Economic Insights – Trends and Challenges	Vol.V(LXVIII) No. 3/2016	31 - 38

Forecast of the Evolution of Employment in Romanian Agriculture Using Fuzzy Time Series

Cristian Marinoiu

Petroleum-Gas University of Ploieşti, Bd. Bucureşti 39, 100680, Ploieşti, Romania
e-mail:marinoiu_c@yahoo.com

Abstract

In terms of agricultural indicators Romania occupies a special place among EU countries: it is the sixth EU country in terms of utilized agricultural area, it has the largest number of farms and last but not least the highest degree of labour occupancy in agriculture. Given the major differences compared to the averages recorded in EU countries, the future evolution of these indicators is a legitimate concern. In this paper we propose a forecast of the employment evolution of the labour force in Romania's agriculture using fuzzy time series. We also show a comparison between the forecast accuracy achieved on the basis of fuzzy time series and the one achieved by using neural networks, using the Diebold-Mariano's statistic test.

Keywords: *fuzzy time series; forecasting; accuracy*

JEL Classification: *C22; Q10*

Introduction

In a dynamic system as complex as agriculture, the number of employees in the system is a variable strongly affected by uncertainty. Referring to the EU, in (European Commission, 2013) four particular characteristics of the agricultural sector are highlighted, that make the determination of the exact number of employees in the sector more difficult than in other sectors:

- in most EU countries agriculture is dominated by the presence of family farms in which the members work in agriculture at different times of the year;
- many farm workers usually have other sources of income and practice the agricultural work as a part-time activity;
- the agricultural sector is characterized by seasonal peaks of activity, where, for relatively short periods of time, a large numbers of workers are employed;
- statistical data sources, that have methodologies and different purposes, reflect differently the problems of the agricultural sector;

A modern way to deal with uncertainty in data, with consequences in all areas of science, was initiated by Zadeh L.A. in (Zadeh,1965) where he introduced the concept of fuzzy set. In (Abbasov and Mamedova, 2013) it is argued that the data which reflect the evolution of demographic processes for a certain period of time and for a certain area are inevitably affected by uncertainty. Here, the authors use the term of fuzzy set in order to define the fuzzy time series and they propose a methodology to obtain forecasts for these time series using the example of a case study.

Given these considerations, we believe that a suitable method for the forecast of the number of employees in Romania's agricultural sector is Abbasov-Mamedova method which we present, briefly, in the following sections.

The EA Indicator

In this paper we used data describing the evolution of the number of workers in Romania's agriculture available in the database of the (World Bank Group). The data consist of the annual number of employees in the Romanian agriculture during the period 1980-2014. The graph corresponding to this data time series, which for brevity will be referred to as EA, is shown in Figure 1. Also for clarity, we specify that „employment is defined as persons of working age who were engaged in any activity to produce goods or provide services for pay or profit, whether at work during the reference period or not at work due to temporary absence from a job, or to working-time arrangement. The agriculture sector consists of activities in agriculture, hunting, forestry and fishing, in accordance with division 1 (ISIC 2) or categories A-B (ISIC 3) or category A (ISIC 4)". (IndexMundi)

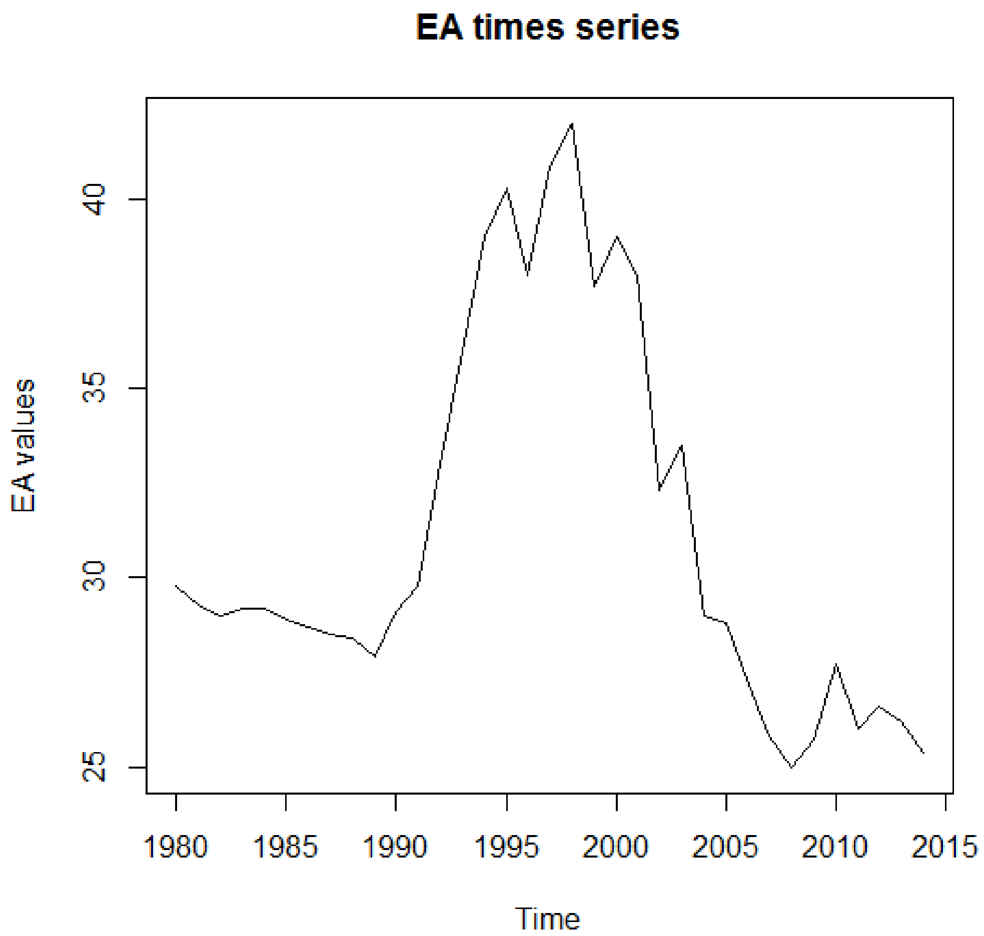

Fig. 1. The graph of the time series EA values

Source: made by the author in R, using data from (World Bank Group)

Fuzzy Sets and Fuzzy Time Series

The notion of fuzzy set was introduced in (Zadeh, 1965) as an extension of the concept of set defined in the classical sense. Be $X = \{x\}$ the set made up of elements that we generically note by x and A a subset of this set. In the classical approach the set A is characterized by its characteristic function $\mu_A(x)$, which takes only two values, i.e.,

$$\mu_A(x) = 1 \text{ if } x \in A \text{ and } \mu_A(x) = 0 \text{ if } x \notin A. \tag{1}$$

A fuzzy set A is characterized by a characteristic function (or membership function) $\mu_A : X -> [0,1]$. Its interpretation is the following one: the closer the value of the characteristic function $\mu_A(x)$ is to 1, the higher the belonging degree of element $x \in X$ with respect to the set A. Mathematically, the fuzzy set A is defined as $A = \{(x, \mu_A(x) / x \in X\}$. If $X = \{x_1, x_2, \cdots x_n\}$ is a finite set, then an usual notation for the fuzzy set A is the following one:

$$A = \{\mu_1 / x_1 + \mu_2 / x_2 + \cdots + \mu_n / x_n\}, \tag{2}$$

where the notation μ_i / x_i refers to the element x_i which has the belonging degree μ_i to A, and the sign $+$ refers here to the union operation.

A time series is "*an ordered sequence of values of a variable at equally spaced time intervals*" (Nist/Sematech). Mathematically, such a sequence of n values from the real numbers set can be represented by $y(t)\ t \in \{0,1,2,..n\}$, meaning that $y(t)$ represents the variable value at the moment t. At the same time we consider that the real numbers set $y(t)\ t \in \{0,1,2,..n\}$ is the universal set of fuzzy sets $f_i(t)\ (i = 1,2\ldots)$. In (Qiang and Chison, p. 270) it is mentioned that the collection of fuzzy sets $f_i(t)\ (i = 1,2\ldots)$ which we note $F(t)$ is, by definition, the fuzzy time series $F(t)$ of the time series $y(t)$.

The Abbasov-Mamedova Method for Forecast of Fuzzy Time Series

In (Abbasov and Mamedova, 2013) it is investigated the opportunity to make forecasts of demographic evolution of the population using fuzzy time series. The authors justify the robustness of their approach and present a methodology which is applied step by step on a concrete example. In essence, the main steps of this methodology are the following (Abbasov and Mamedova, 2013, p. 547):

1. The definition of the universal set U as an interval whose endpoints represent the smallest and the largest variation between the observed values for two consecutive years;
2. Dividing the interval of the universal set U in a number of disjoint intervals which reflect different growth rates of the analysed population;
3. The interpretation of the total population variation analysed as a linguistic variable and the association between linguistic values and the obtained intervals (we obtain the $F(t)$ fuzzy set);
4. The fuzzyfication of the entry data, which means the conversion of the numeric values into fuzzy values;
5. The selection of a parameter $w > 1$ which represents the time period before the year based on which the fuzzy prediction for the next year is realized;

6. Defuzzification of the obtained results, which means the conversion of fuzzy values into numerical values or crisp values.

The Abbasov-Mamedova method is implemented in (The R Project for Statistical Computing, 2014) through the function *fuzzy.ts2* from the package *AnalizeTS*. The following paragraphs present the intermediate and final results obtained by using the *fuzzy.ts2* function in order to forecast the evolution of EA series for the years 2015-2024.

Using Abbasov-Mamedova Method for Forecasting the EA Time series

Table 1 presents the number of agricultural workers between 1980 - 2014 (as a percentage of the total number of employees in Romania) and the differences identified between any two consecutive years of this period

Table 1. Employment in agriculture (% of total employment) and differences between the consecutive years

Year	**Number of employees** (% from total number of employees)	**Differences**	**Year**	**Number of employees** (% from total number of employees)	**Differences**
1980	29.8	-	1998	42.0	1.1
1981	29.3	-0.5	1999	37.7	-4.3
1982	29.0	-0.3	2000	39.0	1.3
1983	29.2	0.2	2001	38.0	-1.0
1984	29.2	0.0	2002	32.3	-5.7
1985	28.9	-0.3	2003	33.5	1.2
1986	28.7	-0.2	2004	29.0	-4.5
1987	28.5	-0.2	2005	28.8	-0.2
1988	28.4	-0.1	2006	27.2	-1.6
1989	27.9	-0.5	2007	25.8	-1.4
1990	29.1	1.2	2008	25.0	-0.8
1991	29.8	0.7	2009	25.7	0.7
1992	33.0	3.2	2010	27.7	2.0
1993	36.0	3.0	2011	26.0	-1.7
1994	39.0	3.0	2012	26.6	0.6
1995	40.3	1.3	2013	26.2	-0.4
1996	38.0	-2.3	2014	25.4	-0.8
1997	40.9	2.9			

Source: made by the author using data from (World Bank Group) and functions from R

We can notice that that the smallest difference between two consecutive years is -5.70 and it is occurring between 2001 and 2002, and the largest difference is 3.20, between 1991 and 1992. Consequently, the universal set U (required by step 1 of the method) has the form $U = [-5.70, 3.20]$ and it was divided (according to step 2 of the method) into five equal intervals: $u_1 = [-5.70 - 3.92]$, $u_2 = [-3.92, -2.14]$, $u_3 = [-2.14, -036]$, $u_4 = [-0.36, 1.42]$, $u_5 = [1.42, 3.20]$. The variable U, as total variation in the number of employees in agriculture during 1989-2014 is interpreted as a linguistic variable (step 3 of the method). In our case we proposed that these variables take the following linguistic values: VLLEG (Very Low Level Employment Growth), LLEG (Low Level Employment Growth) MLEG (Moderate Level Employment Growth), HLEG (High Level Employment Growth), VLEG (Very High Level Employment Growth), which correspond to the intervals $u_1, u_2, \dots u_5$, respectively. To each linguistic value k it is attached a fuzzy set (k , u_k, A_k), where A_k is the fuzzy set defined by the interval u_k included in the universal set U . (Abbasov and Mamedova (2003, p.548) define the membership function of the fuzzy set A_k as:

$$\mu_{A_k}(u) = \frac{1}{1+[C(u-u_m^k)]^2} \quad k \in 1,2,\cdots 5 \tag{3}$$

where: $u \in U$, u_m^k is the middle of the u_k interval and C is a constant.

It is also mentioned that the C constant „is chosen in such a way that it ensures the conversion of definite quantitative values into fuzzy values or their belonging to the interval" (Abbasov and Mamedova, 2013, p.548). In Table 2 we present the MAE (Mean Absolute Error) values for different values of C constant. We chose the C constant so that the MAE value is minimum. We observe that the MAE minimum value is 0.043 and we obtain it for C = 0.3.

Table 2. MAE values for different values of C constant

C	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
MAE	0.357	0.185	0.043	0.055	0.12	0.165	0.196	0.218	0.234	0.246

Source: made by the author using functions from R

In Figure 2 we represent the graphs for the membership functions $\mu_{A_k}(u)\ k = 1,2,\cdots 5$ for the selected value C=0.3

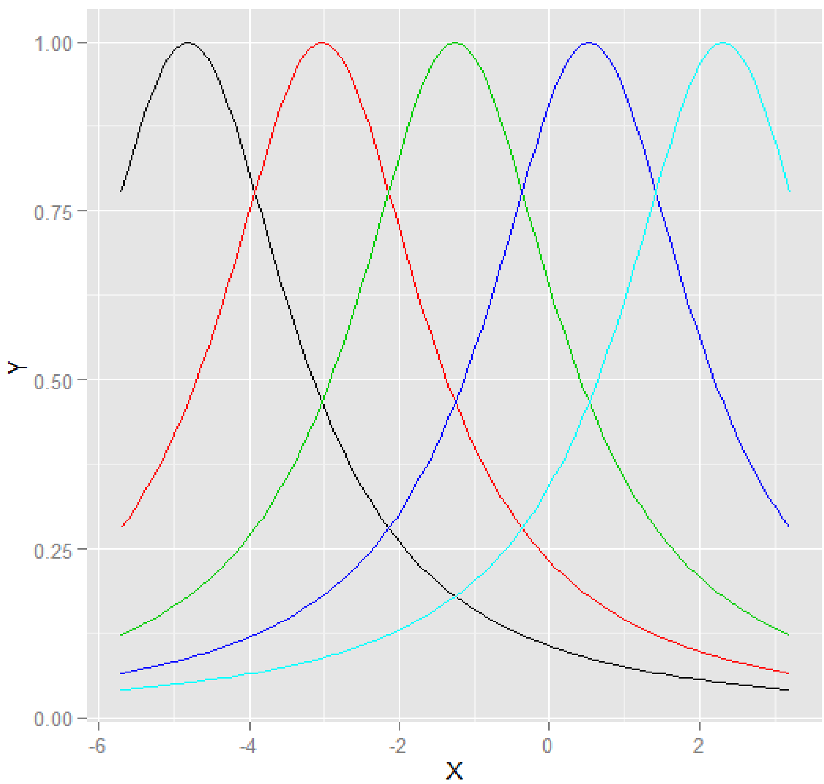

Fig. 2. The graphs of the membership functions $\mu_{A_k}(u)\ k = 1,2,\cdots 5$

Source: made by the author in R

With the help of the membership function defined by relation (3) the objective of step 4 of Abbasov-Mamedova's method was achieved, namely the fuzzification of the differences between the values of the consecutive years calculated in Table 1. For example, the fuzzy variants for these differences (calculated using the function *fuzzy.ts2* from the package *AnalzyeTS* from *R*) for the years 1981-1992 are the following:

A[1981]={(0.37/u1),(0.63/u2),(0.95/u3),(0.91/u4),(0.58/u5)}

A[1982]={(0.35/u1),(0.6/u2),(0.92/u3),(0.94/u4),(0.62/u5)}

A[1983]={(0.31/u1),(0.52/u2),(0.84/u3),(0.99/u4),(0.71/u5)}

A[1984]={(0.32/u1),(0.55/u2),(0.88/u3),(0.98/u4),(0.68/u5)}

A[1985]={(0.35/u1),(0.6/u2),(0.92/u3),(0.94/u4),(0.62/u5)}

A[1986]={(0.34/u1),(0.58/u2),(0.91/u3),(0.95/u4),(0.64/u5)}

A[1987]={(0.34/u1),(0.58/u2),(0.91/u3),(0.95/u4),(0.64/u5)}

A[1988]={(0.33/u1),(0.56/u2),(0.89/u3),(0.97/u4),(0.66/u5)}

A[1989]={(0.37/u1),(0.63/u2),(0.95/u3),(0.91/u4),(0.58/u5)}

A[1990]={(0.24/u1),(0.38/u2),(0.65/u3),(0.96/u4),(0.9/u5)}

A[1991]={(0.27/u1),(0.44/u2),(0.75/u3),(1/u4),(0.81/u5)}

A[1992]={(0.15/u1),(0.22/u2),(0.36/u3),(0.61/u4),(0.93/u5)}

The objectives of the steps 5 and 6 of the method (getting fuzzy forecast values and their defuzzification) were achieved on the basis of the proposed methods in (Abbasov and Mamedova, 2003, pp.549-550) that are also implemented in *fuzzy.ts2* function from R. The parameter *w* asked to step 5 was set to 33. Thus, fuzzy forecast for 2015 was calculated based on fuzzy differences for the period 1982-2014, fuzzy forecast for 2016 was based on fuzzy differences for the period 1983-2015 and so on.

Results and Discussions

Forecasted values for the years 2015-2024 using the Abbasov- Mamedova method are presented in Table 3 and the graph of the EA series and of the forecasted values using this method is shown in Figure 3.

Table 3. The forecasted values for the years 2015-2024, using Abbasov-Mamedova method

Years	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Forecasted values	24.32	23.18	21.98	20.75	19.51	18.27	17.02	15.77	14.52	13.28

Source: made by the author using functions from R

If we consider using neural networks for forecasting as an alternative to the fuzzy method used in this paper, forecast values are as follows (Table 4):

Table 4 . The forecasted values for the years 2015-2024, using neural network method

Years	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Forecasted values	29.27	29.60	30.00	30.50	31.06	31.70	32.40	33.10	33.80	34.45

Source: made by the author using functions from R

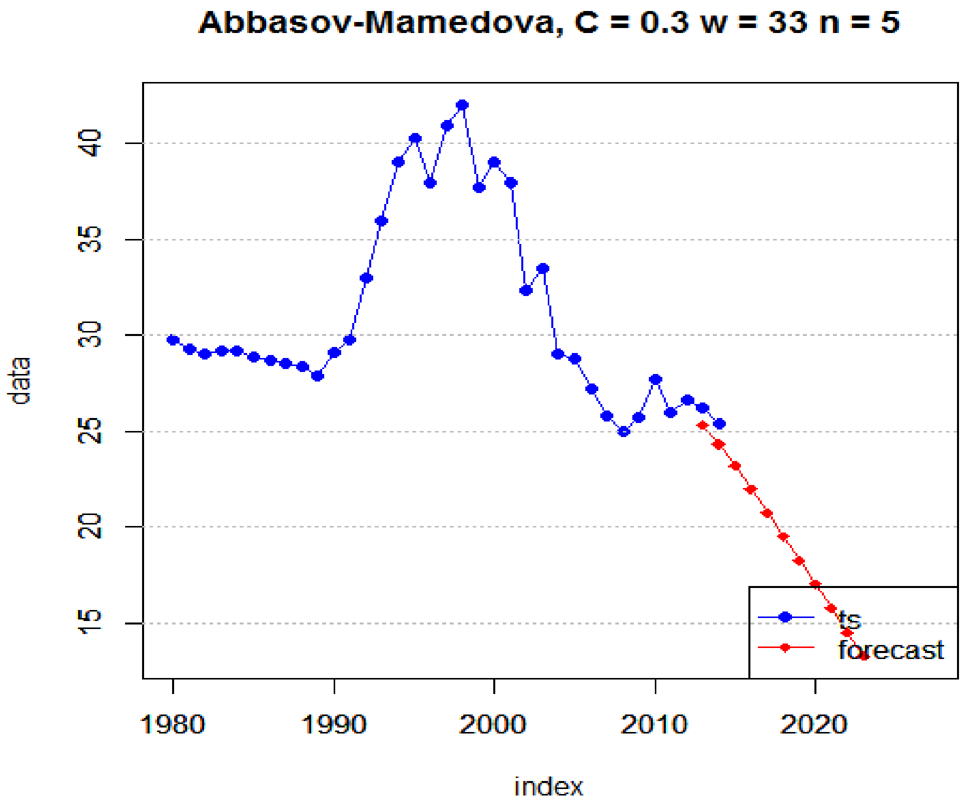

Fig. 3. Graph of the EA time series and forecasted values for the years 2015-2024 using the Abbasov- Mamedova method

Source: made by the author using functions from R.

We estimated the accuracy of the obtained forecasts using the methodology proposed in (Hydman and Athanasopoulos). We basically divided EA series values into two parts: the first part refers to the values in the period 1980-2004 and are considered as training data, and the remaining values (for the period 2005-2014) are used as test data to estimate the accuracy of forecasts obtained using the model built on the training data. By using the function *accuracy* from *R* we obtained the values 4.15, 3.50, 13.25, for RMSE (Root Mean Squared Error), MAE (Mean Absolute Error) and MAPE (Mean Absolute Percentage Error), respectively. By using the neural networks method the values obtained for RMSE, MAE and MAPE are 5.67, 5.15 and 19.77, respectively. The centralization of the results obtained with both methods, *method 1* based on the use of fuzzy sets and *method 2* based on the use of neural networks, is made in Table 5. The results show that RMSE values, MAE and MAPE values are lower when using *method 1*.

Table 5. Estimated values of the forecast error

Models\Forecast accuracy methods	**RMSE**	**MAE**	**MAPE**
Fuzzy model	4.15	3.50	13.25
Neural Network model	5.67	5.15	19.77

Source: made by the author using functions from R

In order to see if the difference in accuracy between the two forecasts is statistically significant we used the Diebold- Mariano test (Diebold and Mariano, 1995). This test, implemented in *dm.test* function from *R*, allowed us to test the null hypothesis:

H_0: the two methods have the same forecast accuracy

Against:

H_1: *method 2* is less accurate than *method 1*.

Applying the test to the significance level 0.05 provided the value p-value=0.05. Because p-value is equal to the significance level we reject the hypothesis H_0, and we accept the hypothesis H_1 *(method 2 is* less accurate than *method 1).*

Conclusions

In this paper we proposed forecasts of the labour force employed in agriculture in Romania. Annual values of number of employees in agriculture during 1980-2014 were modelled as a fuzzy time series due to the uncertainty objectively affecting the determination of the exact number of workers in agriculture. The methodology followed for obtaining forecasts was proposed by Abbasov and Mamedova (*method 1*). We also showed a comparison between the accuracy of forecasts obtained using this method and the accuracy of forecasts obtained by applying a method based on feed forward neural networks *(method 2*). Lower values of the MAE, RMSE and MAPE accuracy indicators obtained with *method 1* as well as the final result given by the Diebold- Mariano test suggest that the forecasts obtained using *method 2* have a lower accuracy than those obtained using *method 1*.

References

1. Abbasov A.M., Mamedova M.H., 2003. Application of fuzzy time series to population forecasting, CORP 2003, *Proceedings of 8th Symposion on Information Technology in Urban and Spatial Planning*,Vienna University of Technology, 25.2.-28.2.2003, pp. 545-552. Available at <http://papers.cumincad.org/cgi-bin/works/Show?_id=50b1&sort=DEFAULT&search=Abbasov&hits=1> [Accessed on 19.06.2016].
2. Diebold, X. F. and Mariano, *S. R.*, *Comparing predictive Accuracy*, *Journal of Business & Economic Statistics*, July 1995, Vol. 13, No 3.
3. European Commission, 2013. How *many people work in agriculture in the European Union? An answer based on Eurostat data sources*, EU agricultural Economics Briefs [online], no. 8, July, 2013. Available at <http://ec.europa.eu/agriculture/rural-area-economics/briefs/index_en.htm> [Accessed on 18.06.2016]].
4. Hydman, R. J. and Athanasopoulos, G. *Forecasting: principles and practice*, [online]. Available at <https://www.otexts.org/fpp/2/5> [Accessed on 22.06.2016]]
5. IndexMundi, *Employment in agriculture (% of total employment)*,[online]. Available at <http://www.indexmundi.com/facts/indicators/SL.AGR.EMPL.ZS> [Accessed on 18.06.2016].
6. Nist/Sematech, *e-Handbook of Statistical Methods*, [online]. Available at <http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc41.htm> [Accessed on 20.06.2016].
7. Qiang S. and Chison B., 1993. *Fuzzy time series and its models*, Fuzzy Sets and Systems (54) 1993, pp.269-277, North-Holland. Available at <https://www.researchgate.net/publication/256410485_Fuzzy_time_series_and_its_model> [Accessed on 15.06.2016].
8. The R Project for Statistical Computing, 2014. R for Windows (3.1.2), [computer program]. Available at <https://cran.r-project.org/bin/windows/base/old/3.1.2/>,[accessed on June 2016].
9. World Bank Group, *API_ROU_DS2_en_EXCEL_v2*. Available at <http://data.worldbank.org/country/romania>
10. Zadeh L. A., 1965. Fuzzy sets, *Information and Control*, 8, pp. 338-353. Available at <https://bi.snu.ac.kr/Info/AI/Fuzzy%20Sets.pdf> [Accessed on 15.06.2016].