Michal Holčapek; Tomaš Tichý

A probability density function estimation using F-transform

Kybernetika, Vol. 46 (2010), No. 3, 447--458

Persistent URL: http://dml.cz/dmlcz/140760

KYBERNETIKA — VOLUME *46* (2010), NUMBER 3, PAGES 447–458

A PROBABILITY DENSITY FUNCTION ESTIMATION USING F-TRANSFORM

Michal Holčapek and Tomaš Tichý

The aim of this paper is to propose a new approach to probability density function (PDF) estimation which is based on the fuzzy transform (F-transform) introduced by Perfilieva in [10]. Firstly, a smoothing filter based on the combination of the discrete direct and continuous inverse F-transform is introduced and some of the basic properties are investigated. Next, an alternative approach to PDF estimation based on the proposed smoothing filter is established and compared with the most used method of Parzen windows. Such an approach can be of a great value mainly when dealing with financial data, i. e. large samples of observations.

Keywords: fuzzy transform, probability density function estimation, smoothing filter, financial returns

Classification: 60E99, 91G80

1. INTRODUCTION

Without loss of generality we can say that any economic variable or, more precisely, its state (e. g. the price), follows a stochastic process, i. e. it moves randomly in time. In order to describe such a randomness, a suitable probability distribution can be used. Doing that, either characteristic function, or cumulative distribution function, or probability density function (PDF) can be used. Since the data are generally observable only in a discrete time, some approach of PDF estimation must be applied.

In general, one can distinguish three distinct approaches to PDF estimation: a parametric, a semi-parametric and a non-parametric one. Since the a priori choice of the PDF may often provide a false representation of the true PDF, here, we focus on non-parametric methods. Although there is a wide variety of non-parametric methods estimating PDF (for a survey, see e. g. [4, 12]), only a few of them are used in practice. Among the most used methods are the histograms, Parzen windows, vector quantization based Parzen, and finite Gaussian mixtures. A comparison of these methods can be found in [2]. In this paper, we would like to propose an approach to PDF estimation based on the fuzzy (F-)transform, introduced by Perfilieva in [10].

The idea is very simple. A larger sample size allows us to build up a good estimation of an unknown PDF using the histogram technique. Unfortunately, the histogram is not continuous. Hence, we should use a smoothing technique to obtain a continuous representation of the histogram and then the unknown PDF. It is well known that the smoothing techniques, such as stochastic processes, kernel regressions, integral transforms, wavelet transforms, or even fuzzy filters, for more details see e. g. [3, 5, 6, 8], often give a noise reduction contained in the histogram, which is very profitable in cases when the samples are used.

A relatively new approach to the smoothing of data seems to be the F-transform method, which is based on a fuzzy partitioning of a universe into fuzzy subsets. Analogously to the standard smoothing approaches, the F-transform offers a noise reduction that has been illustrated on continuous functions in [11].

The aim of this paper is to introduce a filter based on the F-transform that would allow us an efficient smoothing of PDF of financial market returns, i. e. a data set of a large length.

We proceed as follows. The following preliminary section provides the important facts relating to the discrete F-transform. Further, we will propose an FT-smoothing filter which is based on the discrete F-transform and the inverse continuous F-transform. Some of the basic FT-smoothing filter properties are presented. The fourth section is devoted to the PDF estimation for the financial data, where the basic approaches are mentioned and an approach based on the FT-smoothing filter is introduced. The last section is a conclusion.

2. DISCRETE F-TRANSFORM

The key idea on which the F-transform is based is a fuzzy partition of the universe into fuzzy subsets (factors, clusters, granules etc.). A fuzzy partition may be understood as a system of neighborhoods of some chosen nodes. For a sufficient representation of a function we may consider its average values over fuzzy subsets from the partition. Then, a function can be associated with a mapping from a set of fuzzy subsets to the set of thus obtained average function values. We take an interval $[a, b]$ as a universe. That is, all (real-valued) functions considered in this section have this interval as a common domain. The fuzzy partition of the universe is given by fuzzy subsets of the universe $[a, b]$ (determined by their membership functions) which must have properties described in the following definition.

Definition 2.1. Let $x_1 < \cdots < x_n$ be fixed nodes within $[a, b]$, such that $x_1 = a$, $x_n = b$ and $n \geq 2$. We say that fuzzy sets $A_1, \dots, A_n$, identified with their membership functions $A_1(x), \dots, A_n(x)$ defined on $[a, b]$, form a fuzzy partition of $[a, b]$, if they fulfill the following conditions for $k = 1, \dots, n$:

(1) $A_k : [a, b] \to [0, 1]$, $A_k(x_k) = 1$;

(2) $A_k(x) = 0$, if $x \notin (x_{k-1}, x_{k+1})$, where for the uniformity of denotation, we put $x_0 = a$ and $x_{n+1} = b$;

(3) $A_k(x)$ is continuous;

(4) $A_k(x)$, $k = 2, \dots, n$, strictly increases on $[x_{k-1}, x_k]$ and $A_k(x)$, $k = 1, \dots, n-1$, strictly decreases on $[x_k, x_{k+1}]$;

(5) for all $x \in [a, b]$

$$\sum_{k=1}^{n} A_k(x) = 1. \tag{1}$$

The fuzzy sets $A_1, \ldots, A_n$ are called *basic functions.*

Note that an interesting generalization of the fuzzy partition which can lead to a greater smoothness of functions can be found in [14]. A simple consequence of the definition of fuzzy partitions is the following lemma saying that it is sufficient to define only left or right parts of basic functions, where, e. g. the left part of A_k is the restriction of function A_k on the interval $[x_{k-1}, x_k]$.

Lemma 2.2. Let $x_1 < \cdots < x_n$ be fixed nodes within $[a, b]$ and $\mathcal{A} = \{A_1, \ldots, A_n\}$ define a fuzzy partition of $[a, b]$. Then

$$A_k(x) = 1 - A_{k+1}(x) \tag{2}$$

holds for any $k = 1, \ldots, n-1$ and $x \in [x_k, x_{k+1}]$.

Definition 2.3. Let $\mathcal{A} = \{A_1, \ldots, A_n\}$ define a fuzzy partition of $[a, b]$. We say that $\mathcal{A}$ forms a *uniform* fuzzy partition of $[a, b]$, if $x_{k+1} - x_k = h$ for any $k = 1, \ldots, n-1$, where $h > 0$ is a constant, and there is a fuzzy set A identified with its membership function $A(x)$ defined on $(-\infty, \infty)$ satisfying the following two properties:

(i) A is an even function,

(ii) $A(x_k - x) = A_k(x)$ for any $x \in [a, b]$ and $k = 1, \ldots, n$.

From the definition of fuzzy partitions and Lemma 2.2, we simply obtain $A(x) = 0$ for any $x \notin [-h, h]$, $A(0) = 1$ and $A(x) + A(x + h) = 1$ for any $x \in [-h, 0]$. The determination of basic functions from a one fuzzy set seems to be profitable from an optimization point of view. More precisely, only the parameter h of the uniform fuzzy partitions is optimized, contrary to the non-uniform fuzzy partitions, where each basic function is determined by at least two parameters, i. e. a center x_k and a width h_k of, for example, the left part of basic function A_k (cf. Lemma 2.2). Nevertheless, one could imagine that the non-uniform fuzzy partitions may be very important in some cases when the number of basic functions of uniform fuzzy partition is too large[1] and the function (on which the F-transform is applied) over some basic functions has no important changes. In this work, we will mainly deal with the uniform fuzzy partitions.

The following lemma shows a useful property holding for the uniform fuzzy partitions and which will be used latter in the text.

[1]Note that the number of basic functions has a primary importance in the rate of computational algorithms.

Lemma 2.4. Let $\mathcal{A} = \{A_1, \dots, A_n\}$ define a uniform fuzzy partition of $[a, b]$ such that $n \geq 3$ and $x_{k+1} - x_k = h$. Then

$$\int_a^b A_k(x)^r \,\mathrm{d}x = \alpha_{rk} h, \tag{3}$$

$$\int_a^b A_k(x) A_l(x) \,\mathrm{d}x = \begin{cases} \frac{\alpha}{2} h, & \text{if } k = l = 1 \text{ or } k = l = n, \\ \alpha h, & \text{if } k = l \neq 1 \text{ and } k = l \neq n, \\ \beta h, & \text{if } |k - l| = 1, \\ 0, & \text{otherwise,} \end{cases} \tag{4}$$

where $r \geq 1$ and $\alpha_{rk}, \alpha, \beta \in [0, 1]$ are suitable constants, hold for any $k, l = 1, \dots, n$. Especially, we have

$$\alpha_{1k} = \begin{cases} \frac{1}{2}, & \text{if } k = 1 \text{ or } k = n \\ 1, & \text{otherwise.} \end{cases} \tag{5}$$

Proof. This is a simple consequence of the fact that all basic functions are determined by the same function A and the area under A is less than or equal to h. In fact, we may write

$$\int_{-\infty}^{\infty} A(x) \,\mathrm{d}x = \int_{-h}^{h} A(x) \,\mathrm{d}x = \int_{-h}^{0} A(x) \,\mathrm{d}x + \int_{0}^{h} A(x) \,\mathrm{d}x$$
$$= \int_{-h}^{0} A(x) \,\mathrm{d}x + \int_{-h}^{0} A(x + h) \,\mathrm{d}x = \int_{-h}^{0} (A(x) + A(x + h)) \,\mathrm{d}x = \int_{-h}^{0} \mathrm{d}x = h.$$

□

Let us show two basic examples of uniform fuzzy partitions.

Example 2.5. Let $A, B : [-h, h] \to [0, 1]$ be fuzzy sets defined as follows:

$$A(x) = \begin{cases} \frac{h - |x|}{h}, & x \in [-h, h], \\ 0, & \text{otherwise,} \end{cases} \qquad B(x) = \begin{cases} 0.5 \cos \frac{\pi}{h} x + 0.5, & x \in [-h, h], \\ 0, & \text{otherwise.} \end{cases}$$

If $x_1 < \cdots < x_n$ are nodes within $[a, b]$ such that $x_1 = a$, $x_n = b$ and $x_k = x_1 + (i - 1)h$, then $\mathcal{A} = \{A_k \mid A(x_k - x) = A_k(x),\ k = 1, \dots, n\}$ and $\mathcal{B} = \{B_k \mid B(x_k - x) = B_k(x),\ i = 1, \dots, n\}$ define a *triangle* and a *cosine uniform fuzzy partition of* $[a, b]$, respectively. In Figure 1, we can see the triangle and cosine uniform fuzzy partitions of the interval $[2, 7]$ for $h = 0.5$.

Now, we can introduce the definition of the discrete (direct) F-transform which assigns, using basic functions, to each function f defined in a finite number of nodes a vector of real numbers. This vector is then a representation of the function f. Let $t_1, \dots, t_l \in [a, b]$ be nodes and $\mathcal{A}$ be a set of basic functions. In the following text we assume that each set of nodes $t_1, \dots, t_l$ is *sufficiently dense with respect to* $\mathcal{A}$, i. e. for each $A_k \in \mathcal{A}$ there is a node t_j such that $A_k(t_j) > 0$.

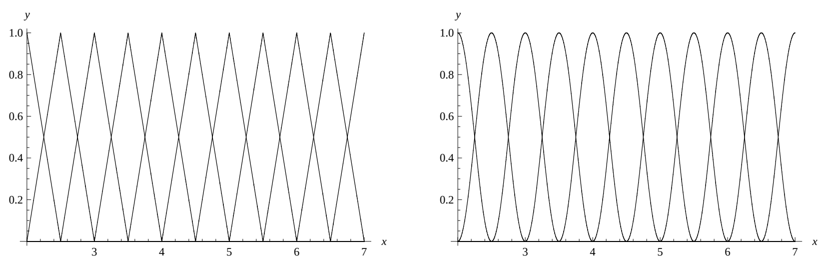

Fig. 1. Triangle and cosine uniform fuzzy partitions of $[2, 7]$.

Definition 2.6. Let a function f be given at nodes $t_1, \dots, t_l \in [a, b]$ and $\mathcal{A} = \{A_1, \dots, A_n\}$, $n < l$, be a set of basic functions which form a fuzzy partition of $[a, b]$. We say that the n-tuple of real numbers $[F_{\mathcal{A},1}, \dots, F_{\mathcal{A},n}]$ is the *discrete F-transform of f with respect to $\mathcal{A}$*, if

$$F_{\mathcal{A},k} = \frac{\sum_{j=1}^{l} f(t_j) A_k(t_j)}{\sum_{j=1}^{l} A_k(t_j)}. \tag{6}$$

For more information about the discrete F-transform, we refer to [10].

3. FT-SMOOTHING FILTER

We will denote $\mathbb{R}$ the set of all real numbers and $\mathrm{D}(f)$ the domain of f. We say that a function f is *finite*, if its domain is a finite set. Let $\mathcal{A}$ be a set of basic functions which form a fuzzy partition of $[a, b]$. We will denote $\mathcal{D}([a, b], \mathcal{A})$ the set of all finite real functions f such that $\mathrm{D}(f) \subseteq [a, b]$, $\mathrm{D}(f)$ is sufficiently dense with respect to $\mathcal{A}$ and $|\mathcal{A}| < |\mathrm{D}(f)|$. Obviously, the set $\mathcal{D}_{\mathcal{A}}$ contains all functions on which the discrete F-transform may be applied. Finally, we will denote $\mathcal{C}([a, b])$ the set of all continuous real functions f with $\mathrm{D}(f) = [a, b]$.

Definition 3.1. Let $\mathcal{A} = \{A_1, \dots, A_n\}$ define a fuzzy partition of $[a, b]$. An *FT-smoothing filter determined by $\mathcal{A}$* is a mapping $\mathcal{F}_{\mathcal{A}} : \mathcal{D}([a, b], \mathcal{A}) \to \mathcal{C}([a, b])$ defined by

$$\mathcal{F}_{\mathcal{A}}(f)(x) = \sum_{k=1}^{n} F_{\mathcal{A},k} A_k(x) \tag{7}$$

for any $x \in [a, b]$, where $F_{\mathcal{A},k}$, $k = 1, \dots, n$, are the components of the discrete F-transform.

Remark 3.2. One can check easily that the linear combination of continuous functions is a continuous function. Hence, our definition is correct and $\mathcal{F}_{\mathcal{A}}$ is really a mapping to the set of all continuous functions.

Remark 3.3. Let us note that our definition of the FT-smoothing filter is the inverse F-transform for the continuous case. For the discrete case the resulted function of the inverse F-transform is again a finite function, more precisely, if f is a finite function with $\mathrm{D}(f) = \{t_1, \dots, t_l\}$, then

$$f_{F_{\mathcal{A}}}(t_j) = \mathcal{F}_{\mathcal{A}}(f)(t_j), \tag{8}$$

$j = 1, \dots, l$, defines the finite function which is the inverse F-transform of f in the sense of Definition 5 in [10].

In Figure 2, there is an illustration of FT-smoothing filters determined by the triangle and cosine uniform partitions introduced in Example 2.5 applied on a finite function. One can see the effect of F-transform which computes the average values from the function values over fuzzy sets defining fuzzy partition. The linear combination used in the FT-smoothing filter defines a type of approximation and the result of this procedure is a smooth function.

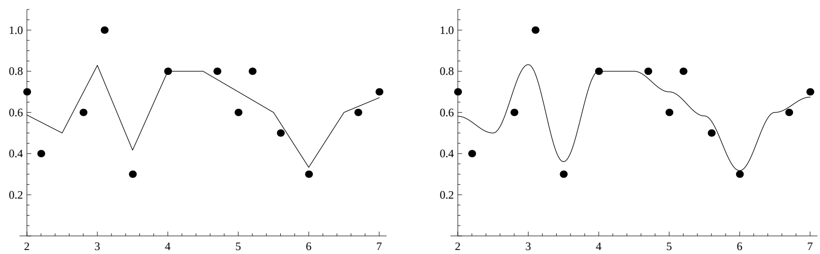

Fig. 2. Smoothed finite function by FT-smoothing filters based on the triangle and cosine uniform fuzzy partitions.

Let us define the *partial addition* on $\mathcal{D}([a, b], \mathcal{A})$ (or on $\mathcal{C}([a, b])$) by

$$(f + g)(t) = f(t) + g(t), \tag{9}$$

for any $f, g \in \mathcal{D}([a, b], \mathcal{A})$ such that $\mathrm{D}(f) = \mathrm{D}(g)$ and the *multiplication by a real number* on $\mathcal{D}([a, b], \mathcal{A})$ (or on $\mathcal{C}([a, b])$) by

$$(\alpha f)(t) = \alpha(f(t)) \tag{10}$$

for any $f \in \mathcal{D}([a, b], \mathcal{A})$ and $\alpha \in \mathbb{R}$. Let $f, g \in \mathcal{D}([a, b], \mathcal{A})$ (or $f, g \in \mathcal{C}([a, b])$). We shall say that f is *less than or equal to* g and write $f \leq g$, if $\mathrm{D}(f) = \mathrm{D}(g)$ and $f(x) \leq g(x)$ for any $x \in \mathrm{D}(f)$. Obviously, the relation $\leq$ is a *partial ordering* on $\mathcal{D}([a, b], \mathcal{A})$ (or on $\mathcal{C}([a, b])$). The following lemmas are straightforward conclusions of the previous definitions.

Lemma 3.4. Let $f, g \in \mathcal{D}([a, b], \mathcal{A})$ such that $\mathrm{D}(f) = \mathrm{D}(g)$ and $\alpha, \beta \in \mathbb{R}$. Then

$$\mathcal{F}_{\mathcal{A}}(\alpha f + \beta g) = \alpha \mathcal{F}_{\mathcal{A}}(f) + \beta \mathcal{F}_{\mathcal{A}}(g) \tag{11}$$

If $f \leq g$, then $\mathcal{F}_{\mathcal{A}}(f) \leq \mathcal{F}_{\mathcal{A}}(g)$.

Lemma 3.5. Let $f \in \mathcal{D}([a,b],\mathcal{A})$ and $f(x) = \alpha$, where $\alpha \in \mathbb{R}$, for any $x \in \mathrm{D}(f)$. Then $\mathcal{F}_\mathcal{A}(f)(x) = \alpha$ for any $x \in [a,b]$.

Let f be a finite function. To investigate an approximation power of FT-smoothing filter, let us consider the following (local) modulus of continuity

$$\omega_f(x,h) = \sup_{\substack{y\in \mathrm{D}(f)\\ |x-y|\le h}} |f(x) - f(y)| \tag{12}$$

defined for any $x \in \mathrm{D}(f)$. Now we can state the following theorem[2] showing how the approximation of f by an FT-smoothing filter is good.

Theorem 3.6. Let $\mathcal{A}$ be a uniform fuzzy partition of $[a,b]$ with $x_{k+1} - x_k = h$ and $f \in \mathcal{D}([a,b],\mathcal{A})$. Then

$$|f(t_i) - \mathcal{F}_\mathcal{A}(f)(t_i)| \le \omega_f(t_i, 2h) \tag{13}$$

holds for any $t_i \in \mathrm{D}(f)$.

Proof. Let $t_i \in \mathrm{D}(f)$, $|\mathrm{D}(f)| = n$. According to Lemma 2.2, there is a natural number k such that either $A_k(t_i) = 1$ (i.e. $A_l(t_i) = 0$ for $k \ne l$), or $A_k(t_i) > 0$ and $A_{k+1}(t_i) > 0$ (i.e. $A_l(t_i) = 0$ for $k \ne l \ne k+1$). Let us consider the second case, the first one may be proved analogously. We can write

$$\begin{aligned}
|F_k - f(t_i)| &= \left| \frac{\sum_{j=1}^n f(t_j)A_k(t_j)}{\sum_{j=1}^n A_k(t_j)} - f(t_i) \right| = \left| \frac{\sum_{j=1}^n (f(t_j) - f(t_i))A_k(t_j)}{\sum_{j=1}^n A_k(t_j)} \right| \\
&\le \frac{\sum_{j=1}^n |f(t_j) - f(t_i)|A_k(t_j)}{\sum_{j=1}^n A_k(t_j)} = \frac{\sum_{t_j\in \mathrm{D}(f), |t_j - t_i|\le 2h} |f(t_i) - f(t)|A_k(t_i)}{\sum_{t_j\in \mathrm{D}(f), |t_j - t_i|\le 2h} A_k(t_i)} \\
&\le \frac{\sum_{t_j\in \mathrm{D}(f), |t_j - t_i|\le 2h} \omega_f(t_i,2h)A_k(t_i)}{\sum_{t_j\in \mathrm{D}(f), |t_j - t_i|\le 2h} A_k(t_i)} = \omega_f(t_i, 2h).
\end{aligned}$$

Analogously, we can prove $|F_{k+1} - f(t_i)| \le \omega_f(t_i, 2h)$. Hence, using Lemma 2.2 (recall that $A_k(t) + A_{k+1}(t) = 1$), we obtain

$$\begin{aligned}
|f(t) - \mathcal{F}_\mathcal{A}(f)(t)| &= |f(t) - (F_k A_k(t) + F_{k+1}A_{k+1}(t))| \\
&= |(f(t) - F_k)A_k(t) + (f(t) - F_{k+1})A_{k+1}(t)| \\
&\le |f(t) - F_k|A_k(t) + |f(t) - F_{k+1}|A_{k+1}(t) \le \omega_f(t_i, 2h)A_k(t) + \omega_f(t_i,2h)A_{k+1}(t) \\
&= \omega_f(t_i, 2h)
\end{aligned}$$

and the proof is finished. □

One can easily see that smaller values of h give smaller values of modulus of continuity and thus a better approximation in general.

[2]Note that the approximation power of the (direct-inverse) F-transform using modulus of continuity for continuous functions has been investigated in [10].

4. PROBABILITY DENSITY ESTIMATION FOR FINANCIAL DATA

Let us assume a time series of financial data $\{x_t \mid t = 1, \ldots, T\}$. A general model for x_t can look as follows:

$$x_t = f(x_1, \ldots, x_{t-1}) + \varepsilon_t, \tag{14}$$

where $f(x_1, \ldots, x_{t-1})$ can be any function (linear, non-linear) of preceding values or simply a given constant and ε_t describes a white noise, i. e. ε_t is a value of random variable with zero mean and no autocorrelation. Further, assume that $\{x_t \mid t = 1, \ldots, T\}$ is a set of values of a continuous random variable X. To make a deeper analysis of data, we are often interested in the unknown probability density function of the random variable X. In practice, we usually use some of non-parametric methods for a probability density function estimation. For our purpose, let us mention basic methods.

4.1. Basic approaches

The simplest method how to approximate PDF is a method called a *histogram*. Let the suitable interval be divided in M mutually disjoint bins B_j of a bin width h covering the considered interval. The PDF is then given by

$$\hat{f}_h(x) = \frac{1}{Th} \sum_{t=1}^{T} \sum_{j=1}^{M} I(x_t \in B_j) I(x \in B_j), \tag{15}$$

where T is the length of the series of data and $I(\cdot)$ is the indicator function

$$I(A) = \begin{cases} 1, & \text{if } A \text{ holds,} \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

It is easy to see that the histogram appears to be strongly dependent on h and the PDF is neither smooth nor continuous. Here, however, it is important to emphasize that we can obtain very precise (of course, non-continuous) image of PDF, if we have a large sample.

Another and also the most popular approach to the estimation of PDF is a *Parzen windows estimator* (established in [9]) based on a kernel function $K(y)$:

$$\hat{f}_h(x) = \frac{1}{Th} \sum_{t=1}^{T} K\left(\frac{x_t - x}{h}\right), \tag{17}$$

where h is a bandwidth parameter for the window size description. A natural choice for the kernel function $K(y)$ is the Gaussian kernel $N(x|x_n, h)$ or any kernel from the symmetric Beta family (for details we refer to [4, 12, 13]). A criterion of goodness of estimation for a given bandwidth h may be given by the *integrated square error*

$$\mathrm{ISE}(h) = \int_{-\infty}^{\infty} (f(x) - \hat{f}_h(x))^2 \,\mathrm{d}x, \tag{18}$$

where $f(x)$ is the unknown PDF and $\hat{f}_h(x)$ its estimation. Obviously, the best choice of h is in which ISE(h) has the minimal value. Since $f(x)$ is unknown, there are several methods how to find an optimal choice of h. For example, Silverman proposed in [12] an optimal value of h for the Gaussian kernel derived from (18) by the following rule of thumb:

$$h_{\mathrm{SIL}} = 0.9AT^{-\frac{1}{5}}, \quad A = \min(s, \frac{R}{1.34}), \tag{19}$$

where s is the empirical standard deviation and R is the sample interquartile range. It is known that the Parzen windows estimator usually gives very good results, but the model complexity is proportional to the number of data samples which can rapidly lead to storage problems (see e. g. [2]). It motivated to develop some methods which circumvent this disadvantage. Two very popular methods are the *Vector quantization based Parzen* and the *Finite Gaussian mixtures* (see e. g. [1, 7]) that goal is to decrease the number of used Gaussian kernels in the procedure.

4.2. An approach based on FT-smoothing filter

As we have mentioned in the previous section, the histogram gives very precise image of PDF for samples with larger sizes. Then the PDF estimation can be obtained by a smoothing procedure, in our case by the FT-smoothing filter. Thus, our approach has two steps:

Step 1. we create a finite function $\{(x_i, \hat{f}_r(x_i)) \mid i = 1, \ldots, m\}$, where x_i are the centers of bins with the bin width r and $\hat{f}_r(x_i)$ are the values of the histogram (15);

Step 2. we apply the FT-smoothing filter determined by a uniform fuzzy partition $\mathcal{A} = \{A_1, \ldots, A_n\}$ with $h = x_{k+1} - x_k$ and the PDF estimation $\hat{f}_h$ has the form

$$\hat{f}_h(x) = \frac{\mathcal{F}_{\mathcal{A}}(\hat{f}_r)(x)}{S}, \tag{20}$$

where $S = h(\frac{1}{2}(F_{\mathcal{A},1} + F_{\mathcal{A},n}) + \sum_{k=2}^{n-1} F_{\mathcal{A},k})$.

Note that S is the area (integral) under the function $\mathcal{F}_{\mathcal{A}}(\hat{f}_r)(x)$ to obtain a PDF and it is computed using Lemma 2.4.

To investigate the goodness of PDF estimation for a given bandwidth h, let us consider the integrated square error ISE(h) defined above. Our aim is to find a value for h in which ISE(h) is as small as possible. Rewriting this criterion, we obtain

$$\mathrm{ISE}(h) = \int_{-\infty}^{\infty} f(x)^2 \,\mathrm{d}x - 2\int_{-\infty}^{\infty} f(x)\hat{f}_h(x)\,\mathrm{d}x + \int_{-\infty}^{\infty} \hat{f}_h(x)^2 \,\mathrm{d}x. \tag{21}$$

Since the first term does not dependent on h, we can ignore it here. Then an error criterion may by given by

$$E(h) = \int_{-\infty}^{\infty} \hat{f}_h(x)^2 \,\mathrm{d}x - 2\int_{-\infty}^{\infty} f(x)\hat{f}_h(x)\,\mathrm{d}x = \int_{-\infty} \hat{f}_h(x)^2 \,\mathrm{d}x - 2E(\hat{f}_h(X)), \tag{22}$$

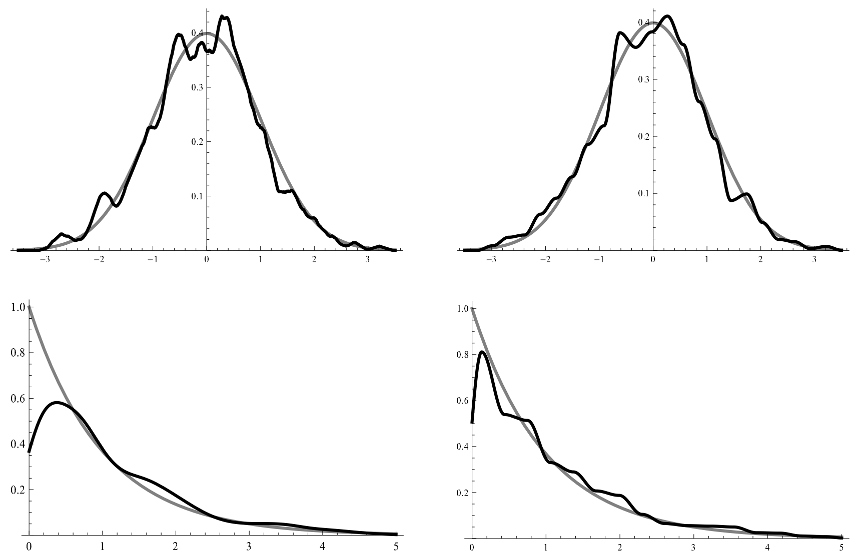

Fig. 3. Comparison of Gaussian kernel (left) and FT-smoothing filter (right) approach applied on the Gaussian and Exponential probability distribution.

where $E(\hat{f}_h(X)))$ denotes the expected value of $\hat{f}_h(X)$. Hence, we can establish a *leave-one-out cross-validation criterion* as follows

$$E_{LOO}(h) = \int_{-\infty}^{\infty} \hat{f}_h(x)^2 \,\mathrm{d}x - 2\sum_{i=1}^{m} \hat{f}_h(x_i) p_h(x_i), \tag{23}$$

$$p_h(x_i) = \frac{\hat{f}_h^{-i}(x_i)}{\sum_{j=1}^{m} \hat{f}_h^{-j}(x_j)}, \tag{24}$$

where $\hat{f}^{-i}(x)$ is the PDF (leave-one-out) estimation of $f(x)$ for the sample without x_i and m denotes the sample size. Intuitively, $p_h(x_i)$ is a discrete probability function which estimates the unknown $f(x)$. The value h minimizing the error E_{LOO} may be understood as an optimal bandwidth (denote it by h_{opt}) for the FT-smoothing filter. If n denotes the number of basic functions, it is easy to show that the integral in E_{LOO} can be expressed by

$$\int_{-\infty}^{\infty} \hat{f}_h(x)^2 \,\mathrm{d}x = \frac{\frac{\alpha}{2}(F_{\mathcal{A},1}^2 + F_{\mathcal{A},n}^2) + \alpha \sum_{k=2}^{n-1} F_{\mathcal{A},k}^2 + 2\beta \sum_{k=1}^{n-1} F_{\mathcal{A},k} F_{\mathcal{A},k+1}}{h(\frac{1}{2}(F_{\mathcal{A},1} + F_{\mathcal{A},n}) + \sum_{k=2}^{n-1} F_{\mathcal{A},k})^2}, \tag{25}$$

where α and β are the suitable constants obtained by the integral (4) in Lemma 2.4. Hence, we can see that the computation of the integral is very simple.

In Figure 3, we can see the results of the Gaussian kernel and the FT-smoothing filter applied on the standard normal distribution and the exponential distribution with $\lambda = 1$, when we use the Silverman's estimation of h for Gaussian kernel

($h_{\mathrm{SIL}} = 0.254$ for the Gaussian and $h_{\mathrm{SIL}} = 0.235$ for the Exponential PDF) and the estimation of the optimal value of h for the FT-smoothing filter based on the uniform cosine fuzzy partition ($h_{\mathrm{opt}} = 0.295$ and $h_{\mathrm{opt}} = 0.265$, respectively). Note that the number of data is 500 and the number of bins is 30 (the bin width is 0.259 for Gaussian and 0.260 for Exponential PDF). Comparing the results it is evident that both estimations of the Gaussian PDF are very similar and there is a slight overfitting here. A better PDF estimation respecting the smoothness of obtained PDF seems to be obtained by a slight increasing of the bandwidth. On the other hand, a better result for the Exponential PDF seems to be for the FT-smoothing filter which is again slightly overfitted. The bandwidth obtained by Silverman's rule of thumb appears too big for a good approximation of values close to 0 and thus an advanced method for estimation of h_{opt} would be used here to ensure a better result.

In Figure 4, we can see a comparison of both approaches on a data set of continuous financial returns, i. e. a natural logarithm of discretely observed prices is considered here. We choose continuous financial returns since they are (relatively) scale-free and since their statistical properties allow us easy handling. The data set consists of 2017 daily returns of CZK/EUR exchange rate over last eight years. The data exhibit very high kurtosis (13.01) and slightly negative skewness (-0.13). Thus, a relatively high probability of extreme returns constitute a challenging task for any filter. We set $h_{\mathrm{SIL}} = 0.00053$ and $h_{\mathrm{opt}} = 0.0025$. Obviously, both results seems to be very similar, perhaps, the FT-smoothing filter gives a generally less smoothed PDF contrary to the Gaussian kernel. By contrast, we believe that the FT-smoothing filter works slightly better when the tails are to be smoothed.

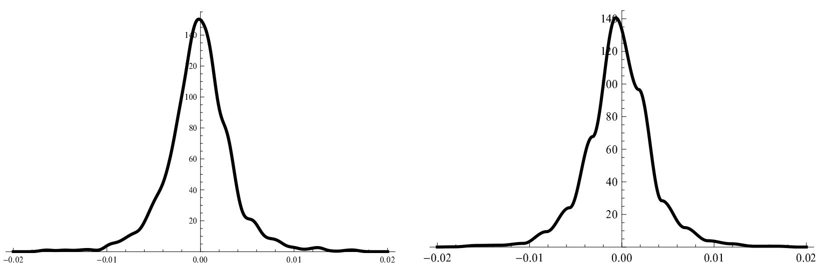

Fig. 4. Comparison of Gaussian kernel (left) and FT-smoothing filter (right) approach applied on CZK/EUR exchange rate.

5. CONCLUSIONS

In this paper, we proposed an alternative approach to the PDF estimation based on the F-transform that is more suitable for samples with larger sizes and non-standard distributions. The approach is based on the discrete F-transform filtering and it can be of a great value mainly for a financial modeling and forecasting.

ACKNOWLEDGEMENT

The support of the Czech Science Foundation under Grant 402/08/1237 and the SGS project of VSB-TU Ostrava SP/20104 is kindly acknowledged.

(Received April 14, 2010)

REFERENCES

[1] S. C. Ahalt, A. K. Krishnamurthy, P. K. Chen, and D. E. Melton: Competative learning algorithms for vector quantization. Neural Networks *4* (1993), 3, 277 – 290.

[2] A. Assenza, M. Valle, and M. Verleysen: A comparative study of various probabilty density estimation methods for data analysis. Internat. J. Comput. Intell. Systems *1* (2009), 2, 188–201.

[3] J. Fan and Q. Yao: Nonlinear Time Series: Nonparametric and Parametric Methods. (Springer Series in Statistics.) Springer-Verlag, Berlin 2005.

[4] W. Härdle, M. Müller, S. Sperlich, and A. Werwatz: Nonparametric and Semiparametric Models. Springer-Verlag, New York 2004.

[5] E. J. Kostelich and J. A. Yorke: Noise reduction: Finding the simplest dynamical system consistent with the data. Physica *D 41* (1990), 183–196.

[6] Hui-Hsiung Kuo: White Noise Distribution Theory. CRC Press, 1996.

[7] G. J. McLachlan and D. Peel: Finite Mixture Models. Wiley, New York 2000.

[8] M. Nachtegael, D. Van der Weken, D. Van De Ville, and E. E. Kerre, eds.: Fuzzy Filters for Image Processing. (Studies in Fuzziness and Soft Computing.) Springer-Verlag, 2003.

[9] E. Parzen: On estimation of a probabilty density function and mode. Ann. Math. Statist. *33* (1962), 1065–1076.

[10] I. Perfilieva: Fuzzy transforms: Theory and applications. Fuzzy Sets and Systems *157* (2006), 8, 993–1023.

[11] I. Perfilieva and R. Valášek: Fuzzy transforms in removing noise. In: Innovation in Hybrid Intelligent Systems. Springer-Verlag, Berlin – Heidelberg 2005.

[12] B. W. Silverman: Density Estimation for Statistics and Data Analysis. Chapman & Hall/CRC, London 1986.

[13] J. S. Simonoff: Smoothing Methods in Statistics. Springer-Verlag, New York 1996.

[14] L. Stefanini: Fuzzy transforms and smooth function. In: Proc. IFSA/EUSFLAT 2009, Lisabon 2009.

Michal Holčapek, Deartment of Mathematical Methods in Economics, Faculty of Economics VŠB-TU Ostrava, Sokolská 33, 701 21 Ostrava. Czech Republic.
e-mail: michal.holcapek@vsb.cz

Tomaš Tichý, Department of Finance, Faculty of Economics VŠB-TU Ostrava, Sokolská 33, 701 21 Ostrava. Czech Republic.
e-mail: tomas.tichy@vsb.cz