The Analysis of Unemployment Degree using Econometrics Methods

■

Elena Bugudui
Candidate Ph.D. Lecturer
University ARTIFEX, București

Abstract. *In the shown material I will analyze the evolution of the unemployment rate registered in the time period during January 1. 2001 and March 30. 2007 by using the Box Jenkins model.*

The model is based on the ARIMA processes and with its help a large number of time series met in economy can be pattern.

The steps that have to be follow are shown to transform a time series and then they have to be covered, in the single case shown by the registered unemployment rate and ending with a prognoses for a future period.

Key words: time series; methodology; corelograma; estimation; unemployment; rate.

■

The Box –Jenkins Methodology is being applied for the stationery time series (average, dispersion and the self-correlation function doesn't modify itself significantly in time).

Most of the times, an encountered time series in economics is not a stationary series. In this case, a number of correspondent modifications will be applied to it, to become stationary.

The easiest way for a stationary analysis is the behavior analysis of the values for the self-correlation function. The self-correlation function (FAC) and the part self-correlation function (FACP) measure the correlation degree between the observation pairs separated through various time periods $\{z_t, z_{t+k}\}$, k = 1, 2, 3,...

If the time series is known, $(z_i)_{i=\overline{1,T}}$ the self-correlation coefficients are estimated through the following relationship:

$$\hat{\rho}(h) = \frac{\sum_{i=1}^{T-h}(z_i - \bar{z})(z_{i+h} - \bar{z})}{\sum_{i=1}^{T}(z_i - \bar{z})^2}, h = 0,1,2,...H,$$

For a time series with T terms, it's recommended the calculation of an H number of self correlation coefficients smaller than T/4.

If in the frame of the self-correlation series are values which differ significantly from zero or from following sensitively equal values, then the date series is not stationary. For this series a first range differences operator is being applied. For the resulted series, expressed from the differentiation operation, is being afterward applied the above operation. Through successive steps a stationary series is being obtained.

To test if the coefficient of the self-correlation differs significantly from zero or from an estimated value, a test Student will be applied.

To establish starting with which order the self-correlation coefficient significantly differs from 1, the following space will be defined:

$$[-t_{T-2;\alpha/2}\hat{\sigma}(\hat{\rho}(h)), t_{T-2;\alpha/2}\hat{\sigma}(\hat{\rho}(h))].$$

If the order is being identify as equal with d, then, to obtain the stationary series, a differential operator of the respectively order is applied. The stationary series is obtained:

$$X_t = (1-L)^d Z_t$$
$$X_{t+1} = (1-L)^d Z_{t+1}$$
$$...$$
$$X_{T-d} = (1-L)^d Z_{T-d}.$$

By applying the differential operator of d order, a stationary time series is obtained but with a smaller d terms number.

Another test used to find out if a series is stationary or not is the Unit Test (Unit Root) - *Dickey - Fuller test.*

Let the model $Z_t = \rho Z_{t-1} + \varepsilon_t$ be. If the coefficient of Z_{t-1} equals 1, then we have what is called "random walk". In this case, the $\rho = 1$ hypothesis must be tested.

An easier way for testing the stationer, based on the unit - root problem, is the one of the model that uses the first order differential operator $\Delta Z_t = Z_t - Z_{t-1}$.

The equivalent model is considered: $\Delta Z_t = (\rho - 1)Z_{t-1} + \varepsilon_t$ or $\Delta Z_t = \delta Z_{t-1} + \varepsilon_t$. For which the $\delta = 0$ hypothesis will be tested.

The testing of will be done with the help of Dickey – Fuller test. This test is build after the model on a statistic of Student type but the theoretical values are not given by the Student distribution instead they were generated by Dickey and Fuller through simulation based on the Monte Carlo method.

The Dickey – Fuller test is for the self regressive predictors rang 1. Dickey – Fuller test uses the general equation:

$Z_t = c + \gamma t + aZ_{t-1} + \varepsilon_t$ or, to ease to process, the equation:

$\Delta Z_t = c + \gamma t + \alpha Z_{t-1} + \varepsilon_t$ for which the following elements are checked:

- the meaning of c constant: With the help of Student test, the following hypothesis are tested:

$$H_0 : c = 0$$
$$H_1 : c \neq 0$$

- the presence of a trend. With the help of Student Test the following hypothesis are tested:

$$H_0 : \gamma = 0$$
$$H_1 : \gamma \neq 0$$

- the presence of an unitary root. With the help of Student test the following hypothesis are tested:

$$H_0 : \alpha = 0$$
$$H_1 : \alpha < 0$$

If the variable follows a self regressive example higher than 1, then the test Augmented Dickey – Fuller is used. The Augmented Dickey – Fuller test uses the equation:

$$\Delta Z_t = c + \gamma t + \alpha Z_{t-1} + \beta_1 \Delta Z_{t-1} + ... + \beta_n \Delta Z_{t-n} \varepsilon_t$$

The Phillips Perron test uses the same equation as the Dickey – Fuller (DF) test, $\Delta Z_t = c + \gamma t + \alpha Z_{t-1} + \varepsilon_t$ and the t test are being calculated to check the shown hypotheses at the DF test not only under the errors independent hypothesis but also under the hypothesis of an eventual self correlation.

After testing, the model that minimizes the information criteria will be remembered and the parameters for the check of non stationary will be tested. The information criteria's remembered are:

- Akaike criteria $AIK = -\frac{2LnL}{T} + \frac{2k}{T}$, where:

T – observations number;
k – parameters number.
LnL – verisimilitude log

$$LnL = -\frac{T}{2}(1 + Ln2\pi) - \frac{T}{2} Ln \frac{\Sigma \hat{\varepsilon}_t^2}{T}$$

- Schwartz criteria, $SC = -\frac{2LnL}{T} + \frac{2LnT}{T}$

The Box-Jenkins procedure assumes getting over various steps to identify the most suitable self regressive analysis model for a time series. Those steps are:

Step I: Identification of the estimated model

1. The self-correlation function (ACF) and the partial self-correlation function will be calculated to establish if the series is stationary. If it is stationary, on goes to the next step, if the data series will be stationar itself through several adequate changes.
2. Identification in advance of the model. The estimated values of the self-corralled and partial self-correlation functions are compared for the structure of the brought in model (p, d, q orders), resulting as a model structure ARIMA for that series for which theoretical ACF AND PACF approximate with enough precision ACF and PACF estimated from the given realization, under a minimum number of the model parameters.

Step II: Estimation

In the estimation phase it will be determined, for the chosen model, the efficient estimations from the statistic point of view of the model parameters, of which structure and preliminary values of the parameters have been established in the above step. Also in this step the stationery, the contrarily of the model, the statistic meaning and other quality indicators for the model modulus of the series will be analyzed.

Step III: Diagnosis ratification

The diagnosis ratification step consists mainly in the analysis of the model wastes for the establishment of the statistic independence of those. In this way, it is applied to:

- establish if the model parameters differ significantly from zero.
- check the conformation of the waste hypothesis of nocorrelation of homoscedasticitate and of the normal allocation of the waste.

In the situation in which this hypothesis does not check itself, it will return to the identification phase for the selection of a new model, adequate to the given accomplishment.

In the case of the model identification phase, the association of some theoretical functions of partial self-correlation and of processes partial self-correlation, which are analyzed, do not ensure the determination of the “best” model for the given outcome, specially at the first identification attempt.

Step IV: The prognosis

The prognoses assume the determination of an appraisal for the studied function, at a t+k moment. To establish the prognoses, it is taken into consideration the differential procedure, which the model has taken in an identification and appraisal step.

Use of the model

For illustration, we will analyze the unemployment rate registered in Romania based on monthly dates reported in the time period 01.01.2001 – 30.03.2007 with the help of Eviews program.

The values of this indicator are represented through time series and although the dates are independently analyzed by other macro-economics indicators, the influence of these indicators at the economics processes is observed through them.

Phase I: Identification of the apriorical model

Step I: The analysis of the data series corelogram

As a result of the accomplishment of the data series corelogram formed out of the 75 observations the results presented in figure 1 are obtained:

Correlogram of RATA_SOM

Sample: 2001:01 2007:03
Included observations: 75

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.926	0.926	66.931	0.000
		2	0.812	-0.324	119.04	0.000
		3	0.690	-0.041	157.21	0.000
		4	0.593	0.129	185.83	0.000
		5	0.522	0.034	208.27	0.000
		6	0.476	0.068	227.24	0.000
		7	0.456	0.107	244.93	0.000
		8	0.461	0.120	263.27	0.000
		9	0.486	0.112	283.92	0.000
		10	0.521	0.099	308.02	0.000
		11	0.551	0.035	335.44	0.000
		12	0.557	-0.073	363.83	0.000
		13	0.498	-0.350	386.90	0.000
		14	0.412	0.024	402.97	0.000
		15	0.321	-0.003	412.88	0.000
		16	0.256	0.070	419.29	0.000
		17	0.205	-0.086	423.46	0.000
		18	0.169	-0.031	426.33	0.000
		19	0.148	0.004	428.59	0.000
		20	0.143	-0.017	430.75	0.000
		21	0.145	-0.075	432.99	0.000
		22	0.143	-0.063	435.23	0.000
		23	0.133	-0.039	437.19	0.000
		24	0.107	-0.032	438.49	0.000
		25	0.065	0.068	438.98	0.000
		26	0.009	-0.059	438.99	0.000
		27	-0.048	-0.037	439.26	0.000
		28	-0.096	-0.085	440.40	0.000
		29	-0.130	0.039	442.51	0.000
		30	-0.152	-0.027	445.49	0.000
		31	-0.165	-0.020	449.06	0.000
		32	-0.167	-0.012	452.81	0.000

Figure 1. *Unemployment rate corelogram*

The corelogram shows that the series is seasonal (not stationary), also confirmed by the graphic

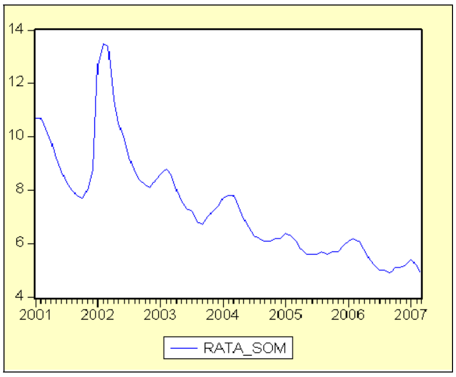


Figure 2. *Graphic representation of the unemployment rate*

The series is being unseasonable, with the help of season coefficients. Those are:

Date: 05/05/07 Time: 14:16
Sample: 2001:01 2007:03
Included observations: 75
Ratio to Moving Average
Original Series: RATA_SOM
Adjusted Series: RATA_SOSA

Scaling Factors:

1	1.107056
2	1.139231
3	1.129881
4	1.050093
5	0.993398
6	0.965536
7	0.939740
8	0.926575
9	0.917069
10	0.926567
11	0.948235
12	0.993077

Figure 3. *Monthly season coefficients*

The unseasonable corelogram series is showed in figure 4:

Correlogram of RATA_SOSA

Sample: 2001:01 2007:03
Included observations: 75

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.947	0.947	69.914	0.000
		2	0.882	-0.138	131.39	0.000
		3	0.817	-0.023	184.88	0.000
		4	0.763	0.075	232.25	0.000
		5	0.715	0.001	274.45	0.000
		6	0.675	0.042	312.62	0.000
		7	0.640	0.017	347.43	0.000
		8	0.612	0.049	379.75	0.000
		9	0.587	0.005	409.90	0.000
		10	0.563	0.009	438.11	0.000
		11	0.543	0.032	464.72	0.000
		12	0.522	-0.019	489.66	0.000
		13	0.476	-0.231	510.80	0.000
		14	0.425	-0.031	527.90	0.000
		15	0.371	-0.035	541.18	0.000
		16	0.332	0.072	551.95	0.000
		17	0.291	-0.076	560.40	0.000
		18	0.254	-0.026	566.93	0.000
		19	0.218	-0.009	571.83	0.000
		20	0.189	0.005	575.57	0.000
		21	0.158	-0.053	578.23	0.000
		22	0.123	-0.060	579.89	0.000
		23	0.087	-0.043	580.73	0.000
		24	0.049	-0.051	581.00	0.000
		25	0.018	0.087	581.04	0.000
		26	-0.014	-0.028	581.06	0.000
		27	-0.044	-0.019	581.29	0.000
		28	-0.072	-0.053	581.94	0.000
		29	-0.098	0.008	583.14	0.000
		30	-0.123	-0.030	585.09	0.000
		31	-0.150	-0.038	588.05	0.000
		32	-0.174	-0.023	592.14	0.000

Figure 4. *The unemployment rate unseasonable corelogram*

The corelogram points out that between all terms important gaps exist. We estimate three models with the help of Dickey – Fuller test (with a gap of 4 periods) obtaining the following results:

Augmented Dickey-Fuller Unit Root Test on RATA_SOM

Null Hypothesis: RATA_SOM has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic based on SIC, MAXLAG=2)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.606221	0.0021
Test critical values:	1% level	-4.088713	
	5% level	-3.472558	
	10% level	-3.163450	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(RATA_SOM)
Method: Least Squares
Date: 05/05/07 Time: 14:52
Sample(adjusted): 2001:03 2007:03
Included observations: 73 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RATA_SOM(-1)	-0.247614	0.053756	-4.606221	0.0000
D(RATA_SOM(-1))	0.595379	0.096896	6.144547	0.0000
C	2.506918	0.569632	4.400945	0.0000
@TREND(2001:01)	-0.018868	0.004966	-3.799292	0.0003

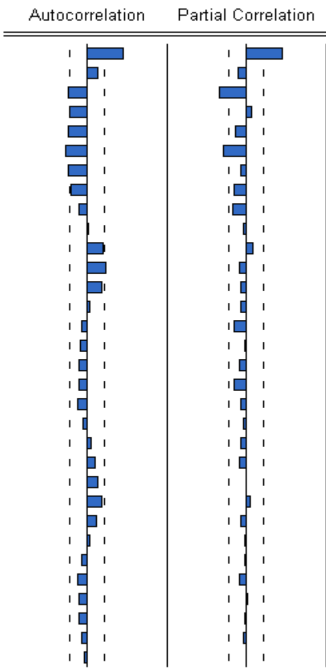
R-squared	0.404970	Mean dependent var	-0.079452
Adjusted R-squared	0.379099	S.D. dependent var	0.604373
S.E. of regression	0.476229	Akaike info criterion	1.407401
Sum squared resid	15.64880	Schwarz criterion	1.532906
Log likelihood	-47.37014	F-statistic	15.65350
Durbin-Watson stat	2.184145	Prob(F-statistic)	0.000000

Figure 5. *The Dickey – Fuller test*

Step II: Identification of the model type:

Correlogram of D(RATA_SOM)

Date: 05/05/07 Time: 14:56
Sample: 2001:01 2007:03
Included observations: 74



	AC	PAC	Q-Stat	Prob
1	0.470	0.470	17.044	0.000
2	0.135	-0.111	18.469	0.000
3	-0.251	-0.351	23.469	0.000
4	-0.232	0.067	27.784	0.000
5	-0.243	-0.132	32.613	0.000
6	-0.271	-0.290	38.678	0.000
7	-0.244	-0.072	43.659	0.000
8	-0.203	-0.155	47.169	0.000
9	-0.099	-0.175	48.015	0.000
10	0.023	-0.031	48.061	0.000
11	0.209	0.084	51.960	0.000
12	0.239	-0.091	57.144	0.000
13	0.187	-0.061	60.367	0.000
14	0.038	-0.058	60.503	0.000
15	-0.078	-0.155	61.084	0.000
16	-0.083	-0.016	61.752	0.000
17	-0.110	-0.094	62.946	0.000
18	-0.111	-0.149	64.195	0.000
19	-0.122	-0.072	65.709	0.000
20	-0.050	-0.040	65.969	0.000
21	0.043	-0.060	66.164	0.000
22	0.099	-0.084	67.219	0.000
23	0.144	-0.007	69.514	0.000
24	0.181	0.041	73.192	0.000
25	0.126	-0.070	75.021	0.000
26	0.039	-0.010	75.203	0.000
27	-0.060	-0.014	75.633	0.000
28	-0.120	-0.086	77.399	0.000
29	-0.107	0.023	78.838	0.000
30	-0.110	-0.018	80.392	0.000
31	-0.068	-0.029	80.996	0.000
32	-0.033	0.006	81.138	0.000

Figure 6. *The differential unemployment rate corelogram of I ordinal*

The self-correlation functions are being calculated based on the differential series of I order.

The simple corelogram presents a decreasing of its terms and the partial corelogram has only the first term different from 0, which makes us anticipate a model type AR (1) MA (1).

Phase II. Estimation

The estimation of the parameters can be done based on the unseasonable differentiate series order I.

Dependent Variable: D(RATA_SOSA)
Method: Least Squares
Date: 05/05/07 Time: 15:04
Sample(adjusted): 2001:03 2007:03
Included observations: 73 after adjusting endpoints
Convergence achieved after 2 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.358027	0.109914	3.257343	0.0017

R-squared	0.100855	Mean dependent var	-0.069254
Adjusted R-squared	0.100855	S.D. dependent var	0.391989
S.E. of regression	0.371697	Akaike info criterion	0.872128
Sum squared resid	9.947423	Schwarz criterion	0.903505
Log likelihood	-30.83269	Durbin-Watson stat	2.014858
Inverted AR Roots	.36		

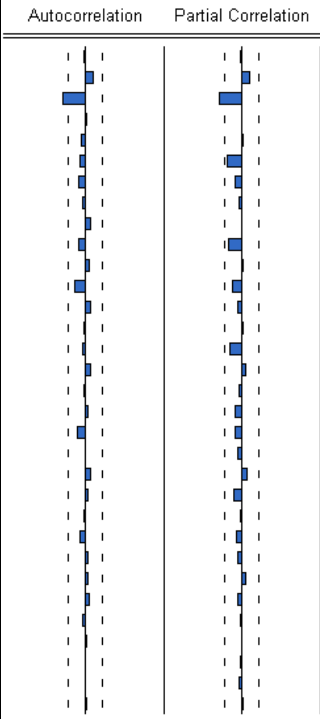
Figure 7. *Parameters estimation*

Phase III:

The process coefficients AR (1) are eloquently different from 0. The waste analysis is done beginning with the self-correlation function.

Correlogram of RESID

Date: 05/05/07 Time: 15:06
Sample: 2001:01 2007:03
Included observations: 73



	AC	PAC	Q-Stat	Prob
1	-0.023	-0.023	0.0412	0.839
2	0.108	0.108	0.9482	0.622
3	-0.298	-0.297	7.9155	0.048
4	0.013	-0.001	7.9298	0.094
5	-0.053	0.010	8.1586	0.148
6	-0.074	-0.182	8.6055	0.197
7	-0.085	-0.085	9.1994	0.239
8	-0.036	-0.027	9.3106	0.317
9	0.069	0.007	9.7210	0.374
10	-0.096	-0.165	10.527	0.395
11	0.054	0.019	10.781	0.462
12	-0.141	-0.128	12.565	0.401
13	0.071	-0.052	13.024	0.446
14	-0.009	0.020	13.032	0.524
15	-0.036	-0.156	13.157	0.590
16	0.064	0.056	13.550	0.632
17	-0.016	-0.027	13.575	0.697
18	0.037	-0.087	13.708	0.748
19	-0.100	-0.088	14.726	0.740
20	0.002	-0.044	14.726	0.792
21	0.062	0.076	15.132	0.816
22	0.026	-0.101	15.205	0.853
23	-0.011	-0.024	15.218	0.887
24	-0.080	-0.072	15.927	0.891
25	0.025	-0.045	15.997	0.915
26	0.038	0.050	16.162	0.932
27	0.057	-0.043	16.552	0.942
28	-0.038	-0.011	16.731	0.954
29	0.016	-0.003	16.764	0.966
30	-0.001	-0.020	16.764	0.975
31	0.002	-0.032	16.765	0.982
32	0.014	0.009	16.793	0.988

Figure 8. *The waste corelogram*

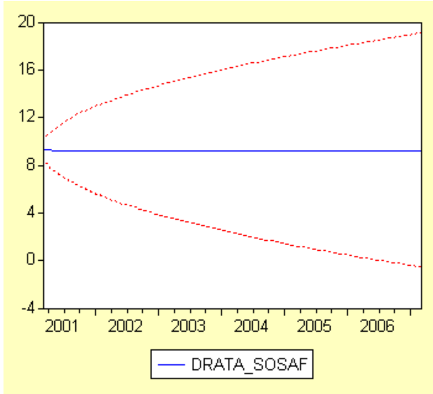
From the above corelogram results that no term is outside the two trust and statistic intervals. Q has a critical probability close to the value 1. We can say that we can assimilate the wastes through a process white noise type.

The estimation of ARIMA (1, 1, O) model is validated, in conclusion the series can be represented through a process ARIMA (1,1,0) type.

The series formed from the unseasonable and differential order I unemployment rates is represented through the process:

DRATA_SOSA = (1-0, 358D) $\times \varepsilon_t = \varepsilon_t - 0,358 \times \varepsilon_{t-1}$

Phase IV: The foreknowledge



The foreknowledge can be calculated with the help from the table below:

	e_t	DRATA_SOSA	RATE_SOSA	CS	UNEMPLOYMENT RATE
03. 06	- 0, 001		4,336738		
04.07		0,000358	4,33736	1,050	4,553933
05.07			4,337716	0,993	4,306739
06.07			4,338074	0,965	4,18531

where:

e_t is the waste value;

DRATE_SOSA is the unseasoned and differential series rang I;

RATE_SOSA is the unseasoned series;

CS are the seasoned coefficients;

UNEMPLOYMENT RATE is the brute series of the inflation rate.

DRATA_SOSA$_{07:04} = e_{07:04} - 0,34 \times e_{07:03} = 0 - 0,358 \times$ (- 0, 001) = 0,000358

RATA_SOSA$_{07:04}$ = RATA_SOSA$_{07:03}$ + DRATA_SOSA$_{07:04}$ = 4,337 + 0,000358= 4,33736

This result from the column RATE_SOSA, if we multiple it with the seasoned coefficients, we will obtain the brute forecasting series.

From the prognoses done with the help from the Box Jenkins method, a downfall of the unemployment rate is being observed for the following months.

It's interesting to apply this method and the time series obtained from the registration done through AMIGO, the BIM unemployed, but having an insufficient number of registered dates (< 50), it is necessarily to do a time extension to complete the series with a sufficient number of observations. It can be analyzed and forecasts can be made at national level but mostly locally to capture the given influences and by other specific regional development factors.

References

Anghelache, C-tin. (2005). *Statistică teoretică și economică*, Editura Economică, București

Andrei, T. (2003). *Statistică și econometrie*, Editura Economică, București

Andrei, T., Pârțachi, I., „Metodologia aplicării procedurii B.J la estimarea inflației în R. Moldova și România", *Economica*, Chișinău, 2007

Bourbonnais, R. (2002). *Économétrie*, 4e édition, Editura Dunod, Paris

Bugudui, E. (2006). *Presiuni asupra pieței muncii din România*, Editura Artifex, București, 2006

Turtureanu, C.I. (2002). Analiza seriilor de timp utilizând metodologia B.J, Editura Libris, Iași

Anuarul Statistic al României, 1990-2006