THE EVOLUTION OF SNP PETROM STOCK LIST - STUDY THROUGH AUTOREGRESSIVE MODELS

Marian Zaharia,
Ioana Zaheu, and
Elena Roxana Stan *

Abstract

Stock exchange market is one of the most dynamic and unpredictable markets. In this context, this work intends to analyze the SNP Petrom shares on the REGS market, based on the chronological series.

The economic series are often not stationary, but they can be stationarized by different data transformations. The simplest method used for stationarizing a series is to apply differentiating operators of various classes on the series. After applying this operator, a stationary series that can be modified by an ARIMA (p.q) process is usually obtained.

Most time series with economic content include a seasonal component besides the trend and random component.

The purpose of this work is to estimate the parameters of an ARIMA (p,d,q) model for SNP Petrom shares, where p is the number of autoregressive terms, d is the integration level of the series (how many times the series must be differentiated in order to become stationary) and q is the number of moving average terms (MA).

Keywords**:** list, economic series, autoregressive models

Introduction

In literature the determination of the best ARIMA(p,d,q) sample in order to shape certain remarks for a series of time entails an assembly of techniques and methods, better known as the Box-Jenkins methodology.

A process $\{Y_t\}$, t belongs to Z, it admits a representation ARIMA(p,d,q) should this meet the subsequent equality: $\Phi(L)(1-L)^d Y_t=\Theta(L)\varepsilon_t$, whereas ε_t is a white noise, the two polinomes $\Phi(L)= 1-\sum\varphi_i L^i$, $\Theta(L) = 1-\sum\theta_i L^i$ have roots larger than one, as the initial conditions $y_{-p-d}, \ldots y_{-1}, \varepsilon_{-q}, \ldots, \varepsilon_{-1}$ are not correlated with the random variables $\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_t, \ldots$

---

* Marian Zaharia, Ioana Zaheu and Elena Roxana Stan are at the Romanian American University in Bucharest.

Building the model with box-jenkins methodology

The Box-Jenkins methodology comprises three main aspects:
- ♣ identification;
- ♣ estimate;
- ♣ checking.

Sample identification

Having available the sample of remarks on the evolution of SNP Petrom share quotation, a series of transformations must be brought to these so as to induce stationarity.

In case of time series describing the processes on the financial market, a scale transformation appears necessary, whereas most of the time the initial i series is being applied a logarithmic filter, in order to have a stationary series.

The next step is the elimination of the determinist component, after finding the possible oscillations present in the evolution of the series (Figure 1.).

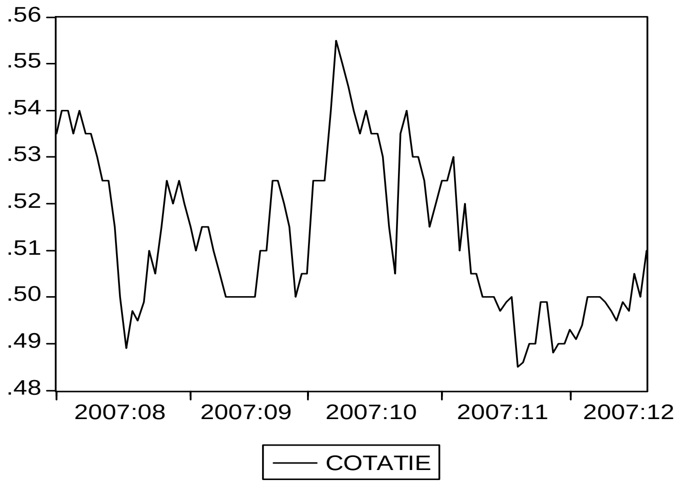

Figure 1 – Average price evolution of
Petrom SA shares on the market

Currently we are able to determine for which values of the parameters p and q the ARMA(p,q) process shape to the best in the stationary series obtained. A criterion in this regard is the behaviour of the autocorrelation (ACF) and of the partial autocorrelation (PACF) functions.

Corelograma p_RRC

Included observations: 489

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
.|********	.|********	1	0.978	0.978	470.55	0.000
.|*******|	*|. |	2	0.951	-0.126	916.31	0.000
.|*******|	.|. |	3	0.924	0.001	1338.0	0.000
.|*******|	.|. |	4	0.900	0.048	1738.6	0.000
.|*******|	.|. |	5	0.874	-0.054	2117.6	0.000
.|*******|	.|. |	6	0.848	-0.020	2475.0	0.000
.|****** |	.|. |	7	0.824	0.045	2813.3	0.000
.|****** |	*|. |	8	0.798	-0.076	3131.5	0.000
.|****** |	.|. |	9	0.771	-0.031	3429.1	0.000
.|****** |	.|. |	10	0.745	0.011	3707.1	0.000
.|****** |	.|. |	11	0.720	0.018	3967.8	0.000
.|***** |	.|. |	12	0.697	0.005	4212.5	0.000
.|***** |	.|. |	13	0.676	0.025	4442.7	0.000
.|***** |	.|. |	14	0.653	-0.038	4658.3	0.000
.|***** |	.|. |	15	0.633	0.050	4861.6	0.000
.|***** |	.|. |	16	0.613	-0.046	5052.1	0.000
.|***** |	.|. |	17	0.592	0.005	5230.5	0.000
.|**** |	.|. |	18	0.572	-0.009	5397.3	0.000
.|**** |	.|. |	19	0.553	0.004	5553.5	0.000
.|**** |	.|* |	20	0.539	0.092	5702.0	0.000
.|**** |	.|. |	21	0.525	-0.018	5843.2	0.000
.|**** |	.|* |	22	0.514	0.079	5979.2	0.000
.|**** |	.|. |	23	0.504	-0.016	6110.1	0.000
.|**** |	.|. |	24	0.495	0.020	6236.8	0.000
.|**** |	.|. |	25	0.485	-0.030	6358.5	0.000
.|**** |	.|. |	26	0.476	0.014	6475.8	0.000
.|**** |	.|. |	27	0.466	-0.012	6588.7	0.000
.|*** |	*|. |	28	0.455	-0.058	6696.6	0.000
.|*** |	.|. |	29	0.443	-0.003	6799.1	0.000
.|*** |	.|. |	30	0.433	0.020	6897.1	0.000
.|*** |	.|. |	31	0.424	0.024	6991.2	0.000
.|*** |	.|. |	32	0.417	0.051	7082.5	0.000
.|*** |	.|. |	33	0.409	-0.029	7170.7	0.000
.|*** |	.|. |	34	0.403	0.045	7256.4	0.000
.|*** |	.|. |	35	0.397	-0.005	7339.9	0.000
.|*** |	.|* |	36	0.394	0.075	7422.3	0.000

We can see that ACF decreases very slowly (up to 36 lags are statistically significant), as PACF dramatically decreases after the first lag. ACF suggests that the series of prices is not stationary, and it must be differentiated before applying the Box-Jenkins methodology. The test for the unit-root Dickey Fuller set out below proves that our series is actually integrated of order 1 (and not more).

Null Hypothesis: P_RRC has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic based on SIC, MAXLAG=17)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.719685	0.0714
Test critical values:	1% level	-3.443551	
	5% level	-2.867255	
	10% level	-2.569876	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(P_RRC)
Method: Least Squares
Sample (adjusted): 1/04/2006 11/15/2007
Included observations: 487 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
P_RRC(-1)	-0.024636	0.009059	-2.719685	0.0068
D(P_RRC(-1))	0.130416	0.044974	2.899791	0.0039
C	0.002333	0.000881	2.648450	0.0084

R-squared	0.029038	Mean dependent var	-4.52E-05
Adjusted R-squared	0.025026	S.D. dependent var	0.002704
S.E. of regression	0.002670	Akaike info criterion	-9.007141
Sum squared resid	0.003451	Schwarz criterion	-8.981341
Log likelihood	2196.239	F-statistic	7.237291
Durbin-Watson stat	2.000354	Prob(F-statistic)	0.000800

Null Hypothesis: D(P_RRC) has a unit root
Exogenous: Constant

Lag Length: 0 (Automatic based on SIC, MAXLAG=17)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-19.54109	0.0000
Test critical values:	1% level	-3.443551	
	5% level	-2.867255	
	10% level	-2.569876	

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

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(P_RRC,2)

Method: Least Squares

Sample (adjusted): 1/04/2006 11/15/2007

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(P_RRC(-1))	-0.880858	0.045077	-19.54109	0.0000
C	-4.00E-05	0.000122	-0.328679	0.7425

R-squared	0.440506	Mean dependent var	-2.05E-06
Adjusted R-squared	0.439352	S.D. dependent var	0.003590
S.E. of regression	0.002688	Akaike info criterion	-8.996081
Sum squared resid	0.003504	Schwarz criterion	-8.978881
Log likelihood	2192.546	F-statistic	381.8541
Durbin-Watson stat	1.996914	Prob(F-statistic)	0.000000

After having established that the series is integrated of order 1, we are interested in ACF and PACF for the first difference d(p_RRC).

Sample: 1/02/2006 1/18/2008
Included observations: 488

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
.|* |	.|* |	1	0.119	0.119	6.9697	0.008
.|. |	.|. |	2	0.002	-0.012	6.9716	0.031
*|. |	*|. |	3	-0.062	-0.062	8.8961	0.031

.|. |	.|. |	4	0.021	0.036	9.1067	0.058
.|. |	.|. |	5	0.021	0.014	9.3220	0.097
.|. |	*|. |	6	-0.049	-0.059	10.529	0.104
.|. |	.|. |	7	0.045	0.063	11.523	0.117
.|. |	.|. |	8	0.039	0.029	12.283	0.139
.|. |	.|. |	9	-0.000	-0.018	12.283	0.198
.|. |	.|. |	10	-0.054	-0.043	13.744	0.185
.|. |	.|. |	11	-0.029	-0.013	14.159	0.224
.|. |	.|. |	12	-0.047	-0.052	15.292	0.226
.|. |	.|. |	13	0.010	0.020	15.338	0.287
*|. |	*|. |	14	-0.077	-0.081	18.311	0.193
.|. |	.|. |	15	0.009	0.021	18.352	0.245
.|. |	.|. |	16	-0.009	-0.013	18.392	0.301
.|. |	.|. |	17	0.007	0.004	18.419	0.363
.|. |	.|. |	18	-0.035	-0.033	19.053	0.389
*|. |	*|. |	19	-0.107	-0.093	24.885	0.164
.|. |	.|. |	20	-0.024	-0.010	25.182	0.195
*|. |	*|. |	21	-0.091	-0.089	29.418	0.104
.|. |	.|. |	22	-0.011	-0.005	29.480	0.132
.|. |	.|. |	23	-0.013	-0.008	29.568	0.162
.|. |	.|. |	24	0.012	-0.006	29.638	0.197
.|. |	.|. |	25	0.004	-0.002	29.648	0.238
.|. |	.|. |	26	0.011	0.016	29.716	0.280
.|. |	.|. |	27	0.027	0.027	30.096	0.310
.|. |	.|. |	28	0.009	-0.001	30.143	0.356
.|. |	.|. |	29	-0.010	-0.012	30.192	0.404
.|. |	.|. |	30	-0.041	-0.049	31.081	0.411
.|. |	*|. |	31	-0.056	-0.066	32.702	0.383
.|. |	.|. |	32	0.033	0.040	33.260	0.406
.|. |	.|. |	33	0.000	-0.039	33.260	0.455
.|. |	.|. |	34	0.035	0.037	33.916	0.472
*|. |	*|. |	35	-0.078	-0.102	37.098	0.372
.|. |	.|. |	36	-0.015	0.007	37.211	0.413

The new correlogram has by far less statistically significant terms, therefore we should search for a sample of ARIMA (3,1,3) type, and even if we take into account how separate are the significant terms, it is possible that this sample be actually ARIMA (1,1,1).

2.2 Sample estimation

The stage of sample estimation includes the effective use of data to do parameter inferences according to the soundness of the sample. In order to estimate parameters the method of maximum probability also known as the method of maximum likelihood or the method of the least squares can be used.

By using least squares, we have estimated the following model in Eviews:

d(p_rrc) c ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)

Dependent Variable: D(P_RRC)
Method: Least Squares
Sample (adjusted): 1/06/2006 11/15/2007
Included observations: 485 after adjustments
Convergence achieved after 78 iterations
Backcast: 1/03/2006 1/05/2006

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.78E-05	3.50E-05	1.364149	0.1732
AR(1)	0.410036	0.217294	1.887007	0.0598
AR(2)	-0.030056	0.259614	-0.115771	0.9079
AR(3)	0.547195	0.171668	3.187527	0.0015
MA(1)	-0.309835	0.203756	-1.520619	0.1290
MA(2)	-0.022812	0.224538	-0.101597	0.9191
MA(3)	-0.656745	0.155067	-4.235223	0.0000

R-squared	0.040818	Mean dependent var	-4.33E-05
Adjusted R-squared	0.028778	S.D. dependent var	0.002710
S.E. of regression	0.002670	Akaike info criterion	-8.998945
Sum squared resid	0.003408	Schwarz criterion	-8.938555
Log likelihood	2189.244	F-statistic	3.390230
Durbin-Watson stat	1.979233	Prob(F-statistic)	0.002766

Inverted AR Roots	.97	-.28-.70i	-.28+.70i
Inverted MA Roots	1.00	-.34+.74i	-.34-.74i

Taking into account that the terms AR (2) and MA (2) are statistically non-significant, we re-estimate the sample without these:

Dependent Variable: D(P_RRC)
Method: Least Squares
Sample (adjusted): 1/06/2006 11/15/2007
Included observations: 485 after adjustments
Convergence achieved after 56 iterations
Backcast: 1/03/2006 1/05/2006

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.84E-05	3.79E-05	1.277387	0.2021
AR(1)	0.364003	0.130325	2.793049	0.0054
AR(3)	0.566562	0.125066	4.530114	0.0000
MA(1)	-0.291995	0.113841	-2.564937	0.0106
MA(3)	-0.696925	0.114552	-6.083924	0.0000

R-squared	0.037552	Mean dependent var	-4.33E-05
Adjusted R-squared	0.029532	S.D. dependent var	0.002710
S.E. of regression	0.002669	Akaike info criterion	-9.003793
Sum squared resid	0.003420	Schwarz criterion	-8.960658
Log likelihood	2188.420	F-statistic	4.682116
Durbin-Watson stat	1.927134	Prob(F-statistic)	0.001026

Inverted AR Roots	97	-.30+.70i	-.30-.70i
Inverted MA Roots	1.00	-.35-.76i	-.35+.76i

In this sample, all coefficients except the constant are statistically significant.

2.3 Sample Checking

This last stage of the Box-Jenkins methodology is at least equally important as identification or estimate stage. The purpose is seeing in what extent the sample built complies with the available observations dealing with the stochastic process studied.

The stage implies testing the sample adjusted in its relation with data in order to discover the inadequacies of the sample and to obtain its improvement.

Taking into account that we have estimated an ARIMA(3,1,3) sample, we are in the first instance interested in knowing if we have eliminated autocorrelation of residuals. The correlogram of residuals (in the object equation -> view -> residual tests -> correlogram Q statistic) proves that there are no more autoregressive statistically significant terms. For verify this assumption we can used the Breusch-Godfrey test.

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.796514	Prob. F(2,478)	0.451496
Obs*R-squared	1.413281	Prob. Chi-Square(2)	0.493299

Test Equation:
Dependent Variable: RESID
Method: Least Squares

Sample: 1/06/2006 11/15/2007
Included observations: 485
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.26E-07	3.79E-05	0.003333	0.9973
AR(1)	-0.082179	0.181701	-0.452276	0.6513
AR(3)	0.070815	0.168008	0.421497	0.6736
MA(1)	0.039016	0.136551	0.285722	0.7752
MA(3)	-0.039871	0.137752	-0.289442	0.7724
RESID(-1)	0.080796	0.078024	1.035525	0.3009
RESID(-2)	-0.014196	0.054266	-0.261598	0.7937

R-squared	0.002914	Mean dependent var	5.37E-05
Adjusted R-squared	-0.009602	S.D. dependent var	0.002658
S.E. of regression	0.002670	Akaike info criterion	-8.998873
Sum squared resid	0.003409	Schwarz criterion	-8.938483
Log likelihood	2189.227	F-statistic	0.232826
Durbin-Watson stat	1.999891	Prob(F-statistic)	0.965813

The assumption can be accepted. Nevertheless, residuals are relatively far from normality, with both excess kurtosis and skewness positive (figure 2).

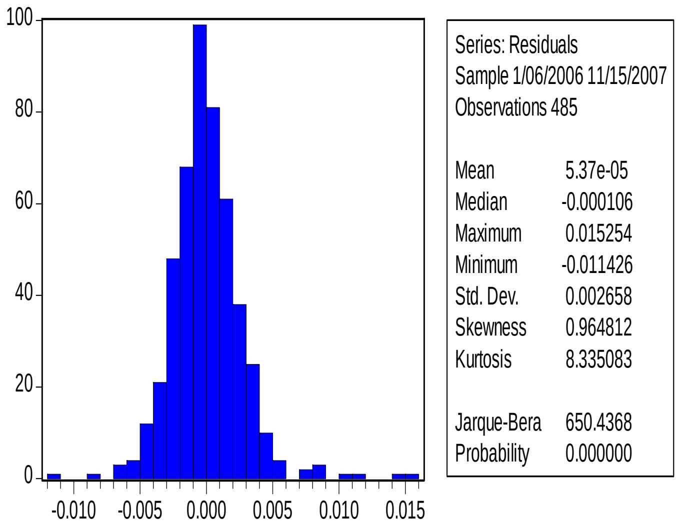

Figure 2 – The residual distribution

The test of double residual autocorrelation (squared residuals) also suggests that the heteroskedasticity hypothesis is not verified, and the ARIMA (3,1,3) sample should be estimated with a ARCH sample for variant, not at all simple least squares.

If we estimate the ARIMA (3,1,3) sample by means of a GARCH (1,1) sample for a variant, results are more encouraging:

Dependent Variable: D(P_RRC)
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 1/06/2006 11/15/2007
Included observations: 485 after adjustments
Convergence achieved after 72 iterations
MA backcast: OFF (Roots of MA process too large), Variance backcast: ON
GARCH = C(6) + C(7)*RESID(-1)^2 + C(8)*GARCH(-1)

	Coefficient	Std. Error	z-Statistic	Prob.
C	-1.18E-06	1.42E-05	-0.083107	0.9338
AR(1)	-0.334103	4.52E-05	-7383.989	0.0000
AR(3)	0.805044	0.000128	6272.457	0.0000
MA(1)	0.377929	0.000473	798.9506	0.0000
MA(3)	-0.883694	0.000149	-5921.544	0.0000
	Variance Equation			
C	5.56E-07	1.75E-07	3.179919	0.0015
RESID(-1)^2	0.256283	0.046331	5.531570	0.0000
GARCH(-1)	0.680985	0.056834	11.98200	0.0000

R-squared	0.076118	Mean dependent var	-4.33E-05
Adjusted R-squared	0.062560	S.D. dependent var	0.002710
S.E. of regression	0.002623	Akaike info criterion	-9.276988
Sum squared resid	0.003283	Schwarz criterion	-9.207971
Log likelihood	2257.670	F-statistic	5.614269
Durbin-Watson stat	1.745316	Prob(F-statistic)	0.000003

Inverted AR Roots	.83	-.58+.79i	-.58-.79i
Inverted MA Roots	.85	-.61+.82i	-.61-.82i
	Estimated MA process is noninvertible		

Now, the residuals distribution is presented in figure 3

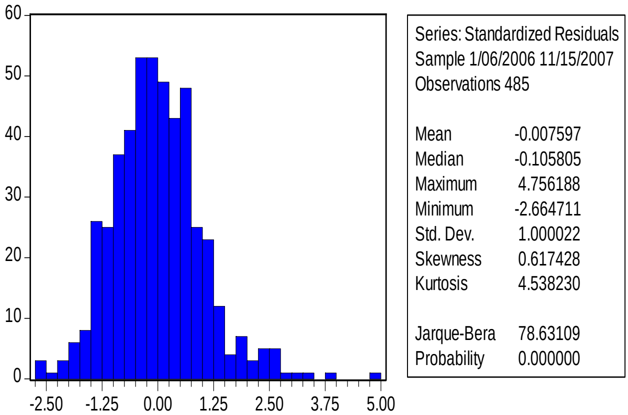

Figure 3 – The residual distribution

Conclusions

ARIMA(3,1,3) sample, possibly with a GARCH (1,1) sample for the variant of residuals, adequately describes the structure of autocorrelation in the field of Rompetrol share prices.

References

Enders W., 1995, Applied Econometric Time Series, John Wiley&Sons Inc

Voineagu V. (coord), 2006, Econometric theory and practice Ed. Meteor Press, Bucuresti

www.bvb.ro