Quantile and expectile smoothing based on L1-norm and L2-norm fuzzy transforms

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:
Quantile and expectile smoothing based on L1-norm and L2-norm fuzzy transforms / Guerra, Maria Letizia; Sorini, Laerte; Stefanini, Luciano. - In: INTERNATIONAL JOURNAL OF APPROXIMATE REASONING. - ISSN 0888-613X. - STAMPA. - 107:(2019), pp. 17-43. [10.1016/j.ijar.2019.01.011]

Availability:
This version is available at: https://hdl.handle.net/11585/663247 since: 2019-12-19

Published:
DOI: http://doi.org/10.1016/j.ijar.2019.01.011

International Journal of Approximate Reasoning ••• (••••) •••–•••

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Quantile and expectile smoothing based on L_1-norm and L_2-norm fuzzy transforms

Maria Letizia Guerra [a,*], Laerte Sorini [b], Luciano Stefanini [b]

[a] Department of Statistical Sciences "Paolo Fortunati", University of Bologna, Italy
[b] Department of Economics, Society, Politics, University of Urbino Carlo Bo, Italy

ARTICLE INFO

Article history:
Received 11 July 2018
Received in revised form 23 January 2019
Accepted 23 January 2019
Available online xxxx

Keywords:
Fuzzy transform
Expectile smoothing
Quantile smoothing
Financial time series

ABSTRACT

The fuzzy transform (F-transform), introduced by I. Perfilieva, is a powerful tool for the construction of fuzzy approximation models; it is based on generalized fuzzy partitions and it is obtained by minimizing a quadratic (L_2-norm) error function. In this paper, within the discrete setting, we describe an analogous construction by minimizing an L_1-norm error function, so obtaining the L_1-norm F-transform, which is again a general approximation tool. The L_1-norm and L_2-norm settings are then used to construct two types of fuzzy-valued F-transforms, by defining expectile (L_2-norm) and quantile (L_1-norm) extensions of the transforms. This allows to model an observed time series in terms of fuzzy-valued functions, whose level-cuts can be interpreted in the setting of expectile and quantile regression. The proposed methodology is illustrated on some financial daily time series.

1. Introduction

The Fuzzy transform (F-transform) has been introduced by I. Perfilieva [32]; the books [31] and [35] contain recent and complete analysis of its possible implications for fuzzy modeling. The recent literature on F-transform is very rich, ranging from theoretical aspects [5,17,29,39,40,49,61] to applications in several fields of computational intelligence (see [10–12,30,33,34,36–38,47,48]).

In this paper we will consider the discrete F-transform. It is assumed that m points $(t_i, f_i) \in \mathbb{R}^2$, $i = 1, ..., m$, are given values $f_i = f(t_i)$ of a function f defined on a subset T of a real interval $[a, b]$ and a decomposition $\mathbb{P} = \{a = x_1 < ... < x_n = b\}$ of $[a, b]$ with n knots x_k, $k = 1, ..., n$ is given, together with a family $\mathbb{A} = \{A_1, ... A_n\}$ of n appropriate basic functions A_k forming a fuzzy partition $(\mathbb{P}, \mathbb{A})$ of $[a, b]$ (see section 2 for details).

In its simplest form, the *direct* F-transform is a vector $(F_1, ..., F_n)$ of real numbers such that each F_k is a weighted average of the function f at the points t_i belonging to the subintervals of $\mathbb{P}$ containing x_k as an extreme vertex; the value $A_k(t_i)$ of the basic function A_k is the weight assigned to f_i when computing F_k, where F_k minimizes a quadratic (L_2-norm based) weighted error function.

The *inverse* F-transform function is then obtained using an inversion formula: it is the linear combination of the basic functions A_k with coefficients given by the components F_k of the direct F-transform. The resulting function represents a good approximation of the values $f(t_i)$ and the reconstruction can be improved to arbitrary precision by refining the fuzzy partition $(\mathbb{P}, \mathbb{A})$ of $[a, b]$.

* Corresponding author.
E-mail addresses: mletizia.guerra@unibo.it (M.L. Guerra), laerte.sorini@uniurb.it (L. Sorini), lucste@uniurb.it (L. Stefanini).

https://doi.org/10.1016/j.ijar.2019.01.011
0888-613X/

By generalizing the setting of discrete F-transform, we describe the quantile (based on L_1-norm error minimization) and the expectile (based on L_2-norm error minimization) F-Transforms. The suggested methodology is similar to the asymmetric estimation technique recently adopted in Statistics to define quantile and expectile regression models. Based on some properties of the obtained (direct) F-transforms, we show that in both cases we can define, in a natural way, fuzzy-valued versions of the direct and inverse F-transforms.

Using quantile and expectile estimation models, we can obtain the intervals giving all level sets (α-cuts with $\alpha > 0$), respectively, of fuzzy-valued quantile and expectile direct F-transforms.

In the case of expectiles, based on the minimization of L_2-norm error functions, a vector of fuzzy numbers (where each F_k is now a membership function) is then obtained as our *direct* fuzzy-valued F-transform; the corresponding *inverse* fuzzy-valued F-transform is the linear combination of the basic functions A_k with the fuzzy coefficients F_k. By an analogous construction, based on the minimization of an L_1-norm error function, we obtain the quantile direct and the corresponding inverse F-transform, which is again a general approximation tool; finally, fuzzy-valued versions of the direct and inverse quantile F-transforms are obtained.

Concerning time series, smoothing is a very powerful technique to analyze their patterns and underlying trends; in this paper we illustrate the discrete F-transform obtained on a generalized fuzzy partition and we analyze its properties as a general smoothing tool, as it has been introduced in [48,49,17].

The F-transform setting appears to be a valid non-parametric methodology to describe movements in a time series; its basic properties are similar to the well-known kernel smoothing [44] or spline-based regression techniques using quantiles or expectiles ([4], [8] and [24]) or interval valued and fuzzy approaches to model time series (as in [6,23,25] and [26]). Based on the classical F-transform with generalized fuzzy partitions, expectile smoothing is obtained immediately (some preliminary results are described in [16] and specified in [54]). By the two types of quantile and expectile F-transforms, we define fuzzy-valued approximations of a time series.

In general, it is also possible to obtain fuzzy-valued approximations of a time series from quantile and expectile smoothing procedures existing in the very extended literature on quantile or expectile regression and estimation, coming from different fields such as Robust Statistics, Generalized Quantile Regression for functional data, Statistical Learning Theory, Non-parametric Smoothing and Regularization, Adaptive Semi-parametric Estimation.

We discuss on some advantages of the proposed fuzzy-valued reconstruction of time series and we show the results obtained on three financial time series, in comparison with other efficient and well tested procedures recently developed in literature.

The paper is organized into seven sections: after the introduction we recall in section 2 the basic concepts about Fuzzy transform. In section 3 we show how to use the F-transform in expectile smoothing while in section 4 we show the use of F-transform in quantile smoothing. In section 5 we show that both expectile and quantile inverse F-transform functions satisfy the non-crossing property. Section 6 presents a detailed comparison of quantile and expectile F-transforms with existing methods in Robust Statistical Regression and recent Support Vector Machine (SVM) Regression, exploring evidence from financial data. Conclusions and future research are shortly highlighted in the final section.

2. Basic elements about F-transform

A fuzzy set on the field of real numbers $\mathbb{R}$, as introduced in [60], is a mapping $u : \mathbb{R} \longrightarrow [0, 1]$. A fuzzy interval is a fuzzy set on $\mathbb{R}$ with the properties that the mapping u is *(i)* normal ($\exists \widehat{x} \in \mathbb{R}$ with $u(\widehat{x}) = 1$), *(ii)* upper semi-continuous, *(iii)* fuzzy convex ($u(\lambda x' + (1 - \lambda)x'') \geq \min\{u(x'), u(x'')\}$ for all $\lambda \in [0, 1]$), *(iv)* $cl\{x|u(x) > 0\}$ is a compact interval. A consequence of (ii) and (iii) is that the α-cuts $[u]_\alpha = \{x|u(x) \geq \alpha\} = [u^-_\alpha, u^+_\alpha]$ are compact intervals for all $\alpha \in]0, 1]$. The 1-cut is the core $[u]_1 = \{x|u(x) = 1\}$ of u; the interval $[u]_0 = cl(\{x|u(x) > 0\})$ is the 0-cut of u (some authors call it the *support* of u).

We denote by $\mathbb{R}_\mathcal{F}$ the space of real fuzzy intervals. The fundamental relationship between the mapping $u \in \mathbb{R}_\mathcal{F}$ and its α-cuts, for $\alpha \in [0, 1]$, is

$$u(x) = \begin{cases} 0 & \text{if} \quad x \notin [u]_0 \\ \sup\{\alpha | x \in [u]_\alpha\} & \text{if} \quad x \in [u]_0 \end{cases}.$$

For additional definitions and results on fuzzy numbers and intervals we will refer to the recent book [2].

Given a real interval $[a, b]$ and a decomposition of $[a, b]$ with $n \geq 2$ points, say $\mathbb{P} = \{a = x_1 < ... < x_n = b\}$, and given a finite family of fuzzy sets (in particular fuzzy numbers) $\mathbb{A} = \{A_1, ..., A_n\}$, we firstly define a fuzzy partition of $[a, b]$ by the pair $(\mathbb{P}, \mathbb{A})$; the standard F-transform (see [32]) of a continuous function $f : [a, b] \longrightarrow \mathbb{R}$ is defined in terms of a vector of real numbers $\mathbf{F} = (F_1, ..., F_n)$ (called the *direct* F-transform) where the components F_k are averages of f on the subintervals of $\mathbb{P}$, namely on $[a, x_1]$ if $k = 1$, $[x_{k-1}, x_{k+1}]$ if $k = 2, ..., n - 1$ (with $n > 2$) and $[x_{n-1}, b]$ if $k = n$, obtained by minimizing a weighted squared error (deviation) between $f(x)$ and F_k on each subinterval. The direct F-transform $\mathbf{F}$ is then used to define the iF-transform (*inverse* F-transform) function $\widehat{f} : [a, b] \longrightarrow \mathbb{R}$ and the main result is that $\widehat{f}$ is an approximating function of f on $[a, b]$ (see [32] for details).

In [47–49] a *generalized fuzzy r-partition* of $[a, b]$ is introduced, for any integer $r \geq 1$, and corresponding direct and inverse F-transforms are defined. If $r = 1$, a 1-partition coincides with the pair $(\mathbb{P}, \mathbb{A})$ of a (standard) fuzzy partition introduced in [32].

An r-partition of $[a,b]$ is defined by the following two steps (A)–(B):

(A) Choose a decomposition $\mathbb{P} = \{a = x_1 < \ldots < x_n = b\}$ of $[a,b]$ with $n \geq 2$; introduce r additional nodes $x_{-r+1} < \ldots < x_0 < a$ on the left side of $[a,b]$ and r new nodes $b < x_{n+1} < \ldots < x_{n+r}$ on the right; remark that the resulting subintervals $[x_k, x_{k+1}]$, for $k = -r+1, \ldots, n+r-1$, need not have the same length.

(B) Then, $n+2r-2$ continuous basic functions $A_k : [a,b] \longrightarrow [0,1]$ are chosen, for $k = -r+2, \ldots, n+r-1$, with the following properties:

B. 1) if $r > 1$, for $k = -r+2, \ldots, 0$, A_k is non-increasing on $\left[a, \min\left\{x_{k+r}, b\right\}\right]$, with $A_k(x_k) = 1$ and $A_k(x) = 0$ for $x \geq \min\left\{b, x_{k+r}\right\}$;

B. 2) for $k = 1, 2, \ldots, n$, A_k is obtained by eventually restricting to $[a,b]$ the membership function of a continuous fuzzy number with core $\{x_k\}$ and support $[x_{k-r}, x_{k+r}]$; in particular $A_k(x_k) = 1$ and $A_k(x) = 0$ for all $x \notin \left([x_{k-r}, x_{k+r}] \bigcap [a,b]\right)$;

B. 3) if $r > 1$, for $k = n+1, \ldots, n+r-1$, A_k is non-decreasing on $[\max\{a, x_k\}]$ with $A_k(x_k) = 1$ and $A_k(x) = 0$ for $x \leq \max\left\{a, x_{k+r}\right\}$;

B. 4) for all $x \in [a,b]$ the following condition holds

$$\sum_{k=-r+2}^{n+r-1} A_k(x) = r.$$

Remark that, on each subinterval $]x_{k-1}, x_k[$ of the decomposition $\mathbb{P}$, for $k = 2, \ldots, n$, only $2r$ basic functions $A_{k-r}(x), \ldots, A_{k+r-1}(x)$ are non-zero.

We denote a fuzzy r-partition by $(\mathbb{P}, \mathbb{A}, r)$, without explicit reference to the added nodes on the left and the right sides of interval $[a,b]$.

A family $\mathbb{A}$ satisfying the conditions in (B) can be obtained by choosing $n+r-2$ continuous functions $L_k(x)$, $k = 2, \ldots, n+r-1$ such that each $L_k(x)$ is increasing on $[x_{k-r}, x_k]$ with $L(x_{k-r}) = 0$, $L(x_k) = 1$. Then, each A_k for $x \in [a,b]$ is obtained as follows:

$$\begin{aligned}
A_k(x) &= 1 - L_{k+r}(x) \quad \text{if } x \in [a, x_{k+r}] \textstyle\bigcap [a,b] \quad \text{for } k = -r+2, \ldots, 1, \\
A_k(x) &= \begin{cases} L_k(x) & \text{if } x \in [x_{k-r}, x_k] \bigcap [a,b] \\ 1 - L_{k+r}(x) & \text{if } x \in [x_k, x_{k+r}] \bigcap [a,b] \end{cases} \quad \text{for } k = 2, \ldots, n-1, \\
A_k(x) &= L_k(x) \quad \text{if } x \in [x_{k-r}, b] \textstyle\bigcap [a,b] \quad \text{for } k = n, \ldots, n+r-1.
\end{aligned} \tag{1}$$

In forthcoming illustrative examples, we construct the basic functions A_k above by choosing a (standardized) increasing function $L : [0,1] \to [0,1]$ such that $L(0) = 0$, $L(1) = 1$ and defining each L_k by translating and rescaling L as

$$L_k(x) = L\left(\frac{x - x_{k-r}}{x_k - x_{k-r}}\right) \text{ for } x \in [x_{k-r}, x_k]. \tag{2}$$

A family of parametric standardized functions is, e.g., the following increasing rational spline (more details in [52] and [53], where also other monotonic functions L are considered)

$$L(\tau) = \frac{\tau^2 + \beta_0 \tau(1-\tau)}{1 + (\beta_0 + \beta_1 - 2)\tau(1-\tau)} \text{ for } \tau \in [0,1] \tag{3}$$

where β_0, β_1 are non-negative real numbers representing the derivatives of $L(\tau)$ at $\tau = 0$ and $\tau = 1$, respectively. By any non-negative values of the two parameters β_0, β_1, we can generate a large number of basic functions A_k as in (1) with L_k given in (2) corresponding to L as in (3). For example, the pair of parameters $\beta_0 = 2$, $\beta_1 = 0$ results in the parabolic function $L(\tau) = 2\tau - \tau^2$.

2.1. L_2-norm F-transform

Consider firstly the simple case where $r = 1$ and $(\mathbb{P}, \mathbb{A}, 1)$ is a standard partition of $[a,b]$ with n basic functions $A_1, \ldots, A_n$ and nodes $a = x_1 < x_2 < \ldots < x_n = b$. We will denote a partition simply by $(\mathbb{P}, \mathbb{A})$.

Definition 1. (from [32]) Given a continuous function $f : [a,b] \longrightarrow \mathbb{R}$ and a fuzzy partition $(\mathbb{P}, \mathbb{A})$ of $[a,b]$, the direct fuzzy transform (*F-transform*) of f with respect to $(\mathbb{P}, \mathbb{A})$ is the n-tuple of real numbers $(F_1, \ldots, F_n)$ given by

$$F_k = \frac{\int_a^b f(x) A_k(x) dx}{\int_a^b A_k(x) dx}, \quad k = 1, \ldots, n \tag{4}$$

Please cite this article in press as: M.L. Guerra et al., Quantile and expectile smoothing based on L_1-norm and L_2-norm fuzzy transforms, Int. J. Approx. Reason. (2019), https://doi.org/10.1016/j.ijar.2019.01.011

Definition 2. (from [32]) Given the direct F-transform $(F_1, \ldots, F_n)$ of a continuous function $f : [a, b] \longrightarrow \mathbb{R}$ on a fuzzy partition $(\mathbb{P}, \mathbb{A})$, the inverse F-transform (iF-transform) is the continuous function $\widehat{f}_{(\mathbb{P},\mathbb{A})} : [a, b] \longrightarrow \mathbb{R}$ given by

$$\widehat{f}_{(\mathbb{P},\mathbb{A})}(x) = \sum_{k=1}^{n} F_k A_k(x) \text{ for } x \in [a, b]. \tag{5}$$

The following approximation property is relevant in the F-transform setting.

Theorem 3. *(from [32]) If $f : [a, b] \longrightarrow \mathbb{R}$ is a continuous function then, for any positive real ε, there exists a fuzzy partition $(\mathbb{P}_\varepsilon, \mathbb{A}_\varepsilon)$ such that the associated F-transform $(F_{1,\varepsilon}, F_{2,\varepsilon}, \ldots, F_{n_\varepsilon,\varepsilon})$ and the corresponding iF-transform $\widehat{f}_{(\mathbb{P}_\varepsilon,\mathbb{A}_\varepsilon)} : [a, b] \longrightarrow \mathbb{R}$ satisfy*

$$\left| f(x) - \widehat{f}_{(\mathbb{P}_\varepsilon,\mathbb{A}_\varepsilon)}(x) \right| < \varepsilon \text{ for all } x \in [a, b].$$

The components of the F-transform solve a weighted L_2-norm minimization problem.

Theorem 4. *(from [32]) If $f : [a, b] \longrightarrow \mathbb{R}$ is a continuous function and a fuzzy partition $(\mathbb{P}, \mathbb{A})$ each component F_k of the associated F-transform $(F_1, F_2, \ldots, F_n)$ is obtained by minimizing the following quadratic function with respect to the single real variable y:*

$$\Phi_k(y) = \int_a^b |f(x) - y|^2 A_k(x)dx$$

In many applications, the function f is known or sampled at $m \in \mathbb{N}$ distinct points $t_i \in [a, b]$, $i = 1, 2, \ldots, m$ or, more generally, m observations (t_i, f_i) are available such that $f(t_i) = f_i$; a discrete version of F-transform was introduced by Perfilieva in her original paper.

Definition 5. (from [32]) Given m values $(t_i, f(t_i))$, $t_i \in [a, b]$, $i = 1, \ldots, m$, of a function $f : T \longrightarrow \mathbb{R}$ defined on a subset $T \subseteq [a, b]$ and a fuzzy partition $(\mathbb{P}, \mathbb{A})$ of $[a, b]$ such that each subinterval $[x_{k-1}, x_{k+1}]$ contains at least one point t_i in its interior (so that $\sum_{i=1}^m A_k(t_i) > 0$ for all k), the discrete direct F-transform of f with respect to $(\mathbb{P}, \mathbb{A})$ is the n-tuple of real numbers $(F_1, \ldots, F_n)$ given by

$$F_k = \frac{\sum\limits_{i=1}^{m} f(t_i) A_k(t_i)}{\sum\limits_{i=1}^{m} A_k(t_i)}, \quad k = 1, \ldots, n. \tag{6}$$

Each F_k minimizes the function $\Phi_{m,k}(y) = \sum\limits_{i=1}^{m} |f(t_i) - y|^2 A_k(t_i)$.

In the following, in view of the applications to expectile and quantile smoothing for time series, we always consider the discrete version of F-transform and its generalizations or extensions.

A first extension of F-transform is suggested in [32] and analyzed in [39,61]: the constant components F_k, representing a weighted average of the function f on $[x_{k-1}, x_{k+1}]$, are substituted by (local) polynomials of fixed degree $p \geq 1$

$$\varphi_{p,k}(x; F_{k,0}, \ldots, F_{k,p}) = F_{k,0} + F_{k,1}(x - x_k) + \frac{F_{k,2}}{2!}(x - x_k)^2 + \ldots + \frac{F_{k,p}}{p!}(x - x_k)^p. \tag{7}$$

The parameters $F_{k,j}$, $j = 0, \ldots, p$, for each k, are obtained, in analogy to Theorem 4, by minimizing the L_2-norm error function, with respect to $y_0, \ldots, y_p$,

$$\overline{\Phi}_{m,k}(y_0, \ldots, y_p) = \sum_{i=1}^{m} \left| f(t_i) - \varphi_{p,k}(t_i; y_0, \ldots, y_p) \right|^2 A_k(t_i), \text{ where} \tag{8}$$

$$\varphi_{p,k}(x; y_0, \ldots, y_p) = y_0 + y_1(x - x_k) + \frac{y_2}{2!}(x - x_k)^2 + \ldots + \frac{y_p}{p!}(x - x_k)^p.$$

The optimal solution $y_0^*, \ldots, y_p^*$ will give the components $F_{k,j}$, $j = 0, \ldots, p$.

The direct F-transform polynomials $\varphi_{p,k}(x; F_{k,0}, \ldots, F_{k,p})$ and the corresponding inverse F-transform, defined by

$$\widehat{f}^{(p)}_{(\mathbb{P},\mathbb{A})}(x) = \sum_{k=1}^{n} A_k(x)\varphi_{p,k}(x; F_{k,0}, \ldots, F_{k,p}) \text{ for } x \in [a, b], \tag{9}$$

are called F-transform of order p (the basic F-transform is then of order zero).

Starting with a generalized r-partition $(\mathbb{P}, \mathbb{A}, r)$, introduced in [49], and following the same ideas as in [32], the L_2-norm F-transform of order $p \geq 0$ on $(\mathbb{P}, \mathbb{A}, r)$ is defined in a similar way: the direct F-transform is an $(n+2r-2)$-tuple of polynomials $(\varphi_{p,2-r}(x), \ldots, \varphi_{p,n+r-1}(x))$ and the corresponding inverse F-transform is given by the expression

$$\widehat{f}^{(p)}_{(\mathbb{P},\mathbb{A},r)}(x) = \frac{1}{r} \sum_{k=2-r}^{n+r-1} A_k(x)\varphi_{p,k}(x; F_{k,0}, \ldots, F_{k,p}) \text{ for } x \in [a,b], \tag{10}$$

where the parameters $F_{k,j}$, $j = 0, \ldots, p$, for $k = 2-r, \ldots, n+r-1$, are obtained as above, and if the order is $p = 0$ simply by

$$\widehat{f}_{(\mathbb{P},\mathbb{A},r)}(x) = \frac{1}{r} \sum_{k=2-r}^{n+r-1} F_k A_k(x) \text{ for } x \in [a,b] \tag{11}$$

with the $(n+2r-2)$-tuple of the direct F-transform $(F_{2-r}, \ldots, F_{n+r-1})$.

In subsection 2.3 we will describe a general matrix form of the minimization problems that are required to be solved in order to compute the coefficients $F_{k,j}$ $(j = 0, 1, \ldots, p)$ of the polynomial components $\varphi_{p,k}(x)$ $(k = 2-r, \ldots, n+r-1)$.

2.2. L_1-norm F-transform

For some fuzzy r-partition $(\mathbb{P}, \mathbb{A}, r)$ of $[a,b]$ and a function $f : [a,b] \longrightarrow \mathbb{R}$, we have seen that the direct F-transform of f, as in the paper [32], is such that each F_k minimizes the function $\Phi_k(y) = \int_a^b |f(x) - y|^2 A_k(x)dx$ with $y \in \mathbb{R}$; $\Phi_k(y)$ can be interpreted as the integral (weighted) L_2-norm of the error $f(x) - y$, restricted on the support of the basic function A_k (assuming that the integrals exist).

We are now interested in generalizing this construction by considering the L_1-norm; in particular we obtain the components of the L_1-norm direct F-transform, denoted them by G_k, as the minimizers of the integral (weighted) L_1-norm of the error $f(x) - y$, given by

$$\Psi_k(y) = \int_a^b |f(x) - y|\, A_k(x)dx \text{ with } y \in \mathbb{R}. \tag{12}$$

Correspondingly, the L_1-norm inverse F-transform is obtained as the analogous inverse F-transform with the components F_k substituted by the new components G_k.

Remark that, in general, G_k may not be unique as in fact there may exist an interval of values $y^* \in \mathbb{R}$ with the same minimal value $\Psi_k(y^*)$.

According to the ideas above, the L_1-norm direct and inverse F-transform are defined as follows.

Definition 6. An L_1-norm direct F-transform of a function $f : [a,b] \to \mathbb{R}$ on the r-partition $(\mathbb{P}, \mathbb{A}, r)$ is any $(n+2r-2)$-tuple of real numbers $(G_{2-r}, \ldots, G_{n+r-1})$, where each G_k minimizes the function Ψ_k in (12).

Definition 7. Given the L_1-norm direct F-transform $(G_{2-r}, \ldots, G_{n+r-1})$ of a function $f : [a,b] \to \mathbb{R}$ on the r-partition $(\mathbb{P}, \mathbb{A}, r)$, the corresponding L_1-norm inverse F-transform of f is the function $\widetilde{f}_{(\mathbb{P},\mathbb{A},r)} : [a,b] \to \mathbb{R}$, given by

$$\widetilde{f}_{(\mathbb{P},\mathbb{A},r)}(x) = \frac{1}{r} \sum_{k=2-r}^{n+r-1} G_k A_k(x), x \in [a,b]. \tag{13}$$

We can see that the same approximation property in Theorem 3 for the L_2-norm F-transform, is valid also for the L_1-norm F-transform. For the sake of simplicity, we consider a uniform partition $(\mathbb{P}, \mathbb{A})$, i.e., $x_k - x_{k-1} = h$ for all k, and $r = 1$.

Theorem 8. *If $f : [a,b] \longrightarrow \mathbb{R}$ is a continuous function then, for any positive real ε, there exists a fuzzy partition $(\mathbb{P}_\varepsilon, \mathbb{A}_\varepsilon)$ such that the associated L_1-norm F-transform $(G_{1,\varepsilon}, G_{2,\varepsilon}, \ldots, G_{n_\varepsilon,\varepsilon})$ and the corresponding inverse L_1-norm F-transform $\widetilde{f}_{(\mathbb{P}_\varepsilon,\mathbb{A}_\varepsilon)} : [a,b] \longrightarrow \mathbb{R}$ satisfy*

$$\left| f(x) - \widetilde{f}_{(\mathbb{P}_\varepsilon,\mathbb{A}_\varepsilon)}(x) \right| < \varepsilon \text{ for all } x \in [a,b].$$

Please cite this article in press as: M.L. Guerra et al., Quantile and expectile smoothing based on L_1-norm and L_2-norm fuzzy transforms, Int. J. Approx. Reason. (2019), https://doi.org/10.1016/j.ijar.2019.01.011

Proof. Consider a uniform partition $(\mathbb{P},\mathbb{A})$ of $[a,b]$ such that $x_{j+1}-x_j=h$ for $j=1,\dots,n-1$. Considering the continuity of f on $[a,b]$, define

$$m_1=\min\{f(x)\mid x\in[a,x_2]\},\ M_1=\max\{f(x)\mid x\in[a,x_2]\}$$

$$m_j=\min\left\{f(x)\mid x\in[x_{j-1},x_{j+1}]\right\}\ \textit{for}\ j=2,\dots,n-1$$

$$M_j=\max\left\{f(x)\mid x\in[x_{j-1},x_{j+1}]\right\}\textit{for}\ j=2,\dots,n-1$$

$$m_n=\min\{f(x)\mid x\in[x_{n-1},b]\},\ M_n=\max\{f(x)\mid x\in[x_{n-1},b]\}.$$

Consider $x\in[x_k,x_{k+1}]$ for a given $k=1,\dots,n-1$; from Definition 6 we have that $G_k\in[m_k,M_k]$ and $G_{k+1}\in[m_{k+1},M_{k+1}]$ so that

$$|f(x)-G_k|\le M_k-m_k \text{ and } \left|f(x)-G_{k+1}\right|\le M_{k+1}-m_{k+1}$$

From Definition 7, we have that $\widetilde{f}_{(\mathbb{P},\mathbb{A})}(x)=G_kA_k(x)+G_{k+1}A_{k+1}(x)$ and, considering that $A_k(x)\ge 0$, $A_{k+1}(x)\ge 0$ and $A_k(x)+A_{k+1}(x)=1$ for all x, we have

$$\begin{aligned}\left|f(x)-\widetilde{f}_{(\mathbb{P},\mathbb{A})}\right| &= \left|f(x)\left(A_k(x)+A_{k+1}(x)\right)-G_kA_k(x)-G_{k+1}A_{k+1}(x)\right|\\ &\le |G_k-f(x)|\,A_k(x)+\left|G_{k+1}-f(x)\right|A_{k+1}(x)\\ &\le |f(x)-G_k|+\left|f(x)-G_{k+1}\right|\le M_k-m_k+M_{k+1}-m_{k+1}.\end{aligned}$$

We can choose $h=h_\varepsilon$ and $n=n_\varepsilon$ such that $\max\left\{M_j-m_j\mid j=1,\dots,n\right\}<\frac{\varepsilon}{2}$. For the uniform partition $(\mathbb{P}_\varepsilon,\mathbb{A}_\varepsilon)$ obtained in the mentioned way, it follows that the inequality

$$\left|f(x)-\widetilde{f}_{(\mathbb{P}_\varepsilon,\mathbb{A}_\varepsilon)}(x)\right|<\varepsilon$$

is satisfied for all $x\in[a,b]$. □

Starting with a generalized r-partition $(\mathbb{P},\mathbb{A},r)$ and following the same ideas as in subsection 2.1, the L_1-norm F-transform of order $p\ge 0$ on $(\mathbb{P},\mathbb{A},r)$ can be defined. The constant components G_k are substituted by polynomials of fixed degree $p\ge 1$, $k=2-r,\dots,n+r-1$,

$$\varphi_{p,k}(x;G_{k,0},\dots,G_{k,p})=G_{k,0}+G_{k,1}(x-x_k)+\frac{G_{k,2}}{2!}(x-x_k)^2+\dots+\frac{G_{k,p}}{p!}(x-x_k)^p \tag{14}$$

and the parameters $G_{k,j}$, $j=0,\dots,p$ for each k, are estimated by minimizing the L_1-norm error function

$$\Psi_k^{(p)}(y_0,\dots,y_p)=\int_a^b\left|f(x)-\varphi_{p,k}(x;y_0,\dots,y_p)\right|A_k(x)dx,$$

or, in the discrete case with values $(t_i,f(t_i))$, $t_i\in[a,b]$, $i=1,\dots,m$, by minimizing the absolute deviation

$$\overline{\Psi}_{m,k}(y_0,\dots,y_p)=\sum_{i=1}^{m}\left|f(t_i)-\varphi_{p,k}(t_i;y_0,\dots,y_p)\right|A_k(t_i),$$

where $\varphi_{p,k}(x;y_0,\dots,y_p)=y_0+y_1(x-x_k)+\frac{y_2}{2!}(x-x_k)^2+\dots+\frac{y_p}{p!}(x-x_k)^p$ is our (local) polynomial of order p with coefficients $y_0,\dots,y_p$.

The optimal solution $y_0^*,\dots,y_p^*$ gives the components $G_{k,j}$, $j=0,\dots,p$.

The direct L_1-norm F-transform polynomials $\varphi_{p,k}(x;G_{k,0},\dots,G_{k,p})$ and the corresponding inverse F-transform, defined by

$$\widetilde{f}^{(p)}_{(\mathbb{P},\mathbb{A},r)}(x)=\frac{1}{r}\sum_{k=2-r}^{n+r-1}A_k(x)\varphi_{p,k}(x;G_{k,0},\dots,G_{k,p})\ \textit{for}\ x\in[a,b],$$

are called L_1-norm F-transform of order $p\ge 0$.

2.3. Computation of L_2-norm and L_1-norm F-transforms

In this section we describe a matrix notation for the direct and inverse F-transforms in the discrete case.

As in the previous subsections, a function f, defined on a subset T of a compact interval $[a,b]$ is given, and a fuzzy r-partition $(\mathbb{P},\mathbb{A},r)$ of $[a,b]$ is selected with the usual notation. Given m distinct points $t_j \in [a,b]$, $j=1,\dots,m$, such that each set $\mathbb{T}_k = \{t_j | A_k(t_j) > 0\}$, $k = 2-r,\dots,n+r-1$, is nonempty (assuming $t_1 < t_2 < \dots < t_m$, we say in this case that $\mathbb{T} = \{t_1,t_2,\dots,t_m\}$ is *sufficiently dense* with respect to $(\mathbb{P},\mathbb{A},r)$), the discrete direct L_1-norm F-transform of order p is obtained, for all k, by minimizing

$$\overline{\Psi}_{m,k}(y_0,\dots,y_p) = \sum_{j=1}^{m} \left| f(t_j) - \varphi_{p,k}(t_j; y_0,\dots,y_p) \right| A_k(t_j) \tag{15}$$

and the discrete direct L_2-norm F-transform of order p is obtained by minimizing

$$\overline{\Phi}_{m,k}(y_0,\dots,y_p) = \sum_{j=1}^{m} \left| f(t_j) - \varphi_{p,k}(t_j; y_0,\dots,y_p) \right|^2 A_k(t_j). \tag{16}$$

Remark that the form of the (local) polynomials of order $p \geq 0$ to generate the components of the direct F-transform, is the same $\varphi_{p,k}(x; y_0,\dots,y_p)$ (for all x and all $y_0,\dots,y_p$) for both the L_1-norm and the L_2-norm cases.

The two optimization problems (15) or (16) can be usefully formulated by using the same matrix notation, as follows.

We introduce the m-dimensional column vector ($(\cdot)^T$ denotes transposition) $\mathbf{f} = (f_1,\dots,f_m)^T$ with components $f_j = f(t_j)$ obtained by the observed values of function f at points $t_j \in [a,b]$; the values of each basic function A_k of the fuzzy r-partition $(\mathbb{P},\mathbb{A},r)$, evaluated at the points $t_j \in [a,b]$, are arranged into a diagonal matrix $\mathbf{W}_k$ of order m

$$\mathbf{W}_k = diag[A_k(t_1), A_k(t_2), \dots, A_k(t_m)];$$

the $p+1$ variables to be determined for each k, are arranged into the column vector $\mathbf{y} = (y_0,\dots,y_p)^T$ and the following matrices with m rows and $p+1$ columns are generated for each k:

$$\mathbf{X}_k = \begin{bmatrix} 1 & t_1 - x_k & \frac{1}{2}(t_1-x_k)^2 & \dots & \frac{1}{p!}(t_1-x_k)^p \\ 1 & t_2 - x_k & \frac{1}{2}(t_2-x_k)^2 & \dots & \frac{1}{p!}(t_2-x_k)^p \\ . & . & . & \dots & . \\ . & . & . & \dots & . \\ 1 & t_m - x_k & \frac{1}{2}(t_m-x_k)^2 & \dots & \frac{1}{p!}(t_m-x_k)^p \end{bmatrix}.$$

According to the notation, the m terms $\varphi_{p,k}(t_i;\mathbf{y})$ for fixed k appearing in $\overline{\Psi}_{m,k}(\mathbf{y})$ or in $\overline{\Phi}_{m,k}(\mathbf{y})$ are the components of the product $\mathbf{X}_k\mathbf{y}$, i.e.,

$$\overline{\Psi}_{m,k}(\mathbf{y}) = \|\mathbf{W}_k(\mathbf{f} - \mathbf{X}_k\mathbf{y})\|_{L_1}, \text{ and}$$
$$\overline{\Phi}_{m,k}(\mathbf{y}) = \left\|\mathbf{W}_k^{1/2}(\mathbf{f} - \mathbf{X}_k\mathbf{y})\right\|_{L_2}$$

where $\mathbf{W}_k^{1/2}$ is the square root of the (non-negative) diagonal matrix $\mathbf{W}_k$ and $\|\mathbf{g}\|_{L_1} = \sum_{j=1}^m |g_j|$, $\|\mathbf{g}\|_{L_2} = \sum_{j=1}^m |g_j|^2$ are the usual norms for real vectors.

Remark 9. For implementation of $\mathbf{W}_k$ and $\mathbf{X}_k$ it is sufficient to consider elements t_j where $A_k(t_j) > 0$, i.e. only $t_j \in]x_{k-r}, x_{k+r}[$. If m_k denotes the cardinality of set $\mathbb{T}_k$ defined above, then $\mathbf{W}_k$ is a square matrix of order m_k and $\mathbf{X}_k$ is a rectangular matrix with m_k rows and $p+1$ columns.

The minimization of $\overline{\Psi}_{m,k}(\mathbf{y})$ and $\overline{\Phi}_{m,k}(\mathbf{y})$ are well known problems in regression analysis: the first is a (weighted) linear Least Absolute Deviation (LAD) problem (see e.g., [42]) and the second is a linear Least Squares (LS) problem. From their solution, there exist available several very efficient computational procedures; some details are given in sections 3 and 4.

We end this section with an example, to show a first comparison of L_1-norm and L_2-norm F-transform.

Example 10. Consider the function (as in [5]) $f : [0,10] \to R$ defined by $f(x) = x(10-x)sin(x^2)$, $x \in [0,10]$. The runs are executed with $m = 2001$ perturbed values $f_j = f(t_j) + e_j$ where the points t_j are uniform on $[0,10]$ and e_j are random numbers from the standard normal distribution $N(0,1)$; we compare two cases for the number n of basic functions $n = 81$ and $n = 161$; in both cases, the value of the bandwidth, obtained by GCV (see [49]), is $r = 1$. The mean absolute variation of the data $\{f_j | j = 1,\dots,m\}$, defined by $MAV(f) = \frac{1}{m-1}\sum_{j=1}^{m-1}|f_{j+1} - f_j|$, is $MAV(f) = 1.2612$. For the two values of n,

Table 1
Results for L_1-norm and L_2-norm F-transform.

Cases		$\widetilde{f_{L_1}}$			$\widehat{f_{L_2}}$		
		MAV	MSE	$MAVR$	MAV	MSE	$MAVR$
$n = 81$	$p = 1$	0.5507	2.43	0.437	0.5101	2.40	0.404
$n = 81$	$p = 3$	0.5566	0.91	0.441	0.5360	0.88	0.425
$n = 161$	$p = 1$	0.5516	0.98	0.437	0.5443	0.94	0.432
$n = 161$	$p = 3$	0.5510	0.85	0.437	0.5457	0.79	0.433

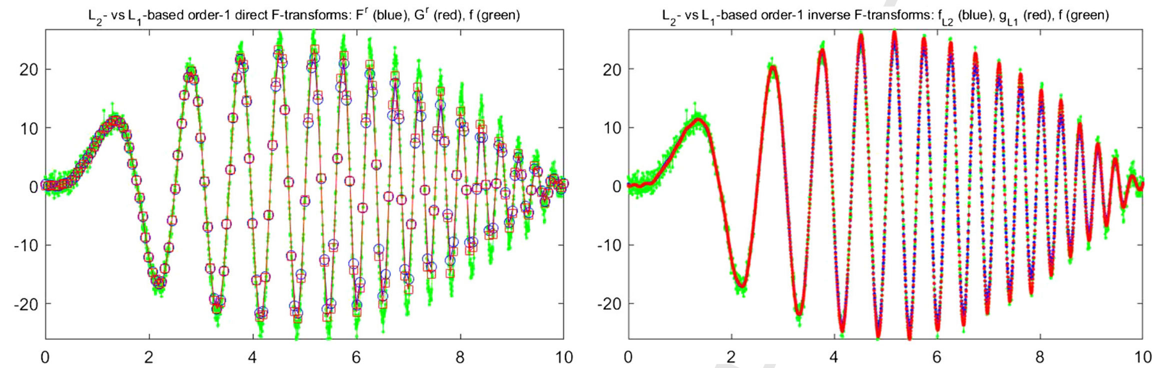

Fig. 1. (Example 10) Left: L_1 (red, squares) and L_2 (blue, circles) direct F-transform components $G_{k,0}$ and $F_{k,0}$; Right: L_1 (red) and L_2 (blue) inverse F-transform functions. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

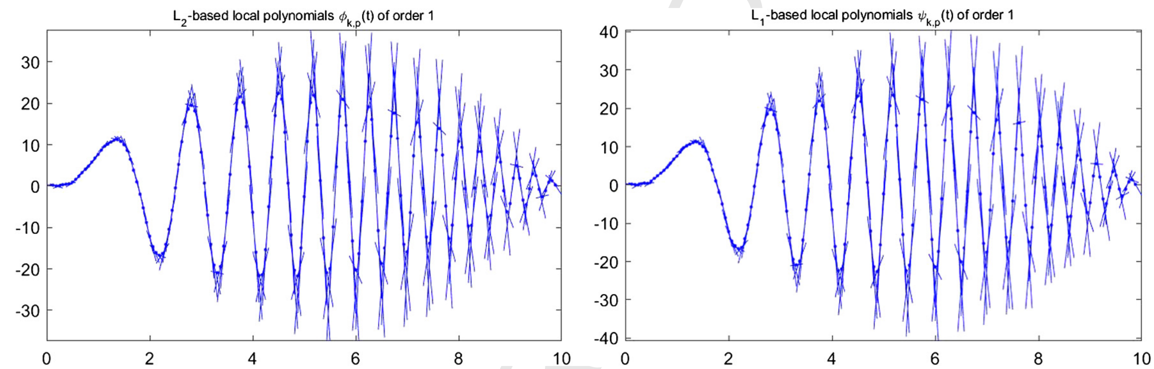

Fig. 2. (Example 10) Right: Local 1st order polynomials (lines) of the L_1-norm based direct F-transform; Left: Local 1st order polynomials (lines) of the L_2-norm based direct F-transform.

the L_1-norm iF-transform $\widetilde{f}$ and the L_2-norm iF-transform $\widehat{f}$ are obtained for two orders $p = 1$ and $p = 3$. In Table 1 we report the average variation $V_m(\widetilde{f_{L_1}})$ and the mean squared error (residual) $MSE(\widetilde{f_{L_1}}) = \frac{1}{m}\sum_{j=1}^{m} |f_j - \widetilde{f}_j|^2$ (similarly for the L_2-norm iF-transform $\widehat{f}$). The degree of smoothness, obtained by the iF-transforms, can be measured, e.g., by the mean absolute variation ratio $MAVR(\widetilde{f_{L_1}}) = \frac{MAV(\widetilde{f_{L_1}})}{MAV(f)}$ (similarly by $MAVR(\widehat{f_{L_2}}) = \frac{MAV(\widehat{f_{L_2}})}{MAV(f)}$).

In terms of the reported indicators, the L_1-norm iF-transform $\widetilde{f}$ and the L_2-norm iF-transform $\widehat{f}$ perform similarly. Fig. 1 pictures the components of direct F-transform and the inverse F-transform of types L_1 and L_2 and order $p = 1$ for $n = 81$; Fig. 2 plots the local polynomials of first order (lines) corresponding to the L_1-norm and L_2-norm direct F-transforms, according to equation (7).

The two figures show a very similar behavior for the L_1 and the L_2 F-transform smoothing approximations (at least for this example).

3. L_2-norm fuzzy-valued F-transform in expectile smoothing

In order to investigate the role of F-transform in smoothing, let us first introduce some basic facts on expectile regression, a recent interesting field in non-parametric regression (see e.g. [8], [43], [58], [59] and the references therein).

In applied statistics, given a set of m observations (values of a function $f : [a, b] \to \mathbb{R}$) $f_j = f(t_j)$ with $j = 1, ..., m$, where $t_j \in [a, b]$, the expectiles $\mu(\omega)$, for $\omega \in]0, 1]$, are considered with respect to the set of values $\{f_j \mid j = 1, .., m\}$ and defined

Please cite this article in press as: M.L. Guerra et al., Quantile and expectile smoothing based on L_1-norm and L_2-norm fuzzy transforms, Int. J. Approx. Reason. (2019), https://doi.org/10.1016/j.ijar.2019.01.011

by tail expectations: for a given value of $\omega \in]0,1]$, the sample expectile $\mu(\omega)$ is obtained by minimizing the following so-called least asymmetrical weighted squares (LAWS) function

$$S_\omega(\mu) = \sum_{j=1}^{m} w_j(\omega;\mu)\left(f_j - \mu\right)^2, \tag{17}$$

where the weights are

$$w_j(\omega;\mu) = \begin{cases} \omega & \text{if} \quad f_j > \mu \\ 1-\omega & \text{if} \quad f_j \le \mu \end{cases}.$$

If $\omega = \frac{1}{2}$ we obtain the mean value μ_e of the observations

$$\mu_e = \arg\min_\mu \left(S_{\frac{1}{2}}(\mu) = \frac{1}{2}\sum_{j=1}^{m}\left(f_j - \mu\right)^2 \right).$$

The value $\mu = \mu(\omega)$ (depending on ω) is the population expectile for different values of the asymmetry parameter $\omega \in]0,1]$.

Consider now the F-transform of order zero ($p = 0$). The expectile fuzzy-valued F-transform, for a fixed r-partition $(\mathbb{P}, \mathbb{A}, r)$ and according to the expectiles setting described above, is defined by using the minimizers of the following strictly convex functions, for $k = -r+2, \ldots, n+r-1$ and $\omega \in]0,1[$,

$$\Phi_{k,\omega}(\mu) = \sum_{j=1}^{m} w_j(\omega;\mu)\left(f_j - \mu\right)^2 A_k\left(t_j\right). \tag{18}$$

The minimization in (18) can be solved by applying the following iterated least asymmetrical weighted squares algorithm (see [28,59]).

Iterated LAWS algorithm for expectiles: *Given: m observations (t_j, f_j) with $t_j \in T \subseteq [a,b]$, $j = 1, \ldots, m$; a generalized r-partition $(\mathbb{P}, \mathbb{A}, r)$ of $[a,b]$; a value $k \in \{-r+2, \ldots, n+r-1\}$ and a value $\omega \in]0,1[$, find the real value $\mu_k^*(\omega)$ that minimizes the function $\Phi_{k,\omega}$ in equation (18).*

- **Step 0 (Initialization)**
  Choose a positive small tolerance $\delta > 0$ to use as a convergence test (e.g., $\delta = 0.00001$).
  Choose a positive integer *mit* to use as a maximum number of iterations (e.g., $mit = 100$).
  Set $l = 0$ to count the number of performed iterations.
  Choose $w_j^{(0)}$, $j = 1, \ldots, m$ as initial estimates of weights (e.g., $w_j^{(0)} = \frac{1}{2}$ for all j).
- **Step 1 (Solve minimization at iteration *l*)**
  Compute the value $\mu^{(l)}$ that minimizes $\sum_{j=1}^{m} w_j^{(l)}\left(f_j - \mu\right)^2 A_k(t_j)$ with respect to μ, i.e.,

$$\mu^{(l)} = \frac{\sum_{j=1}^{m} w_j^{(l)} A_k(t_j) f_j}{\sum_{j=1}^{m} w_j^{(l)} A_k(t_j)}.$$

- **Step 2 (Update weights for next iteration)**
  Compute the new weights as

$$w_j^{(l+1)} = \begin{cases} \omega & \text{if} \quad f_j > \mu^{(l)} \\ 1-\omega & \text{if} \quad f_j \le \mu^{(l)} \end{cases} \text{ for } j = 1, \ldots, m$$

- **Step 3 (Test if weights are stable)**
  Compare $w_j^{(l+1)}$ with $w_j^{(l)}$; if

$$\left| w_j^{(l+1)} - w_j^{(l)} \right| < \delta \text{ for all } j = 1, \ldots, m$$

  then declare the set of current weights $\left\{ w_j^{(l+1)} | j = 1, \ldots, m \right\}$ as stable.

- **Step 4 (Test convergence or termination)**
  If the weights $\left\{w_j^{(l+1)} | j=1,\dots,m\right\}$ are stable or if $l=mit$, then compute the final solution as

$$\mu_k^*(\omega)=\frac{\sum_{j=1}^m w_j^{(l+1)}A_k(t_j)f_j}{\sum_{j=1}^m w_j^{(l+1)}A_k(t_j)}$$

  and stop the procedure.
  Otherwise, increase counter l by one and continue with steps 1 to 4.

The iterated LAWS algorithm alternates weighted least squares minimization (Step 1) and weights updating (Step 2) until the weights in two consecutive iterations do not change significantly (within the chosen tolerance δ). The loss function (18) is continuously differentiable and convex with respect to μ and the algorithm is guaranteed to converge to the existing unique solution (see e.g. [59], Section 2.1). In practice, five to ten iterations are usually sufficient (see also [43], Section 3).

Remark that if $\omega=0.5$, the minimization of $\Phi_{k,0.5}(\mu)$ with respect to μ is obtained in closed form as

$$\mu_k^*(0.5)=\frac{\sum_{j=1}^m A_k(t_j)f_j}{\sum_{j=1}^m A_k(t_j)}.$$

The following well known result (see, e.g., [59]) allows the achievement of the construction.

Proposition 11. *Consider the (unique) minimizing values $\mu_k^*(\omega)$ of $\Phi_{k,\omega}(\mu)$ for $\omega\in]0,1[$; then $\mu_k^*(\cdot)$, as a function of ω, is non-decreasing, i.e.,*

$$\omega'>\omega''\Longrightarrow\mu_k^*(\omega')\geq\mu_k^*(\omega''). \tag{19}$$

The monotonicity of functions $\mu_k^*:]0,1[\longrightarrow\mathbb{R}$ for all $k=-r+2,\dots,n+r-1$ ensures the existence of the following functions $\mu_k:]0,1[\longrightarrow\mathbb{R}$, defined by the left limits

$$\mu_k(\omega)=\lim_{\delta\downarrow 0}\mu_k^*(\omega-\delta)\text{ for all }\omega\in]0,1[\,. \tag{20}$$

It is well known that each function μ_k is left-continuous and non-decreasing on $]0,1[$ (see [41], Ch. 4).

Proposition 12. *Let $\{\mu_k(\omega)\,|\omega\in]0,1[\}$ be the set of values (from the minimizers of $\Phi_{k,\omega}(\mu)$) as in (20); consider $\alpha\in[0,1]$ and define the following compact intervals*

$$U_{k,\alpha}=\begin{cases}\left\{\mu_k\left(\frac{1}{2}\right)\right\} & \text{if}\quad \alpha=1\\ \left[\mu_k\left(\frac{\alpha}{2}\right),\mu_k\left(1-\frac{\alpha}{2}\right)\right] & \text{if}\quad \alpha\in]0,1[\\ cl\left(\bigcup\limits_{\beta>0}U_{k,\beta}\right) & \text{if}\quad \alpha=0\end{cases}. \tag{21}$$

Then, for each $k=2-r,\dots,n+r-1$, the family of intervals $\{U_{k,\alpha};\alpha\in[0,1]\}$ defines the α-cuts of a fuzzy number $F_k\in\mathbb{R}_{\mathcal{F}}$ having membership function

$$F_k(x)=\begin{cases}\sup\{\alpha|x\in U_{k,\alpha}\} & \text{if}\quad x\in U_{k,0}\\ 0 & \text{if}\quad x\notin U_{k,0}\end{cases}. \tag{22}$$

Proof. Consider a fixed value of k. We apply the characterization theorem of Negoita–Ralescu (see [2], Theorem 4.8). Denote for convenience $U_{k,\alpha}=[\underline{u}_{k,\alpha},\overline{u}_{k,\alpha}]$.

(i) $U_{k,\alpha}$ is a closed interval for all $\alpha\in[0,1]$: this is obvious from (21).

(ii) $\alpha'<\alpha''\Longrightarrow U_{k,\alpha''}\subseteq U_{k,\alpha'}$ for all $\alpha',\alpha''\in[0,1]$: this follows from (19) in Proposition 11. Indeed, with respect to α, $\mu_k\left(\frac{\alpha}{2}\right)$ is increasing and $\mu_k\left(1-\frac{\alpha}{2}\right)$ is decreasing; consequently, if $\alpha'<\alpha''$ we have $\underline{u}_{k,\alpha'}\leq\underline{u}_{k,\alpha''}$ and $\overline{u}_{k,\alpha''}\leq\overline{u}_{k,\alpha'}$, i.e., $U_{k,\alpha''}\subseteq U_{k,\alpha'}$.

(iii) Let $\alpha\in]0,1]$ be fixed and let α_n be any increasing sequence with $\lim\limits_{n\longrightarrow\infty}\alpha_n=\alpha$; from $\alpha_{n+1}\geq\alpha_n$ it follows that $U_{k,\alpha_{n+1}}\subseteq U_{k,\alpha_n}$ and the sequence of nested intervals $(U_{k,\alpha_n})_{n\in\mathbb{N}}$ is decreasing and consequently (section 6.3 in [27]) it has a limit and $\lim\limits_{n\to\infty}U_{k,\alpha_n}=\bigcap\limits_{n=1}^{\infty}U_{k,\alpha_n}$; from the left continuity of μ_k we then also have that $\lim\limits_{n\to\infty}U_{k,\alpha_n}=U_{k,\alpha}$.

(iv) From the definition of U_0, we have that for any convergent decreasing sequence β_n with $\lim_{n\longrightarrow\infty}\beta_n = 0$, $\beta_{n+1} \le \beta_n$ (so that $U_{\beta_n} \subseteq U_{\beta_{n+1}}$) the sequence of intervals $(U_{\beta_n})_{n\in\mathbb{N}}$ is increasing and $cl\left(\bigcup_{n=1}^{\infty} U_{\beta_n}\right) = U_0$.
Clearly, F_k is normal with $core(F_k) = U_{k,1}$ and F_k is compactly supported, indeed $cl(supp(F_k)) = U_{k,0}$. □

Based on the last proposition, the discrete L_2-norm iF-transform of f can be fuzzified to obtain a fuzzy-valued function.

Definition 13. *Given a set of m points $Y_m = \{(t_i, f(t_i));\ i = 1, \dots, m\}$ of a function $f : T \longrightarrow \mathbb{R}$ with $t_i \in T \subseteq [a, b]$ and given a fuzzy r-partition $(\mathbb{P}, \mathbb{A}, r)$ of $[a, b]$, the $(n + 2r - 2)$-vector of fuzzy numbers*

$$\mathbf{F}_{(\mathbb{P},\mathbb{A},r)} = (F_{-r+2}, \dots, F_{n+r-1}),$$

where each F_k is given by (22) in Proposition 12, is called the discrete *direct expectile fuzzy transform of* f with respect to $(\mathbb{P}, \mathbb{A}, r)$, based on the data-set Y_m.
The corresponding *inverse expectile fuzzy transform* of f is the fuzzy-valued function defined by

$$\widehat{f}_{(\mathbb{P},\mathbb{A},r)}(x) = \frac{1}{r}\sum_{k=2-r}^{n+r-1} F_k A_k(x) \text{ for } x \in T. \tag{23}$$

In Definition 13 and in the rest of the paper we always assume that the points $t_i \in [a, b]$, $i = 1, 2, \dots, m$, are sufficiently dense with respect to the fuzzy partition $(\mathbb{P}, \mathbb{A}, r)$.

The α-cuts of F_k will be denoted by

$$[F_k]_\alpha = \left[F^-_{k,\alpha}, F^+_{k,\alpha}\right];$$

then, considering that each basic function A_k has non-negative values on $[a, b]$, it follows that the α-cuts of $\widehat{f}_{(\mathbb{P},\mathbb{A},r)}(x)$ are

$$\left[\widehat{f}_{(\mathbb{P},\mathbb{A},r)}(x)\right]_\alpha = \left[\frac{1}{r}\sum_{k=2-r}^{n+r-1} F^-_{k,\alpha} A_k(x), \frac{1}{r}\sum_{k=2-r}^{n+r-1} F^+_{k,\alpha} A_k(x)\right]. \tag{24}$$

When $\alpha = 1$ we obtain the standard (crisp) iF-transform, corresponding to the core of the expectile fuzzy-valued iF-transform (24); indeed, we have, by construction, $F^-_{k,1} = F^+_{k,1}$ for all k.

4. L_1-norm fuzzy-valued F-transform in quantile smoothing

Given a set of m observations $f_j = f(t_j)$ with $j = 1, \dots, m$, where $t_j \in [a, b]$, the quantiles $q(\omega)$, for $\omega \in]0, 1]$, are considered with respect to the population $\{f_j \mid j = 1, .., m\}$ and can be obtained as the solution to the minimization (with respect to q) of the function (see, e.g., [7,19,21,22])

$$Q_\omega(q) = \sum_{j=1}^{m} p_j(\omega; q)\,|f_j - q|, \tag{25}$$

where the weights are

$$p_j(\omega; q) = \begin{cases} \omega & \text{if } f_j > q \\ 1-\omega & \text{if } f_j \le q \end{cases}.$$

We also have

$$Q_\omega(q) = (1-\omega)\sum_{f_j<q}(q - f_j) + \omega\sum_{f_j>q}(f_j - q)$$

and it is immediate to see that $Q_\omega(q) \ge 0$ for all real q.

If $\omega = \frac{1}{2}$, then the quantile gives the median m_e of the values $\{f_j | j = 1, \dots, m\}$, i.e. the following minimizer

$$m_e = \arg\min_{q\in\mathbb{R}} \sum_{j=1}^{m} |f_j - q|.$$

Please cite this article in press as: M.L. Guerra et al., Quantile and expectile smoothing based on L_1-norm and L_2-norm fuzzy transforms, Int. J. Approx. Reason. (2019), https://doi.org/10.1016/j.ijar.2019.01.011

Remark 14. As it is well known, function $Q_\omega(q)$ is convex with respect to q but, in general, not strictly convex; as a consequence (see [1], Ch. 8), the optimal set $\arg\min_{q\in\mathbb{R}} Q_\omega(q)$ of all minimizers of Q_ω is a nonempty closed and convex set, namely an interval $[q^L_\omega, q^R_\omega]$, with $q^L_\omega \le q^R_\omega$.

The quantile fuzzy-valued F-transform, for a fixed fuzzy r-partition $(\mathbb{P}, \mathbb{A}, r)$ can be defined, according to the quantile setting, by using the minimizers of the following convex functions, for $k = -r+2, \dots, n+r-1$ and $\omega \in]0,1]$,

$$\Psi_{k,\omega}(\eta) = \sum_{j=1}^{m} p_j(\omega;\eta)\,|f_j - \eta| A_k(t_j)\,. \tag{26}$$

The minimization of $\Psi_{k,\omega}(\eta)$ with respect to η, for fixed k and ω, can be obtained by solving the linear programming problem described with the following three steps.

LP minimization algorithm for quantiles: *Given m observations (t_j, f_j) with $t_j \in T \subseteq [a,b]$, $j = 1, \dots, m$; a generalized r-partition $(\mathbb{P}, \mathbb{A}, r)$ of $[a,b]$; a value $k \in \{-r+2, \dots, n+r-1\}$ and a value $\omega \in]0,1[$, find a real value $\eta^*_k(\omega)$ minimizes the function $\Psi_{k,\omega}$ in equation (26).*

Step 1 (Define the variables and the constraints of the LP problem)
Introduce $2m$ non-negative new variables y^-_j, y^+_j, $j = 1, 2, \dots, m$, related to η and the set $\{f_j | j = 1, \dots, m\}$ by

$$y^-_j = \begin{cases} 0 & \text{if } f_j \ge \eta \\ \eta - f_j & \text{if } f_j < \eta \end{cases} \text{ and } y^+_j = \begin{cases} 0 & \text{if } f_j \le \eta \\ f_j - \eta & \text{if } f_j > \eta \end{cases};$$

then, the following identities $y^-_j - y^+_j = \eta - f_j$ will hold for all $j = 1, 2, \dots, m$.
The $2m+1$ variables of the LP problem will be y^-_j and y^+_j, $j = 1, 2, \dots, m$ (to be non-negative), and η (to be unrestricted in sign).
Additionally, the constraints $\eta - y^-_j + y^+_j = f_j$ for all $j = 1, \dots, m$ are required to relate all the variables to the values f_j.

Step 2 (Formulate the LP objective)
Considering that, from their definition, the variables y^-_j and y^+_j satisfy

$$y^-_j + y^+_j = |f_j - \eta|,\ j = 1, 2, \dots, m,$$

we can express the objective function (26) as

$$\Psi_{k,\omega}(\eta) = (1-\omega)\sum_{f_j<\eta}(\eta - f_j)\,A_k(t_j) + \omega\sum_{f_j>\eta}(f_j - \eta)\,A_k(t_j)$$

and, from the non-negativity of all terms $(1-\omega)\,A_k(t_j)$ and $\omega A_k(t_j)$,

$$\Psi_{k,\omega}(\eta) = (1-\omega)\sum_{j=1}^{m} y^-_j A_k(t_j) + \omega\sum_{j=1}^{m} y^+_j A_k(t_j).$$

Step 3 (Solve the LP problem)
Solve the LP problem with $2m+1$ variables y^-_j, y^+_j and η (the cost coefficient of variable η is equal to zero)

$$\min\left(\sum_{j=1}^{m} A_k(t_j)\,(1-\omega)\,y^-_j + \sum_{j=1}^{m} A_k(t_j)\omega y^+_j\right) \tag{27}$$

s.t.

$$\begin{cases} \eta - y^-_j + y^+_j = f_j, & j = 1, \dots, m \\ y^-_j \ge 0, & j = 1, \dots, m \\ y^+_i \ge 0, & j = 1, \dots, m \\ \eta \text{ unconstrained} \end{cases} \tag{28}$$

The component η of the solution found by solving (27)–(28) is our value $\eta^*_k(\omega)$.

The minimization of function (26), for fixed k and ω, is obtained by solving a linear programming problem with $2m+1$ variables and linear constraints: any standard LP solver can be used.

As we have remarked, the values of η that minimize $\Psi_{k,\omega}(\eta)$ form a closed real interval, say $[\eta_k^L(\omega), \eta_k^R(\omega)]$ with $\eta_k^L(\omega) \leq \eta_k^R(\omega)$ and with $\Psi_{k,\omega}(\eta') = \min_{\eta\in\mathbb{R}} \Psi_{k,\omega}(\eta)$ for all $\eta' \in [\eta_k^L(\omega), \eta_k^R(\omega)]$. On the other hand, for any fixed $\omega \in]0,1[$, we need to choose a single element from the optimal interval and several selection criteria have been proposed and implemented in the available statistical packages (see [18] for a short survey and discussion); we adopt the most frequently used selection, i.e. the midpoint of the interval.

The following property, given e.g. in [7,21], is analogous to Proposition 11.

Proposition 15. *For any fixed value of k, consider the intervals $[\eta_k^L(\omega), \eta_k^R(\omega)]$ corresponding to the minimizers of $\Psi_{k,\omega}(\eta)$, for $\omega \in]0,1[$; let $\eta_k^*(\omega) = \frac{1}{2}(\eta_k^L(\omega) + \eta_k^R(\omega))$ denote their midpoint values. Then $\eta_k^*(\cdot)$, as a function of ω, is non-decreasing, i.e.,*

$$\omega' > \omega'' \Longrightarrow \eta_k^*(\omega') \geq \eta_k^*(\omega''). \tag{29}$$

The monotonicity of functions $\eta_k^* :]0,1[\longrightarrow \mathbb{R}$ for all $k = -r+2, ..., n+r-1$, similarly to equation (20), ensures that the functions $\eta_k :]0,1[\longrightarrow \mathbb{R}$ defined in (30) are left-continuous and non-decreasing

$$\eta_k(\omega) = \lim_{\delta\downarrow 0} \eta_k^*(\omega - \delta) \text{ for all } \omega \in]0,1[\,. \tag{30}$$

Proposition 16. *Let $\eta_k(\omega)$ for $\omega \in]0,1[$ be given as in equation (30); consider $\alpha \in [0,1]$ and define the following compact intervals*

$$V_{k,\alpha} = \begin{cases} \left\{\eta_k\left(\frac{1}{2}\right)\right\} & \text{if} \quad \alpha = 1 \\ \left[\eta_k\left(\frac{\alpha}{2}\right), \eta_k\left(1-\frac{\alpha}{2}\right)\right] & \text{if} \quad \alpha \in]0,1[\\ cl\left(\bigcup_{\beta>0} V_{k,\beta}\right) & \text{if} \quad \alpha = 0; \end{cases} \tag{31}$$

then, for each $k = 2-r, ..., n+r-1$, the family of intervals $\{V_{k,\alpha}; \alpha \in [0,1]\}$ forms the α-cuts of a fuzzy number $G_k \in \mathbb{R}_{\mathcal{F}}$ having membership function

$$G_k(x) = \begin{cases} \sup\{\alpha | x \in V_{k,\alpha}\} & \text{if} \quad x \in V_{k,0} \\ 0 & \text{if} \quad x \notin V_{k,0}. \end{cases}$$

Proof. The proof is the same as for Proposition 12. □

Definition 17. Given a set of m points $Y_m = \{(t_i, f(t_i));\ i = 1, ..., m\}$ of a function $f : T \longrightarrow \mathbb{R}$ with $t_i \in T \subseteq [a,b]$ and given a fuzzy r-partition $(\mathbb{P}, \mathbb{A}, r)$ of $[a,b]$, the $(n+2r-2)$-vector of fuzzy numbers

$$\mathbf{G}_{(\mathbb{P},\mathbb{A},r)} = (G_{-r+2}, ..., G_{n+r-1}),$$

where each fuzzy interval G_k has α-cuts $V_{k,\alpha}$ given by (31) in Proposition 16, is called the discrete *direct quantile fuzzy transform* of f with respect to $(\mathbb{P}, \mathbb{A}, r)$, based on the data-set Y_m.

The corresponding *inverse quantile fuzzy transform* of f is the fuzzy-valued function defined by

$$\widetilde{f}_{(\mathbb{P},\mathbb{A},r)}(x) = \frac{1}{r} \sum_{k=2-r}^{n+r-1} G_k A_k(x) \text{ for } x \in T. \tag{32}$$

Denoting the α-cuts of G_k by

$$[G_k]_\alpha = \left[G_{k,\alpha}^-, G_{k,\alpha}^+\right]$$

and considering that each function A_k is non-negative on $[a,b]$, the α-cuts of the fuzzy-valued function $\widetilde{f}_{(\mathbb{P},\mathbb{A},r)}(x)$ will be

$$\left[\widetilde{f}_{(\mathbb{P},\mathbb{A},r)}(x)\right]_\alpha = \left[\frac{1}{r} \sum_{k=2-r}^{n+r-1} G_{k,\alpha}^- A_k(x)\ ,\ \frac{1}{r} \sum_{k=2-r}^{n+r-1} G_{k,\alpha}^+ A_k(x)\right]. \tag{33}$$

When $\alpha = 1$ we obtain the crisp L_1-norm iF-transform, corresponding to the core of the quantile fuzzy-valued iF-transform (33); indeed, we have, by construction, $G_{k,1}^- = G_{k,1}^+$ for all k.

5. Non-crossing property of F-transform smoothing

In this section we will describe the computational way to apply L_1 and L_2 inverse F-transforms in expectile and quantile smoothing of observed time series and we will show that both expectile and quantile reconstructions have the important *non-crossing* property.

As we have seen, both L_1 and L_2 fuzzy-valued (discrete) F-transforms are composed of two steps:

Step 1) given m data points (t_j, f_j), $t_j \in [a,b]$, $j=1,...,m$, and a fuzzy r-partition $(\mathbb{P},\mathbb{A},r)$ of $[a,b]$, the fuzzy-valued expectile $\mathbf{F}_{(\mathbb{P},\mathbb{A},r)}$ or quantile $\mathbf{G}_{(\mathbb{P},\mathbb{A},r)}$ direct F-transforms are computed, according to Definitions 13 or 17, respectively;

Step 2) from the direct F-transform obtained in Step 1), the fuzzy-valued L_1 or L_2 (inverse) iF-transforms $\widehat{f}_{(\mathbb{P},\mathbb{A},r)}$ or $\widetilde{f}_{(\mathbb{P},\mathbb{A},r)}$ are then given with α*-cuts* as in (24) or (33), respectively; clearly, the iF-transforms are fuzzy-valued because so are both direct F-transforms, and because all basic functions A_k are non-negative on $[a,b]$.

In forthcoming experiments, the decomposition $\mathbb{P}$ of $[a,b]$ is uniform with knots $x_k = a+(k-1)h$ and $h=(b-a)/(n-1)$ for $k=2-r,...,n+r-1$; the basic functions $A_k(x)$ are obtained as in equation (1) by translating and rescaling to the subintervals $[x_{k-r}, x_{k+r}]$ the same symmetric fuzzy number $U \in \mathbb{R}_{\mathcal{F}}$, with support $[U]_0=[-1,1]$, core $[U]_1=\{0\}$ and membership

$$U(\tau)=\begin{cases} L(1+\tau) & \text{if} \quad \tau\in[-1,0] \\ 1-L(\tau) & \text{if} \quad \tau\in[0,1] \\ 0 & \text{otherwise} \end{cases} \tag{34}$$

where function L, given by

$$L(\tau)=\frac{2\tau^2+\tau(1-\tau)}{2-2\tau(1-\tau)},$$

is of type (3) with parameters $\beta_0=0.5$, $\beta_1=0.5$.

We remark that, in general, the smoothing effect is not so much depending on the choice of the membership function U, or, more generally, of the basic functions A_k, as documented by papers on applications of F-transform (see, among others, [11,17,33,34,37]).

Instead, the number n of subintervals in the decomposition $\mathbb{P}$ and the integer bandwidth r strongly impact the smoothing effect (see [48,49]). In general, the increase of $n\geq 2$ and $r\geq 1$ produce opposite effects: when $n=2$ and $\mathbb{P}=\{a,b\}$ the direct F-transform of order $p=0$ has $2r$ components $F_{2-r},...,F_{r+1}$ and the inverse F-transform functions are flat; on the contrary, if $n=m$, $r=1$ and $\mathbb{P}=\{t_j; j=1,...,m\}$ (assuming $a=t_1<...<t_m=b$) then the inverse F-transforms (with $\alpha=1$) are interpolating (see [47,48]).

Within the implementation, the best combination of n and r is chosen by a generalized cross validation (GCV) approach (see [49]) in order to balance the smoothing and the fitting (interpolation) effects.

The crossing phenomenon in quantile and expectile smoothing frequently appears when several curves are computed, corresponding to different values of $\omega\in]0,1[$: the estimated functions can cross or overlap at different places in the interval $[a,b]$ (see, e.g., [19,57,58]).

On the other hand, expectile and quantile smoothing curves corresponding to specified values of $\omega\in]0,1[$, are easily obtained from the α-cuts of the fuzzy-valued expectile and quantile iF-transforms $\widehat{f}_{(\mathbb{P},\mathbb{A},r)}(x)$ and $\widetilde{f}_{(\mathbb{P},\mathbb{A},r)}(x)$ given, respectively, in equations (24) and (33): denoting by $\mathcal{E}_\omega(x)$ and $\mathcal{Q}_\omega(x)$ the ω-expectile and the ω-quantile curves we have

$$\mathcal{E}_\omega(x)=\begin{cases} \frac{1}{r}\sum\limits_{k=2-r}^{n+r-1} F^-_{k,2\omega}A_k(x) & \text{if} \quad \omega\leq\frac{1}{2} \\ \frac{1}{r}\sum\limits_{k=2-r}^{n+r-1} F^+_{k,2(1-\omega)}A_k(x) & \text{if} \quad \omega\geq\frac{1}{2} \end{cases} \tag{35}$$

and

$$\mathcal{Q}_\omega(x)=\begin{cases} \frac{1}{r}\sum\limits_{k=2-r}^{n+r-1} G^-_{k,2\omega}A_k(x) & \text{if} \quad \omega\leq\frac{1}{2} \\ \frac{1}{r}\sum\limits_{k=2-r}^{n+r-1} G^+_{k,2(1-\omega)}A_k(x) & \text{if} \quad \omega\geq\frac{1}{2}. \end{cases} \tag{36}$$

In all cases, we apply the expectile and quantile F-transforms of order $p=0$ in equations (7) and (14); this means that the direct L_2 and L_1 transforms are locally constant and the shapes of the inverse F-transforms are modeled by the form of the basic functions A_k, $k=2-r,...,n+r-1$. The choice of $p=0$ has a motivation related to an important property of both expectile and quantile F-transforms.

Please cite this article in press as: M.L. Guerra et al., Quantile and expectile smoothing based on L_1-norm and L_2-norm fuzzy transforms, Int. J. Approx. Reason. (2019), https://doi.org/10.1016/j.ijar.2019.01.011

Proposition 18. *The expectile and quantile functions $\mathcal{E}_\omega(x)$ and $\mathcal{Q}_\omega(x)$ of order $p=0$, defined in (35)–(36), have the non-crossing property, i.e., for all values of $\omega', \omega'' \in]0,1]$, we have, for each $x \in [a,b]$,*

$$\omega' < \omega'' \Longrightarrow \mathcal{E}_{\omega'}(x) \le \mathcal{E}_{\omega''}(x) \text{ and } \mathcal{Q}_{\omega'}(x) \le \mathcal{Q}_{\omega''}(x).$$

Proof. The proof follows immediately from Propositions 12 and 16 and the definitions of $\mathcal{E}_\omega(x)$, $\mathcal{Q}_\omega(x)$ in (35)–(36). □

6. Comparison with other expectile–quantile procedures

In this section we compare the fuzzy-valued iF-transform with other tools available within the scientific community, i.e., 1) statistical expectile and quantile regression routines: the **expectreg** package (see [43,45,46]) and the well known **quantreg** package (see [20]), both implemented in the **R** language and available at the CRAN repository, and

2) SVM-type (Support Vector Machine) non-parametric learning algorithms, for which a very efficient package exists for the **R** language at the CRAN repository, called **liquidSVM**; it is remarkable that this package implements both the expectile solver, routine **exSVM**, and the quantile solver, routine **qtSVM** (see [55], [56] and [13]).

We apply the described procedures to three well known daily time series from the financial market.

Clearly, fuzzy-valued approximations of a time series can be easily obtained also from the estimations based on the packages **expectreg**, **quantreg** and **liquidSVM**: the rule is that an α-cut, for $\alpha \in]0,1]$ has lower and upper functions given by the smoothed time series obtained with $\omega = \alpha/2$ and $\omega = 1-\alpha/2$, respectively.

More precisely, let A_t denote the t-th observed value of a time series, $t = 1, 2, \ldots, m$ and let $S_t(\omega)$ be the ω-expectile (or the ω-quantile) obtained by one of the procedures above with a given value of $\omega \in]0,1[$. Let $\mathcal{A}_t$ denote the fuzzy-valued smoothing series of A_t; in order to compute the α-cuts of $\mathcal{A}_t$ corresponding to a set $\mathcal{L} = \{0 < \alpha_1 < \alpha_2 < \ldots < \alpha_N = 1\}$ on $N > 2$ values of $\alpha \in]0,1]$, we generate the $2N-1$ curves

$$S_t\left(\frac{\alpha_l}{2}\right), S_t\left(1-\frac{\alpha_l}{2}\right) \text{ for } l = 1, \ldots, N-1, \text{ and}$$
$$S_t\left(\frac{\alpha_N}{2}\right) \text{ for } l = N,$$

corresponding to the values of ω in the set $\Omega = \{\frac{\alpha_l}{2}, 1-\frac{\alpha_l}{2} | l = 1, 2, \ldots, N\}$.

If the series $S_t(\omega)$, $\omega \in \Omega$ do not cross, i.e. $\omega' < \omega'' \Longrightarrow S_t(\omega') \le S_t(\omega'')$ for all t, then the α-cuts of the fuzzy-valued series $\widehat{A}_t$ are simply

$$[\mathcal{A}_t]_{\alpha_l} = \left[S_t\left(\frac{\alpha_l}{2}\right), S_t\left(1-\frac{\alpha_l}{2}\right)\right] \text{ for } l = 1, \ldots, N.$$

If the curves $S_t(\omega)$, for different $\omega \in \Omega$, are crossing at some t, then the α-cuts of $\mathcal{A}_t$ can be approximated as follows (see e.g. Definition 2 in [50]): the core is the singleton $[\mathcal{A}_t]_1 = \left\{S_t\left(\frac{1}{2}\right)\right\}$ and the remaining α-cuts are adjusted, for $l = N-1, N-2, \ldots, 1$, by

$$[\mathcal{A}_t]_{\alpha_l} = \left[\min\left\{S_t\left(\frac{\alpha_{l'}}{2}\right) | l' \ge l\right\}, \max\left\{S_t\left(1-\frac{\alpha_{l'}}{2}\right) | l' \ge l\right\}\right].$$

The comparison is performed by considering the estimations denoted as follows, with the distinction between the corresponding expectile and quantile procedures:

A: *for the expectile estimations,*
- **A.1**) **expFT**: expectile iF-transform series corresponding to the α-cuts $\left[\widehat{f}_{(\mathbb{P},\mathbb{A},r)}(t)\right]_\alpha$.
- **A.2**) **expRS**: expectile regression smoothed series, obtained by routine **expectreg.ls** of package **expectreg**, corresponding to the α-cuts $\left[\widehat{f}_{RS}(t)\right]_\alpha$.
- **A.3**) **exSVM**: expectile regression smoothed series, obtained by routine **exSVM** of package **liquidSVM**, corresponding to the α-cuts $\left[\widehat{f}_{SVM}(t)\right]_\alpha$.

B: *for the quantile estimations,*
- **B.1**) **quaFT**: quantile iF-transform series corresponding to the α-cuts $\left[\widetilde{f}_{(\mathbb{P},\mathbb{A},r)}(t)\right]_\alpha$.
- **B.2**) **quaRS**: quantile regression smoothed series, obtained by routine **rq** of package **quantreg**, corresponding to the α-cuts $\left[\widetilde{f}_{RS}(t)\right]_\alpha$.
- **B.3**) **qtSVM**: quantile regression smoothed series, obtained by routine **qtSVM** of package **liquidSVM**, corresponding to the α-cuts $\left[\widetilde{f}_{SVM}(t)\right]_\alpha$.

We distinguish between the comparison of the fuzzy-valued time series with respect to the *core*, corresponding to $\alpha = 1$ or equivalently to $\omega = 0.5$ and to the fuzzy-valued series in terms of all their α-cuts for $\alpha \in]0,1]$.

The three time series are the following, well known from the financial market.

Please cite this article in press as: M.L. Guerra et al., Quantile and expectile smoothing based on L_1-norm and L_2-norm fuzzy transforms, Int. J. Approx. Reason. (2019), https://doi.org/10.1016/j.ijar.2019.01.011

Series 1: Silver prices in US dollars

The first time series contains the silver prices in US dollars. The price is set once a day by three London Bullion Market Association (LBMA) market makers that listen to customers' purpose as buyers or sellers: the silver fixing price is then set by collating bids and offers until the supply and demand are matched. The role of silver in recent years has been different from the gold one because it's considered a tangible asset rather than a store of value, that is why generally silver prices are more volatile than gold prices.

Series 2: Apple daily stock in US dollars

The second time series is the well-known Apple stock that has an high volatility in the short term, a shared property for all the stocks. As is visible from the graphical representations, something particular happened in June 2014: a share of Apple varied from $645.57 (as of Friday's closing price) to $92.44, because the company issued more shares to existing investors in order to put down the price of the stock. Current shareholders received seven shares of Apple for each one they owned. As a result, the stock price is one-seventh of where it used to be.

Series 3: S&P500 index

The third historical data series is the S&P500 index, that is probably the most accurate quantifier of the US economy, measuring the cumulative float-adjusted market capitalization of 500 of the nation's largest corporations; due to its definition it is considered a low volatility stock. Practitioners remember very well the milestones of S&P500 index: on 11 October 2007, S&P500 index reached its all-time intra-day high of 1,576.09; on 28 March 2013, the S&P500 finally surpassed its closing high level of 1,565.15, recovering all its losses from the financial crisis and on 26 August 2014 it closed a hair above 2000 points.

It is generally known that the three considered time series are deduced by assets that behave in different ways in the financial market. The time period covers from October 9th 2007 to October 8th 2015 for a total amount of $m = 2016$ observations and it includes the most recent financial crisis.

6.1. Core comparison with three financial time series

To show the results for the *core* of the fuzzy-valued expectile and quantile estimations, we provide a set of (standard) performance measures; hereafter, we denote by $\mathbf{A} = \{A_t, t = 1, \ldots, m\}$ a given time series of observed (actual) data A_t and by $\mathbf{S} = \{S_t, t = 1, \ldots, m\}$ the smoothed time series obtained by one of the used methods, i.e., S_t is the forecast value at time t.

1. **MAV** (Mean Absolute Variation of time series **A**):

$$MAV(\mathbf{A}) = \frac{1}{m-1}\sum_{t=1}^{m-1}|A_{t+1} - A_t|$$

   and we will also use the MAV ratio of smoothed **S** with respect to **A**:

$$MAVR(\mathbf{S},\mathbf{A}) = \frac{MAV(\mathbf{S})}{MAV(\mathbf{A})};$$

   Its percentage version is denoted by $MAV\%(\mathbf{S},\mathbf{A}) = 100MAVR(\mathbf{S},\mathbf{A})$.

2. **sqrtMSE** (square root of Mean Square Error or Deviation of smoothed **S** with respect to **A**; the Mean Square Error, without taking the square root, is denoted by **MSE**):

$$sqrtMSE(\mathbf{S},\mathbf{A}) = \sqrt{\frac{1}{m}\sum_{t=1}^{m}|A_t - S_t|^2}.$$

3. **sqrtRMSE** (square root of Relative Mean Square Error or Deviation of smoothed **S** with **A**):

$$sqrtRMSE(\mathbf{S},\mathbf{A}) = \sqrt{\frac{1}{m}\sum_{t=1}^{m}\left|\frac{A_t - S_t}{A_t}\right|^2}.$$

4. **MAD** (Mean Absolute Deviation of **S** with **A**, sometimes denoted by AAD – Average Absolute Deviation, also called *median pinball loss* [57]):

$$MAD(\mathbf{S},\mathbf{A}) = \frac{1}{m}\sum_{t=1}^{m}|A_t - S_t|.$$

5. **MAPE** (Mean Absolute Percentage Error or Deviation of **S** with **A**):

$$MAPE(\mathbf{S},\mathbf{A}) = \frac{100}{m}\sum_{t=1}^{m}\left|\frac{A_t - S_t}{A_t}\right|.$$

The measures reported in all the tables are computed with A_t being the actual value of the time series at all times $t = 1, ..., m$, and S_t being the corresponding smoothed value obtained by one of the six smoothing methods (expFT, expRS, exSVM, quaFT, quaRS and qtSVM) for the central (median) quantile, i.e., in our notation, for the core (α-cut with $\alpha = 1$) of the fuzzy-valued smoothed series, corresponding to the quantile parameter $\omega = \frac{1}{2}$ in packages **expectreg**, **quantreg** and **liquidSVM**.

Consider that the measures above can be computed also for time sub-periods of the series; in particular the indicator MAV can be interpreted as a measure of local variability, in addition to the well known volatility obtained using local variances.

The mean absolute variations of the three time series are $MAV(Silver) = 0.2205$, $MAV(Apple) = 4.295$ and $MAV(S\&P) = 11.759$.

We launch the comparison by executing the six methods according to their own (internal) best combination of smoothing and fitting effects; in particular, neither over-fitting nor under-fitting is required to happen.

As just underlined, the best combinations of n and r for the F-transforms have been chosen, for the different series, by a generalized cross validation (GCV) approach. In the **expectreg** and in the **liquidSVM** packages, similar GCV schemes are implemented; in particular, **liquidSVM**, based on a kernel smoothing technique with regularization, determines the best combination of two parameters, the bandwidth $h > 0$ of the kernel and the regularization parameter $\lambda \geq 0$, based on a 10×10 grid of possible pairs (h, λ). The package **quantreg** adopts a more sophisticated procedure, based on the regularization of the total variation and a combination of criteria including GCV, the Akaike information index (AIC) and other extractor methods for the best selection.

In Table 2a we report some of the measures, obtained by running the six methods with their default selection strategies, as suggested by the authors of the corresponding packages.

Table 2a
MAV%, MSE and MAD for all time series and methods.

	expFT	*expRS*	*exSVM*	*quaFT*	*quaRS*	*qtSVM*
MAV%-Silver	9.2%	9.1%	12.0%	10.6%	8.9%	16.1%
MAV%-Apple	17.1%	20.3%	22.9%	17.9%	18.2%	22.9%
MAV%-S&P	9.5%	11.6%	11.0%	10.7%	11.3%	14.7%
MSE-Silver	0.64	0.97	0.72	0.60	1.07	0.51
MSE-Apple	1158.4	1711.1	932.4	957.3	2696.8	898.2
MSE-S&P	1205.2	1594.6	1412.3	1115.9	1655.8	780.0
MAD-Silver	0.56	0.73	0.65	0.52	0.75	0.51
MAD-Apple	15.6	23.1	16.9	14.0	23.9	11.8
MAD-S&P	26.3	30.0	30.4	24.9	29.7	19.9

As we can see from the first three rows, the six methods behave differently in terms of the MAV% measure; recall that this measure gives the percentage of total variation MAV in the smoothed series with respect to the MAV of the observed one, so that, in some sense, the quantity (100-MAV%) gives the percentage of total variation *removed* by the smoothing effect and this clearly depends on the adopted selection strategy.

Not only the MAV% is different for the three series, but also across the methods. This is an important fact to take into account, because all measures strongly depend on how the balancing between smoothing and interpolating levels is performed: at least qualitatively, a higher MAV% value (i.e., less reduction in total variation) will imply a lower value of error-based measures like MSE and MAD. For example, the quaRS and qtSVM methods for Silver series have very different MAV% measures and the quaRS method gives a much more rigid smoothed series than qtSVM; it is then not surprising that they have very different MSE and MAD.

Remark 19. It is to be remarked that, for all methods, the MSE obtained by expectile smoothing (where an L_2 measure is minimized) is not necessarily smaller than the one resulting from quantile smoothing (where an L_1 measure is minimized); and the MAD measure with quantiles may be not smaller than the one with expectiles (see, e.g., [19,44]). In particular, in the F-transform setting we have two levels of approximation, the first obtained by the components of the direct F-transforms and the second by the inverse F-transforms. A component F_k in the expectile direct F-transform is a local (weighted) *average* of the time series around the node x_k of the fuzzy partition, obtained by minimizing an L_2-norm error; analogously, each component G_k in the quantile direct F-transform represents locally a weighted *median* of the series, obtained by minimizing an L_1-norm error. But this does not imply, in general, that the reconstruction obtained by the inverse expectile iF-transform has globally an average L_2-norm error smaller than the inverse quantile reconstruction. To be more precise, if we consider a linear approximating function

$$f(x) = \sum_{k=2-r}^{n+r-1} \vartheta_k A_k(x),$$

where A_k are basic functions of an r-partition and the parameters ϑ_k are obtained by minimizing the L_2-norm (or the L_1-norm) of the errors $f(t_j) - \sum_{k=2-r}^{n+r-1} \vartheta_k A_k(t_j)$, then the estimated coefficients ϑ_k have no relationship with the expectile components F_k (or the quantile components G_k); in general, the average L_2-norm (or L_1-norm) error for the ϑ_k will be much smaller than for the F_k and the smoothing effect will have a much greater MAV%. For example, the L_2-norm estimation of the linear parameters ϑ_k for the Silver series with the same pair (n, r) as for expFT in first column of Table 2a, produces MAV%= 27.1%, MSE= 0.182 and MAD= 0.29, a very different result with respect to all smoothing methods.

In order to make the comparison effective, we have first executed our F-transform procedures expFT and quaFT with the purpose to determine the pairs of values (n, r) that reproduce the same (or the nearest) MAV% value as for the other procedures expRS and exSVM for the expectiles and quaRS and qtSVM for the quantiles. The results are given in Table 2b. For example, the pairs of values n and r, using expFT, that reproduced the *nearest* results to the MAV% of routines expRS and exSVM for the *Silver* series have been: $(151, 4)$ and $(181, 3)$.

Table 2b
Values of pair (n, r) used with F-transform.

	expRS	*exSVM*	*quaRS*	*qtSVM*
(n, r)-Silver	$(151, 4)$	$(181, 3)$	$(127, 4)$	$(281, 3)$
(n, r)-Apple	$(301, 4)$	$(317, 3)$	$(181, 4)$	$(285, 3)$
(n, r)-S&P	$(227, 4)$	$(143, 3)$	$(201, 4)$	$(307, 3)$
MAV%-Silver	09.05%	11.94%	08.86%	15.89%
MAV%-Apple	20.30%	22.85%	18.13%	23.05%
MAV%-S&P	11.64%	11.03%	11.34%	14.74%

The results obtained by expFT and quaFT with those pairs (n, r) have then (almost) the same smoothing effects, measured by MAV, as the other routines and we can now compare the F-transform results with the regression smoothing and the SVM smoothing ones.

Clearly, the comparison can be performed only pairwise, i.e., by splitting the comparison of the other measures under the condition of uniform values of MAV(**S**), i.e. by considering four separate cases:

1) expFT with expRS and the (n, r) pairs in first column of Table 2b;
2) expFT with exSVM and the (n, r) pairs in second column of Table 2b;
3) quaFT with quaRS and the (n, r) pairs in third column of Table 2b;
4) quaFT with qtSVM and the (n, r) pairs in fourth column of Table 2b.

Without having the same MAV values, the error-based measures sqrtMSE, sqrtRMSE, MAPE and MAD resulted essentially incomparable.

For the **Silver time series**, expFT and expRS smoothing have (almost) the same MAV% (9.05% and 9.1%, respectively) and similarly in the other cases.

To have a deeper knowledge of this series, we compute the volatility on annual basis and it goes from the highest $\sigma_{Silver,4th}$ in the fourth year (Oct. 2010 to Oct. 2011) to the smallest $\sigma_{Silver,7th}$ in the seventh year (Oct. 2013 to Oct. 2014); as it is expected, Silver is characterized by volatility values that can be very high. On the other hand, the MAV measure has a direct relationship with volatility; for example it holds that $\sigma_{Silver,4th} = 58.78\%$ and $MAV(Silver, 4th) = 0.87$ while in the seventh year it holds $\sigma_{Silver,7th} = 21.14\%$ and $MAV(Silver, 7th) = 0.21$. The main difference between the two measures is that volatility is expressed as a percentage based on intra-day returns, whereas MAV takes into account intra-day price variations and is not in percentage form.

Fig. 3 represents the core (1-cut) obtained by the expectile and quantile iF-transforms (23), (32), compared with the expectile regression smoothing obtained by the expRS and the quantile regression smoothing with quaRS. Fig. 4 gives the core of the expectile and quantile iF-transforms, compared with the corresponding smoothing obtained by methods exSVM and qtSVM.

From all figures, we see a better capability of the iF-transform smoothed series to capture local variations of the real time series; in particular, this happens at the portion of time where the values have local peaks or falls. This is confirmed by the performance measures as reported in Tables 3a and 3b.

Table 3a gives the performance measures for the comparison of F-transform and Regression-type expectile and quantile smoothing.

In Table 3b the same measures compare F-transform and SVM-type expectile and quantile smoothing; the values of the measures for expectile and quantile iF-transform are again better than the corresponding SVM smoothed series.

We see that, at the same level of MAV(**S**), the expectile and quantile F-transforms have a significantly smaller value for the four measures MSE, RMSE MAD and MAPE, with respect to the counterparts based on regression smoothing (Table 3a) and with respect to the SVM-based series (Table 3b).

Please cite this article in press as: M.L. Guerra et al., Quantile and expectile smoothing based on L_1-norm and L_2-norm fuzzy transforms, Int. J. Approx. Reason. (2019), https://doi.org/10.1016/j.ijar.2019.01.011

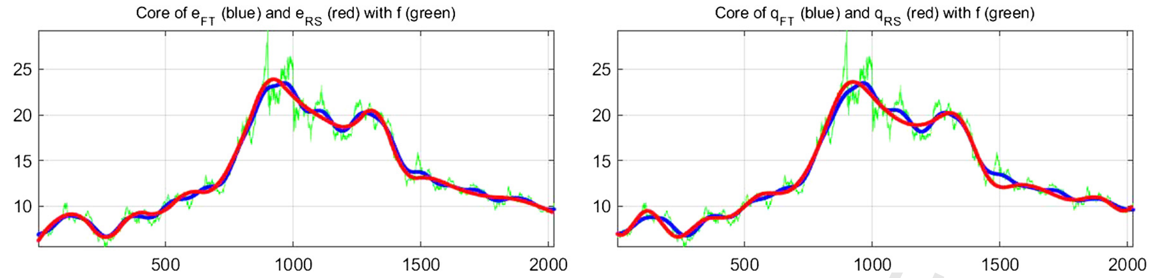

Fig. 3. (Silver time series) Left picture: expectile smoothing obtained by methods expFT (blue) and expRS (red); right picture: quantile smoothing obtained by quaFT (blue) and quaRS (red).

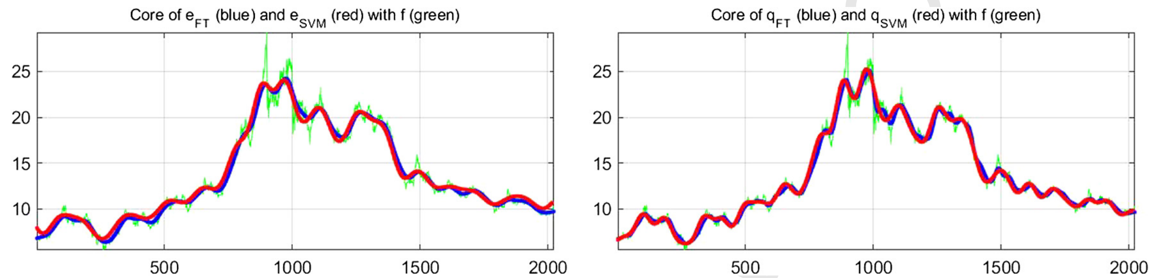

Fig. 4. (Silver time series) Left picture: expectile smoothing obtained by methods expFT (blue) and exSVM (red); right picture: quantile smoothing obtained by quaFT (blue) and qtSVM (red).

Table 3a
Comparison FT-RS for silver time series $MAV(\mathbf{A}) = 0.2205$.

. . Measure . .	expFT	expRS	quaFT	quaRS
MAV(**S**)	0.0200	0.0206	0.0195	0.0204
sqrtMSE	0.8353	0.9872	0.8688	1.0330
sqrtRMSE	0.0548	0.0656	0.0570	0.0712
MAD	0.5912	0.7250	0.6102	0.7540
MAPE	4.2905	5.2667	4.4349	5.5567

Table 3b
Comparison FT-SVM for silver time series $MAV(\mathbf{A}) = 0.2205$.

. . Measure . .	expFT	exSVM	quaFT	qtSVM
MAV(**S**)	0.0263	0.0265	0.0351	0.0354
sqrtMSE	0.6714	0.8478	0.5215	0.7161
sqrtRMSE	0.0431	0.0642	0.0340	0.0467
MAD	0.4561	0.6511	0.3384	0.5052
MAPE	3.2824	5.1082	2.4683	3.6481

Observe that, from Table 2a, the F-transform series and the RS-based or SVM-based series have very different smoothing levels; e.g., for the FT-SVM comparison, $MAV\%(expFT) = 9.2\%$ and $MAV\%(qtSVM) = 16.19.2\%$, but the MSE and MAD measures have essentially similar values.

It is well-known that the **Apple time series**, representing a stock value, has a high volatility in the short term, a property shared with the totality of stock prices. Figs. 5 and 6 that in June 2014, a share of Apple varied from \$645.57 (as of Friday's closing price) to \$92.44, because the company issued more shares to existing investors in order to put down the price of the stock. Current shareholders received seven shares of Apple for each one they owned. As a result, the stock price became one-seventh of where it used to be. Examining this series offers a very interesting case of the possible problems raised by a rapid instantaneous big change in the subsequent values (levels).

We observe that, in this situation, the F-transform expectile and quantile smoothing have a significantly better capability to follow the variations in level without requiring to change the degree of smoothing (measured qualitatively, e.g., by MAV% index).

From Figs. 5 and 6 we see that expFT and expRS are able to follow the real values of the series (even in June 2014) significantly better than all other considered methods (the four error measures in Tables 4a and 4b are all smaller for F-transform methods); the best case is obtained with expectile F-transform, corresponding to the pair of values $n = 317$, $r = 3$ (see Table 4b and left picture of Fig. 6).

Please cite this article in press as: M.L. Guerra et al., Quantile and expectile smoothing based on L_1-norm and L_2-norm fuzzy transforms, Int. J. Approx. Reason. (2019), https://doi.org/10.1016/j.ijar.2019.01.011

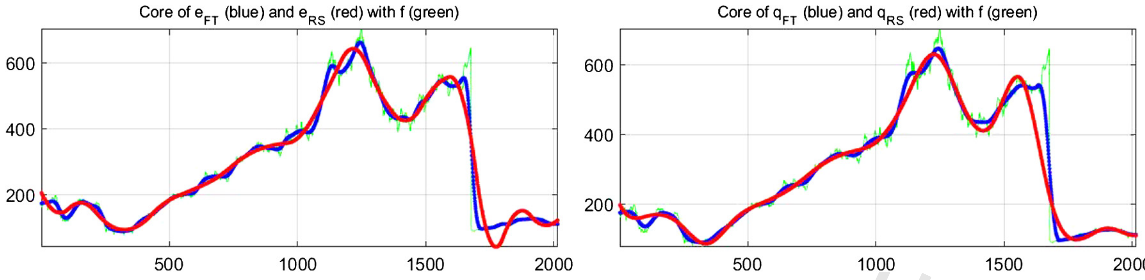

Fig. 5. (Apple time series) Left picture: expectile smoothing obtained by methods expFT (blue) and expRS (red); right picture: quantile smoothing obtained by quaFT (blue) and quaRS (red).

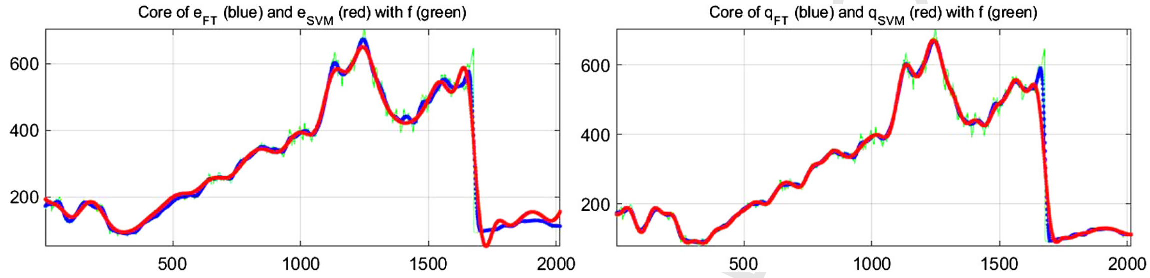

Fig. 6. (Apple time series) Left picture: expectile smoothing obtained by methods expFT (blue) and exSVM (red); right picture: quantile smoothing obtained by quaFT (blue) and qtSVM (red).

Table 4a
Comparison FT-RS for Apple time series $MAV(\mathbf{A}) = 4.2949$.

. . Measure . .	expFT	expRS	quaFT	quaRS
MAV(**S**)	0.8718	0.8727	0.7788	0.7793
sqrtMSE	25.6313	41.3656	30.1436	51.9308
sqrtRMSE	0.1746	0.3029	0.1501	0.1748
MAD	10.8842	23.1364	13.0787	23.9099
MAPE	4.7755	11.7144	5.2191	8.8218

Table 4b
Comparison FT-SVM for Apple time series $MAV(\mathbf{A}) = 4.2949$.

. . Measure . .	expFT	exSVM	quaFT	qtSVM
MAV(**S**)	0.9812	0.9847	0.9900	0.9854
sqrtMSE	21.7176	30.5351	20.0894	29.9708
sqrtRMSE	0.1488	0.2106	0.1375	0.1617
MAD	8.8300	16.9497	8.3866	11.8088
MAPE	3.7912	8.4619	3.5133	4.9581

Observe that the MAV% values in Table 2a range, for the six methods, from 17.1% to 22.9% and the MAD value ranges from 11.8 to 23.9. With the values of the parameters (n, r) as in Table 2b (see the rows with label *Apple*), for each of the four cases 1)–4), the MAV% values are very similar. Tables 4a and 4b contain the results for the cases FT vs RS and FT vs SVM, respectively.

The measures show that the quantile iF-transform has a better fitting with respect to the expectile iF-transform (true also in Table 2a): the main reason is that the fitting performance may not depend on the L_1- or L_2-norm minimization to estimate the direct components of the F-transform, but on the property that both expectile and quantile iF-transform exhibit uniform convergence as given by Theorems 3 and 8.

Also for the second time series, the pair of measures MAD and MSE produce a strong reduction in the errors of both quantile and expectile iF-transform smoothed series, confirming the strength of our proposed methodology.

The **S&P500 time series** (Figs. 7 and 8) contains the historical values of the Standard & Poor's 500 Index, that is probably the most accurate quantifier of the U.S. economy, measuring the cumulative float-adjusted market capitalization of 500 large U.S. publicly traded companies; due to its definition it is considered a low volatility stock. Practitioners remember very well the milestones of S&P500 index: on October 2007, S&P500 index reached its all-time intra-day high of 1,576.09; on March 2013, the index finally surpassed its closing high level of 1,565.15, recovering all its losses from the financial crisis and on August 2014 it closed a hair above 2000 points. The average variation is $MAV(S\&P) = 11.759$.

Please cite this article in press as: M.L. Guerra et al., Quantile and expectile smoothing based on L_1-norm and L_2-norm fuzzy transforms, Int. J. Approx. Reason. (2019), https://doi.org/10.1016/j.ijar.2019.01.011

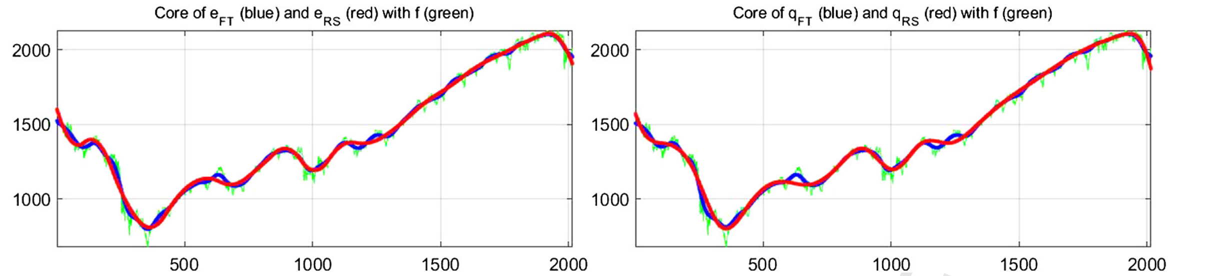

Fig. 7. (S&P500 time series) Left picture: expectile smoothing obtained by methods expFT (blue) and expRS (red); right picture: quantile smoothing obtained by quaFT (blue) and quaRS (red).

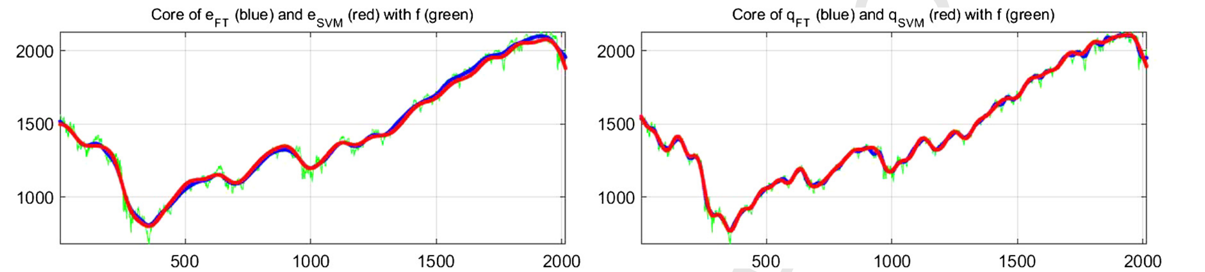

Fig. 8. (S&P500 time series) Left picture: expectile smoothing obtained by methods expFT (blue) and exSVM (red); right picture: quantile smoothing obtained by quaFT (blue) and qtSVM (red).

Table 5a
Comparison FT-RS for S&P500 time series $MAV(\mathbf{A}) = 11.759$.

. . Measure . .	*expFT*	*expRS*	*quaFT*	*quaRS*
*MAV(**S**)*	1.3692	1.3703	1.3332	1.3320
sqrtMSE	30.5665	39.9330	31.0307	40.6918
sqrtRMSE	0.0258	0.0340	0.0265	0.0352
MAD	22.9776	30.0181	23.0007	29.6685
MAPE	1.7556	2.3289	1.7693	2.3193

Table 5b
Comparison FT-SVM for S&P500 time series $MAV(\mathbf{A}) = 11.759$.

. . Measure . .	*expFT*	*exSVM*	*quaFT*	*qtSVM*
*MAV(**S**)*	1.2967	1.2969	1.7322	1.7338
sqrtMSE	32.9268	37.5801	23.3289	27.9291
sqrtRMSE	0.0280	0.0305	0.0197	0.0225
MAD	24.8522	30.3933	16.6673	19.8776
MAPE	1.9057	2.2752	1.2811	1.5065

Also in this third case, we see that, at the same level of MAV(**S**), the expectile and quantile F-transforms perform better in the four average error-bases measures (Tables 5a and 5b). On the other hand, from Table 2a we can remark that the expectile and quantile F-transforms have a smaller value of $MAV\%$ than the RS-based or SVM-based ones but also a smaller value of the error-based measures, with an exception: expFT performs better than expRS and exSVM, quaFT is better than quaRS, but qtSVM is more precise than quaFT, possibly due to the difference in the $MAV\%$ value.

To summarize the results for the three time series, we can refer to Table 6, containing the ratios between the values (as reported in Tables 3a, 3b, 4a, 4b, 5a and 5b) of the measures for the F-transform series to the ones with RS-based and SVM-based series: e.g., the value in column *exp:FT/RS* and row *sqrtMSE*, contains the ratio of the *sqrtMSE* measure obtained by *expFT* over the *sqrtMSE* one by *expRS*, similarly for the other entries.

What emerges clearly is that the F-transform expectile and quantile smoothing have error-based measures significantly smaller than the other methods. The best performances of F-transform are more evident in the case of the *Apple* series, in particular in the time period where the values have the great decrease as we described above.

6.2. Comparison results for fuzzy-valued smoothing

To show the results for the fuzzy-valued series corresponding with respect to all the α-cuts with $\alpha \in]0, 1]$, we choose three measures well known in the fuzzy literature: the ambiguity, the average fuzzy distance and the length of 0.5-cut.

Table 6
Summary of Performance ratios.

. Silver Series .	exp:FT/RS	exp:FT/SVM	qua:FT/RS	qua:FT/SVM
MAVR	0.9709	0.9925	0.9559	0.9915
sqrtMSE	0.8461	0.7919	0.8410	0.7283
sqrtRMSE	0.8354	0.6713	0.8006	0.7281
MAD	0.8154	0.7005	0.8093	0.6698
MAPE	0.8146	0.6426	0.7979	0.6766
. Apple Series .	exp:FT/RS	exp:FT/SVM	qua:FT/RS	qua:FT/SVM
MAVR	0.9990	0.9964	0.9994	1.0047
sqrtMSE	0.6196	0.7112	0.5805	0.6703
sqrtRMSE	0.5764	0.7066	0.8587	0.8503
MAD	0.4704	0.5210	0.5470	0.7102
MAPE	0.4077	0.4480	0.5916	0.7086
. S&P Series .	exp:FT/RS	exp:FT/SVM	qua:FT/RS	qua:FT/SVM
MAVR	0.9992	0.9998	1.0009	0.9991
sqrtMSE	0.7654	0.8762	0.7626	0.8353
sqrtRMSE	0.7588	0.9180	0.7528	0.8756
MAD	0.7655	0.8177	0.7753	0.8385
MAPE	0.7538	0.8376	0.7629	0.8504

For a fuzzy number $u \in \mathbb{R}_{\mathcal{F}}$, with α-cuts $[u]_\alpha = [u^-_\alpha, u^+_\alpha]$, the (fuzzy) *ambiguity* is defined (see [9]) as

$$amb(u) = \int_0^1 \alpha(u^+_\alpha - u^-_\alpha)d\alpha$$

and the *central interval* is simply its (1/2)-cut $[u]_{1/2} = [u^-_{1/2}, u^+_{1/2}]$. For two fuzzy numbers $u, v \in \mathbb{R}_{\mathcal{F}}$ the (integral) L_2*-distance* is

$$dist(u,v) = \left(\int_0^1 \alpha \left((u^+_\alpha - v^+_\alpha)^2 + (u^-_\alpha - v^-_\alpha)^2 \right) d\alpha \right)^{1/2}.$$

The ambiguity $amb(u) \geq 0$ can be seen as a measure of how much vagueness or spread is present in the ill-defined magnitude which underlies the fuzzy number $u \in \mathbb{R}_{\mathcal{F}}$; if u is crisp, then $amb(u) = 0$.

The central interval is frequently used as an easy-to-obtain alternative to the mean interval; in terms of a quantile interpretation of the membership function of $u \in \mathbb{R}_{\mathcal{F}}$ (see [51]) it coincides with the inter-quartile interval and its length is the well known *inter-quartile range*: approximately 50% of observed points with positive membership values is contained in the interval, remaining 25% is on its left and 25% on its right (see [7,19]).

The (weighted) L_2-distance (with some of its variants) is a standard metric in the space of fuzzy numbers (see [2]); it can be used also to measure the distance between a fuzzy and a crisp number.

A fuzzy-valued time series is denoted by $\mathbf{U} = \{U_t; t = 1, ..., m\}$ where each U_t is a fuzzy number with α-cuts $\left[(U_t)^+_\alpha, (U_t)^-_\alpha\right]$. Accordingly, the three reported measures are

1. AMB (Mean Ambiguity measure of fuzzy-valued time series $\mathbf{U}$):

$$AMB(\mathbf{U}) = \frac{1}{m}\sum_{t=1}^m amb(U_t);$$

2. CIR (Mean Central Interval Range (Length) of fuzzy-valued time series $\mathbf{U}$):

$$CIR(\mathbf{U}) = \frac{1}{m}\sum_{t=1}^m \left((U_t)^+_{1/2} - (U_t)^-_{1/2} \right);$$

3. DIS (Mean Distance measure between fuzzy-valued time series $\mathbf{U}$ and observed values f_t, where the crisp data are considered as fuzzy numbers with concentrated membership value 1 at f_t and 0 elsewhere):

$$DIS(\mathbf{U}) = \frac{1}{m}\sum_{t=1}^m dist(U_t, f_t).$$

Please cite this article in press as: M.L. Guerra et al., Quantile and expectile smoothing based on L_1-norm and L_2-norm fuzzy transforms, Int. J. Approx. Reason. (2019), https://doi.org/10.1016/j.ijar.2019.01.011

In all figures of this section, the reported α-cuts are obtained for the $N = 11$ values of $\alpha \in \mathcal{L} = \{0.001, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$, corresponding to the 21 quantile values $\omega \in \{\frac{\alpha}{2}, 1 - \frac{\alpha}{2} | \alpha \in \mathcal{L}\}$. The curves with $\omega = \frac{\alpha}{2}$, $\alpha \in \mathcal{L}$ are red-colored (corresponding to the lower branches of the fuzzy-valued functions) and the curves with $\omega = 1 - \frac{\alpha}{2}$, $\alpha \in \mathcal{L}$ are blue-colored (corresponding to the upper branches of the fuzzy-valued functions); the core is black-colored.

Remark 20. In the estimation by exSVM or qtSVM (see Figs. 10, 12, 14) and for small α, i.e. values of ω near to 0 or 1, the smoothed curves are constant and the support of the fuzzy-valued functions does not change for all t_j; the same did not happen for all other methods. For this reason, in computing measures AMB and DIS we have not considered the 0.001-cut (this is easy to obtain by approximating the integrals by trapezoidal rule).

For the four pairs of comparisons (*expFT, expRS*), (*quaFT, quaRS*), (*expFT, exSVM*), and (*quaFT, qtSVM*), the three performance measures are collected in Table 7 for the *Silver* time series, in Table 8 for the *Apple* time series and in Table 9 for the *S&P500* time series.

Table 7
Silver time series.

Fuzzy-valued FT and RS				
... Measure ...	*expFT*	*expRS*	*quaFT*	*quaRS*
AMB	1.0017	0.9634	1.5546	1.4830
CIR	0.7818	0.7757	1.3645	1.3890
DIS	1.3768	1.4491	1.7435	1.7336
Fuzzy-valued FT and SVM				
... Measure ...	*expFT*	*exSVM*	*quaFT*	*qtSVM*
AMB	0.5504	0.8038	0.6610	0.6469
CIR	0.6184	0.7390	0.7717	0.7750
DIS	0.8481	1.3537	0.8049	0.9664

Table 8
Apple time series.

Fuzzy-valued FT and RS				
... Measure ...	*expFT*	*expRS*	*quaFT*	*quaRS*
AMB	18.0861	31.1668	34.4456	47.9194
CIR	14.3794	25.4447	29.4278	44.8778
DIS	25.1900	46.5326	39.3149	54.9424
Fuzzy-valued FT and SVM				
... Measure ...	*expFT*	*exSVM*	*quaFT*	*qtSVM*
AMB	10.4197	25.1443	16.2968	18.5887
CIR	11.7270	23.4582	19.5498	21.8254
DIS	16.3738	39.8699	20.3668	25.1815

Table 9
S&P500 time series.

Fuzzy-valued FT and RS				
... Measure ...	*expFT*	*expRS*	*quaFT*	*quaRS*
AMB	40.0614	41.5572	60.5002	59.6647
CIR	31.2139	32.7700	52.1620	53.5968
DIS	54.2838	61.9662	67.2433	71.1619
Fuzzy-valued FT and SVM				
... Measure ...	*expFT*	*exSVM*	*quaFT*	*qtSVM*
AMB	30.3988	46.7578	32.3657	30.2987
CIR	34.0324	50.1652	36.8682	35.4048
DIS	46.3936	71.5752	39.4386	42.1775

Figs. 9, 10 for *Silver*, Figs. 11, 12 for *Apple* and Figs. 13, 14 for *S&P500* represent the α-cuts obtained by the repeated application of F-transform (23), (32) for the indicated values of $\alpha \in \mathcal{L}$.

After a preliminary inspection of the figures, we see that the F-transform series are significantly more adherent to the observed values, not only for the core of the fuzzy-valued series. In some portions of time, expectile and quantile curves based on RS and on SVM may have (locally) a value of crossing (see particularly the *Apple* series).

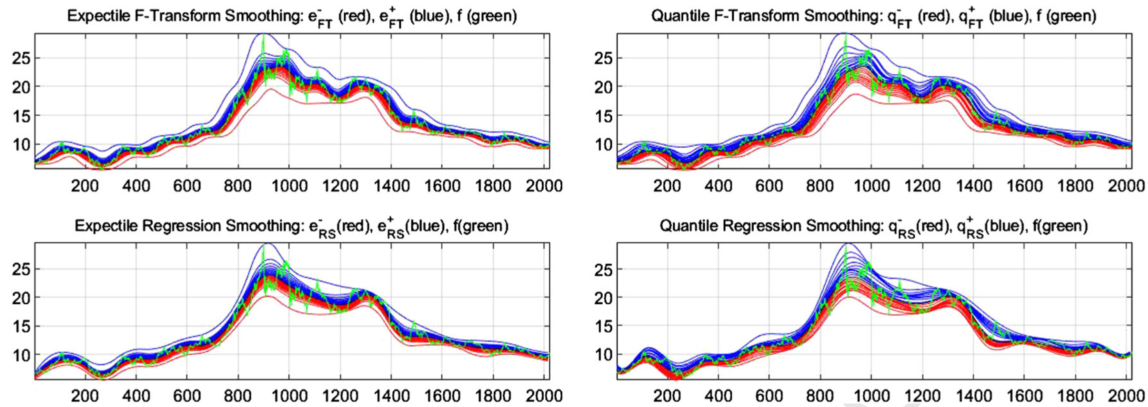

Fig. 9. (Silver time series) Left: fuzzy-valued expectile smoothing by methods expFT and expRS. Right: fuzzy-valued quantile smoothing by methods quaFT and quaRS.

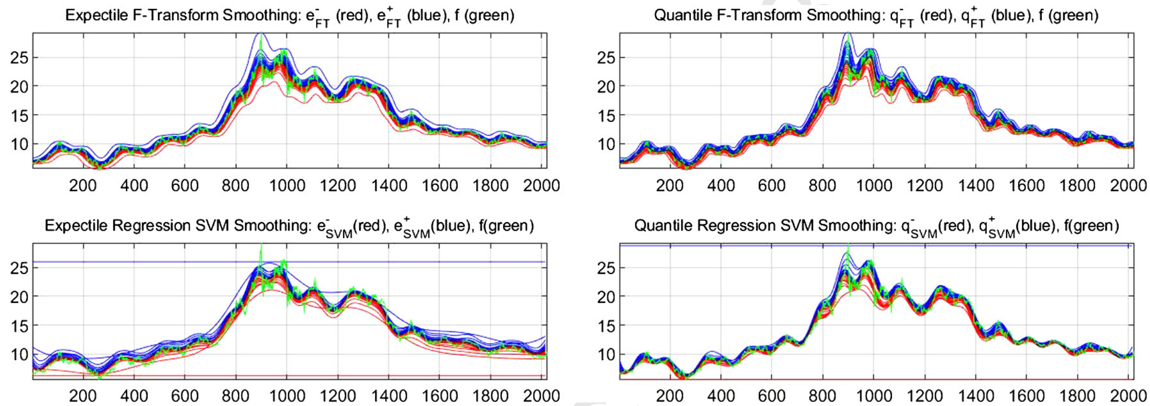

Fig. 10. (Silver time series) Left: fuzzy-valued expectile smoothing by methods expFT and exSVM. Right: fuzzy-valued quantile smoothing by methods quaFT and qtSVM.

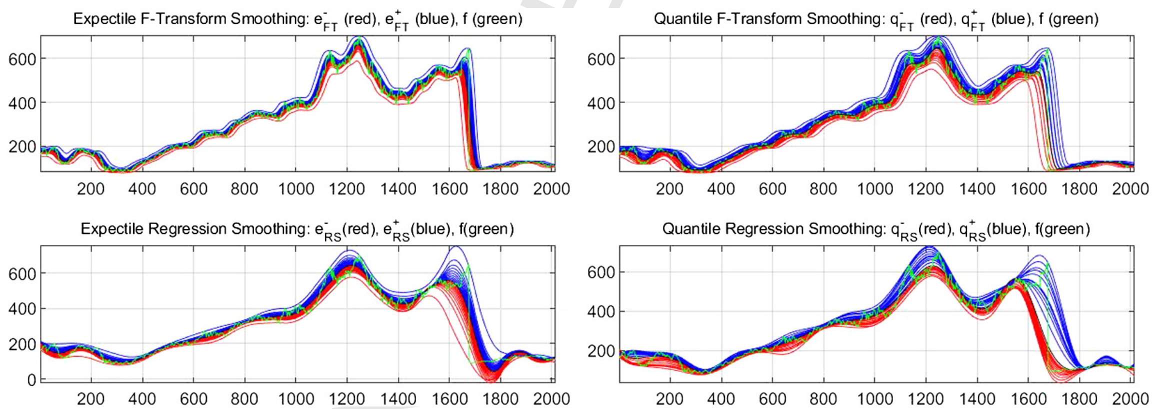

Fig. 11. (Apple time series) Left: fuzzy-valued expectile smoothing by methods expFT and expRS. Right: fuzzy-valued quantile smoothing by methods quaFT and quaRS.

With respect to the distance measure DIS, a measure of how much the fuzzy-valued series is adherent to the observed values, the F-transform performs significantly better than the other methods. E.g., in expectile smoothing we have $DIS(FT) = 1.38$ vs $DIS(RS) = 1.45$ for the *Silver* series (Table 7), $DIS(FT) = 25.19$ vs $DIS(RS) = 46.53$ for the *Apple* series (Table 8) and $DIS(FT) = 54.28$ vs $DIS(RS) = 61.97$ for the *S&P500* series (Table 9). In the quantile cases the behavior is similar, except for the *Silver* series where *quaRS* has a small advantage: $DIS(quaRS) = 1.73 < DIS(quaFT) = 1.74$.

Please cite this article in press as: M.L. Guerra et al., Quantile and expectile smoothing based on L_1-norm and L_2-norm fuzzy transforms, Int. J. Approx. Reason. (2019), https://doi.org/10.1016/j.ijar.2019.01.011

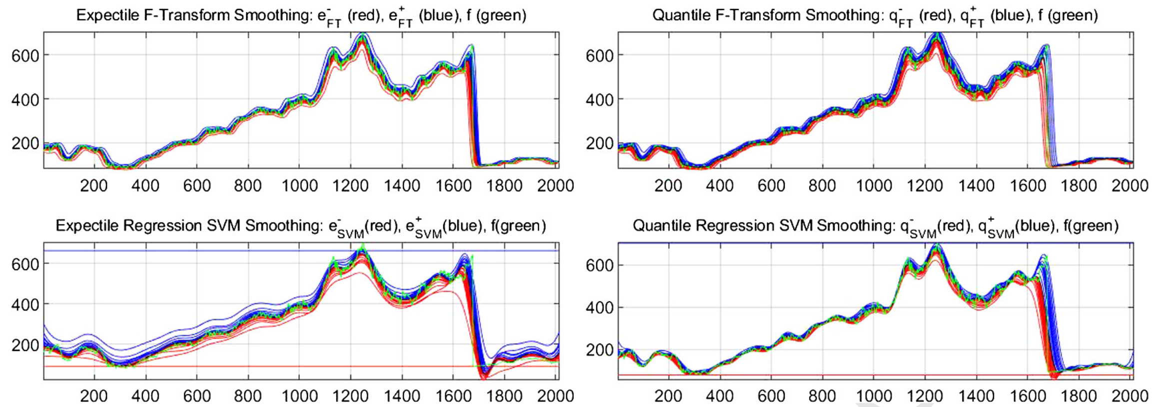

Fig. 12. (Apple time series) Left: fuzzy-valued expectile smoothing by methods expFT and exSVM. Right: fuzzy-valued quantile smoothing by methods quaFT and qtSVM.

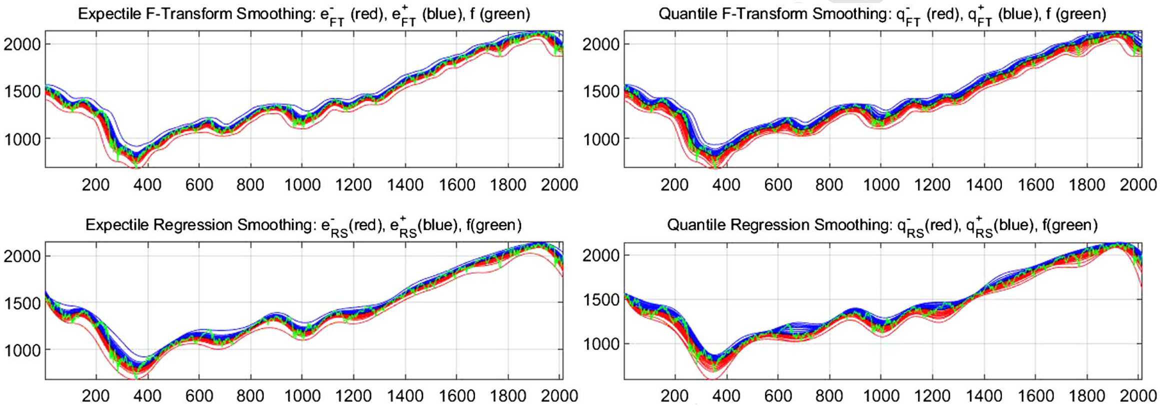

Fig. 13. (S&P500 time series) Left: fuzzy-valued expectile smoothing by methods expFT and expRS. Right: fuzzy-valued quantile smoothing by methods quaFT and quaRS.

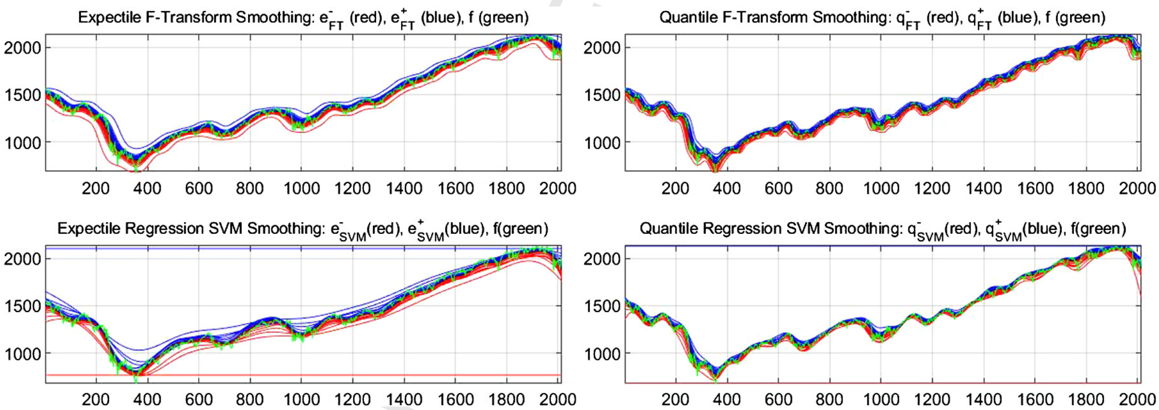

Fig. 14. (S&P500 time series) Left: fuzzy-valued expectile smoothing by methods expFT and exSVM. Right: fuzzy-valued quantile smoothing by methods quaFT and qtSVM.

Finally, with respect to the fuzzy ambiguity *AMB* and the central interval range *CIR*, which measure the spread or dispersion of the fuzzy-valued series, we can say that the F-transforms tend to construct less dispersed approximations than the other methods; this is particularly evident in the *Apple* case where both expectile and quantile F-transforms have the AMB and CIR smallest values.

As a conclusion of the computational comparison we can assert that expectile and quantile fuzzy-valued smoothing based on F-transform represent very promising tools, having good theoretical approximation properties and excellent empirical performance.

Please cite this article in press as: M.L. Guerra et al., Quantile and expectile smoothing based on L_1-norm and L_2-norm fuzzy transforms, Int. J. Approx. Reason. (2019), https://doi.org/10.1016/j.ijar.2019.01.011

7. Concluding remarks and further work

We introduced two new non-parametric smoothing methodologies, called expectile and quantile fuzzy-valued direct and inverse F-transform, the first one based on the classical direct F-transform obtained by minimizing a least squares (L_2-norm) operator, the second one based on the L_1-type direct F-transform, obtained by minimizing an L_1-norm operator.

We model three different time series belonging to disjoint classes of assets (one commodity, one stock and one index) in terms of fuzzy-valued functions supposing that level-cuts are deduced in the setting of expectile and quantile smoothing. Taking into account some standard measures to compare the smoothing techniques, we deduce that the expectile and quantile F-transform smoothed series perform very well in all the examined situations. As we have seen, the F-transforms of order $p = 0$ have the non-crossing property, but this is in general not true for $p > 0$, where the direct F-transform components F_k and G_k are functions (polynomials) instead of constants. An efficient procedure to preserve the non-crossing property for quantile and expectile F-transforms of order $p > 0$ can take advantage of additional non-crossing constraints (usually linear) introduced into the optimization problems outlined in sections 3 and 4, as is done e.g. in [57] for the quantile case.

Future research involves the investigation of relationships between probabilistic and fuzzy financial instruments for derivatives pricing and for risk management. More generally, the F-transform expectiles and quantiles contribute to the setting and analysis of fuzzy time series. In particular, it will be interesting to study possible connections of fuzzy-valued smoothing with volatility in markets, commonly measured as a standard deviation. We studied volatility models since the contribution [15], and in [14] we show their effects on financial options pricing. By adopting a fuzzy-valued approximation we can hopefully apply fuzzy logic and possibility theory to formulate or revisit volatility-like concepts; this idea is not new in the literature (see [3] and the references therein) and this area of investigation is of great interest.

Acknowledgements

The authors would like to thank the editors and the anonymous reviewers for their insightful and constructive comments and suggestions that have led to the actual improved version of the paper.

References

[1] A. Beck, Introduction to Nonlinear Optimization: Theory, Algorithms and Applications with Matlab, MOS-SIAM Series on Optimization, SIAM, 2014.
[2] B. Bede, Mathematics of Fuzzy Sets and Fuzzy Logic, Springer, 2013.
[3] A. Capotorti, G. Figà-Talamanca, On an implicit assessment of fuzzy volatility in the Black and Scholes environment, Fuzzy Sets Syst. 223 (2013) 59–71.
[4] C. Chesneau, I. Dewan, H. Doosti, Nonparametric estimation of a quantile density function by wavelet methods, Comput. Stat. Data Anal. 94 (2016) 161–174.
[5] J.F. Crouzet, Fuzzy projection versus inverse fuzzy transform as sampling/interpolation schemes, Fuzzy Sets Syst. 193 (2012) 108–121.
[6] J. Dan, F. Dong, K. Hirota, Fuzzy local trend transform based fuzzy time series forecasting model, Int. J. Comput. Commun. Control 4 (2011) 603–614.
[7] C. Davino, M. Furno, D. Vistocco, Quantile Regression, Theory and Applications, J. Wiley Series in Probability and Statistics, 2014.
[8] G. De Rossi, A. Harvey, Quantiles, expectiles and splines, J. Econom. 152 (2009) 179–185.
[9] M. Delgado, M.A. Vila, W. Voxman, On a canonical representation of fuzzy numbers, Fuzzy Sets Syst. 93 (1998) 125–135.
[10] F. Di Martino, V. Loia, S. Sessa, Fuzzy transforms method and attribute dependency in data analysis, Inf. Sci. 180 (2010) 493–505.
[11] F. Di Martino, V. Loia, S. Sessa, Fuzzy transforms for compression and decompression of color videos, Inf. Sci. 180 (2010) 3914–3931.
[12] F. Di Martino, S. Sessa, Time series seasonal analysis based on fuzzy transform, Symmetry 9 (281) (2017) 1–21, https://doi.org/10.3390/sym9110281.
[13] M. Farook, I. Steinwart, An SVM-like approach for expectile regression, Comput. Stat. Data Anal. 109 (2017) 159–181.
[14] G. Figà-Talamanca, M.L. Guerra, Fuzzy option value with stochastic volatility models, in: ISDA 2009 – 9th International Conference on Intelligent Systems Design and Applications, 2009, pp. 306–311, art. no. 5364838.
[15] M.L. Guerra, L. Sorini, Testing robustness in calibration of stochastic volatility models, Eur. J. Oper. Res. 163 (1) (2005) 145–153.
[16] M.L. Guerra, L. Stefanini, Expectile smoothing of time series using F-transform, in: Conference of the European Society for Fuzzy Logic and Technology, in: Advances in Intelligent Systems Research, vol. 32, ISBN 978-162993219-4, 2013, pp. 559–564.
[17] M. Holcapek, T. Tichy, A smoothing filter based on fuzzy transform, Fuzzy Sets Syst. 180 (2011) 69–97.
[18] R.J. Hyndman, Y. Fan, Sample quantiles in statistical packages, Am. Stat. 50 (4) (1996) 361–365.
[19] R.W. Koenker, Quantile Regression, Cambridge University Press, Cambridge, 2005.
[20] R.W. Koenker, R., Quantreg: quantile regression R package, CRAN.R-project.org repository, package quantreg, 2008.
[21] R.W. Koenker, V. d'Oray, Computing regression quantiles, Algorithm AS 229, Stat. Algorithms (1987) 383–393.
[22] R.W. Koenker, Z. Xiao, Quantile autoregression, J. Am. Stat. Assoc. 101 (2006) 980–990.
[23] H. Lee, H. Tanaka, Upper and lower approximation models in interval regression using regression quantile techniques, Eur. J. Oper. Res. 116 (3) (1999) 653–666.
[24] D. Li, L. Simar, V. Zelenyuk, Generalized nonparametric smoothing with mixed discrete and continuous data, Comput. Stat. Data Anal. 100 (2016) 424–444.
[25] W. Lin, G. González-Rivera, Interval-valued time series models: estimation based on order statistics exploring the Agriculture Marketing Service data, Comput. Stat. Data Anal. 100 (2016) 694–711.
[26] W. Lu, W. Pedrycz, X. Liu, J. Yang, P. Li, The modeling of time series based on fuzzy information granules, Expert Syst. Appl. 41 (8) (2014) 3799–3808.
[27] R.E. Moore, R.B. Kearfott, M.J. Cloud, Introduction to Interval Analysis, SIAM, 2009.
[28] W.K. Newey, J.L. Powell, Asymmetric least squares estimation and testing, Econometrica 55 (4) (1987) 819–847.
[29] V. Novák, I. Perfilieva, M. Holcapek, V. Kreinovich, Filtering out high frequencies in time series using F-transform, Inf. Sci. 274 (2014) 192–209.
[30] V. Novák, Fuzzy vs probabilistic techniques in time series analysis, in: L.H. Anh, L.S. Dong, V. Kreinovich, N. Thach (Eds.), Econometrics for Financial Applications, in: Studies in Computational Intelligence, vol. 760, Springer, 2018, pp. 213–234.
[31] V. Novák, I. Perfilieva, A. Dvorak, Insight into Fuzzy Modelling, J. Wiley & Sons, Hoboken, New Jersey, 2016.
[32] I. Perfilieva, Fuzzy transforms: theory and applications, Fuzzy Sets Syst. 157 (2006) 993–1023.

Please cite this article in press as: M.L. Guerra et al., Quantile and expectile smoothing based on L_1-norm and L_2-norm fuzzy transforms, Int. J. Approx. Reason. (2019), https://doi.org/10.1016/j.ijar.2019.01.011

[33] I. Perfilieva, Fuzzy transforms: a challenge to conventional transforms, in: P.W. Hawkes (Ed.), Advances in Images and Electron Physics, vol. 147, Elsevier Academic Press, 2007, pp. 137–196.
[34] I. Perfilieva, Fuzzy transforms and their applications to data compression, in: I. Bloch, et al. (Eds.), Fuzzy Logic and Applications, in: LNAI, vol. 3849, Springer, 2006, pp. 19–31.
[35] I. Perfilieva, F-transform, in: J. Kacprzyk, W. Pedrycz (Eds.), Springer Handbook of Computational Intelligence, 2015, pp. 113–130 (Chapter 7).
[36] I. Perfilieva, M. Daňková, Image fusion on the basis of fuzzy transform, in: D. Ruan, et al. (Eds.), Computational Intelligence in Decision and Control, World Scientific, 2008, pp. 471–476.
[37] I. Perfilieva, V. Novák, A. Dvorak, Fuzzy transform in the analysis of data, Int. J. Approx. Reason. 48 (2008) 36–46.
[38] I. Perfilieva, B. De Baets, Fuzzy transforms of monotone functions with applications to image compression, Inf. Sci. 180 (2010) 3304–3315.
[39] I. Perfilieva, M. Daňková, B. Bede, Towards a higher degree F-transform, Fuzzy Sets Syst. 180 (2011) 3–19.
[40] I. Perfilieva, M. Holcapek, V. Kreinovich, A new reconstruction from the F-transform components, Fuzzy Sets Syst. 288 (2016) 3–25.
[41] W. Rudin, Principles of Mathematical Analysis, third edition, McGraw-Hill, New York, 1976.
[42] K. Sabo, R. Scitovski, The best least absolute deviation line – properties and two different methods for its derivation, ANZIAM J. 50 (2008) 185–198.
[43] S.K. Schnabel, P.H.C. Eilers, Optimal expectile smoothing, Comput. Stat. Data Anal. 53 (2009) 4168–4177.
[44] J.S. Simonoff, Smoothing Methods in Statistics, Springer Series in Statistics, Springer Verlag, 1996.
[45] F. Sobotka, T. Kneib, Geoadditive expectile regression, Comput. Stat. Data Anal. 56 (2012) 755–767.
[46] F. Sobotka, T. Kneib, S. Schnabel, Expectreg: Expectile Regression R package, CRAN.R-project.org repository, package expectreg, 2010.
[47] L. Stefanini, Fuzzy transform with parametric LU-fuzzy partitions, in: D. Ruan, et al. (Eds.), Computational Intelligence in Decision and Control, World Scientific, 2008, pp. 399–404.
[48] L. Stefanini, Fuzzy transform and smooth functions, in: J.P. Carvalho, D. Dubois, U. Kaymak, J.M.C. Sousa (Eds.), Proceedings of the IFSA-EUSFLAT 2009 Conference, Lisbon, July 2009, pp. 579–584.
[49] L. Stefanini, F-transform with parametric generalized fuzzy partitions, Fuzzy Sets Syst. 180 (2011) 98–120.
[50] L. Stefanini, B. Bede, Generalized fuzzy differentiability with LU-parametric representation, Fuzzy Sets Syst. 257 (2014) 184–203.
[51] L. Stefanini, M.L. Guerra, On possibilistic representations of fuzzy intervals, Inf. Sci. 405 (2017) 33–54.
[52] L. Stefanini, L. Sorini, M.L. Guerra, Parametric representations of fuzzy numbers and application to fuzzy calculus, Fuzzy Sets Syst. 157 (2006) 2423–2455.
[53] L. Stefanini, L. Sorini, M.L. Guerra, Fuzzy numbers and fuzzy arithmetic, in: W. Pedrycz, A. Skowron, V. Kreynovich (Eds.), Handbook of Granular Computing, J. Wiley & Sons, 2009, Chapter 12.
[54] L. Stefanini, L. Sorini, M.L. Guerra, Time series modeling based on fuzzy transform, in: M.B. Ferraro, et al. (Eds.), Soft Methods for Data Science, in: Advances in Intelligent Systems and Computing, vol. 456, Springer International Publishing, 2017, pp. 463–470.
[55] I. Steinwart, P. Thomann, liquidSVM: a fast and versatile SVM package, arXiv:1702.06899v1, 2016.
[56] I. Steinwart, P. Thomann, Package liquidSVM, version 1.2.1, publication date 2017-07-12, CRAN Repository, 2017.
[57] I. Takeuchi, Q.V. Le, T.D. Sears, A.J. Smola, Nonparametric quantile estimation, J. Mach. Learn. Res. 7 (2006) 1231–1264.
[58] L.S. Waltrup, F. Sobotka, T. Kneib, G. Kauermann, Expectile and quantile regression – David and Goliath? Statistical Modelling, SAGE J. 15 (5) (2015) 433–456.
[59] Q. Yao, H. Tong, Asymmetric least squares regression estimation: a nonparametric approach, J. Nonparametr. Stat. 6 (1996) 273–292.
[60] L.A. Zadeh, Fuzzy sets, Inf. Control 8 (1965) 338–353.
[61] M. Zeinali, R. Alikhani, S. Shahmorad, F. Bahrami, I. Perfilieva, On the structural properties of F^m-transform with applications, Fuzzy Sets Syst. 342 (2018) 32–52.

Please cite this article in press as: M.L. Guerra et al., Quantile and expectile smoothing based on L_1-norm and L_2-norm fuzzy transforms, Int. J. Approx. Reason. (2019), https://doi.org/10.1016/j.ijar.2019.01.011