International Journal on Emerging Technologies **10**(2a): 232-236(2019)

ISSN No. (Print) : 0975-8364
ISSN No. (Online) : 2249-3255

Forecasting the Enterprise Tax Base through Regression of One-Dimensional Time Series

EkaterinaIvanovna Kadochnikova[1], Tatiana Valeryevna Erina[2] and AnnaIgorevna Margushova[3]
[1]Associate Professor, Head University/Institute of Management, Economics and Finance/Department of Economic Theory and Econometrics, Kazan Federal University, Russia.
[2,3]Institute of Management, Economics and Finance/Department of Economic Theory and Econometrics, Kazan Federal University, Russia.

(Corresponding author: EkaterinaIvanovna Kadochnikova)
(Received 05 May 2019, Revised 16 July 2019 Accepted 25 July 2019)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: The authors obtained a forecast of the enterprise tax base. In this paper, the authors evaluated ARIMA models according to the Box-Jenkins method and regression models with dummy variables to account for additive and multiplicative seasonality. On a sample of 48 observations on the tax basis of the estimated model ARMA (1;0), ARMA (1;1), SARMA (1, 1) x (0,1) 6, a model with seasonal dummy variables. The authors focused on the method of selecting the most valid model according to the criteria RMSE and AIC used the method of selection of the designated circle of the most simple model with the fewest parameters. The reliability of the results is confirmed by the information criterion of Akaike, the mean square error of the forecast, the diagnosis of residues on the normal distribution using a special test and the absence of autocorrelation using the Ljung-Box test. The statistical significance of regression models with dummy variables for seasonality was not confirmed. A promising direction of development of this study can be a combination of forecasts, as well as the use of polynomial trends.

Keywords: Tax base, time series, stationarity, forecasting, ARIMA model, model with dummy seasonal variables.

I. INTRODUCTION

In Economics, there are often processes in which the use of statistical methods does not always give adequate results. In these conditions, it is necessary to be able to analyze the dynamics of the predicted indicator, to understand what this dynamics is, and to select a suitable forecast model. A number of studies in the field of forecasting have shown that the use of complex, statistically based methods does not necessarily lead to an increase in the accuracy of forecasts. Model autoregressive regressions, which is considered statistically reasonable, may not be sufficiently accurate [1, 2]. Therefore, a practicing forecaster should have at his disposal much more tools of forecasting methods and models than mathematical statistics offers him. Forecasting is one of the most popular tasks of business analysts in terms of prospective assessments of the tax base. The purpose of tax forecasting is to identify for a certain time period the tax payments of the enterprise, which are due to the dynamics of the company's income [3, 4]. Tax forecasting, closely interacting with such management functions as marketing, Finance, accounting, personnel policy, supply, at the same time is one of the basic tools for generating performance indicators of the enterprise. Therefore, tax forecasting should become a mandatory tool in the management kit of a Russian enterprise when making a management decision [5]. The basic prerequisite for forecasting a one-dimensional time series of the tax base (income) of the enterprise is its stationarity. The dynamics of the tax base (income) of enterprises with seasonal production is often non-stationary and contains a trend and seasonal fluctuations. Therefore, the forecast of the tax base is convenient to obtain with the help of ARIMA model, additive or multiplicative trend-seasonal model, regression model with fictitious seasonal variables [6], and then choose the most accurate by the minimum value of the mean square error of the forecast. The aim of the study was to develop methodological and practical recommendations to improve the process of tax forecasting in the enterprise. The hypothesis of the study –the tax base of the enterprise has seasonality and a tendency to increase.

II. METHODS

The study used monthly data on the tax base of the company from 2015 to 2018, obtained from the open financial statements. Estimation of ARIMA models performed with the software Gretl–GNU Regression, Econometrics and Time-series Library.

The highly flexible software products, a model of type ARIMA is a classic in the predictive estimates. To build ARIMA-type models, we used the Box-Jenkins approach [7; 8], which consists of the following procedures:

1. Graphical analysis of the time series (is the time series stationary?);
2. Construction of ACF and PACF diagrams for the initial time series (with a slow decrease in the ACF correlogram, there is reason to believe that the time series is not stationary);
3. Conducting the Dickey-Fuller test (ADF test, null hypothesis of unsteadiness [9]) for the initial time series (if the series is not stationary, then taking the first difference between the levels (d=1));
4. Conducting the Dickey-Fuller test (ADF test) for the first time series difference;
5. Charting ACF and PACF for the first difference of the time series to identify the order p (for PACF), and q (for ACF);
6. Evaluation of ARIMA model (p, d, q) or several models by the first time series difference;
7. Diagnostics of ARIMA model residues (p, d, q) for autocorrelation and normality (Ljung-Box test [10]);
8. The choice of the model ARIMA (p, d, q) according to the minimum AIC criterion.
9. Using ARIMA model (p, d, q) for prediction;
10. RMSE forecast error calculation.

Modeling of seasonal variations using fictitious variables is possible by including fictitious variables in a constant (corresponds to the construction of a model with additive seasonality), or by including fictitious variables in the slope angle (corresponds to the construction of a model with multiplicative seasonality). In the case of quarterly data, a dummy variable s-1 can be entered to account for additive seasonality in the model:

$$y_t = \bar{y}_t + b_1 season_1 + b_2 season_2 + b_3 season_3 + \varepsilon_t$$

The coefficients of such a model can be estimated by the usual least squares method. The standard in this case is the fourth quarter, for which the model takes the form: $y_t = \bar{y}_t + \varepsilon$. To account for multiplicative seasonality, you must add dummy variables to the slope of a simple linear model:

$$y_t = a_0 + a_1 t + b_1 season_1 \cdot t + b_2 season_2 \cdot t + b_3 season_3 \cdot t + \varepsilon$$

In this case, with the growth of the value of t, the value of y will grow seasonally. For example, for quarterly data on the first observation in the first season, y will be higher than the reference by b1*1, and on the fifth (a year later)-higher than the standard by b1*5 . This specification adequately reflects the economically intuitive to the growing trend of tax base successfully developing company.

III. RESULTS AND DISCUSSION

Dynamics of the tax base of the enterprise (Fig.1) most likely demonstrates a stationary time series.

The result of the extended Dickey-Fuller test with a constant (p-value was 0.0054) allows us to formulate a conclusion about the stationarity of the initial time series of the tax base of the enterprise at all possible levels of significance. Autocorrelation function and private autocorrelation function (Fig. 2) time series contain significant correlation coefficients on the first lag.

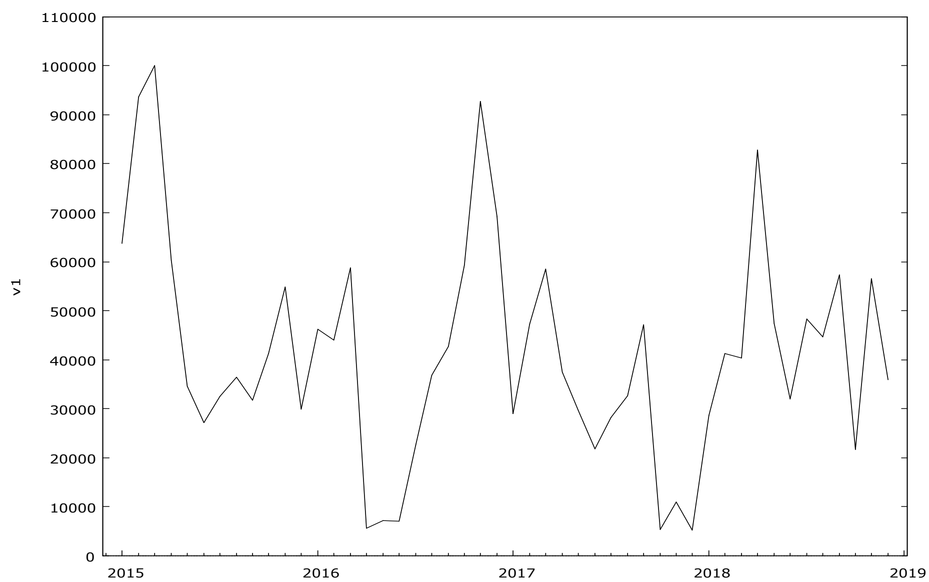

Fig. 1. Dynamics of the enterprise tax base.

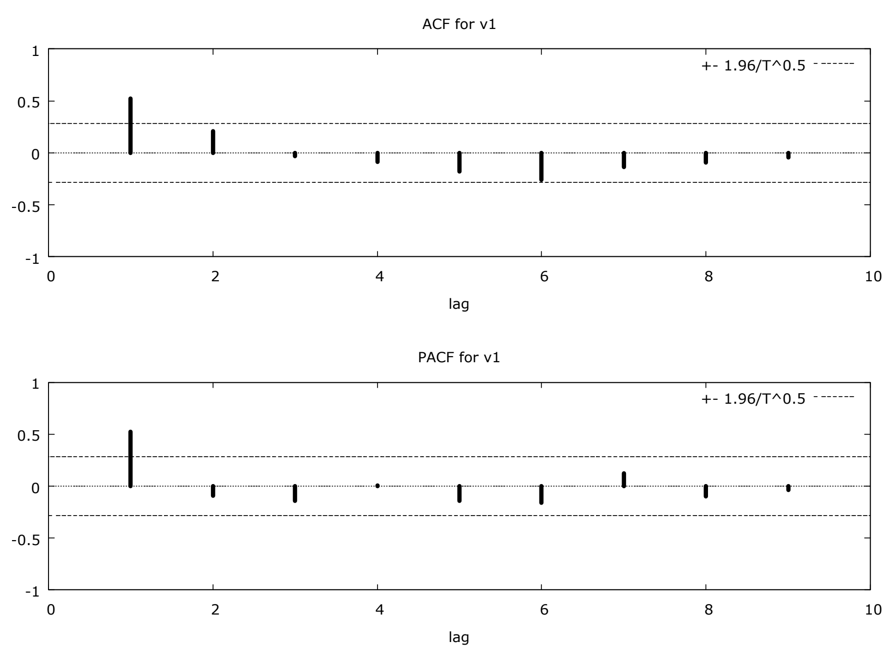

Fig. 2. Correlogram of initial levels of the tax base time series.

Therefore, we estimate the ARIMA model (1, 0, 1), where p = 1, q = 1, which is equivalent to the ARMA model (1,1) (Fig. 3).

In ARMA (1,1) model the parameter theta_1 is not significant, so we estimate ARMA (1,0). Also on the ACF chart (Fig. 2) we see a significant coefficient on the 6 lag, which makes it possible to evaluate the model c seasonal MA-component: SARMA (1,1) x $(0,1)_6$.

Diagnosis of the residues of the models pointed to the observance of the null hypothesis of normal distribution of residuals and absence of autocorrelation for all possible levels of significance (Table 1). Therefore, the choice of the most correct model from Table 1 for forecasting accounts payable is feasible by the least Akaike criterion (AIC), which contains a penalty for the complexity of the model, and the least mean square error (RMSE), which measures the accuracy of the model. The table shows that the RMSE and AIC criteria for the model SARMA (1,1)x$(0,1)_6$ are the maximum, so the choice of the best model for forecasting the tax base will be carried out between ARMA (1,1) and ARMA (1,0). According to the generally accepted practice of choosing between complexity and accuracy, the model with the least number of parameters should be chosen from these models, provided that its accuracy is not significantly different from another model (the deviation between the RMSE ratio of the models should not exceed 10%). Compare between average quadratic error of the models: 18833/18802 = 1,002 (i.e., the deviation is 0.2%). We will give preference to the ARMA model (1;0) with a smaller value of the Akaike information criterion and predict the tax base for the 12 months of 2019.

```
Model 1: ARMA, using observations 2015:01-2018:12 (T = 48)
Estimated using Kalman filter (exact ML)
Dependent variable: v1
Standard errors based on Hessian
coefficient     std. error      z       p-value
----------------------------------------------------------
const      41638.7      5370.29      7.754    8.94e-015 ***
phi_1      0.457176     0.216583     2.111    0.0348     **
theta_1    0.0924766    0.229911     0.4022   0.6875
Mean dependent var   41419.25   S.D. dependent var   22430.81
Mean of innovations−74.8974   S.D. of innovations  18801.65
Log-likelihood−540.6736      Akaike criterion      1089.347
Schwarz criterion    1096.832      Hannan-Quinn        1092.176
```

Fig. 3. The results of the evaluation of the ARIMA model (1 ,0, 1).

Table 1: Evaluation results of ARMA models.

Model	Specification	AIC	RMSE	P-value test for the normal distribution of residues	P-value test Ljung-Box
ARMA (1,1)	Y_t=41638,7$^{(***)}$+0,457$^{(**)}$$Y_{t-1}$ +0,092ε_{t-1}+ε_t	1089,347	18 802	0,51830	0,7339
ARMA (1,0)	Y_t=41786,3+0,524Y_{t-1} +ε_t	1087,504	18 833	0,55754	0,8018
SARMA (1,1)x$(0,1)_6$.	Y_t=42616,3$^{(***)}$+0,501 $^{(**)}$$Y_{t-1}$ + 0,088 ε_{t-1} + 0,224 ε_{t-6} + 0,088*0,224 ε_{t-7}+ε_t	1089,604	18 334	0,38297	0,8174

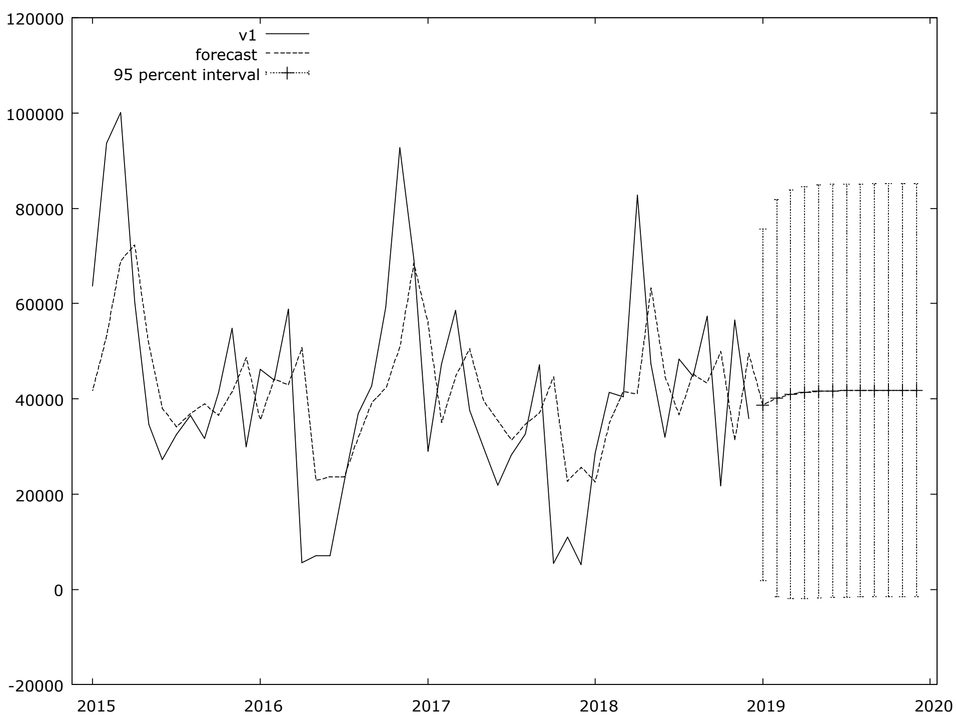

Fig. 4. Forecast of the enterprise tax base in 2019.

Table 2 : Results of the forecast of the enterprise tax base in 2019, thousand rubles. ARMA model (1:0).

Period	Forecast	Standard error	Lower limit	Upper limit
2019:01	38686.21	18833.355	1773.51	75598.90
2019:02	40161.06	21264.596	1516.78	81838.90
2019:03	40934.27	21885.555	1960.63	83829.17
2019:04	41339.63	22053.162	1883.77	84563.03
2019:05	41552.15	22099.006	1761.11	84865.40
2019:06	41663.56	22111.589	1674.36	85001.48
2019:07	41721.97	22115.047	1622.73	85066.66
2019:08	41752.59	22115.997	1593.97	85099.15
2019:09	41768.64	22116.258	1578.42	85115.71
2019:10	41777.06	22116.330	1570.15	85124.27
2019:11	41781.47	22116.349	1565.77	85128.72
2019:12	41783.79	22116.355	1563.47	85131.05

Table 3: Summary table of the seasonal modeling results using dummy variables.

Regressors	Dependent variable – tax base	
	Additive model (1)	Multiplicative model (2)
Trend	-233,866 (232,545)	-343,885 (282,581)
Q1 Dummy	11945,0 (9085,13)	
Q2 Dummy	-8870,69 (8950,21)	
Q3 Dummy	-2501,60 (8868,26)	
Q1 and Trend Dummy intersection		169,658 (346,323)
Q2 and Trend Dummy intersection		-38,385 (324,487)
Q3 and Trend Dummy intersection		138,790 (306,821)
Free ratio	47005,8*** (9195,57)	48314,7*** (6776,44)
P-value (F)	0,136	0,734
R2	0,147	0,045

We use the rule of thumb that simple linear models often produce better predictions. Some researchers believe that the simplicity and compactness of linear models makes them more resistant to incorrect specifications, which is important for long-term forecasting [11]. Therefore, we apply linear regression models with additive seasonality and multiplicative seasonality to the tax base forecast (Table 3).

As can be seen from Table 3, regression models with additive seasonality and multiplicative seasonality turned out to be statistically insignificant, as well as seasonal coefficients of these models are insignificant, which does not allow them to be used to obtain forecast estimates of the tax base.

IV. SUMMARY

In the process of forecasting the tax base of the enterprise, the method of predicting values by autoregressive models and models with a seasonal dummy variable was chosen. The hypothesis of the study on seasonality and tax base of the enterprise has not been confirmed. The obvious advantages of ARIMA models are the following: the presence of a formalized and the most detailed developed methodology, following which you can choose the model that is most suitable for each specific time series. Developed methods for automatic selection of the best ARIMA [12, 13] and is "greatly facilitate life" of the forecaster. In addition, point and interval forecasts follow from the model itself and do not require separate estimation. The advantages of modeling seasonality using dummy variables are the following: the method does not require classical decomposition of the time series and there is no loss of observations; the method allows to include seasonality in any regression model, not only in the trend model, which can be useful if we are trying to predict some indicator based on the values of another known; to estimate the parameters it is enough to use the usual method of least squares.

One of the obvious drawbacks of ARIMA models is the requirement for data series: to build an adequate ARIMA model requires at least 40 observations, and for SARIMA- about 6-10 seasons, which is not always possible in practice. The second serious drawback is the lack of adaptability of autoregression models: when receiving new data, the model should be periodically overestimated, and sometimes re-identified. The very same construction of the model is rather an "art", i.e. requires a lot of experience on the part of the forecaster. At the end of the last century, studies conducted by the International Institute of Forecasters have shown that ARIMA models have shown themselves to be no better than exponential smoothing models, and in each case you need to use your model [14]. Moreover, the use of models AR(1), AR(2) and ARMA(1,1) bypassing the Boxing-Jenkins methodology (i.e. without studying correlograms and estimating residuals) gives no less accurate predictions than the ARIMA models built on the basis of the Box-Jenkins methodology [15]. The construction of ARIMA models is based on the assumption that the time series is generated infinitely in accordance with some function whose parameters need to be identified and evaluated. However, economic processes, as we already know, are essentially irreversible, and therefore such a "technical" attitude to them does not allow to take into account their features and, as a result, does not allow to give accurate forecasts. In evolutionary economic processes, there are constant changes in all the characteristics of the distribution, and therefore the "race" for the best (unbiased, efficient and well-founded) estimates without heteroscedasticity and autocorrelation in the residues is more like a search for what does not exist, where it does not exist in principle. It is possible to point out the following disadvantages of seasonality modeling using dummy variables: the method involves averaging all seasonal coefficients and does not allow for the possibility of evolutionary changes in seasonal components over time; the method does not take into account possible "outliers" of data; it is required to pre-select the type of function that best describes this process (for simple trend models it can be difficult when the trend component evolves over time).

V. CONCLUSIONS

The advantage of linear models is due to their overall "robustness" of resistance to wrong specification of the model, the resistance to bias and inaccuracies in the evaluation of resistance to structural change and the drift of the model parameters. The obvious advantages of using linear models include the fact that it is based on a very clear mathematical and statistical justification, which makes them one of the most scientifically sound models of the whole set of models for predicting trends in time series. Another undeniable advantage is the formalized and most detailed developed method, following which you can choose the model that is most suitable for each specific time series. The perspective direction of development of this research can be obtaining forecast estimates of accounts payable on the basis of modeling of multidimensional time series through the elimination of false regression and analysis of co-integration. Polynomial trends described by discrete polynomials of low orders are also popular. Effective application in the construction of polynomial trends can find algorithms for their evaluation based on discrete transformations [16]. Also of practical interest is forecasting using fuzzy time series models [17]. In conditions where all individual models are imperfect approximations of the true process, a combination of forecasts may be optimal. In a series of papers, Granger and his co-authors developed a technique for obtaining an optimal combination of forecasts, when all individual models are only an approximation of the process that generates data [18,19]. In addition, it is likely that in many situations the predicted process itself is a combination of simpler microprocesses.

ACKNOWLEDGEMENTS

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

REFERENCES

[1]. Makridakis, S., & Hibon, M. (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting*. No. 16. P. 451-476.

[2]. Athanasopoulos George, Hyndman Rob J., Song Haiyan, & Wu Doris C. (2011). Thetourism forecasting competition. *International Journal of Forecasting*. No. 27. P. 822–844.

[3]. Needles, B.E., Powers, M., Frigo, M.L., & Shigaev, A.I. (2018). Integrated Reporting and Sustainability Reporting: A Global Assessment. In Marc J. Epstein, Frank H. M. Verbeeten, Sally K. Widener (ed.) Performance Measurement and Management Control: The Relevance of Performance Measurement and Management Control Research *Studies in Managerial and Financial Accounting*, Vol. *33*. Emerald Publishing Limited, pp. 93-120

[4]. Needles, B.E., Powers, M., Frigo, M.L., & Shigaev A.I. (2016) Integrated reporting and sustainability reporting: An exploratory study of high performance companies//Performance Measurement and Management Control: Contemporary Issues. Studies in Managerial and Financial Accounting, Vol. *31*. Edited by: Marc J. Epstein, Frank Verbeeten, Sally K. Widener. Emerald Group Publishing Limited. p. 41-81.

[5]. Nagumanova, R. (2018). Method of calculation of critical values of financial indicators for developing food security strategy. International Conference Information Technologies in Business and Industry. IOP Conf. Series: *Journal of Physics: Conf*.

[6]. Brooks, C. (2008). Introductory Econometrics for Finance. Cambridge University Press, p. 674.

[7]. Box, G., Jenkins, G. (1970). Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day.

[8]. Hyndman, R. J. Forecasting: Principles & Practice. (2018). Leader: 23-25. University of Western Australia. 138 p.

[9]. Dickey, D. A. & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time-Series with a Unit Root. *Journal of the American Statistical Association,* Vol. *74*. pp. 427-431.

[10]. Ljung, G. M., & Box, G. E. P. (1978). On a Measure of Lack of Fit in Time Series Models. Biometrika, Vol. *65*, pp. 297-303.

[11]. Stock, J, & Watson, M. (1999). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. *R. Engle. H. White. Cointegration, Causality and Forecasting: A Festschrift for Clive W. J. Granger.* Oxford: Oxford University Press, 1-44.

[12]. Hyndman, R.J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: the forecast Package for R. *Journal of Statistical Software,* Vol. *27*(3), pp.1-23.

[13]. Hyndman, R. J, & Athanasopoulos, G. (2013). Forecasting: principles and practice. Melbourne, Australia: O Texts.

[14]. Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. & Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting.* Vol. *1*(2), pp. 111–153.

[15]. Makridakis, S., & Hibon, M. (1997). ARMA models and the Box-Jenkins Methodology. *Journal of Forecasting*. Vol. *16*., pp. 147-163.

[16]. Ismagilov, I.I., & Khasanova, S.F. (2016). Algorithms of parametric estimation ofpolynomial trend models of time series ondiscrete transforms.*Academy of Strategic Management Journal*. Vol. *15* (Special Issue), pp 21-28.

[17]. Ismagilov, I.I., & Khasanova, S.F. (2015). Short-Term Fuzzy Forecasting of Brent oil prices. *Asian Social Science,* Vol. *11*, pp.60-67.

[18]. Bates, J., & Granger, C. (1969). The combination of forecast. *Operations Research Quarterly 20*, pp. 451-468.

[19]. Granger, C., & Newbold, P. (1977). Forecasting Economic Time Series. New York: Academic Press.