2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

Forecasting using F-transform based on Bootstrap technique

Woo-Joo Lee[1], Hye-Young Jung[1], Jin-Hee Yoon[2], Seung-Hoe Choi[3]

[1]Dept. of Mathematics, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, 120-749, South Korea.
[2]School of Mathematics and Statistics, Sejong University, 98 Gunja-dong, Gwangjin-gu, Seoul, 143-747, South Korea.
[3]School of Liberal Arts and Science, Korea Aerospace University, Koyang 411, South Korea.

***Abstract*—A new modified Fuzzy transform (F-transform) method which is combined with the bootstrap technique for forecasting is proposed in this paper. We apply the bootstrap technique to improve the accuracy of the F-transform method. An example is given to show the superior of proposed method.**

I. Introduction

We propose a new method to analyze data based on a combination of two methods : Fuzzy transform (F-transform) and bootstrap technique. The F-transform is first introduced by Perfilieva [9] in 2001. It has been studied by Perfilieva et al. [8-13], and applied to many practical problems such as the construction of approximate models, filtering, solution of differential equations, and data compression. From the practical point of view, it finds the estimated original data using the F-transform with respect to fuzzy sets of given domain identified by their membership functions. In a time series analysis and forecast, the use of the F-transform (in its inverse form) was reported in several studies, e.g. Novák et al. [8]; Di Martino et al. [6]; Troiano and Kriplani [15]; Štěpnička et al. [14]. In Novák et al. [8] and Štěpnička et al. [14], this technique was used to extract a low-frequency trend component, whereas in Di Martino et al. [6] it was used for the modeling of an autoregression function. In Troiano and Kriplani [15], the inverse F-transform was used as a technical indicator in a stock market instead of the commonly used simple and exponential moving averages. In Novák et al. [8] and Štěpnička et al. [14], the inverse F-transform was used in combination with perception-based logical deduction, where the latter provides forecasts of the future F-transform component(s). In this contribution, the bootstrapping technique is combined with F-transform. The bootstrap method is introduced by Efron [2-3], which has been widely used in various fields. It allows estimation of sampling distribution using resampling from observed dataset. This can be implemented by constructing a number of resamples of the observed dataset (and of equal size to the observed dataset), which is obtained by random sampling with replacement from the original dataset. It has been shown by many authors that the accuracy of statistical inference or performance of data analysis can be improved by bootstrap method [3-5,7]. The general fuzzy transform method use the F-transform with respect to fuzzy sets of given domain to obtain inverse F-transform of original data. The main idea of this contribution is applying bootstrap technique to find the cumulative distribution of F-transform by rearrangement of the residuals after we get the original F-transform. The forecasted value can be obtained by inverse F-transform using cumulative distribution of F-transform based on bootstap method. The forecasting values of out-of-sample data can be obtained when the input data is located in domain which is called universe.

This paper is organized as follows. In section II, we introduce some basic concepts and definitions regarding F-transform. In section III, the procedure for F-transform combined with bootstrap technique is proposed. We confirm that the accuracy of data forecasting can be improved by proposed method in section IV through an example.

II. Preliminaries

For any fuzzy set A, the function μ_A represents the membership function for which $\mu_A(x)$ indicates the degree of membership that x, of the universal set X, belongs to set A and is, usually, expressed as a number between 0 and 1, i.e $\mu_A(x) : X \to [0,1]$. The fuzzy sets can be either discrete or continuous.

A. F-transform

Here, we introduce some basic concepts and definitions in [11].

Definition 1
Let $x_1 < \cdots < x_n$ be fixed nodes within $[a,b]$, such that $x_1 = a, x_n = b$ and $n \geq 2$. We say that fuzzy sets $A_1 < \cdots < A_n$, identified with their membership functions $A_1(x) < \cdots < A_n(x)$ defined on $[a,b]$, constitute a fuzzy partition of $[a,b]$ if they fulfill the following conditions for $k = 1, \cdots, n$:
1) $A_k : [a,b] \to [0,1], A_k(x_k) = 1$;
2) $A_k(x) = 0$ if $x \notin (x_{k-1}, x_{k+1})$ where for the uniformity of denotation, we put $x_0 = a$ and $x_{n+1} = b$;
3) $A_k(x)$ is continuous;
4) $A_k(x)$, $k = 2, \cdots, n$, strictly increases on $[x_{k-1}, x_k]$ and $A_k(x)$, $k = 1, \cdots, n-1$, , strictly decreases on $[x_k, x_{k+1}]$
5) For all $x \in [a,b]$,

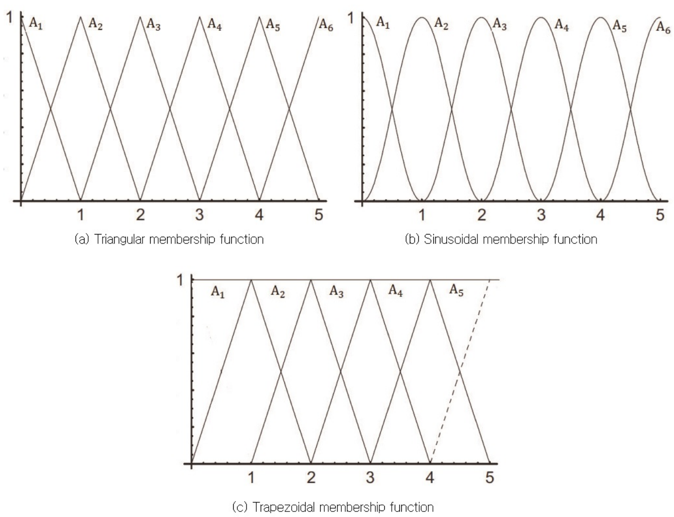
$$\sum_{k=1}^{n} A_k(x) = 1.$$

The membership functions $A_1, \cdots, A_n$ are called basic functions.

Let $[a, b]$ be an interval on $\mathbb{R}$ and $c_1 < \cdots < c_k$ be fixed nodes within $[a, b]$ such that $c_1 = a$, $c_k = b$. The assigned fuzzy sets $\mu_{A_i} : [a, b] \to [0, 1]$, identified with their membership functions(basic functions), fulfill the following conditions for all $l = 1, \cdots, k$

$$\text{(i) } \mu_{A_l}(c_l) = 1, \qquad \text{(ii) } \sum_{l=1}^{k} \mu_{A_l}(x) > 0.$$

F-transform can use several different basic functions μ_{A_l}. In contrast to Fourier and wavelet transforms, the basic functions μ_{A_l} in a fuzzy come from linguistic terms.



(a) Triangular membership function

(b) Sinusoidal membership function

(c) Trapezoidal membership function

Fig.1. Examples of basic functions

The example of fuzzy sets $A_1, \cdots, A_k$ with symmetric triangular membership functions on the interval $[a, b]$ is given below $(l = 1, \cdots, k)$:

$$\mu_{A_l}(x) = \begin{cases} 1 - \left| \dfrac{x - c_l}{h_l} \right| & x \in [c_{l-1}, c_{l+1}] \\ 0 & \text{otherwise} \end{cases}$$

where $h_l = c_{l+1} - c_l$, $c_0 = c_1$ and $c_{k+1} = c_k$.

Definition 2
Let a discrete function $f : X \to \mathbb{R}$ be given at a finite set of points $X = \{x_t : t = 1 \cdots, n\} \subseteq [a, b]$. The F-transform of a discrete function f with respect to $A_1, \cdots, A_k$ define the numerical vector $F_k[f] = [F_1, F_2, \cdots, F_k]$, where each F_l is given by

$$F_l = \frac{\sum_{t=1}^{n} f(x_t)\mu_{A_l}(x_t)}{\sum_{t=1}^{n} \mu_{A_l}(x_t)}, \qquad l = 1, \cdots, k. \tag{1}$$

The F_l are weighted mean values of f, where the weights are determined by the membership values. The F_l are called components of the discrete F-transform.

Definition 3
Let $F_k[f] = [F_1, \cdots, F_k]$ be the F-transform of f with respect to $A_1, \cdots, A_k$. Then the function

$$f_{F,k}(x_t) = \frac{\sum_{l=1}^{k} \mu_{A_l}(x_t)F_l}{\sum_{l=1}^{k} \mu_{A_l}(x_t)}, \qquad t = 1, \cdots, n. \tag{2}$$

is called the inverse F-transform of f.

The inverse F-transform $f_{F,k}$ can approximate f with an arbitrary precision. For various properties of the F-transform and detailed proofs, see [11].

B. Performance measures

The accuracy of the forecast can be evaluated on the basis of the average forecasting error percentage (AFEP) and the index of agreement(d) suggested in [16], which are defined as follows:

$$\text{AFEP} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{P_i - O_i}{O_i} \right| \tag{3}$$

$$d = 1 - \left[\frac{\Sigma(P_i - O_i)^2}{\Sigma(|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \right], \qquad 0 \le d \le 1, \tag{4}$$

where, N is the total number of data and O_i, and P_i are the observed and predicted data, respectively. $\bar{O}_i$ is the mean value of the observed load. The metric d quantifies the relative contribution of systematic error to random error and has a value of 1 in a perfect model [16].

III. F-TRANSFORM BASED ON BOOTSTRAP TECHNIQUE

The new procedure uses to fit original data not a fixed F-transform with respect to fuzzy set which constitute a fuzzy partition of domain but a cumulative distribution of F-transforms generated by bootstrap samples. The procedures of F-transform based on bootstrap technique is as follows.

Step 1. For a fixed decomposition $\mathbb{R}$ on the universe [a,b], generate fuzzy sets$(A_1, A_2, \cdots, A_k)$ by assigning the membership function.

Step 2. Calculate the F-transform $F_k[f] = [F_1, \cdots, F_k]$ with respect to $A_1, A_2, \cdots, A_k$, which is defined in (1).

Step 3. Calculate the inverse F-transform $f_{F,k}(x_t)(t = 1, \cdots, n)$ which is definded in (2).

Step 4. Generate bootstrap samples using the residuals $(\hat{\epsilon}_t = f(x_t) - f_{F,k}(x_t))$ of F-transform as follows: $\hat{f}^b(x_t) = f(x_t) + \hat{e}_t^b$ $(b = 1, \cdots, B)$ where $\hat{e}_t = \hat{\epsilon}_t - \bar{\hat{\epsilon}}_t$.

Step 5. Find $F_k^b[f] = [F_1^b, \cdots, F_k^b]$ with respect to $A_1, A_2, \cdots, A_k$. $(b = 1, \cdots, B)$.

$$F_l^b = \frac{\sum_{t=1}^{n} f^b(x_t)\mu_{A_l}(x_t)}{\sum_{t=1}^{n} \mu_{A_l}(x_t)}, \qquad l = 1, \cdots, k. \tag{5}$$

Step 6. Calculate the inverse F-transform $\hat{f}(x_t)$.

$$\hat{f}(x_t) = \frac{\sum_{l=1}^{k} \mu_{A_l}(x_t) F_l(\mu_{A_l}(x_t))}{\sum_{l=1}^{k} \mu_{A_l}(x_t)}, \qquad t = 1, \cdots, n. \tag{6}$$

We take $F_l(\mu_{A_l}(x_t))$ among one of the values of $F_l^b(b = 1, \cdots, B)$. For the choice of $F_l(\mu_{A_l}(x_t))$, after ordering $F_l^b(b = 1, \cdots, B)$, we take F_l^b when its cumulative ordered ratio coincides with $\mu_{A_l}(x_t)$.

In *Step 4*, the boostrap samples are obtained from rearrangements of residuals. For B bootstrap samples, rearrangements of residuals are performed B times.
In general F-transform method, when a new data $(x_t^*, f(x_t^*))$ is observed, the forecasted value $f_{F,k}(x_t^*)$ will be obtained from k F-transformed values $F_k[f] = [F_1, \cdots, F_k]$. But proposed method doesn't use fixed k values $F_k[f] = [F_1, \cdots, F_k]$, it finds a empirical distribution of F_k^b for each k, then it uses the membership degrees of A_l of the new data x_t^* to find the forecasted value $f_{F,k}(x_t^*)$.

IV. Example

Jet fuel production and flight distances data are used to compare the accuracy of the proposed F-transform method and the general F-transform method. The data used are monthly aviation fuel production of Turkey and Turkish airlines monthly flight distances, which are taken from [1]. Both time series data is given in table I.

TABLE I
JET FUEL PRODUCTION AND FLIGHT DISTANCES DATA

2005	Actual Productions	Fight Distance	2006	Actual Productions	Fight Distance	2007	Actual Productions	Fight Distance
1	109638	13095	1	112841	14710	1	158228	18020
2	117833	11250	2	111585	12637	2	169590	15026
3	129249	12868	3	140647	14456	3	196069	17270
4	155930	12789	4	175648	15913	4	209676	18745
5	170577	13410	5	185818	16685	5	199312	19378
6	176257	14649	6	213993	18226	6	218337	20287
7	202266	15909	7	215639	20316	7	262617	22123
8	204692	16646	8	237061	21256	8	231397	22013
9	198061	15609	9	202908	19932	9	236127	20828
10	190251	19994	10	193121	18494	10	213711	20253
11	176401	17775	11	172281	16442	11	227243	18252
12	165861	19607	12	157865	18135	12	235386	19951

For the fuzzy partition, 12 triangular basic functions (i.e., k=12) are used in given universe. For each k, 1000 iteration (i.e., B=1000) for F-transform based on bootstrap technique is repeated 1000 times.

AFEP and d are obtained taking average of 1000 repetitions. The proposed method performed better 787 times out of 1000 repetitions with respect to AFEP, and 1000 times out of 1000 repetitions performed better than the general method with respect to d. Table II shows a comparison of the average forecasting accuracy of the proposed method and the general method.

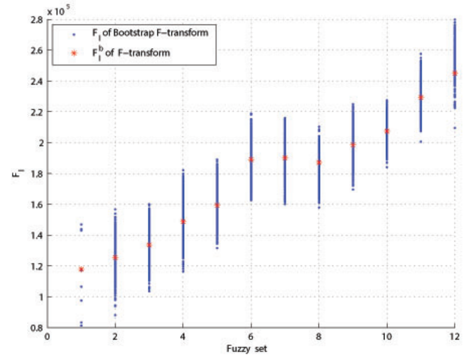


Fig.3. F-transformed values based on bootstrap technique

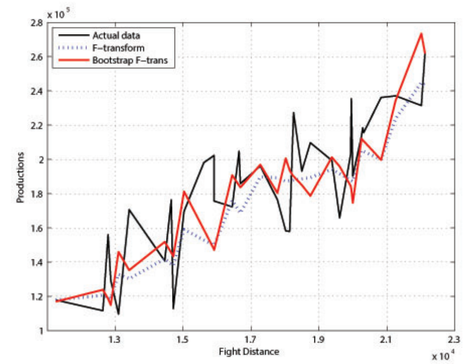


Fig.4. Actual data and estimated values

We can see that the forecasting results of the proposed method performs better than the forecasting results of the general method. The 1000 iterated F-transform based on bootstrap technique are shown in Figure 3. It is repeated 1000 times for each k, and the F-transform values by general method are marked in each iteration in Figure 3. Figure 4 shows the actual data and estimated values using general F-transform and F-transform based on bootstrap technique.
Next, we perform validations to ensure superiority of the proposed method. 35 data out of 36 are used as a training dataset for validation, and the removed data is used as a new data. Each data is removed when it is used as a new data and the other 35 data are used as a training dataset. Using the performance measures suggested in section II, AFEP and d with these 36 dataset, each of which excludes one data, for validation are repeated 1000 times, respectively. The average of 1000 AFEP, d are shown in Table III. AFEP(T) and d(T) shows the average of 1000 AFEP and d using 36 training dataset, each of which excludes one data. For AFEP in the last cell, it is AFEP of a new input as a new data which was

TABLE II
THE COMPARISON OF ACCURACY

	AFEP	d
The proposed method	0.1123995	0.8806755
F-transform method	0.1146653	0.8709050

excluded in each 36 training dataset. After taking average of 36 dataset, it is repeated 1000 times to get average of 1000 AFEP, which is is shown in Table II.

TABLE III
VALIDATION

	AFEP (T)	d (T)	AFEP
The proposed method	0.1127435	0.8833037	0.1198886
F-transform method	0.1143231	0.8714363	0.1252016

For the performances of the proposed model in table III, AFEP(T), the average of 1000 repetition of AFEP, is 0.113 and AFEP of a new input data is 0.120. Thus, we see that our new procedure of F-transform is superior to the general F-transform. In addition we can ensure that the proposed F-transform based on bootstrap technique have the better forecasting accuracy for new input data than the general F-transform method.

V. Conclusions

In this paper, F-transform combined bootstrap technique is proposed. The cumulative distributions are used to find the membership grade of input data. An example is suggested to show the performance of the proposed method. We confirm the superiority of proposed method through the average forecasting error percentage (AFEP) and the index of agreement (d). The prediction of out-of-sample will be focused in our further research, another combined method to improve accuracy will be proposed as well.

References

[1] K. Ecerkale, T Kücükíz and S. Esnaf, "Comparision of fuzzy time series based on difference parameters and two factor time-variant fuzzy time series models for aviation fuel production forecasting," *J. Aeronautics and space technologies*, Vol. 4, pp.57-63, 2010.

[2] B. Efron, "Bootstrap methods: Another look at the jackknife," *Ann. Statist.*,Vol. 7, pp.1-16, 1979.

[3] B. Efron, "The jackknife, the bootstrap, and other resampling plans," *Society of Industrial and Applied Mathematics CBMS-NSF Monographs*, 1982.

[4] B. Efron and R. Tibshirani, "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy", *Statistical Science*, Vol. 1, pp. 54-75, 1986.

[5] A.C. Davison and D.V. Hinkley, "Bootstrap methods and their application", *Cambridge University Press,* 1997.

[6] F. Di Martino, V. Loia and S. Sessa, "Fuzzy Transforms Method in Prediction Data Analysis," *Fuzzy Sets and Systems*, vol. 180, pp.146-163, 2011.

[7] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," *Proc. International Joint Conference on Artificial Intelligence (IJCAI)*, 1995.

[8] V. Novák, V., M. Štěpnička, A. Dvořák, I. Perfilieva and V. Pavliska, "Analysis of Seasonal Time Series Using Fuzzy Approach," *International Journal of General Systems*, vol. 39, pp.305328, 2010.

[9] I. Perfilieva and E. Haldeeva, "Fuzzy transformation," *Proc. of the Joint 9th IFSA World Congress and 20th NAFIPS Intl Conf.,* pp. 127-130, 2001,

[10] I. Perfilieva and R. Valášek, "Fuzzy transforms in removing noise, in: B.Reusch(Ed.), Computational Intelligence, Theory and Applications," *Advances in Soft Computing*, Springer, Berlin, pp. 221-230, 2005.

[11] I. Perfilieva, "Fuzzy transforms: theory and applications," *Fuzzy Sets and Systems*, Vol. 157, pp. 993-1023, 2006.

[12] I. Perfilieva, V. Novák, Antonín and A. Dvořák, "Fuzzy transform in the analysis of data," *International Journal of Approximate Reasoning*, Vol. 48, pp.36-46, 2008.

[13] I. Perfilieva, N. Yarushkina, T. Afanasieva and A. Romanov, "Time series analysis using soft computing methods," *International Journal of General Systems*, Vol. 42, No. 6, pp. 687-705, 2013.

[14] M. Štěpnička, A. Dvořák, V. Pavliska and L. Vavříčková, "Linguistic Approach to Time Series Modeling with the Help of F-Transform,"*Fuzzy Sets and Systems*, vol. 180, pp. 164184, 2011.

[15] L. Troiano and P. Kriplani, "Supporting Trading Strategies by Inverse Fuzzy Transform," *Fuzzy Sets and Systems*, 180, pp. 121-145, 2011.

[16] C. J. Willmott, "Some comments on the evaluation of model performance," *Bull. Amer. Meteorol. Soc.*, vol. 63, pp. 1309-1313, 1982.