CREATIVE MATH. & INF.
**17** (2008), 78 - 83

# ARIMA models for unemployment

ALINA BĂRBULESCU

ABSTRACT. In a previous paper it was proved that laws of Okun's type can not be determined for Romanian economy, after 1990. In consequence, in this paper we give a description of unemployment evolution after 1990, using ARIMA models.

## 1. INTRODUCTION

**Definition 1.1.** A discrete time process is a sequence of random variables $X_t, t \in \mathbf{Z}$. A discrete time process $X_t, t \in \mathbf{Z}$ is called stationary if:

$$\forall t \in \mathbf{Z}, M(X_t^2) < \infty,$$
$$\exists \mu \in \mathbf{R}, \forall t \in \mathbf{Z}, M(X_t) = \mu,$$
$$\exists \gamma : \mathbf{R}^+ \to \mathbf{R}, \forall t \in \mathbf{Z}, \forall h \in \mathbf{Z}, Corr(X_t,\ X_{t+h}) = \gamma(h).$$

where M($X$) is the expected value of the random variable $X$, $\mu$ is a constant, $Corr(X, Y)$ is the correlation of the random variables $X$ and $Y$ and $\gamma$ is a real function.

**Definition 1.2.** A stationary process $\xi_t, t \in \mathbf{Z}$ is called a white noise if $\gamma(h) = 0$, for $h \neq 0$, $M(\xi_t) = 0$ and $D^2(\xi_t) = \sigma^2 = \gamma(0)$, $\forall t \in \mathbf{Z}$.

**Definition 1.3.** If $X_t, t \in \mathbf{Z}$ is a discrete time process, the function defined by:

$$\rho(h) = \frac{Corr(X_t, X_{t+h})}{\sqrt{D^2(X_t) D^2(X_{t+h})}} = \frac{\gamma(h)}{\gamma(0)}, \forall h \in \mathbf{Z}$$

is called the autocorrelation function.

**Definition 1.4.** If $X_t, t \in \mathbf{Z}$ is a stationary process, the function defined by:

$$\tau(h) = \frac{Corr(X_t - X_t^*, X_{t-h} - X_{t-h}^*)}{D^2(X_t - X_t^*)},\ h \in \mathbf{Z}_+$$

is called the partial autocorrelation function, where $X_t^*$ ($X_{t-h}^*$) is the affine regression of $X_t$ ($X_{t-h}$ ) with respect $X_{t-1}, ..., X_{t-h+1}$.

**Definition 1.5.** Let $p, d, q$ be natural numbers, ($\varphi_1, ..., \varphi_p$) *and* ($\theta_1, ..., \theta_p$) *finite sequences of real coefficients, I the identity function,* $X_t, t \in \mathbf{Z}_+$ *a time process and*:

$$\begin{aligned}
B(X_t) &= X_{t-1}, \\
\mathbf{\Phi}(B) &= I - \varphi_1 B - ... - \varphi_p B^p,\ \varphi_p \neq 0, \\
\mathbf{\Theta}(B) &= I - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q,\ \theta_p \neq 0, \\
\Delta^d X_t &= (1 - B)^d X_t.
\end{aligned}$$

Received: 10.08.2007. In revised form: 24.10.2007.
2000 *Mathematics Subject Classification.* 11M20.
Key words and phrases. *Model, ARIMA, ACF, residual.*

*The time process is called an autoregressive integrated moving average process of* $p, d, q$ *orders and is denoted by* ARIMA $(p, d, q)$ *if*:

$$\mathbf{\Phi}(B)\Delta^d X_t = \mathbf{\Theta}(B)\xi_t,$$

*where* $\xi_t$, $t \in \mathbf{Z}_+$ *is a white noise.*

$\xi_t$, $t \in \mathbf{Z}_+$ *is called the residual in the* ARIMA *process.*

*An autoregressive process of p order, denoted by* AR($p$)*, is an* ARIMA *process of p,* 0, 0 *orders*.

*A moving average process of q order, denoted by* MA($q$)*, is an* ARIMA *process of* 0, 0, *q orders*.

*An autoregressive moving average process of p and q orders, denoted by* ARMA $(p, q)$ *is an* ARIMA *process of p,* 0, *q orders.*

**Remark 1.1.** The autocorrelation function (partial autocorrelation function) calculated for some values of the random variables $X_t$ is called an empirical autocorrelation function (empirical partial autocorrelation function) and is denoted by ACF (PACF).

**Remark 1.2.** In order to determine the type of the process, the form of the ACF and PACF graphs of the process can be used.

i. The ACF of an AR($p$) process is an exponential decreasing or a damped sine wave oscillation. The PACF of an AR($p$) process is vanishing for all $h > p$ and $\tau(p) = \varphi_p$.

ii. The ACF of an MA($q$) process is vanishing for all $h > q$. The PACF of an MA($q$) process is non-vanishing beginning at some lag value.

iii. The ACF graph of an ARMA($p, q$) process is a mixture of exponential decreasing curves and damped sine wave oscillation, when $p > q$; when $p < q$, the it is of the previous type for all $h > q - p$.

## 2. Results

In what follows we shall work with the unemployment rate in Romania, after 1991 and we determine a model for the unemployment evolution.

The first idea was to determine a simple model. The best model determined was a polynomial of 6-th order. The data and the fitted data are represented in Figure 1. The equation of the model is:

$$X_t = 11.4118 - 25.4055t + 18.5711t^2 - 5.1492t^3 + 0.6716t^4 - 0.0415t^5 + 0.001t^6,$$

where: $X_t$ is the unemployment rate and $t$ is the year's number.

It was considered t=1 for 1991 and the unemployment rate 0 in 1991.

For this model, the standard error was $s = 0.7203419$ and the correlation coefficient: $r = 0.9880021$. The graph of the errors is presented in Figure 2 and their ACF in Figure 3.

The values of ACF of the residuals are computed for lags between 1 and 12 (the column 2, Figure 3) and the standard deviations are given in the column 3. The last column of this figure contains the probabilities to reject the hypotheses of the autocorrelation of the residuals. It is clear that the probabilities are very small, so the residuals are dependent.

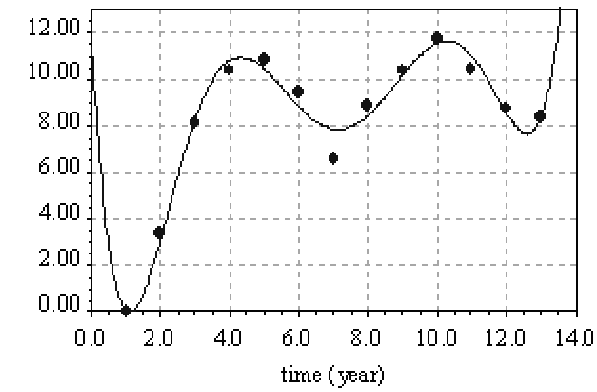


FIGURE 1. The unemployment rate

Also, applying Bartlett test, it results that the residuals are heteroscedastic.

These are some reasons to look for another model of the unemployment. Another one is that the initial data are dependent, as it can be seen in the Figure 4 (there are values of ACF outside the confidence interval).

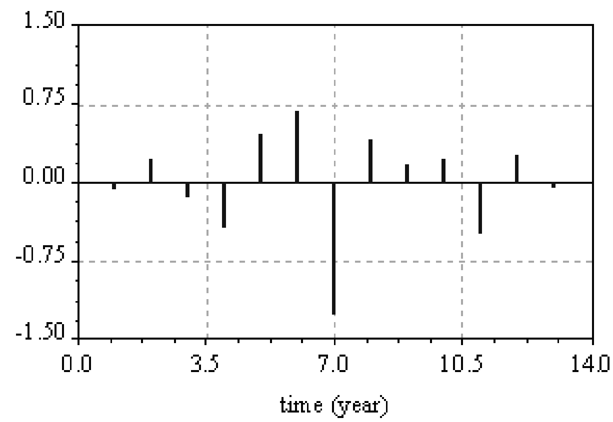


FIGURE 2. The errors graph in the polynomial model

Analyzing the graphs of ACF and PACF of the initial data we can decide that a possible model is of ARMA type. We studied the models AR(1), AR(2), ARMA(1,1) with and without constant.

The residuals in the models AR(1) with and without constant and in AR(2) without constant are correlated; in AR(2) with constant they are heteroscedastic. So, we had to chose between the models ARMA(1,1) with and without constant. The model ARMA(1, 1) was preferred because the values of Schwarz and Akäike tests applied to it were smaller ($61.85 < 62.50$, respectively $60.72 < 60.81$).

In what follows we shall analyze the model ARMA(1,1), without constant.

```
     Auto-  Stand.
Lag  Corr.   Err.  -.5 -.25   0   .25  .5  Box-Ljung  Prob.
                    +----+----+----+----+
  1  -.446   .248  .*********I          .    3.233    .072
  2  -.292   .238  .    ******I         .    4.744    .093
  3   .148   .226   .        I***      .     5.173    .160
  4   .342   .215   .        I*******  .     7.712    .103
  5  -.330   .203    .*******I        .     10.361    .066
  6   .015   .189    .       *        .     10.368    .110
  7   .092   .175     .      I**     .      10.641    .155
  8  -.008   .160      .     *      .       10.644    .223
  9  -.045   .143      .    *I      .       10.741    .294
 10   .030   .124       .    I*    .        10.798    .373
 11  -.008   .101        .   *    .         10.804    .460

Plot Symbols: Autocorrelations * Two Standard Error Limits.
Total cases :   13 Computable first lags:   12
```

FIGURE 3. ACF of residuals in the polynomial model

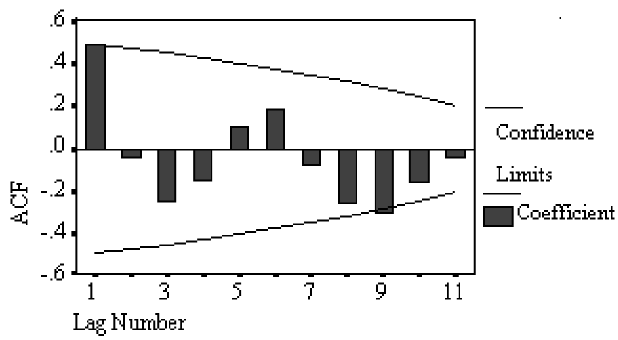


FIGURE 4. ACF of initial data

The coefficients are: $\varphi = 0.9261, \psi = -0.5225$, so the model is:

$$X_t - 0.9261X_{t-1} = \xi_t + 0.5225\xi_{t-1}, t \in \mathbf{N}^*$$

where $\xi_t, t \in \mathbf{N}^*$ is the residual.

I. *Validity test for the estimators of the coefficients $\varphi$ and $\psi$*

First, we shall test the hypothesis: $H_0 : \varphi = 0$ against $H_1 : \varphi \neq 0$, at the significance level $\alpha = 0.05$.

Let us denote by: $n$ the data volume, $p$ the number of variables in model, $t_{1-\alpha/2}$ the value in the Student quartile tables for $n-p$ degrees of freedom.

If $\left|\frac{\varphi}{\sigma_\varphi}\right| \geq t_{n-p,1-\alpha}$, then $H_0$ is rejected.

An analogous, for to test the hypothesis: $H_0 : \psi = 0$.

Since,

$$\left|\frac{\varphi}{\sigma_\varphi}\right| = 9.9035 > 1.762 = t_{11,0.95}\ , \ \left|\frac{\psi}{\sigma_\psi}\right| = 2.032 > 1.762 = t_{11,0.95},$$

we reject the hypotheses that $\varphi$ and $\psi$ are zero.

II. *Test on the residuals*

1. *Independence test*

In order to prove the independence of the residuals, ACF and PACF were used. In Figure 5 it can be seen that the values of ACF and PACF are inside the confidence interval (at the confidence level 0.950).

The computed probabilities to accept the hypothesis that the residuals are independent are close to 1 (for example, 0.946 for the lag 1 and 0.992 for the lag 2), so this hypothesis can be accepted.

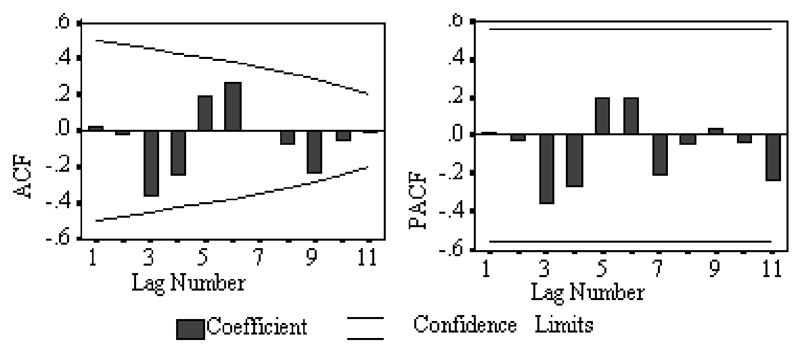


FIGURE 5. ACF and PACF of residuals

2. *Normality test*

In order to prove that the residuals have a normal distribution, the quantile - quantile diagram (Q – Q plot) is drawn (Figure 6). The observed values (the points on diagram) are close to the theoretical distribution (represented by the straight line). Therefore, we can say that the residuals have a normal distribution.

Since generally the graphical methods are subjective, Lilliefors test was also used.

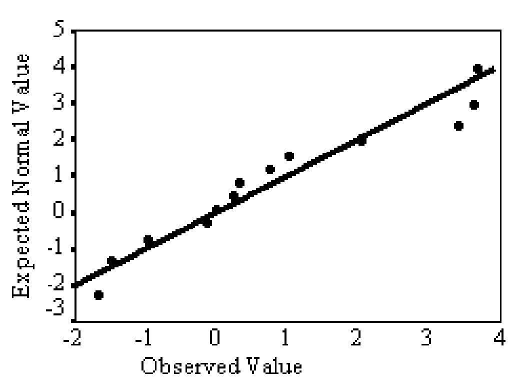


FIGURE 6. ACF and PACF of residuals

Let us consider the selection $(x_1, x_2, ..., x_n)$, $\overline{x}$ the selection mean, $s^2$ the selection variance, $z_i = \dfrac{x_i - \overline{x}}{s}$, $z_1 \leq z_2 \leq ... \leq z_n$ the values of $z_i$, increasingly ordered,

$$F_0(z_i) = \Phi(z_i),\ i = \overline{1,n},$$

$$F_n(z_i) = \begin{cases} 0\ ,\ i \leq 0 \\ \frac{i}{n}\ ,\ 1 \leq i \leq n-1 \\ 1\ ,\ i = n \end{cases},$$

$$D_n = \max_{i=\overline{1,n}} |F_0(z_i) - F_n(z_i)|\,,$$

where $\Phi$ is the normal distribution function.

The hypothesis that the residuals have a normal distribution is accepted, at the significance level $\alpha = 0.05$ if $D_n \leq D_{n,1-\alpha}$, where the values $D_{n,1-\alpha}$ are given in the Lilliefors table.

In our case, $D_{13} = 0.1564 < 0.234 = D_{13,\,0.95}$ , confirming the conclusions derived from Q – Q diagram.

3. *Homoscedasticity test*

To prove that the residuals have the same variance, Bartlett test was used. The errors were divided in $k = 2$ groups, of $n_1 = 7$ respectively $n_2 = 6$ values and the hypothesis $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2$ of the equality of the residual variances was formulated.

Let $s_1^2$, $s_2^2$ and $s^2$ be respectively the selection variances of each group and the total one, $\alpha = 0.05$ the significance level and $\chi^2_{1-\alpha,k-1}$ the quartile of $\chi^2$ distribution with $k-1$ freedom degree, at the significance level $\alpha$. If

$$X^2 = \frac{-\sum\limits_{i=1}^{k} n_i \ln \frac{s_i^2}{s^2}}{1 + \frac{1}{k-1}\sum\limits_{i=1}^{k}(\frac{1}{n_i} - \frac{1}{n})} < \chi^2_{1-\alpha,k-1},$$

the hypothesis $H_0$ is accepted.

In our case, $X^2 = 0.3436 < 3.84 = \chi^2_{0.95,1}$, so the residuals are homoscedastic.

The conclusions of the tests 1–3 is that the residuals form a white noise and the model ARMA(1,1) is well chosen.

## REFERENCES

[1] Barbulescu, A., *The modelling of the unemployment in Romania*, Bull. St. Univ. Baia Mare, Ser. B, Matematica-Informatica, Vol. XVIII **2** (2002), 141–146

[2] Barbulescu, A., *Time series, with applications*, Junimea, Iasi, 2002

[3] Vaté M., *Statistique chronologique et prévision*, Économica, Paris, 1993

OVIDIUS UNIVERSITY
DEPARTMENT OF MATHEMATICS
BD. MAMAIA NO. 124
900527 CONSTANTA, ROMANIA
*E-mail address*: alinadumitriu@yahoo.com