Model of the unemployment rate in the Czech Republic

Radmila Stoklasová[1]

Abstract. This paper deals with the labor market modeling in the Czech Republic. Box-Jenkins or ARIMA model is one of the approaches for modeling time series. Based in the model the proposed equation describes the unemployment rate in the Czech Republic. The unemployment rate is predicted on the basis of the model by the end of 2012. The unemployment rate is the second most watched real indicator of the state of the economy after the real income and is dealt with by most states.

Keywords: unemployment, time series, time series stationarity, autocorrelation function, partial autocorrelation function, Box-Jenkins methodology, ARIMA process, SARIMA.

JEL Classification: C32, C51
AMS Classification: 62-07

1 Introduction

Unemployment rate is one indicators of the state of the economy of the Czech Republic. The source of data of the unemployment rate is the Ministry of Labor and Social Affairs. In conclusion of the article is then by the model predicted the expected rate of unemployment in the Czech Republic by the end of 2012.

The main objectives of time series analysis are: finding patterns over time series, finding deviations from the expected course, prognosis (prediction) of the time series behavior in the future, and continuous determination of appropriate management interventions for influencing the behavior of time series. For the analysis of time series of unemployment rates here is used Microsoft Excel and statistical program SPSS (Statistical Package for the Social Sciences). For more details about this program see [7]. SPSS program allows some operations to perform time series in the standard module (Base), for a deeper analysis of time series is necessary to connect the module Trends. In terms of SPSS time series is a normal data file, and it is assumed that one row of data matrix contains observations at any one time and the rows are sorted by time so that the oldest observation is the first and the youngest observation is the last line of the matrix.

This area is important because the unemployment rate is the second most watched real indicator of the state of the economy. In the literature [3], the unemployment rate is defined as the proportion of unemployed and the labor force. The labor force is the sum of unemployed and employed. For an unemployed person is considered someone who is unemployed and actively looking. Those who don't try to get work actively are outside the labor force. Official unemployment figures underestimate actual unemployment. Sometimes, in times of recession employees shorten time jobs. This unemployment does not show in the official unemployment rate. Unemployment is often divided into frictional, structural and cyclical.

2 Registered unemployment rate in the Czech Republic - 30. 4. 2012

At 30. 4. 2012 Labor Office of the Czech Republic registered a total of 497 322 job seekers at its regional offices and their contact centers. Their number was about 27 858 less than at the end of the previous month, compared to the same period of 2011 decreased by 16 520.

During April there were 39 879 newly registered persons. Compared to the previous month it was more about 1420 persons in the same period last year about 6,800 fewer people. The records in April left a total of 67 737 applicants. It was about 12 773 more than in the previous month and 12 862 persons less than in April 2011.

Registered unemployment rate reached 8.4%. The higher unemployment rate than the national average showed 43 districts, the highest being in Bruntál (15.9%), Most (15.7%), Jesenik (15.2%), Decin (14.0%) and Hodonin (13, 9%). The lowest unemployment rate was recorded in the districts of Prague-East (3.3%), Prague Capital (4.0%), Prague-West (4.2%) and Mlada Boleslav (4.7%). Registered unemployment rate fell to men and women 7.5% to 9.7%. This article analyzes the unemployment rate for the whole Czech Republic, in [1] there are the methods for Measuring regional unemployment disparities in the Czech Republic.

[1] School of Business Administration, Silesian University, Department of Mathematical Methods in Economy, University Square 1934/3, Karviná, Czech Republic, stoklasova@opf.slu.cz

Time series analysis

The following chart shows the time course of registered unemployment rate from January 2004 to April 2012.

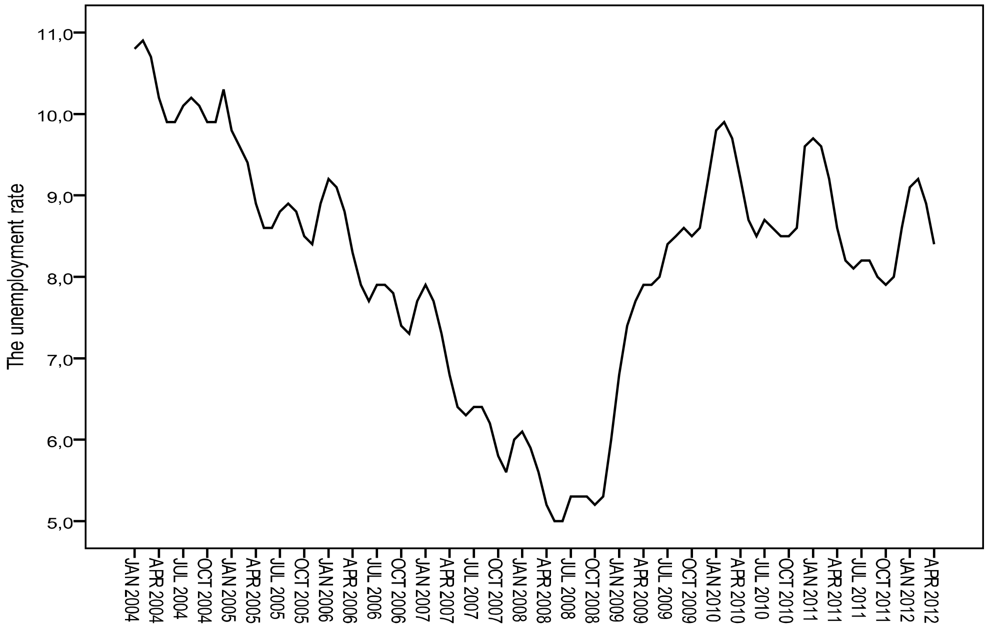

Figure 1 The registered unemployment rate

Source: Ministry of Labor and Social Affairs

The chart suggests that since February 2004 when the registered unemployment rate was highest (10.9%) during the reporting period, the time series trend downward until June 2008. Since mid-2008 to February 2010 (9.9%) had a rising trend line. The analyzed time series is nonstationary. Some methods require stationary model for the construction. Such models include the Box-Jenkins models are described in the next section. Nonstationarity of time series can be removed by an ordinary, respectively seasonal differencing time series.

3 ARIMA model

Behavior and prediction of time series is calculated by using ARIMA model. ARIMA(p, d, q) model is a complex linear model composed of three sub-sections (not always present all three): AR (Autoregressive) - linear combination of the effects of past values, I (Integrative) - random walk (designed to filter out any transient data folder), MA (Moving Average) - linear combination of past mistakes. These models are extremely flexible, they are relatively difficult to calculate and to understand the results, are demanding quality and number of measured data (assumed to be at least 50 measurements or observations). The investigated time series values, the registered unemployment rate has 100 observations, so that the ARIMA model can be used, see [2].

First, it is necessary to identify the model, i.e. determine the type of model and parameter values. This can be done based on the waveform autocorrelation (ACF) and partial autocorrelation (PACF) function of stationary series. Autocorrelation is a correlation between time series and the same time series delayed by a fixed number of measurements. Partial autocorrelation coefficients are correlation between the time series and the same series, delayed by a fixed number of measurements, the fact that the partial correlation technique eliminates the influence of intermediate members of the series. This coefficient reflects only direct link between y_t and y_{t-2} such as transmission through the exclusion of observations y_{t-1}.

In the SPSS output in the ARIMA model for each parameter there are calculated their significance. According to these values it is possible to determine whether the parameters in the model to include or not. When calculating the number of models that describe the behavior of the series about the same, then we select a model in which the AIC (Akaik's information criterion), respectively SBC (Schwartz-Bayesian criterion) is minimal and Log Likelihood (CAF) maximal. Finally, verify that the residual component is white noise, i.e. the sequence of random variables normally distributed with zero mean and constant standard deviation.

The ARIMA model assumes interdependence among variables $y_{t-2}, y_{t-1}, y_t, y_{t+1}, y_{t-2},...$ If the process contains even seasonal fluctuations, as in our case, we can also expect dependencies between variables in different seasons: $y_{t-2s}, y_{t-s}, y_{ts}, y_{t+s}, y_{t-2s},...$, where *s* is length of the period (in our case 12).

This process is called **SARIMA (*p,d,q*) (*P,D,Q*)$_s$**, where
p is order of the process AR, *q* is order of the process MA, *d* is order of the difference,
P is order of the seasonal process AR, *Q* is order of the seasonal process MA, *D* is order of the seasonal difference, *s* is the length of the seasonal period.
Model can be expressed by the formula:

$$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^D y_t = \theta_q(B)\Theta_Q(B^s)a_t, \quad (1)$$

where $\phi_p(B)$ is autoregressive operator, $\theta_q(B)$ is the operator of moving averages, $\Phi_P(B^s)$ is seasonal autoregressive operator, $\Theta_Q(B^s)$ is seasonal operator of moving averages, $\{a_t\}$ is white noise.

3.1 Identification of the model

First, the parameters of the model are estimated using autocorrelation and partial autocorrelation functions. The time series is nonstationary, therefore it must be differentiated by the first order. The time series exhibits seasonality, that is why it is also seasonally differentiated.

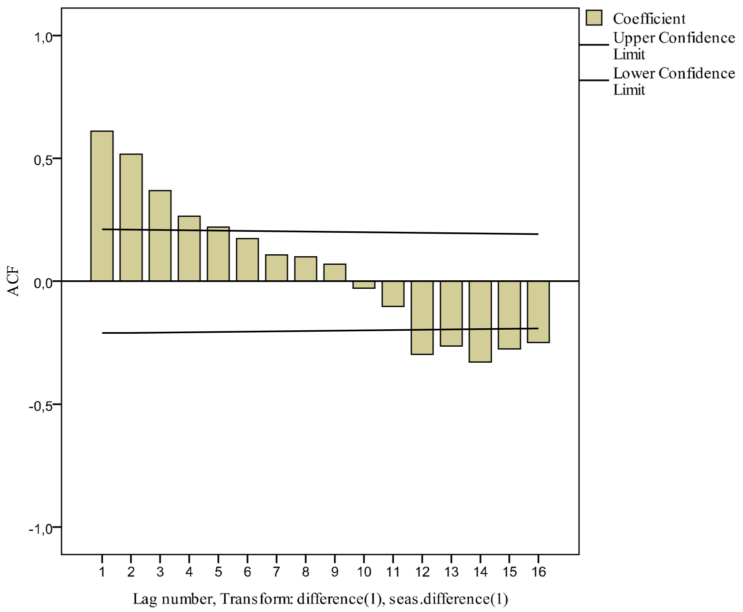

Figure 2 Autocorrelation

Based on the functions we can chose models: SARIMA (1,1,0) $(1,1,0)_{12}$ or model SARIMA (1,1,0) $(0,1,0)_{12}$. Because Akaik‘s information criterion as well as Schwartz-Bayes criterion are less for the model SARIMA (1,1,0) $(1,1,0)_{12}$, we chose that model.

SPSS output for model SARIMA (1,1,0) (1,1,0)$_{12}$

FINAL PARAMETERS:

Log likelihood	23.013138
AIC	-42.026277
SBC	-37.094461

Variables in the Model:

	B	SEB	T-RATIO	APPROX. PROB.
AR1	0.66163944	0.08297210	7.9742404	0.00000000
SAR1	-0.47709376	0.10673931	-4.4697098	0.00002402

General form of the model is: $$\Phi_1\left(B^{12}\right)\phi_1(B)(1-B)^1\left(1-B^{12}\right)^1 y_t = a_t \,. \tag{2}$$

After adjustment equation and substituting the estimated values into (2) we obtain the following equation, which describes the dynamics of the investigated time series:

$$\begin{aligned} y_t = 1.66y_{t-1} - 0.66y_{t-2} + 0.53y_{t-12} - 0.88y_{t-13} + 0.35y_{t-14} + \\ + 0.47y_{t-24} - 0.78y_{t-25} + 0.13y_{t-26} + a_t \end{aligned} \tag{3}$$

It can be inferred from the model equation, the registered unemployment rate in the current month is most affected by the value rates in the previous month and the value 13 months ago.

3.2 Diagnostic model control

Basic empirical diagnostic check is to assess residues $$\hat{a}_t = \left[\hat{\theta}_q(B)\right]^{-1}\hat{\phi}_p(B)y_t \,, \tag{4}$$
where $\hat{\phi}_p(B) = 1 - \hat{\phi}_1 B - \ldots - \hat{\phi}_p B^p$ and $\hat{\theta}_q(B) = 1 - \hat{\theta}_1 B - \ldots - \hat{\theta}_q B^q$.

Autocorrelation of unsystematic components can be tested using the sample autocorrelation fiction
$$r_k = \frac{\sum_t \hat{a}_t \hat{a}_{t-k}}{\sum_t \hat{a}_t^2} \,. \tag{5}$$

If unsystematic components are not autocorrelated, the value of this function should lie within the range $\pm 2/\sqrt{n}$ (95% confidence interval), where $n = 100$.

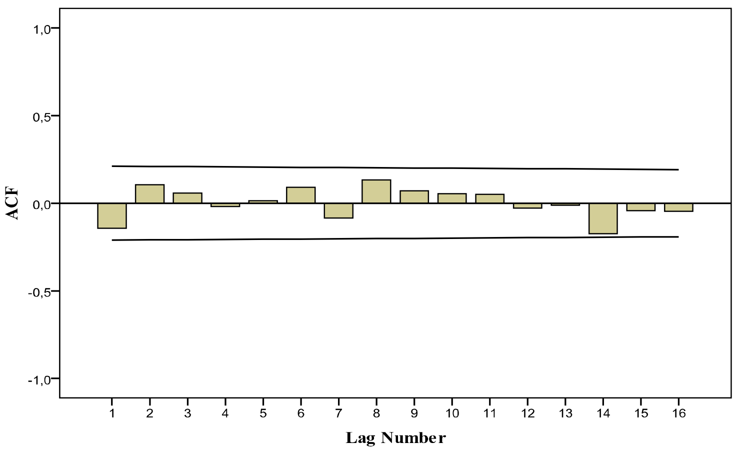

Figure 3 Autocorrelation of residues

Another way to determine whether non-systematic component is autocorrelated, is using of portmanteau test, which suggested Box and Pierce.

There is tested hypothesis H_0: $\rho_1 = \rho_2 = \ldots = \rho_K = 0$ against hypothesis H_1: non H_0, where ρ_k, $k = 1,\ldots, K$ is autocorrelation of unsystematic model components for the delay k. If ARIMA model is properly constructed, then the statistics $$Q = n\sum_{k=1}^{K}\hat{r}_k^2 \,, \tag{6}$$

for high n and K has approximately distribution χ^2 with $(K - p - q)$ degrees of freedom.

Lag	Autocorrelation	Lag	Autocorrelation
1	– 0.143	9	0.071
2	0.106	10	0.055
3	0.057	11	0.05
4	–0.019	12	–0.029
5	0.013	13	–0.012
6	0.091	14	–0.174
7	–0.085	15	–0.043
8	0.132	16	–0.046

Table 1 Autocorrelation of residues

The statistics $Q = 11.41$ and $\chi^2_{0,05}(15) = 24.9$. The hypothesis H_0 can not be reject. There was no evidence of correlation of residues.

3.3 Short-time prediction by estimated model

The course of the original time series of registered unemployment rate and the time series, which was calculated by the above SARIMA model, is presented in the following graph.

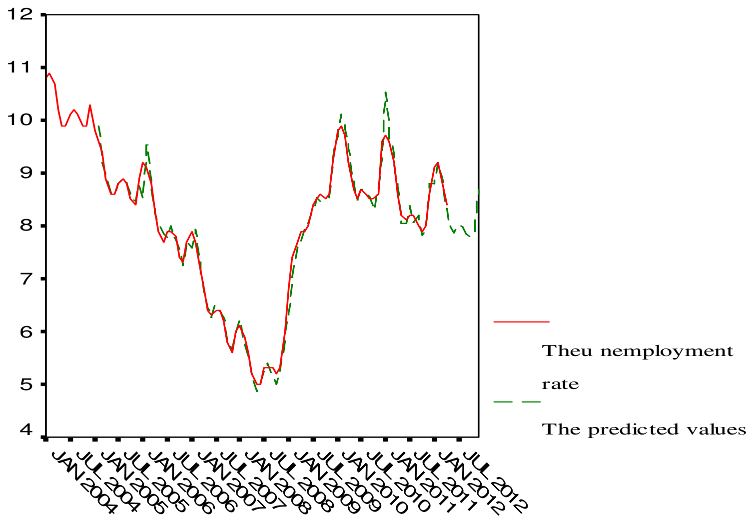

Figure 4 Graph the real and predicted time series

Source: Ministry of Labor and Social Affairs and own calculations

Point and interval forecasts (95%) of the registered unemployment rate are shown in Table 1.

Month	Point prediction	Interval prediction
May	7.99	(7.6 ; 8.4)
June	7.86	(7.2 ; 8.6)
July	8.02	(7.0 ; 9.1)
August	7.99	(6.6 ; 9.3)
September	7.84	(6.2 ; 9.5)
October	7.79	(5.9 ; 9.7)
November	7.90	(5.7 ;10.1)
December	8.69	(6.3 ;11.1)

Table 2 Prediction of the unemployment rate in the Czech Republic

Source: own calculations

4 Conclusions

In 2012 we can expect the highest values of the registered unemployment rate during the holiday period, which may be caused by registration of graduates of secondary schools and universities to the labor offices. The lowest value of the unemployment rate is expected in October 2012 (7.79%). The reason may be seasonal work in construction, as shown for example in [6].

With unemployment facing many countries, therefore it is important to capture the trend in unemployment. ARIMA model used is a highly flexible, so it is highly likely, that unemployment will develop according to the above model if there is absence of government intervention that would significantly change this trend.

References

[1] Antoušková, M.: *Measuring regional unemployment disparities in the Czech Republic.* Ostrava: Vysoká škola báňská, Území, znalosti a rozvoj na počátku 21. století: sborník příspěvků odborné sekce z konference Zvyšování konkurenceschopnosti, aneb, Nové výzvy pro rozvoj regionů, států a mezinárodních trhů, 2007, s. 153-160.

[2] Arlt, J., Arltová, M.: *Ekonomické časové řady*. Praha: Professional Publishing, 2009.

[3] Dornbusch, R., Fischer, S.: *Makroekonomie.* Praha: SPN, 1994.

[4] Mielcová, E.: *Unemployment rate and new cars registration - the case of the Czech republic 2003-2009.* Opava: Slezská univerzita, Obchodně podnikatelská fakulta v Karviné: Acta academica karviniensia, 2010, s. 348-353.

[5] Ministerstvo práce a sociálních věcí. *Informace o počtu nezaměstnaných v České republice*, duben 2012. http://portal.mpsv.cz/sz/stat/nz/mes

[6] Ministry of Labour and Social Affairs. *Basic indicators of labour and social protection in the Czech Republic : time series and graphs 2007*. Prague: Ministry of Labour and Social Affairs, 2008.

[7] Norušic, M. J. : *SPSS Professional Statistics 7.5.* Chicago: SPSS Inc., 1997.