STATISTICAL ANALYSIS OF A COMPANY'S REVENUE USING TIME SERIES

Alexandru-Eugen Stătescu[1]

Abstract: Economic and financial analysis is a method of knowing the mechanism of formation and modification of the economic phenomena through their decomposition into their component elements and by identifying the factors of influence. The object of the decomposition into elements or factors may be a result (structural analysis), or a change in the result from a basis of comparison (causal analysis).

Revenue is the inflow of economic benefits during the reporting period resulted in the ordinary activity of the company as assets increase or decrease in debt that build equity excluding gains from property company contributions.

The purpose of this paper is to analyze, using statistical research methods, the evolution of the income of a state owned company. For this it will be used the data from the Income and Expense Budget of METROREX S.A. on a 10-year horizon. In order to analyze the time evolution of the enterprise's revenue, it will be used the chronological series analysis methodology, the set of data from the mentioned source (namely the Income and Expense Budget of METROREX SA) and will design an econometric model with a trend and residual variable component.

Time series is a form of orderly presentation of statistical data which reflect the manifestation of phenomena in a given moment or time. In other words, the time series is a sequence of values of an economic indicator or other observed over time, reflecting the process of change and development of a statistical sample in successive periods of time.

Also the purpose of this paper is to build an ARMA model that fits in an appropriate way the evolution of the revenue's time series.

JEL Classification Number: C32; **DOI:** http://dx.doi.org/10.12955/cbup.v6.1198

Keywords: Financial and economic analysis, Revenue, Time series, Adjustment function, Adjusting the model parameters

Introduction

Income and expenses are not equal to each other, and as a result, there is a change in the volume of the property, a change that is reflected in the accounting through the result of the year (profit or loss).

Information provided by accounting on the profitability and performance of an enterprise is useful in anticipating the ability of an enterprise to generate cash flows with existing resources. They are also useful for making judgments about the efficiency with which the enterprise can use the new resources.

The balance sheet presents assets, liabilities and equity on the basis of which the financial position of an enterprise is measured. The Profit and Loss Account (Income Statement) presents the income and expenses on the basis of which the enterprise's performance is measured.

The general framework defines the expenses and revenues as follows:

- Expenditure is a decrease in the economic benefits recorded during the accounting period in the form of asset write-offs or decreases, which result in reductions in equity other than those resulting from their distribution to shareholders.
- Revenues are increases in the economic benefits recorded during the accounting period in the form of inflows or increases of assets or decreases in debt, which result in increases in equity other than those resulting from shareholder contributions.

Time series. Stationarity. Statistical tests

The present paper proposes to analyze, using statistical research methods, the evolution in time of the revenues of a state owned company. For this it will be used the data from the Income and Expense Budget of METROREX S.A. on a 10-year horizon. In order to analyze the time evolution of the enterprise's revenue, it will be used the chronological series analysis methodology, the dataset being from the mentioned source (i.e. the Income and Expenses Budget of METROREX S.A.).

A chronological series is a form of orderly presentation of statistical data that reflects the level of manifestation of phenomena at a given time or period of time. In other words, the chronological series is a series of values of an economic or other indicator, observed over time, mirroring the process of changing and developing a statistical collectivity over successive periods of time.

The general form of a chronological series can be as follows:

$$\begin{pmatrix} 1 & 2 & 3 & & t & & T \\ Y_1 & Y_2 & Y_3 & \cdots\cdot\cdot & Y_t & \cdots\cdot\cdot & Y_T \end{pmatrix}$$

[1] Bucharest University of Economic Studies, Economic Cybernetics and Statistics Doctoral School, alexandrustatescu@yahoo.com

where:

t = time or time interval (t = 1, T);

y_t= the level (expressed by absolute or relative data) reached by the Y phenomenon at time t.

The statistical description of the time series starts from the analysis of the factors causing their movement. Generally, the evolution of a phenomenon is generated by the action of some groups of factors:

- the essential factors with long-lasting action, which implies the trend of their evolution; the action of these factors being studied according to the time units for which the phenomenon analyzed was measured;

- seasonal factors, for periods of less than one year, which lead to deviations from the trend of the phenomenon imprinted by the essential factors;
- cyclical factors, with effects over periods of more than one year, which imply an oscillating evolution of the phenomenon in the case of long-time series;
- Random factors (random action), the action of which is compensated if the recorded data refers to a large number of periods of time. Starting from the structure of the factors that determine the evolution of a phenomenon, the statistical description of the time series can be made using the following models:

1. Addition models

   $y_t = f(t) + s(t) + c(t) + u(t)$ (3.1)

2. Multiplicative models

   $yt = f(t)\cdot s(t)\cdot c(t)\cdot u(t)$ (3.2)

where:

$f(t)$ = the trend component, the effect of the action of the key factors;
$s(t)$ = seasonal component, effect of seasonal factors action;
$c(t)$ = the cyclical component generated by the action of cyclic factors;
$u(t)$ = the residual component, which expresses the influence of the random factors on the evolution of the phenomenon.

A time series y_t is stationary if:

-$E(y_t) = \bar{y}\ , \forall\ t = \overline{1,T}$ (3.3) the average of the series does not depend on time;

- $var(y_t) = \sigma_y^2 = constant, \forall\ t = \overline{1,n}$ (3.4) the dispersion of the series is time independent;

- $Cov(y_t, y_{t+k}) = constant, \forall\ t = \overline{1,n}\ , k < n$ (3.5) the covariance of the series does not depend on time. The definition above is the definition of weak stationarity. A series of time is therefore stationary if its media, dispersion and covariance remain constant over time. If any of the above conditions is not satisfied, then the time series is non-stationary.

If the first two conditions above cannot be accepted, but $Cov(y_t, y_{t+k}) = \rho(k), \forall\ t = \overline{1,n}\ , k < n$ (3.10), where $\rho(k)$ represents the k-autocorrelation function, then the assumption of weak stationarity can be accepted.

The time series $\{\varepsilon_t\}_{t=\overline{1,T}}$ consisting of randomly uncorrelated variables, respectively:

-$E(\varepsilon_t) = 0\ , \forall\ t = \overline{1,T}$
- $var(\varepsilon_t) = \sigma_\varepsilon^2, \forall\ t = \overline{1,n}$
- $Cov(\varepsilon_t, \varepsilon_p) = 0, \forall\ t, p = \overline{1,T}\ , t \neq p$

is called the white noise.

This series is stationary with the autocovariance function: $\rho(k) = \begin{cases} \sigma^2, k = 0 \\ 0, \text{otherwise} \end{cases}$

and the autocorrelation function $r(k) = \begin{cases} 1, k = 0 \\ 0, \text{otherwise} \end{cases}$.

It is noted $WN(0, \sigma^2)$ the white noise, namely $\{\varepsilon_t\}_{t=\overline{1,T}} \approx WN(0, \sigma^2)$.

The time series $\{y_t\}_{t=\overline{1,T}}$, where $y_t = \varepsilon_t - \gamma \cdot \varepsilon_{t-1}$, $\gamma \neq 0$ and $\{\varepsilon_t\}_{t=\overline{1,T}} \approx WN(0, \sigma^2)$, is called the first-order moving average $MA(1)$, it is noted $\{y_t\}_{t=\overline{1,T}} \approx MA(1)$. This series is stationary for any $\gamma \neq 0$, has the mean 0 and the autocovariance function: $\rho(k) = \begin{cases} (1+\gamma^2)\cdot\sigma^2, k = 0 \\ -\gamma^2 \cdot \sigma^2, k = \pm 1 \\ 0, \text{otherwise} \end{cases}$ and autocorrelation function $(k) = \begin{cases} 1, k = 0 \\ -\frac{\gamma}{1+\gamma^2}, k = \pm 1 \\ 0, \text{otherwise} \end{cases}$.

In analyzing time series, most of these are non-stationary.

A time series $\{y_t\}_{t=\overline{1,T}}$ is non-stationary if the mean and / or variance is variable over time.

Depending on the characteristics of the average, a non-stationary time series of may be:

- a homogeneous non-stationary time series, in which case:
  - the average of the series is not constant over time, it has a linear trajectory with a positive or negative slope;
  - the series is characterized by constant variations from one period to another.
- a series of non-homogeneous non-stationary time, in which case:
  - both average and variance are variable over time;
  - the series is characterized by non-constant variations from one period to another.

To test the stationarity of a time series, the following statistical tests are used:

- ADF test (Augmented Dickey-Fuller);
- PP test (Phillips-Perron).

The *t* test (Student)
H_0: the series has a unitary root (is non-stationary);
$t_{test\_ADF} < t_{critic(1\%,5\%,10\%)}$ → is rejected H_0(series is stationary)
Prob < Relevance_Level (1%,5%,10%) → is rejected H_0(series is stationary).

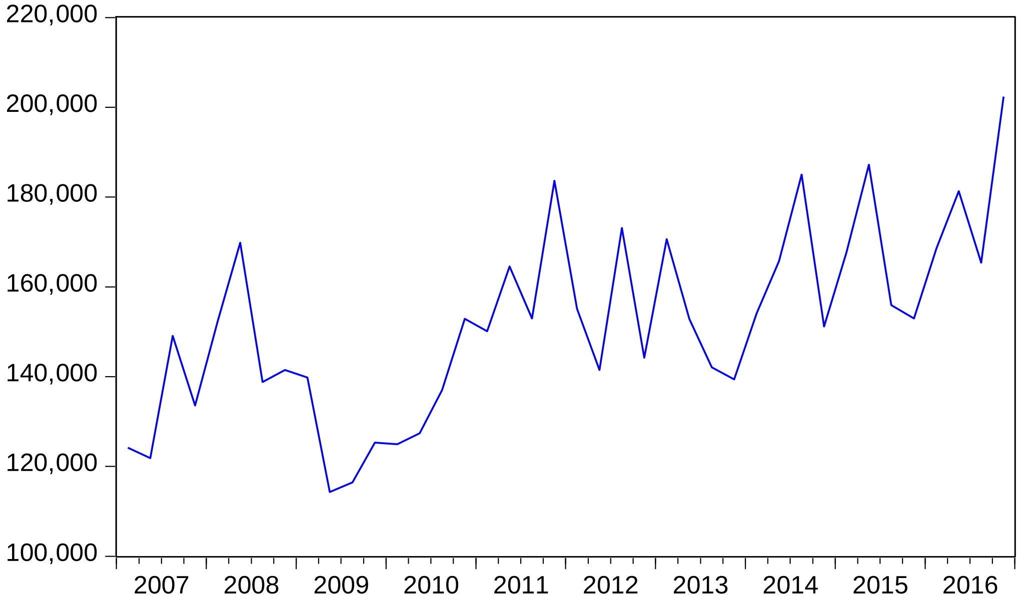
In order to analyze the evolution in time of the revenues of a company with state-owned capital, it will be used the data from the Income and Expense Budget of METROREX S.A. on a 10-year horizon and every quarter.

The Revenue time series is as follows:

Year/Quarter	Revenue (thousands lei)
2007Q1	124162.5824190487
2007Q2	121830.3564447934
2007Q3	149069.5202375271
2007Q4	133564.4408986309
2008Q1	152159.2502677935
2008Q2	169822.9430267043
2008Q3	138791.5493957481
2008Q4	141448.4673097541
2009Q1	139826.4564389078
2009Q2	114276.2584948338
2009Q3	116463.8746693226
2009Q4	125282.7703969358
2010Q1	124966.0547784584
2010Q2	127358.3081327084
2010Q3	137002.1538544534
2010Q4	152906.306567713
2011Q1	150094.2319658395
2011Q2	164550.5501226293
2011Q3	152967.5196799344
2011Q4	183652.7102315969
2012Q1	155116.5414363921
2012Q2	141489.0326087752

Year/Quarter (contd.)	Revenue (thousands lei) (contd.)
2012Q3	173123.536902886
2012Q4	144197.5890519468
2013Q1	170587.9838074871
2013Q2	152844.7173169066
2013Q3	142085.6830118408
2013Q4	139416.7958637656
2014Q1	154083.3072818941
2014Q2	165750.8275678453
2014Q3	184992.3241416051
2014Q4	151189.0610086556
2015Q1	167716.6768345032
2015Q2	187186.3827178873
2015Q3	155910.7766289194
2015Q4	152982.2038186901
2016Q1	168560.2424250711
2016Q2	181323.9874574992
2016Q3	165394.0665293109

The time series graph is as follows:



The results of the ADT (Augmented Dickey-Fuller) test for the presented time series (Revenue) are as follows:

Null Hypothesis: VENIT has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=9)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.535271	0.0043
Test critical values:	1% level	-4.211868	
	5% level	-3.529758	
	10% level	-3.196411	

*MacKinnon (1996) one-sided p-values.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
VENIT(-1)	-0.771840	0.170186	-4.535271	0.0001
C	100225.0	22664.89	4.422039	0.0001
@TREND("2007Q1")	901.7130	288.1934	3.128847	0.0035

R-squared	0.365458	Mean dependent var	2005.404
Adjusted R-squared	0.330206	S.D. dependent var	19427.81
S.E. of regression	15899.90	Akaike info criterion	22.25982
Sum squared resid	9.10E+09	Schwarz criterion	22.38778
Log likelihood	-431.0664	Hannan-Quinn criter.	22.30573
F-statistic	10.36694	Durbin-Watson stat	1.960003
Prob(F-statistic)	0.000278		

As you can see $t_{test\_ADF} = -4.535271$ and the associated value p is 0.0043. Because $t_{test\_ADF} < t_{critic(1\%,5\%,10\%)}$ is rejected H_0, so the time series is stationary.

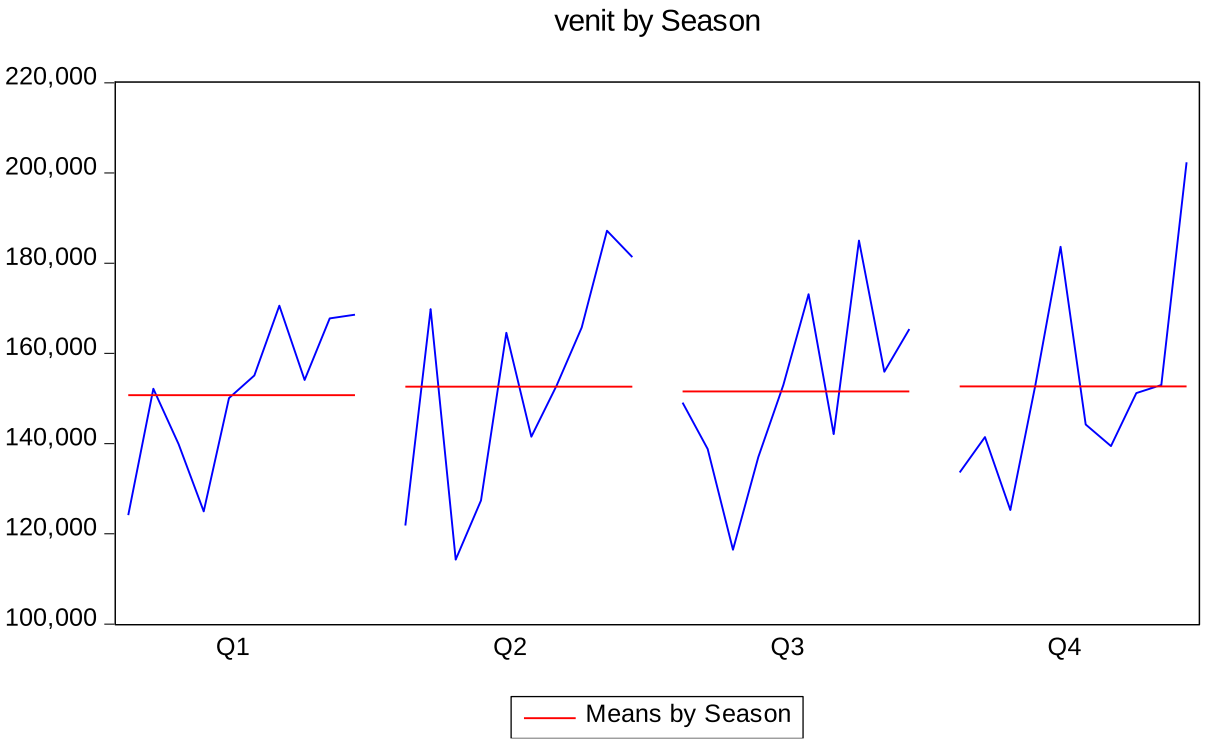
Correlogram errors are used to match the ARMA model specification. In order to analyze autocorrelation coefficients and partial autocorrelation coefficients, select option View → Correlogram, which is available in the window of the series considered.

The correlogram obtained for the studied variable is as follows:

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
. |*** |	. |*** |	1	0.469	0.469	9.4775	0.002
. |*** |	. |** |	2	0.406	0.238	16.753	0.000
. |*** |	. |*. |	3	0.386	0.175	23.504	0.000
. |*. |	.*| . |	4	0.157	-0.166	24.655	0.000
. |*. |	. | . |	5	0.135	-0.018	25.528	0.000
. |*. |	. |*. |	6	0.182	0.126	27.162	0.000
. |*. |	. | . |	7	0.103	0.023	27.708	0.000
. |*. |	. | . |	8	0.094	-0.021	28.170	0.000
. |*. |	. |*. |	9	0.165	0.084	29.636	0.001
. |*. |	. | . |	10	0.075	-0.033	29.955	0.001
. |*. |	. | . |	11	0.123	0.068	30.830	0.001
. | . |	.*| . |	12	0.019	-0.154	30.852	0.002
. | . |	. |*. |	13	0.070	0.101	31.153	0.003
. |*. |	. |*. |	14	0.130	0.122	32.251	0.004
. |*. |	. | . |	15	0.080	-0.001	32.678	0.005
.*| . |	***| . |	16	-0.100	-0.344	33.373	0.007
. | . |	. |*. |	17	0.052	0.155	33.573	0.010
.*| . |	.*| . |	18	-0.098	-0.075	34.308	0.012
.*| . |	. | . |	19	-0.136	-0.007	35.781	0.011
. | . |	.*| . |	20	-0.030	-0.078	35.858	0.016

According to the time series correlogram, it can be seen that the autocorrelation function (AC) has the first 7 significant values different from 0 and decreasing. Also, the first 4 values of the partial autocorrelation function (ACP) differ significantly from 0. It can be concluded that the series will have a MA component with the number of lags (q) equal to at most 7 and that the autoregressive process has an AR component of order p = 4.

Analyzing the seasonality of the time series is done as follows: from the View menu of the time series, select the Graph option, then the Seasonal Graph and the seasonality graph is obtained. For the time series presented, the seasonality graph generated by EViews is as follows:



The method of removing from the time series values the trend values to be used is the moving average method. Thus, to adjust the seasonality of the time series, it is selected from the Proc data series window, then Seasonal Adjustment, Moving Average Methods, Ratio to moving - Multiplicative. The adjusted venity series is obtained.

A first equation could be: revenue ar(1) ar (2) ar (3) ar (4) c. The results of parameter estimation generated with EViews are as follows:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	160068.5	13346.15	11.99361	0.0000
AR(1)	0.386831	0.188422	2.052998	0.0486
AR(2)	0.231909	0.191494	1.211050	0.2350
AR(3)	0.313735	0.191865	1.635180	0.1121
AR(4)	-0.191394	0.185977	-1.029131	0.3114

R-squared	0.374317	Mean dependent var	154108.2
Adjusted R-squared	0.293583	S.D. dependent var	20413.05
S.E. of regression	17156.88	Akaike info criterion	22.46643
Sum squared resid	9.13E+09	Schwarz criterion	22.68637
Log likelihood	-399.3958	Hannan-Quinn criter.	22.54320
F-statistic	4.636455	Durbin-Watson stat	1.881560
Prob(F-statistic)	0.004744		

Inverted AR Roots	.79	.47	-.44+.56i	-.44-.56i

Heteroskedasticity Test: White

F-statistic	0.951225	Prob. F(14,21)	0.5271
Obs*R-squared	13.97020	Prob. Chi-Square(14)	0.4519
Scaled explained SS	4.658408	Prob. Chi-Square(14)	0.9900

As we can see, it can be said that the autoregressive model is valid because the F-statistic = 4.636455 > F critic and Prob(F-statistic)= 0.004744 < 0,05. Because the Durbin-Watson test value is 1.881560 and is close to 2 then it can be said that there is no serial correlation between residues.

The equation thus obtained is:

Revenue=160068.512989+0.3868309333*$Revenue_{t-1}$+0.231909334819*$Revenue_{t-2}$+0.313734553884*$Revenue_{t-3}$-0.191394355907*$Revenue_{t-4}$

> F critic and Prob(F-statistic)= 0.011579 < 0,05. Because the Durbin-Watson test value is 1.881706 and is close to 2 then it can be said that there is no serial correlation between residues.

Another equation could be: revenue ar (1) ma (1) c. The results of estimating the parameters generated with EViews are as follows:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	163524.9	17136.16	9.542679	0.0000
AR(1)	0.881660	0.129015	6.833761	0.0000
MA(1)	-0.506109	0.216197	-2.340956	0.0249

R-squared	0.355531	Mean dependent var	152624.6
Adjusted R-squared	0.319727	S.D. dependent var	20511.67
S.E. of regression	16917.75	Akaike info criterion	22.38392
Sum squared resid	1.03E+10	Schwarz criterion	22.51188
Log likelihood	-433.4864	Hannan-Quinn criter.	22.42983
F-statistic	9.929962	Durbin-Watson stat	1.962302
Prob(F-statistic)	0.000368		
Inverted AR Roots	.88		
Inverted MA Roots	.51		

As we can see, it can be said that the autoregressive model is valid because F-statistic= 9.929962 > F critic and Prob(F-statistic) = 0.000368 < 0,05. Because the Durbin-Watson test value is 1.962302 and is close to 2 then it can be said that there is no serial correlation between residues.
Another equation could be: revenue ar(1) ar (2) ar (3) ma (1) ma (2) c. The results of the estimation of parameters generated with EViews are as follows:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	160287.0	13605.79	11.78079	0.0000
AR(1)	-0.082614	0.379227	-0.217848	0.8290
AR(2)	0.150075	0.339450	0.442111	0.6615
AR(3)	0.503365	0.213167	2.361363	0.0247
MA(1)	0.484129	0.409589	1.181986	0.2462
MA(2)	0.275258	0.386167	0.712796	0.4813

R-squared	0.392751	Mean dependent var	153552.9
Adjusted R-squared	0.294807	S.D. dependent var	20408.93
S.E. of regression	17138.55	Akaike info criterion	22.48344
Sum squared resid	9.11E+09	Schwarz criterion	22.74467
Log likelihood	-409.9437	Hannan-Quinn criter.	22.57554
F-statistic	4.009978	Durbin-Watson stat	1.934535
Prob(F-statistic)	0.006383		
Inverted AR Roots	.83	-.46+.63i	-.46-.63i
Inverted MA Roots	-.24+.47i	-.24-.47i	

As we can see, it can be said that the autoregressive model is valid because F-statistic= 4.009978 > F critic and Prob(F-statistic) = 0.006383 < 0,05. Because the Durbin-Watson test value is 1.934535 and is close to 2 then it can be said that there is no serial correlation between residues.

Another equation could be: revenue ar (1) ar (2) ar (3) ar (4) ma (1) ma (2) ma (3) c. The results of estimating the parameters generated with EViews are as follows:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	159478.5	13768.20	11.58310	0.0000
AR(1)	0.338820	1.128363	0.300276	0.7662
AR(2)	-0.498684	0.431416	-1.155924	0.2575
AR(3)	0.710397	0.501375	1.416898	0.1675
AR(4)	0.005560	0.738372	0.007530	0.9940
MA(1)	0.059988	1.199182	0.050024	0.9605
MA(2)	0.937647	0.061490	15.24875	0.0000
MA(3)	-0.067818	0.931639	-0.072794	0.9425

R-squared	0.453953	Mean dependent var	154108.2
Adjusted R-squared	0.317441	S.D. dependent var	20413.05
S.E. of regression	16864.67	Akaike info criterion	22.49696
Sum squared resid	7.96E+09	Schwarz criterion	22.84885
Log likelihood	-396.9453	Hannan-Quinn criter.	22.61978
F-statistic	3.325378	Durbin-Watson stat	1.955729
Prob(F-statistic)	0.010523		

Inverted AR Roots	.81	-.01	-.23-.91i	-.23+.91i
Inverted MA Roots	.07	-.07-.97i	-.07+.97i	

As we can see, it can be said that the autoregressive model is valid because F-statistic= 3.325378 > F critic and Prob(F-statistic) = 0.010523< 0,05. Because the Durbin-Watson test value is 1.955729 and is close to 2 then it can be said that there is no serial correlation between residues.

Conclusions

According to the time series correlogram, it can be seen that the autocorrelation function (AC) has the first 7 significant values different from 0 and decreasing. Also, the first 4 values of the partial autocorrelation function (ACP) differ significantly from 0. It can be concluded that the series will comprise a MA component with the number of lags (q) equal to at most 7 and that the autoregressive process has an AR component of order p = 4.

Several autoregressive patterns can be built, we have generated 3 such models. From the data generated by EViews following the estimation of the five equation parameters, it can be observed that all autoregressive models are statistically valid (F-statistic> F critical and prob (F-statistic) = <0,05).

To choose the best model, it is chosen according to the classical criterion, namely the highest value of the regression coefficient R^2 (which is 0.453953) for the 5th model, respectively the lattest, but also because it has all the coefficients of the parameters significantly different from 0. The selected model is ARMA (4.3).

The equation thus obtained is:

Revenue=159478.5+0.338820*Revenue_{t-1}-0.498684*Revenue_{t-2}+0.710397*Revenue_{t-3}-0.005560*Revenue_{t-4}+0.059988*ε_t+0.937647*ε_{t-1}-0.067818*ε_{t-2}

References

Voineagu, V. Boboc, C., Ghita, S., Titan, E. (2007). Statistica : baze teoretice si aplicatii, Editura Economică, București, 153-160.

Tudorel, A., Regis, B. (2008). Econometrie, Editura Economică, București, 335-381.

Stancu, S., (2011). Econometrie. Teorie și aplicații utilizând EViews, Editura ASE, București, 377-444.

Tudorel, A., Stancu, S., Iacob, A. I., Tușa, E., (2008). Introducere în econometrie utilizând Eviews, Editura Economică, București, 189-209.