending to perform classifications of big data sets in two-class setups.

Results presented in the paper can be applied to different fields of e-economy, namely in website management, spam filtering, or protection of computer systems against hacking. As modern economy provides a great deal of functional data sets, some non-obvious applications in the economy can be considered. They are connected, i.e. with optimization of electricity production, municipal road traffic management, or optimization of local air-protection policy (see Kosiorowski, Rydlewski, Zawadzki, 2018). Finally, we would like to stress that a classification rule for functional data enables a consideration of not only scalar values of economic quantities but also the trajectories/shapes of functions describing the quantities. Often, the scalar values describe averages, while managers may be more interested in peak or depression areas. This knowledge is summarized as a function describing a process. Looking further, the classification rule for functional data enables discrimination between the possible paths which the process is following.

Acknowledgements

References

Arribas-Gil A., Romo J. (2013), *Shape Outlier Detection and Visualization for Functional Data: the Outliergram*, "Biostatistics", vol. 15, issue 4, pp. 603–619.

Berlinet A., Thomas-Agnan C. (2004), *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Kluwer, Dordrecht.

Białek J. (2012), *Proposition of the general formula for price indices*, "Communications in Statistics: Theory and Methods", vol. 41, issue 5, pp. 943–952.

Bosq D. (2000), *Linear Processes in Function Spaces*, Springer, New York.

Christmann A., Van Messem A. (2008), *Bouligand Derivatives and Robustness of Support Vector Machines for Regression*, "Journal of Machine Learning Research", vol. 9, pp. 915–936.

Christmann A., Salibian-Barrera M., Van Aelst S. (2013), *Qualitative Robustness of Bootstrap Approximations for Kernel Based Methods*, [in:] C. Becker, R. Fried, S. Kuhnt (eds.), *Robustness and Complex Data Structures*, Springer, Berlin–Heidelberg, pp. 263–278.

Cuevas A. (1988), *Qualitative robustness in abstract inference*, "Journal of Statistical Planning and Inference", vol. 18, issue 3, pp. 277–289.

Cuevas A., Fraiman R. (2009), *On depth measures and dual statistics. A methodology for dealing with general data*, "Journal of Multivariate Analysis", vol. 100, issue 4, pp. 753–766.

Cuevas A., Febrero-Bande M., Fraiman R. (2007), *Robust estimation and classification for functional data via projection-based depth notions*, "Computational Statistics", vol. 22, issue 3, pp. 481–496.

Devroye L., Gyorfi L., Lugosi G. (1996), *A Probabilistic Theory of Pattern Recognition*, Springer, New York.

Febrero-Bande M.O., Fuente M. de la (2012), *Statistical computing in functional data analysis: the R package fda.usc,* "Journal of Statistical Software", vol. 51, issue 4, pp. 1–28.

Ferraty F., Vieu P. (2006), *Nonparametric Functional Data Analysis: Theory and Practice*, Springer, Berlin.

Górecki T., Krzyśko M., Wołyński W. (2018), *Independence test and canonical correlation analysis based on the alignment between kernel matrices for multivariate functional data*, "Artificial Intelligence Review", https://doi.org/10.1007/s10462-018-9666-7

Górecki T., Krzyśko M., Waszak Ł., Wołyński W. (2018), *Selected Statistical Methods of Data Analysis for Multivariate Functional Data*, "Statistical Papers", vol. 59, issue 1, pp. 153–182.

Haykin S. (2009), *Neural networks and learning machines*, Prentice Hall, New Jersey.

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

Hubert M., Van Driessen K. (2004), *Fast and robust discriminant analysis*, "Computational Statistics & Data Analysis", vol. 45, issue 2, pp. 301–320.

Hubert M., Rousseeuw P., Segaert P. (2016), *Multivariate and functional classification using depth and distance*, "Advances in Data Analysis and Classification", vol. 11, issue 3, pp. 445–466.

Kosiorowski D. (2016), *Dilemmas of robust analysis of economic data streams*, "Journal of Mathematical Sciences", vol. 218, issue 2, pp. 167–181.

Kosiorowski D., Zawadzki Z. (2019), *DepthProc: An R package for robust exploration of multidimensional economic phenomena*, "Journal of Statistical Software" (forthcoming).

Kosiorowski D., Mielczarek D., Rydlewski J.P. (2017), *SVM classifiers for functional data in monitoring of the Internet users behavior*, [in:] M. Papież, S. Śmiech (eds.), *The 11th Professor*

A. Zeliaś International Conference on Modelling and Forecasting of Socio-Economic Phenomena, Conference Proceedings, Zakopane, pp. 143–152.

Kosiorowski D., Mielczarek D., Rydlewski J.P. (2018), *New proposal of robust classifier for functional data*, [in:] M. Papież, S. Śmiech (eds.), *The 12ᵗʰ Professor A. Zeliaś International Conference on Modelling and Forecasting of Socio-Economic Phenomena*, Conference Proceedings, Zakopane, pp. 200–208.

Kosiorowski D., Rydlewski J.P., Snarska M. (2019), *Detecting a structural change in functional time series using local Wilcoxon statistic*, "Statistical Papers", vol. 60, pp. 1677–1698, http://dx.doi.org/10.1007/s00362-017-0891-y

Kosiorowski D., Rydlewski J.P., Zawadzki Z. (2018), *Functional outliers detection by the example of air quality monitoring*, "Statistical Review", vol. 65, no. 1, pp. 81–98.

Li J., Cuesta-Albertos J.A., Liu R.Y. (2012), *DD-Classifier: Nonparametric Classification Procedure Based on DD-Plot*, "Journal of the American Statistical Association", vol. 107, issue 498, pp. 737–753.

OECD (2018), *Consumer confidence index (CCI) (indicator)*, http://dx.doi.org/10.1787/46434d78-en

Preda C. (2007), *Regression models for functional data by reproducing kernel Hilbert spaces methods*, "Journal of Statistical Planning and Inference", vol. 137, issue 3, pp. 829–840.

Ramsay J.O., Silverman B.W. (2005), *Functional data analysis*, Springer, Berlin.

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

Schölkopf B., Smola A.J. (2002), *Learning with Kernels*, MIT Press, Cambridge.

Steinwart I., Christmann A. (2008), *Support Vector Machines*, Springer, New York.

Tarabelloni N. (2017), *Robust Statistical Methods in Functional Data Analysis, Doctoral thesis and R package roahd*, Politecnico di Milano, Milano.

Vencálek O. (2013), *Depth-based Modification of the k-nearest Neighbour Method*, "SOP Transactions on Statistics and Analysis", vol. 1, no. 2, pp. 131–138.

Vencálek O., Pokotylo O. (2018), *Depth-weighted Bayes classification*, "Computational Statistics and Data Analysis", vol. 123, pp. 1–12.

Krytyczna analiza wybranych klasyfikatorów dla danych funkcjonalnych w kontekście ich zastosowań w ekonomii

Streszczenie: W artykule przeprowadzono krytyczną analizę najbardziej znanych klasyfikatorów dla danych funkcjonalnych. Ponadto zaproponowano nowy klasyfikator dla danych funkcjonalnych. Przedyskutowano pewne, związane z odpornością, własności rozważanych klasyfikatorów. Wypracowane w artykule podejście może zostać użyte do przewidywania stanu gospodarki na podstawie indeksu mierzącego optymizm konsumentów – CCI (*Consumer Confidence Index*) oraz indeksu odzwierciedlającego optymizm w sektorze przemysłowym – IPI (*Industrial Price Index*), przy czym wykorzystuje się nie tylko skalarne wartości indeksu, lecz także całą trajektorię/kształt funkcji opisującej dany indeks. W związku z tym nasze rozważania mogą być pomocne w skonstruowaniu lepszego barometru stanu gospodarki. O ile wiadomo autorom, jest to pierwsze porównanie klasyfikatorów dla danych funkcjonalnych ze względu na kryterium ich użyteczności aplikacyjnej w ekonomii. Głównym celem artykułu jest zaprezentowanie, jak mała frakcja obserwacji nietypowych w próbce uczącej, będących liniowo niezależnymi z próbką uczącą, która z kolei składa się z funkcji prawie liniowo zależnych, jest w stanie poważnie zaburzyć wyniki klasyfikacji dla wszystkich rozpatrywanych klasyfikatorów.

Słowa kluczowe: klasyfikator funkcjonalny, analiza danych funkcjonalnych, metody odporne, barometr optymizmu w ekonomii

JEL: C38, C63, C32