Forecasting of Egypt Wheat Imports Using Multivariate Fuzzy Time Series Model Based on Fuzzy Clustering

Ashraf K. Abd-Elaal, *Member, IAENG,* Hesham A. Hefny, and Ashraf H. Abd-Elwahab

Abstract— **This paper presents Multivariate-Factors fuzzy time series model for improving forecasting accuracy. The proposed model is based on fuzzy clustering and it employs eight main procedures to build the multivariate-factors model. The model is evaluated by studying the Egypt Wheat imports as a forecasting problem. Forecasting Egypt wheat imports depend on three factors: population size, wheat area, and wheat production. This forecasting problem is considered to be a good benchmark for comparing different forecasting techniques since it exhibits highly nonlinearities over a long period of time and it provides important economical indicators needed for national future planning. Experimental results show that the proposed model provides higher forecasting accuracy than ARIMA model, Regression model and neural network model. Therefore, the proposed model can lead to satisfactory high performance for fuzzy time series.**

Index Terms—**ARIMA, Egypt wheat imports, Fuzzy clustering, Multivariate-Factors fuzzy time series.**

I. Introduction

THE uncertainties existed in historical data represent a real challenge for traditional time series forecasting models. The main motivation for using Fuzzy time series forecasting models is their abilities to handle such uncertainties in historical data of real-world forecasting problem. Song and Chissom presented the first concept of fuzzy time series model for forecasting the enrolments of the University of Alabama based on the fuzzy set theory [1]-[2]. Since then, many researchers have contributed to developing and improving fuzzy time series models.

Qiu et al. presented a new method to generalize the conventional models for forecasting process, where the data of the University of Alabama and Shanghai stock index are adopted to illustrate the processes [3]. Bahrepour et al. presented a novel approach for high-order fuzzy time series. Their model was based on two facets. First was the use of a self-organizing map (SOM), as a fast clustering technique, to partition the universe of discourse unequally. The second facet of their model was the adoption of three different agents; voting, statistical and emotional, for estimating the best order of the high-order fuzzy time series model [4]. Abd-Elaal et al. introduced a fuzzy time series model, which depended on fuzzy clustering for partitioning the universe of discourse to forecast the Gold Reserves of Egypt based on official data starting from the first quarter of 2002 up to the first quarter of 2010. The comparison result, with other fuzzy time series models as well as the traditional ARIMA model, showed that the proposed model provided a higher accuracy and an efficient performance [5]. Liu et al. presented an approach to improve the derivation of fuzzy relationships in the fuzzy time series model using rough sets. Their proposed model, not only required no prior knowledge or pre-review dataset to generate heuristic rules, but also, effectively reduced computational effort by decreasing the numbers of fuzzy sets of linguistic variables [6]. Duru presented a study that aimed to improve the fuzzy logical forecasting model by utilizing multivariate inference. The model also allows handling the partitioning problem for an exponentially distributed time series by using a multiplicative clustering approach [7]. Hassan et al. presented a hybrid fuzzy time series model, based on Interval Type-2 Fuzzy Inference System (IT2-FIS) and ARIMA model. The model improved the forecasting result by handling the measurement and parametric uncertainties of ARIMA model by using Fuzzy approach [8]. Egrioglu et al. presented a novel hybrid fuzzy time series approach in which fuzzy c-means (FCM) method and artificial neural networks were employed for fuzzification and defuzzification, respectively. The model has successfully been applied to the well known enrollment data for the University of Alabama [9]. Khiabani et al. proposed combination of the adaptive time-variant model (ATVF) with PSO algorithm to improve Alabama University enrollments forecasting, ATVF model automatically adapts the analysis window size of fuzzy time series based on the predictive accuracy in the training phase and uses heuristic rules to determine forecasting values [10].

In this paper, researchers introduce multivariate-factors fuzzy time series forecasting model based on fuzzy clustering to handle real-world multivariate forecasting problems. The proposed model is examined by the problem of forecasting Egypt Wheat imports based on the official data of Central Agency for Public Mobilization and Statistics (CAPMAS).

II. Background

A. ARIMA Model

For Stationary time series, i.e. time series with constant mean, the ARMA (Autoregressive Moving Average) forecasting model is widely used. The ARMA(p,q) model consists of two models, Autoregressive (AR(p)) and moving average (MA(q)) models. The forecasting of a time series

Ashraf Abd-Elaal is a Senior Lecture at Computer and Information Sciences Department. HICS, Al-Kawser, Sohag, Egypt, fax:+20932280234 (ashrafsohag@yahoo.com).

Hesham A. Hefny is a Professor at Department of Computer and Information Sciences, ISSR, Cairo University, Egypt (hehefny@cu.edu.eg).

Ashraf Abd-Elwahab is a Professor at Department of Computer Sciences, ERI, National Center for Research, Egypt (elazazyashraf@yahoo.com).

(Advance online publication: 29 November 2013)

using the ARMA model is given in (1).

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \xi_t + \theta_1 \xi_{t-1} + \theta_2 \xi_{t-2} + \ldots + \theta_q \xi_{t-q} \quad (1)$$

Where ξ_t is the error, ϕ_p, θ_q are parameters for ARMA model. This model can be used when data are stationary. However, it can't be used in case of non-stationary data, i.e. Stationary means that there is no growth or decline in the data. That means that both mean and variance remain constant over time, then the forecasting model will be ARIMA (autoregressive integrated moving average) mode, ARIMA(p,d,q) where *p* is the order of the autoregressive part, *d* is the degree of the difference and *q* is the order of the moving average part [11]. For instance, ARIMA(4,1,2) model has four autoregressive parameters, two moving-average parameters, computed after the series have a difference of order one. The ARIMA model is a widely used time series model that represents the basis of many fundamental ideas in time-series analysis.

B. Neural Networks

Neural network (NN) is a computational structure inspired by the study of biological neural processing. NN consists of many processing units, called neurons. Each neuron has multi-inputs and a single output. The inputs emulate the external signals received by the biological neurons. The output represents the neuron response according to the accumulated weighted input signals. The weights representing the force of the synaptic union: positive weight representing an excitatory effect, and negative weight which is an inhibitory effect. If the result of the sum of weighted inputs is higher than a certain threshold value, the neuron is activated providing a positive value (normally +1); in the opposite case, the output presents a zero or a negative value (normally -1). Fig. 1 shows the schematic representation of an artificial neuron. [12]. Due to their powerful learning capabilities, neural networks have been adopted by many researchers for solving several forecasting problems [13]-[17].

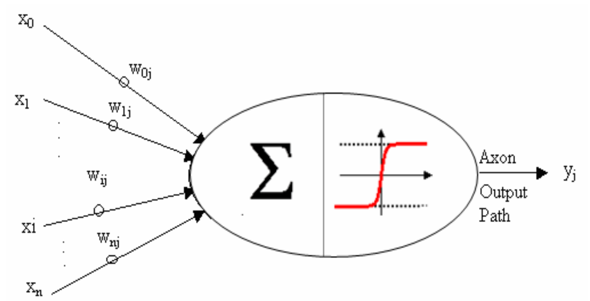

Fig. 1. Schematic representation of an artificial neuron

C. Fuzzy Clustering

Fuzzy clustering aims at partitioning a data set into homogeneous fuzzy clusters [18]. Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to more than one cluster with different degrees of membership values. Fuzzy C-Mean Iterative (FCMI) is a famous implementation of fuzzy clustering algorithms [19]. Assuming a set of m patterns $X=(x_1, x_2, \ldots, x_m)$ is distributed in d-dimensional pattern space and c fuzzy clusters, whose centers have initial values $y_{10}, y_{20}, \ldots, y_{c0}$. At each iteration (K), the distance d_{ij} between each pattern x_i and each cluster c_j is computed by:

$$d_{ij}^{(k)} = \left\| x_j - y_i^{(k)} \right\| \quad (2)$$

Fuzzy clustering is carried out through an iterative optimization of the objective function d_{ij}, with the update of membership u_{ij} and the cluster centers y_i by:

$$y_i^{(k+1)} = \sum_{j=1}^{m} u_{ij}^{(k)} x_j \Big/ \sum_{j=1}^{m} u_{ij}^{(k)} \quad (3)$$

Where $u_{ij}^{(k)} = \left[\sum_{l=1}^{c} \left(d_{ij}^{(k)} \Big/ d_{lj}^{(k)} \right)^{2/(\beta-1)} \right]^{-1}$

The process terminates when the difference between two consecutive clusters centers do not exceed a given tolerance,

$$\left[\sum_{i=1}^{c} \left\| y_i^{(k+1)} - y_i^{(k)} \right\|^2 \right]^{1/2} < \varepsilon \quad (4)$$

D. Fuzzy Time Series

Song and Chissom presented the concept of fuzzy time series (FTS) for allowing solving forecasting problems of considerably short time series with uncertainties. They presented the time-invariant fuzzy time series model and the time-variant fuzzy time series model based on the fuzzy set theory for forecasting the enrollments of the University of Alabama. The following are definitions of basic concepts of the fuzzy time-series [1][5][20]-[24].

Definition 1. Assume Y(t) (t = . . 0, 1, 2, . . .) is a subset of a real numbers. Let Y(t) is the universe of discourse defined by the fuzzy set $f_i(t)$. If F(t) is a collection of $f_1(t)$, $f_2(t)$. . . then F(t) is defined as a FTS on Y(t) (t = . . . , 0, 1, 2, . . .).

Definition 2. If there is a fuzzy logical relationship R(t − 1, t), such that F(t) = F(t − 1) o R(t − 1, t), where "o" represents a max-min composition operation, then F(t) is induced by F(t − 1). The fuzzy logical relationship (FLR) between F(t) and F(t − 1) is denoted by F(t − 1) → F(t).

Definition 3. Suppose F(t − 1) = A_i and F(t) = A_j where A_i and A_j are two fuzzy sets defined on Y(t). The relationship between two consecutive observations, F(t) and F(t − 1), referred to as a FLR, can be denoted by $A_i \rightarrow A_j$, where A_i is called the Left-Hand Side (LHS) and A_j is called the Right-Hand Side (RHS) of the FLR.

Definition 4. All fuzzy logical relationships (FLRs) in the training dataset can be grouped together into different fuzzy logical relationship groups according to the same LHS of the FLR. For example, if there are two fuzzy logical relationships with the same LHS (A_i) as: $A_i \rightarrow A_{j1}$ and $A_i \rightarrow A_{j2}$, then, these two FLRs can be grouped into a fuzzy logical relationship group (FLRG) as: $A_i \rightarrow A_{j1}\ A_{j2}$.

Definition 5. IF F(t) is FTS that is caused by F(t-1), F(t-2), ….and F(t-n), then F(t) is called "$n^{\underline{th}}$ order FTS" with " $n^{\underline{th}}$ order fuzzy logical relation" represented as: F(t − n), ….., F(t-2), F(t-1) → F(t).

Definition 6. Let $F_1(t)$ and $F_2(t)$ be two FTS. If F1(t) is caused by ((F_1(t-1), F_2(t-1)), ((F_1(t-2), F_2(t-2)), ….and ((F_1(t-n), F_2(t-n)), then F(t) is called "two-factors n^{th} order FTS" with "n^{th} order fuzzy logical relationships represented as" ((F_1(t-1), F_2(t-1)), ((F_1(t-2), F_2(t-2)), …. and ((F_1(t-n), F_2(t-n)) → F(t).

III. METHODOLOGY

A. The Proposed Multivariate-Factors Model

If F(t) is caused by (F_1(t-1), F_2(t-1),…F_m(t-1)), (F_1(t-2), F_2(t-2),…F_m(t-2)), … (F_1(t-n), F_2(t-n),…F_m(t-n)), then the FLR is represented by

$$((F_1(t-1), F_2(t-1),\ldots F_m(t-1)),\; (F_1(t-2), F_2(t-2),\ldots F_m(t-2)),\; \ldots\; (F_1(t-n), F_2(t-n),\ldots F_m(t-n))) \rightarrow F(t) \tag{5}$$

This is called Multivariate-factors fuzzy time series model (MFFTS), where (F_1(t-n), F_2(t-n),…F_m(t-n)) are called antecedents factors, and F(t) is called consequent factor.

1) The Eight Steps of the Proposed Model

The stepwise computational procedure of the proposed MFFTS model is explained in the following eight steps:

Step 1. Cluster antecedents and consequent factors data into c_j clusters: Clustering (F_1(t-1), F_2(t-2),…F_m(t-n)), (F_1(t-2), F_2(t-2),…F_m(t-2)), … (F_1(t-n), F_2(t-n),…F_m(t-n)) and F(t) with n observation into c_j $(2 \le c_j \le n)$ clusters.

At iteration k=0, initialize $Y_{i,j} = y_{i,j,0}$, $1 \le i \le c_j$

$$Y_{i,j,0} = D_{j,min} / (a * i) \tag{6}$$

Where, $D_{j,min}$ is the minimum value of j factor, c_j is the number of clusters of j factor and *a* is a positive integer .

Step 2. Determine membership values for each cluster: In this step, the proposed model selected the maximum membership grade of each value for each cluster for antecedents and consequent factors which it belongs to.

Step 3. Define the Universe of Discourse: In this step, the proposed model defines the universe of discourse V_j as.

$$V_j = [V_{j,min} - V_{j,down},\; V_{j,max} + V_{j,up}] \tag{7}$$

Where, $V_{j,min}$ is the minimum value of j factor, $V_{j,max}$ is the maximum value of j factor, and $V_{j,down}$, $V_{j,up}$ are the positive real numbers to divide V_j into c_j intervals.

Step 4. Partition the universe of discourse: According to this step, the proposed model partitions the universe of discourse of antecedents and consequents factors into c_j intervals.

Step 5. Fuzzify the historical data: Proposed model fuzzifies historical data, by determining the best fuzzy cluster to each actual data for antecedents and consequents factors.

Step 6. Calculate the crisp value for each linguistic term: The crisp value that represents each linguistic term is calculated by:

$$Crisp\_value(A_{i,j}) = (mg_{1,j}\, x\, X_{1,j} + mg_{2,j}\, x\, X_{2,j} + \ldots + mg_{n,j}\, x\, X_{n,j}) \Big/ \sum_{r=1}^{n} mg_{r,j} \tag{8}$$

Where $mg_{n,j}$ is the membership grade, and $X_{n,j}$ is the actual value.

Step 7. Re-fuzzify of historical data: Linguistics with highly frequency of occurrence is selected for further partitioning. The interval corresponding to linguistic with highest frequency of occurrence is re-clustered into k clusters, where $2 \le k \le n\text{-}k$. The interval corresponding to the next linguistic term with highest frequency of occurrence is re-clustered to k-1 clusters, and so on. Thus, historical data can be fuzzified again using a larger number of linguistics. The crisp values, which were produced in this step after re-fuzzification of the historical data, are the required forecasting value for antecedents and consequent factors.

Step 8. Determine the forecasted values of consequent factor: The proposed model uses the fuzzy logical relationship between the forecasting values of the antecedent's factors and the forecasting values of consequent factor to calculate the forecasting values.

$$Forecasting\,(A_i) = \left(\sum_{k=1}^{n} \frac{\sum_{r=0}^{n-1} Crisp\_value(X_{k,(i-r)})\, x\, Crisp\_value(Y_{i-r})}{\sum_{r=0}^{n-1} Crisp\_value(X_{k,(i-r)})} \right) \Big/ n \tag{9}$$

Where n is the order degree

IV. EXPERIMENTAL RESULTS

To examine the above proposed model, we use the problem of forecasting "Egypt Wheat imports" as the experimental test problem. Egypt Wheat imports depend on three factors: population size, wheat cultivated area, and wheat production. If population size increases, then wheat imports should be increased too. If wheat production increases or wheat area increases, then wheat imports should be decreased. Thus, there are many relationships among these factors that may be positive or negative relationship. In our experimental study, we use the official data provided by the Central Agency for Public Mobilization and Statistics (CAPMAS), for "Egypt Wheat imports", population size, wheat cultivated area and Wheat production during the period starting from year 1986 up to 2008.

The forecasting accuracy is compared by using the "Normalized Root Mean Squared Error" (NRMSE), and the "Normalized Root Mean Squared Error" (NMSE) performance indices.

NRMSE, in statistics, is the square root of the sum of the squared deviations between actual and predicted values divided by the sum of the square of actual values. The NRMSE is often expressed in units of percent. Smaller values indicate a better agreement between measured and calculated values.

$$NRMSE = \sqrt{\sum_{i=1}^{N} (actual_i - predict_i)^2 \Big/ \sum_{i=1}^{N} (actual_i)^2} \tag{10}$$

NMSE, in statistics, is the sum of the squared deviations between actual and predicted values divided by the sum of the squared deviations between actual and mean of the actual values.

$$NMSE = \sum_{i=1}^{N}(actual_i - predict_i)^2 \Big/ \sum_{i=1}^{N}(actual_i - mean\ actual)^2 \quad (11)$$

A. Testing the Multivariate-Factors Proposed Model

According to Table I. the universe of discourse of wheat imports is found to be U=[3011,8173] with D_{down} =72, and D_{up}=100, then U is partitioned into 9 intervals. The universe of discourse of population size is found to be V_1=[47677,57315] with $V_{1,down}$=74, and $V_{1,up}$=90, then V_1 is partitioned into 9 intervals. The universe of discourse of wheat cultivated area is V_2=[1174,3129] with $V_{2,down}$=32, and $V_{2,up}$=65, then V_2 is partitioned into 9 intervals. Finally, the universe of discourse of wheat production is V_3=[1832,8425] with $V_{3,down}$=97, and $V_{3,up}$=151. Then, V_3 is partitioned into 9 intervals, with see Fig. 2.

TABLE I
EGYPT'S PRODUCTION, IMPORT, AND AREA OF WHEAT AND POPULATION FROM 1986 UP 2008

Years	Wheat Imports (000 tons)	Population (000 people)	Wheat Area (000 feddans)	Wheat Production (000 tons)
1986	6811	47751	1206	1929
1987	6874	48816	1373	2722
1988	6890	49826	1421	2839
1989	7527	50858	1532	3183
1990	7712	51911	1955	4268
1991	5550	52985	2215	4483
1992	5496	54082	2092	4618
1993	6118	55201	2171	4786
1994	4509	56344	2111	4437
1995	6256	57642	2512	5722
1996	5868	58835	2421	5735
1997	5202	60053	2486	5849
1998	5817	61296	2421	6093
1999	4776	62564	2380	6347
2000	3930	63860	2463	6564
2001	4982	65182	2342	6255
2002	4819	66531	2450	6625
2003	5205	67908	2506	6845
2004	3083	69313	2605	7178
2005	6272	70748	2985	8141
2006	5372	72212	3064	8274
2007	8073	73608	2716	7379
2008	7381	75225	2920	7977

Hence, the intervals of Egypt Wheat imports are u_1; u_2; u_3; u_4; u_5; u_6; u_7; u_8, u_9 where:

u_1=[3011.00, 3584.56] u_2=[3584.56, 4158.11],
u_3=[4158.11, 4731.67], u_4=[4731.67, 5305.22],
u_5=[5305.22, 5878.78], u_6=[5878.78, 6452.33],
u_7=[6452.33, 7025.89], u_8=[7025.89, 7599.44],
u_9=[7599.44, 8173.00]

A linguistic variable named "Wheat imports" (Y) is defined to have the following linguistic values: A_1 = (very very very few), A_2 = (very very few), A_3 = (very few), A_4 =

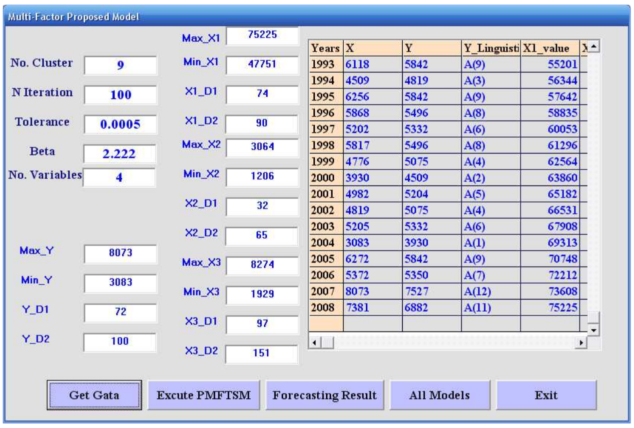

Fig. 2. Execute multi- factors proposed model at Egypt production, import, and area of wheat and population

(few), A_5 = (moderate), A_6 = (many), A_7 = (many many), A_8 = (many many many), A_9 = (too many). The proposed model fuzzifies the historical data by assigning the corresponding linguistic values as shown in Table VI. Table II shows the membership grade of wheat import actual values, which affect on linguistic values. The proposed model selects the maximum membership grade for each cluster. The representing crisp value for each cluster is calculated as follows:

$$Crisp\_value(A_1) = 0.7 \text{ x } (2000)/0.7 = 3930$$

$$Crisp\_value(A_2) = 1.0\text{x } (1994)/1.0 = 4509$$

…

$$Crisp\_value(A_9) = (1.0x\,(1987) + 1.0x(1988))/2.0 = 6882$$

TABLE II
DATA OF WHEAT IMPORT AND ITS MEMBERSHIP GRADES FROM 1986 UP 2008.

Year	Y	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9
1986	6811	0	0	0	0	0	0	0	0	0
1987	6874	0	0	0	0	0	0	0	0	1
1988	6890	0	0	0	0	0	0	0	0	1
1989	7527	0	0	0	0	0	0	0	1	0
1990	7712	0	0	0	0	0	0	0	0	0
1991	5550	0	0	0	0	0	0	0	0	0
1992	5496	0	0	0	0	1	0	0	0	0
1993	6118	0	0	0	0	0	0	0	0	0
1994	4509	0	1	0	0	0	0	0	0	0
1995	6256	0	0	0	0	0	0	1	0	0
1996	5868	0	0	0	0	0	1	0	0	0
1997	5202	0	0	0	1	0	0	0	0	0
1998	5817	0	0	0	0	0	1	0	0	0
1999	4776	0	0	0	0	0	0	0	0	0
2000	3930	0.	0	0	0	0	0	0	0	0
2001	4982	0	0	0	0	0	0	0	0	0
2002	4819	0	0	1	0	0	0	0	0	0
2003	5205	0	0	0	1	0	0	0	0	0
2004	3083	0	0	0	0	0	0	0	0	0
2005	6272	0	0	0	0	0	0	0	0	0
2006	5372	0	0	0	0	0	0	0	0	0
2007	8073	0	0	0	0	0	0	0	0	0
2008	7381	0	0	0	0	0	0	0	0	0

The intervals of population number are v_{11}; v_{12}; v_{13}; v_{14};

v_{15}; v_{16}; v_{17}; v_{18}, v_{19} where:

v_{11}=[47677.00, 50747.89] v_{12}=[50747.89, 53818.78],
v_{13}=[53818.78, 56889.67], v_{14}=[56889.67, 59960.56],
v_{15}=[59960.56, 63031.44], v_{16}=[63031.44, 66102.33],
v_{17}=[66102.33, 69173.22], v_{18}=[69173.22, 72244.11],
v_{19}=[72244.11, 75315.00]

A linguistic variable named "population size" (X_1), is defined to have the following linguistic values which are defined as: B_{11} = (very very very few), B_{12} = (very very few), B_{13} = (very few), B_{14} = (few), B_{15} = (moderate), B_{16} = (many), B_{17} = (many many), B_{18} = (many many many), B_{19} = (too many). The proposed model fuzzifies the historical data by assigning the corresponding linguistic values as shown in Table VI. Table III shows the membership grade of population actual values, which affect on linguistic values. The proposed model selects the maximum membership grade for each cluster. The representing crisp value for each cluster calculating as:

$$Crisp\_value(B_{11}) = (0.9\ x\ (1988) + 0.9\ x\ (1989))/1.8 = 55936$$

$$Crisp\_value(B_{12}) = 1.0x\ (1991)/1.0 = 52985$$

...

$$Crisp\_value(B_{19}) = 1.0x\ (2006)/1.0 = 72212$$

TABLE III
POPULATION AND ITS MEMBERSHIP GRADES FROM 1986 UP 2008.

Year	X_1	B_{11}	B_{12}	B_{13}	B_{14}	B_{15}	B_{16}	B_{17}	B_{18}	B_{19}
1986	47751	0	0	0	0	0	0	0	0	0
1987	48816	0	0	0	0	0	0	0	0	0
1988	49826	0.9	0	0	0	0	0	0	0	0
1989	50858	0.9	0	0	0	0	0	0	0	0
1990	51911	0	0	0	0	0	0	0	0	0
1991	52985	0	1	0	0	0	0	0	0	0
1992	54082	0	0	0	0	0	0	0	0	0
1993	55201	0	0	0.8	0	0	0	0	0	0
1994	56344	0	0	0.8	0	0	0	0	0	0
1995	57642	0	0	0	0	0	0	0	0	0
1996	58835	0	0	0	0.9	0	0	0	0	0
1997	60053	0	0	0	0	0	0	0	0	0
1998	61296	0	0	0	0	0.9	0	0	0	0
1999	62564	0	0	0	0	0	0	0	0	0
2000	63860	0	0	0	0	0	0.9	0	0	0
2001	65182	0	0	0	0	0	0	0	0	0
2002	66531	0	0	0	0	0	0	0	0.9	0
2003	67908	0	0	0	0	0	0	0	0	0
2004	69313	0	0	0	0	0	0	0.9	0	0
2005	70748	0	0	0	0	0	0	0	0	0
2006	72212	0	0	0	0	0	0	0	0	1
2007	73608	0	0	0	0	0	0	0	0	0
2008	75225	0	0	0	0	0	0	0	0	0

The intervals of Wheat area, are v_{21}; v_{22}; v_{23}; v_{24}; v_{25}; v_{26}; v_{27}; v_{28}, v_{29} where:

v_{21}=[1174.00, 1391.22] v_{22}=[1391.22, 1608.44],
v_{23}=[1608.44, 1825.67], v_{24}=[1825.67, 2042.89],
v_{25}=[2042.89, 2260.11], v_{26}=[2260.11, 2477.33],
v_{27}=[2477.33, 2694.56], v_{28}=[2694.56, 2911.78],
v_{29}=[2911.78, 3129.00]

A linguistic variable named "Wheat area" (X_2), is defined to have the following linguistic values which are defined as: B_{21} = (very very very few), B_{22} = (very very few), B_{23} = (very few), B_{24} = (few), B_{25} = (moderate), B_{26} = (many), B_{27} = (many many), B_{28} = (many many many), B_{29} = (too many). The proposed model fuzzifies the historical data by assigning the corresponding linguistic values as shown in Table VI. Table IV shows the membership grade of Wheat area actual values, which affect on linguistic values.

TABLE IV
DATA OF WHEAT AREA AND ITS MEMBERSHIP GRADES FROM 1986 UP 2008.

Year	X_2	B_{21}	B_{22}	B_{23}	B_{24}	B_{25}	B_{26}	B_{27}	B_{28}	B_{29}
1986	1206	0	0	0	0	0	0	0	0	0
1987	1373	1	0	0	0	0	0	0	0	0
1988	1421	1	0	0	0	0	0	0	0	0
1989	1532	0	0	0	0	0	0	0	0	0
1990	1955	0	0.7	0	0	0	0	0	0	0
1991	2215	0	0	0	0	0	0	0	0	0
1992	2092	0	0	1	0	0	0	0	0	0
1993	2171	0	0	0	1	0	0	0	0	0
1994	2111	0	0	0	0	0	0	0	0	0
1995	2512	0	0	0	0	0	0	0	0	0
1996	2421	0	0	0	0	0	1	0	0	0
1997	2486	0	0	0	0	0	0	1	0	0
1998	2421	0	0	0	0	0	1	0	0	0
1999	2380	0	0	0	0	0	0	0	0	0
2000	2463	0	0	0	0	0	0	0	0	0
2001	2342	0	0	0	0	1	0	0	0	0
2002	2450	0	0	0	0	0	0	0	0	0
2003	2506	0	0	0	0	0	0	0	0	0
2004	2605	0	0	0	0	0	0	0	1	0
2005	2985	0	0	0	0	0	0	0	0	0
2006	3064	0	0	0	0	0	0	0	0	0
2007	2716	0	0	0	0	0	0	0	0	0
2008	2920	0	0	0	0	0	0	0	0	1

The proposed model selects the maximum membership grade for each cluster. The representing crisp value for each cluster calculating as:

$$Crisp\_value(B_{21}) = (1.0\ x\ (1987) + 1.0\ x\ (1988))/2.0 = 1397$$

$$Crisp\_value(B_{22}) = 0.7x\ (1990)/0.7 = 1955$$

...

$$Crisp\_value(B_{29}) = 1.0x\ (2008)/1.0 = 2920$$

The intervals of Wheat production are v_{31}; v_{32}; v_{33}; v_{34}; v_{35}; v_{36}; v_{37}; v_{38}, v_{39} where:

v_{31}=[1832.00, 2564.56] v_{32}=[2564.56, 3297.11],
v_{33}=[3297.11, 4029.67], v_{34}=[4029.67, 4762.22],
v_{35}=[4762.22, 5494.78], v_{36}=[5494.78, 6227.33],
v_{37}=[6227.33, 6959.89], v_{38}=[6959.89, 7692.44],
v_{39}=[7692.44, 8425.00]

A linguistic variable named "Wheat production" (X_3), is defined to have the following linguistic values which are defined as: B_{31} = (very very very few), B_{32} = (very very few), B_{33} = (very few), B_{34} = (few), B_{35} = (moderate), B_{36} = (many), B_{37} = (many many), B_{38} = (many many many), B_{39} = (too many).

The proposed model fuzzifies the historical data by assigning the corresponding linguistic values as shown in Table VI. Table V shows the membership grade of Wheat production actual values, which affect on linguistic values.

TABLE V
DATA WHEAT PRODUCTION AND MEMBERSHIP GRADES FROM 1986 UP 2008.

Year	X_3	B_{31}	B_{32}	B_{33}	B_{34}	B_{35}	B_{36}	B_{37}	B_{38}	B_{39}
1986	1929	0	0	0	0	0	0	0	0	0
1987	2722	1	0	0	0	0	0	0	0	0
1988	2839	1	0	0	0	0	0	0	0	0
1989	3183	0	0	0	0	0	0	0	0	0
1990	4268	0	1	0	0	0	0	0	0	0
1991	4483	0	0	1	0	0	0	0	0	0
1992	4618	0	0	0	0	0	0	0	0	0
1993	4786	0	0	0	0	0	0	0	0	0
1994	4437	0	0	0	0	0	0	0	0	0
1995	5722	0	0	0	1	0	0	0	0	0
1996	5735	0	0	0	0	0.6	0	0	0	0
1997	5849	0	0	0	0	0	0	0	0	0
1998	6093	0	0	0	0	0	0	0	0	0
1999	6347	0	0	0	0	0	0	0	0	0
2000	6564	0	0	0	0	0	0	0	1	0
2001	6255	0	0	0	0	0	0	0.9	0	0
2002	6625	0	0	0	0	0	0	0	0	0
2003	6845	0	0	0	0	0	0	0	0	0
2004	7178	0	0	0	0	0	0.8	0	0	0
2005	8141	0	0	0	0	0	0	0	0	0
2006	8274	0	0	0	0	0	0	0	0	0
2007	7379	0	0	0	0	0	0	0	0	0
2008	7977	0	0	0	0	0	0	0	0	1

TABLE VI
WHEAT IMPORTS, POPULATION, WHEAT AREA, WHEAT PRODUCTION AND LINGUISTIC VALUES FROM 1986 UP 2008.

Year	Y	A	X_1	B_1	X_2	B_2	X_3	B_3
1986	6811	A_7	47751	B_{11}	1206	B_{21}	1929	B_{31}
1987	6874	A_7	48816	B_{11}	1373	B_{21}	2722	B_{32}
1988	6890	A_7	49826	B_{11}	1421	B_{22}	2839	B_{32}
1989	7527	A_8	50858	B_{12}	1532	B_{22}	3183	B_{32}
1990	7712	A_9	51911	B_{12}	1955	B_{24}	4268	B_{34}
1991	5550	A_5	52985	B_{12}	2215	B_{25}	4483	B_{34}
1992	5496	A_5	54082	B_{13}	2092	B_{25}	4618	B_{34}
1993	6118	A_6	55201	B_{13}	2171	B_{25}	4786	B_{35}
1994	4509	A_3	56344	B_{13}	2111	B_{25}	4437	B_{34}
1995	6256	A_6	57642	B_{14}	2512	B_{27}	5722	B_{36}
1996	5868	A_5	58835	B_{14}	2421	B_{26}	5735	B_{36}
1997	5202	A_4	60053	B_{15}	2486	B_{27}	5849	B_{36}
1998	5817	A_5	61296	B_{15}	2421	B_{26}	6093	B_{36}
1999	4776	A_4	62564	B_{15}	2380	B_{26}	6347	B_{37}
2000	3930	A_2	63860	B_{16}	2463	B_{26}	6564	B_{37}
2001	4982	A_4	65182	B_{16}	2342	B_{26}	6255	B_{37}
2002	4819	A_4	66531	B_{17}	2450	B_{26}	6625	B_{37}
2003	5205	A_4	67908	B_{17}	2506	B_{27}	6845	B_{37}
2004	3083	A_1	69313	B_{18}	2605	B_{27}	7178	B_{38}
2005	6272	A_6	70748	B_{18}	2985	B_{29}	8141	B_{39}
2006	5372	A_5	72212	B_{18}	3064	B_{29}	8274	B_{39}
2007	8073	A_9	73608	B_{19}	2716	B_{28}	7379	B_{38}
2008	7381	A_8	75225	B_{19}	2920	B_{29}	7977	B_{39}

The proposed model selects the maximum membership grade for each cluster. The representing crisp value for each cluster calculating is:

$$Crisp\_valu\ e(B_{31}) = (1.0\ x\,(1987) + 1.0\ x\,(1988)) / 2.0 = 6882$$

$$Crisp\_valu\,e(B_{32}) = 1.0\text{x }(1990) / 1.0 = 7712$$

$$\ldots$$

$$Crisp\_valu\ e(B_{39}) = 1.0\text{x }(2008) / 1.0 = 7381$$

The proposed model selects linguistics with highly frequency of occurrence then partitions its interval into k sub-partitions. For consequent factor "wheat imports", the proposed model divides u_4=[4731.67, 5305.22] into three partitions and u_5=[5305.22, 5878.78] into two partitions. When proposed model divided u_4 then linguistic A_5 will be converted to be linguistic A_7 and linguistic A_7 to be linguistic A_{10} and so on.

For antecedent's factors: population linguistic remains the same, but for "Wheat area", the proposed model divides v_{26}=[2260.11, 2477.33] into three partitions and v_{25}=[2042.89, 2260.11] into two partitions. When proposed model divided v_{25} then linguistic B_{26} will be converted to be linguistic B_{27} and linguistic B_{27} to be linguistic B_{210} and so on, and for Wheat production, the proposed model divides v_{37}=[6227.33,6959.89] into three partitions and v_{36}=[5494.78, 6227.33] into two partitions. When proposed model divided v_{36} then linguistic B_{37} will be converted to be linguistic B_{38} and linguistic B_{38} to be linguistic B_{311} and so on. By using the fuzzy logical relationship between the forecasting values of the antecedents factors and the forecasting values for consequents factors:

$$Forecasting(A_i) = \left(\sum_{k=1}^{3} \frac{\sum_{r=0}^{2} Crisp\_value(X_{k,(i-r)})\ \text{x}\ Crisp\_value(Y_{i-r})}{\sum_{r=0}^{2} Crisp\_value(X_{k,(i-r)})} \right) \Big/ 3 \tag{12}$$

The forecasting value for year 1998 is found to be 5554 while the actual value was 5817 and the forecasting value for year 2003 is found to be 5204 while the actual value was 5205. Table VI shows linguistic terms and forecasting values deduced by proposed model.

B. Testing Regression Model

The regression equation is:

$$Imprts = -2542 + 0.212\ Population + 2.84\ Area - 1.96\ Production \tag{13}$$

TABLE VII
ANALYSES OF VARIANCE

Source	DF	SS	MS	F	P
Regression	3	6931883	2310628	1.60	0.223
Residual Error	19	27493563	1447030		
Total	22	34425445			

From Table VII, it is found that: the F value is greater than the P value, so this model is accepted.

The R-squared and adjusted R-squared values are estimates of the 'goodness of fit' of the line. They represent the variation percentage of the data explained by the fitted line; the closer the points to the line, the better the fit.

Adjusted R-squared is not sensitive to the number of points within the data. R-squared is derived from:

$$R\text{-}squared = 100 * SSR/SST \tag{14}$$

Adjusted R-squared squared tends to optimistically estimate how well the models fits the real data, is derived from:

$$\text{adjusted R-squared} = 100 * (1 - MSE / MST) \tag{15}$$

R-Sq = 20.1%, indicate that this model is not good to fit the data. The R-S(adj)= 7.5%, indicate that this model can't represent the four factors due to their logical relationships that are sometimes positive and other times negative.

C. *Testing ARIMA Model*

The time series has four seasons (Import, population, Area, Production) by year. The best model is ARIMA(0,0,0)(2,1,2)$_4$ which has the following form:

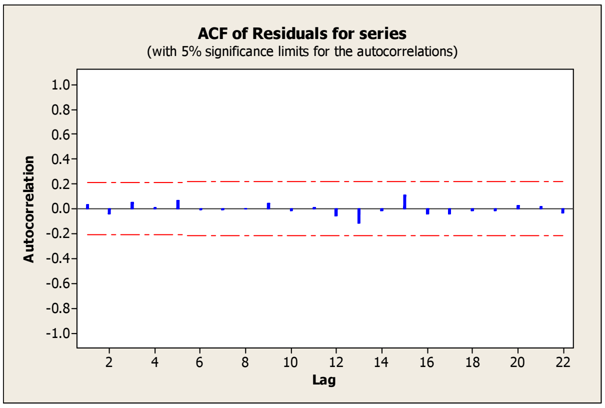

Fig. 3. PACF for errors.

$$Y^1{}_t = 0.6181\ \ Y^1{}_{t-1} \times 0.3832\ \ Y^1{}_{t-2} \times \xi_t - 1.068\ \xi_{t-1} \times -0.1907\ \xi_{t-2} \tag{16}$$

Where $Y_t^1 = Y_t - Y_{t-4}$, Y_t =log(Z_t) and Z_t is the data series.

From Fig. 3 and Fig. 4 note that: the error plots of ACF and PACF, it is clear that there are no lags out of ranges. So the model is the best time series model for representing the data.

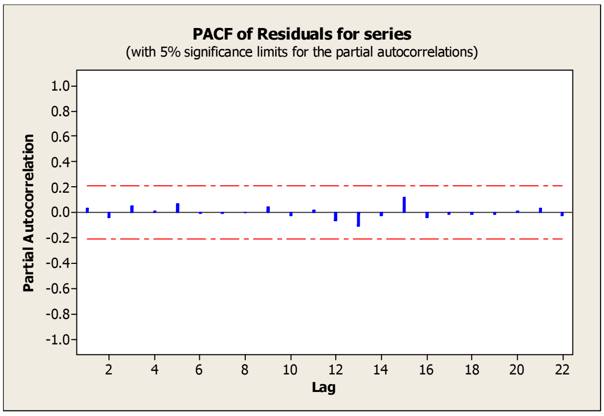

Fig. 4. PACF for errors.

D. *Testing Neural Network Model*

The proposed feed forward neural network, which we call PROPOSEDFF, is designed using the MATLAB neural network toolbox. The PROPOSEDFF creates a feed-forward back-propagation network. It requires three inputs: population, Wheat area, and Wheat production and returns one output: Wheat import. See Fig. 5.

$$\text{Network} = \text{PROPOSEDFF}([0\,1],[10\,5\,3\,1],\{\text{'tansig' 'logsig'} \\ \text{'tansig' 'purelin'}\},\text{'trainrp'},\text{'learngd'},\text{'mse'},\text{nn})$$

The first argument is a matrix [*0 1*] of minimum and maximum values for the input vector. The second argument is an array [*10 5 3 1*] containing the sizes of each layer. The third argument is a cell array,{'*tansig*' '*logsig*' '*tansig*' '*purelin*'} containing the names of the transfer functions to be used in each layer. The fourth argument contains '*trainrp*' the name of the training function to be used. The fifth argument '*learngd*' is back-propagation weight/bias learning function. The sixth argument '*mse*' is the performance functions. And the final input '*nn*' the number of inputs.

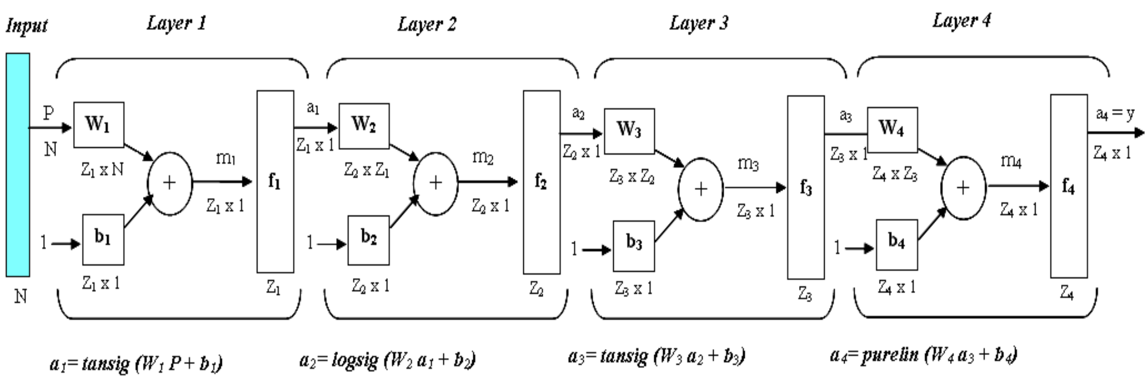

Fig. 5. PROPOSEDFF model network

V. RESULTS

The researchers build four models: regression model, which has R-Sq = 20.1%, that due to the non-linearity in the problem, second model is ARIMA(0,0,0)$(2,1,2)_4$ model, the

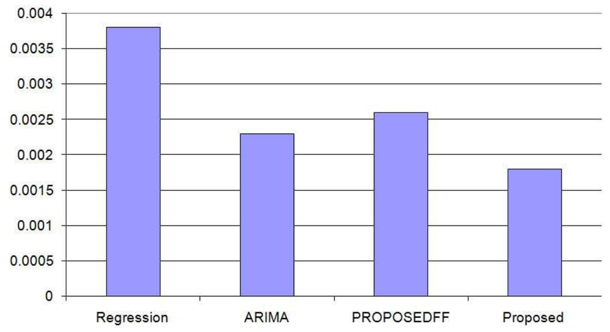

Fig 6. NRMSE accuracy

third model is PROPOSEDFF, which designed using the MATLAB neural network toolbox, and the fourth model is the proposed multi-factors fuzzy time series model (MFFTS). The result of the forecasting accuracy for the models: Regression, ARIMA, PROPOSEDFF, and MFFTS model compared with the actual data by using NRMSE performance indices illustrates in Fig. 6.

It is clear that the multivariate fuzzy time series proposed model provides higher accuracy in forecasting than ARIMA Model, Regression and PROPOSEDFF models, so the proposed model can lead to satisfactory high performance for fuzzy time series.

VI. CONCLUSION

This paper presents a new Multivariate-factors fuzzy time series model (MFFTS). The proposed model has been used successfully to forecast Egypt Wheat imports as consequent factor depending on: population size, Wheat area, and Wheat production as antecedent's factors. The experimental test is based on the official data provided by the Central Agency for Public Mobilization and Statistics (CAPMAS) starting from 1986 up to 2008. The Multivariate-factors proposed model has been evaluated through a comparison with Regression, ARIMA, and neural network model. The result of the comparison ensures the superiority of the proposed model over the other three models in terms of lowest NRMSE.

REFERENCES

[1] Q. Song, and B. S. Chissom, "Forecasting enrollments with fuzzy time series. I", *Fuzzy sets and systems*, vol. 54(1), pp. 1-9, 1993.
[2] L. A. Zadeh, "Fuzzy Sets", *Information Sciences*, vol. 8(3), 338-353, 1965.
[3] W. Qiu, X. Liu and H. Li, "A generalized method for forecasting based on fuzzy time series", *Expert Systems with Applications*, vol. 38(8), pp. 10446-10453, 2011.
[4] M. Bahrepour, R. Akbarzadeh-T, M. Yaghoobi and M.-B. Naghibi, "An adaptive ordered fuzzy time series with application to FOREX", *Expert Systems with Applications* vol. 38, pp. 475–485, 2011.
[5] A. K. Abd-Elaal, H. A. Hefny and A. H. Abd-Elwahab, "Fuzzy time series model for Egypt Gold Reserves forecasting based on fuzzy clustering", *Computer Engineering & Systems (ICCES)*. Cairo, IEEE, pp. 140 – 147, 2011.
[6] T. Liu, Y. Chen and J.H. Chou, "Extracting fuzzy relations in fuzzy time series model based on approximation concepts", *Expert Systems with Applications* vol. 38(9), pp. 11624–11629, 2011.
[7] O. Duru, "A multivariate model of fuzzy integrated logical forecasting method (M-FILF) and multiplicative time series clustering: A model of time-varying volatility for dry cargo freight market", *Expert Systems with Applications* vol. 39(4), pp. 4135–4142, 2012.
[8] S. Hassan, J. Jaafar, B. Samir, and T. Jilani "A Hybrid Fuzzy Time Series Model for Forecasting", *Engineering Letters*, vol. 20(1). pp. 88-93, 2012.
[9] E. Egrioglu, C. H. Aladag, U. Yolcu, " Fuzzy time series forecasting with a novel hybrid approach combining fuzzy c-means and neural networks", *Expert Systems with Applications*, vol. 40(3), pp. 854–857, 2013.
[10] K. Khiabani, M. Yaghoobi, A. Lary, and S. Yousof-Khani, "An Improved Adaptive Time-Variant Model for Fuzzy-Time-Series Forecasting Enrollments based on Particle Swarm Optimization", *Proceedings of The World Congress on Engineering and Computer Science 2011* (*WCECS)*, San Francisco, USA, pp325-330, 2011.
[11] A. K. Palit, and D. Popovic, *Computational intelligence in time series forecasting theory and engineering applications*, London, UK, Springer-Verlag, 2005
[12] A. Konar, *Computational intelligence principles, techniques and applications*. Netherlands, Springer Berlin Heidelberg, 2005.
[13] M. Qi, "Forecasting Real Time Financial Series", *Proceedings of the 2002 International Joint Conference on Neural Networks IJCNN'02*, Honolulu, Hawaii, pp. 377-381, 2002.
[14] W. Huang, K.K. Lai, Y. Nakamori, and S. Wang, "Forecasting foreign exchange rates with artificial neural networks: a review", *International Journal of Information Technology & Decision Making*, vol. 2(3) pp. 145-165, 2004.
[15] Y. Chen, B. Yang, J. Dong, and A. Abraham, "Time-series forecasting using flexible neural tree model", *Information Sciences*, vol.174(4), pp. 219–235, 2005.
[16] K. K. Lai, L. Yu, S. Wang, and W. Huang, "Hybridizing Exponential Smoothing and Neural Network for Financial Time Series Prediction", *Computational Science – ICCS 2006*, Springer-Verlag Berlin Heidelberg, pp. 493-500, 2006.
[17] A. K. Pandey, A.K Sinha, and V.K Srivastava, "A Comparative Study of Neural-Network & Fuzzy Time Series Forecasting Techniques– Case Study: Wheat Production Forecasting", *IJCSNS International Journal of Computer Science and Network Security*, vol. 8(9), pp. 382-387, 2008.
[18] O. Castillo, E. Rubio, J. Soria, and E. Naredo, "Optimization of the Fuzzy C-Means Algorithm using Evolutionary Methods", *Engineering Letters*, vol. 20(1), pp. 61-67, 2012.
[19] M. Friedman, and A. Kandel, *Introduction to Pattern Recognition: Statistical, Structural, Neural And Fuzzy Logic Approaches*, World Scientific, 1999.
[20] H. Chu, T. Chen, C.-H. Cheng and C. Huang, "Fuzzy dual-factor time-series for stock index forecasting", *Expert Systems with Applications*, vol. 36(1), pp. 165–171, 2009.
[21] K. Huarng, "Effective lengths of intervals to improve forecasting in fuzzy time series", *Fuzzy Sets and Systems*, vol. 123(3), pp.:387–394, 2001.
[22] C. Wang, "A comparison study between fuzzy time series model and ARIMA model for forecasting Taiwan export", *Expert Systems with Applications*, vol. 38(8), pp. 9296-9304, 2011.
[23] S-M. Chen and C-d. Chen, "Handling forecasting problems based on high-order fuzzy logical relationships", *Expert Systems with Applications*, vol. 38(8), pp. 3857-3864, 2011.
[24] N. Wang, and S. Chen, "Temperature prediction and TAIFEX forecasting based on automatic clustering techniques and two-factors high-order fuzzy time series", *Expert Systems with Applications* , vol. 36(2), pp. 2143–2154, 2009.