Közgazdász Fórum
Forum on Economics and Business
21(135), 66–78.
2018/2

Forecasting the regional unemployment rate based on the Box-Jenkins methodology vs. the Artificial Neural Network approach. Case study of Brașov and Harghita counties

SZILÁRD MADARAS[1]

This paper presents different methods for the forecasting of unemployment rates in two Romanian counties. The stationarity of the monthly unemployment rate time series between January 2000 and November 2016 was examined using the ADF and KPSS tests. Based on time series, a forecast was estimated using two approaches: the Box-Jenkins methodology and the Artificial Neural Network-based NAR model. Results showed a decreasing trend by the end of the forecasted period in all cases, except for the NAR model of Harghita County. Comparing the forecasted values with the officially registered unemployment rates from the same period, we observed that, by the end of the period, the differences between the real and predicted values became higher in the NAR model than in the ARMA model-based forecasting. These results indicate that, in these particular cases, NAR neuron network model-based forecasts fit well if values are estimated for a short-term period, while for medium-term forecasts the ARMA model-based forecasting is more precise.

Keywords: regional unemployment, time-series models, forecasting and prediction methods, Box-Jenkins methodology, Artificial Neural Network.

JEL codes: C32, C53, R15.

Introduction

Time series analysis is an actual topic in regional studies. Approaches differ in the assumptions and models used for testing, i.e. the regions are studied as unique cases using the Box-Jenkins methodology or the group of regions or counties form a panel data base structure. Both of these are generally used in regional time series forecasting, while Artificial Neural Networks (ANN) currently represent a new approach in economic research.

The present study examines the unemployment rate monthly time series in Brașov and Harghita counties (NUTS3 level territorial units for statistics) from Romania. The differences between the two case studies were verified using the main regional indicators, while the employment and unemployment analysis

[1] PhD, assistant professor, Sapientia Hungarian University of Transylvania, Faculty of Economics, Socio-Human Sciences and Engineering, e-mail: madarasszilard@uni.sapientia.ro.

Publisher: Hungarian Economists' Society of Romania, the Department of Economics and Business Administration in Hungarian Language at the Babeş-Bolyai University, and the Partium Christian University
ISSN: 1582-1986 http://econ.ubbcluj.ro/feb/

Research methodology

Stationarity analysis is one of the primary subjects of time series analysis, while time dataset-based model identification, which represents a forecasting instrument, is another important research topic.

In this paper, we used two of the most commonly known unit root tests, the ADF (Augmented Dickey-Fuller) and the KPSS Lagrange Multiplier test. Those tests present significant differences regarding the null hypothesis: the first one has the null of non-stationarity, while the second one has the null of stationarity (Kirchgässner–Wolters 2007).

The Box-Jenkins methodology has a long and prestigious past in the field of time series research. The p-order auto-regressive model (AR_p) is based on the assumption of a given time t, the endogenous y_t variable depends on its time delayed values of the previous 1, 2, ... p periods (Kirchgässner–Wolters 2007; Pecican 2006):

$$y_t = a_0 + a_1 y_{t-1} + ... + a_p y_{t-p} + u_t,$$

with u_t being the error term.

The q-order moving average process (MA_q) describes the y_t, as

$$y_t = \mu + b_1 u_{t-1} + ... + b_p u_{t-q},$$

where μ is the mean and $u_{t-1},...u_{t-q}$ are pure random processes, for the previous 1, 2, … q periods (Kirchgässner–Wolters 2007; Pecican 2006).

The autoregressive moving average process (ARMA) with AR order p and MA order q are:

$$y_t = a_0 + a_1 y_{t-1} + ... + a_p y_{t-p} + \mu + b_1 u_{t-1} + ... + b_p u_{t-q} + u_t$$

and the ARIMA (p,d,q) autoregressive integrated moving average process refers to an I (integrated) process.

This model, initially developed by Box and Jenkins in the 1970s, is constructed with the following steps: model identification, i.e. the determination of p, q values, using the autocorrelation function (ACF) and the partial autocorrelation function (PACF), and the d of the I(1) process are specified testing the stationarity of the time series. These tests are followed by the estimation of the model coefficients and model validation (Kirchgässner–Wolters 2007; Pecican 2006).

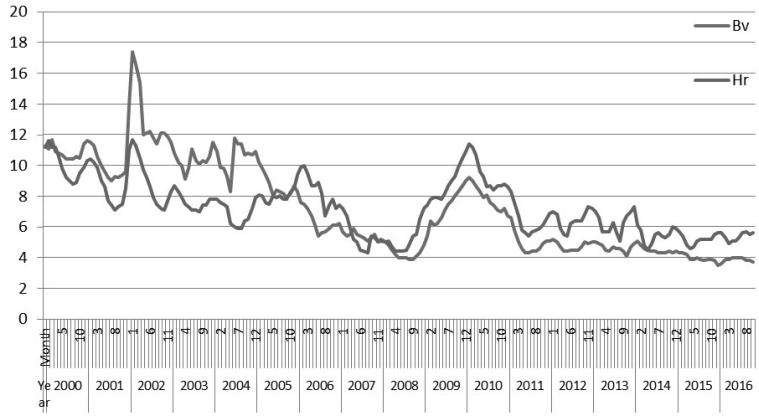
The selection of the best fitting model is usually based on the Akaike Information Criterion (AIC) values (Pecican 2006). And, in the last step, the thus constructed model is used for the short- or medium-term forecasting of the time series.

levels in 2016, there were major differences in the structure of the employed population: a high share of agricultural employment in Harghita (23.67%), as opposed to a high share of employment in services in Brașov (51.54%).

The time series of monthly unemployment rates in Brașov and Harghita counties for the period January 2000 – November 2016, used for the calculations below, were obtained from the Tempo Online database of the Romanian National Institute of Statistics (INSSE 2018).

Results

The time series analysis contains the stationarity tests, the Box-Jenkins method and the Artificial Neural Network analysis, as described below. The evolution of unemployment rates in the two counties followed similar trends in the studied period (Figure 1).



Source: author's own design based on INSSE (2018) data

Figure 1. Evolution of unemployment rates in Brașov and Harghita counties (January 2000 – November 2016)

The stationarity of the series was examined using the ADF and KPSS tests (Table 1). Both of the univariate unit root tests suggest that the unemployment rate monthly time series in Brașov is a non-stationary, eventually I(1) series, while in Harghita the results of the ADF test suggest an AR(p) process. The results of the KPSS test are similar.

The best fitting ARMA model was chosen based on the AIC values: ARMA(1,1) for both counties (Table 2).

Table 2. Akaike information criterion (AIC) values for the estimated models of the unemployment rate time series in Brașov and Harghita counties

Brașov County		Harghita County	
ARIMA	**AIC**	**ARMA**	**AIC**
(2,0,0)	2.891793	(2,0)	2.499080
(1,0,1)	1.899506	(1,1)	1.214645
(1,1,1)	1.920951	(1,2)	1.382784
(2,1,1)	1.966818	(1,3)	1.420065
(2,0,1)	1.961739	(2,1)	1.394078
(2,0,2)	2.900252	(2,2)	2.314901
(2,0,3)	2.862625	(2,3)	2.435688

Source: author's own calculations based on INSSE (2018) data

The ARMA(1,1) model statistics of the unemployment rate time series from Brașov and Harghita counties are presented in Table 3 and we can see that, in the first case, the R-squared is equal to 0.96, while in the second case the R-squared is 0.94, which means that both estimated models explain the time series well. In the next step, the models are used for a medium-term forecast of the time series.

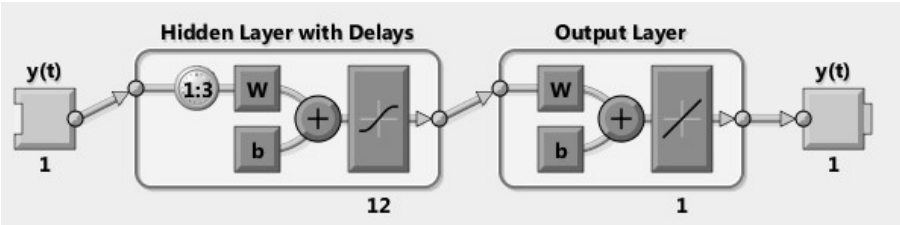
Table 3. ARMA models of monthly unemployment rate time series in Brașov and Harghita counties (January 2000 – November 2016)

Dependent variable	**Unemployment rate in Brașov County**			
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.140487	1.643046	3.737259	0.0002
AR(1)	0.964376	0.018738	51.46694	0.0000
MA(1)	0.260068	0.069814	3.725161	0.0003
R-squared	0.956994			
Adj. R-squared	0.956562			
AIC	1.899506			

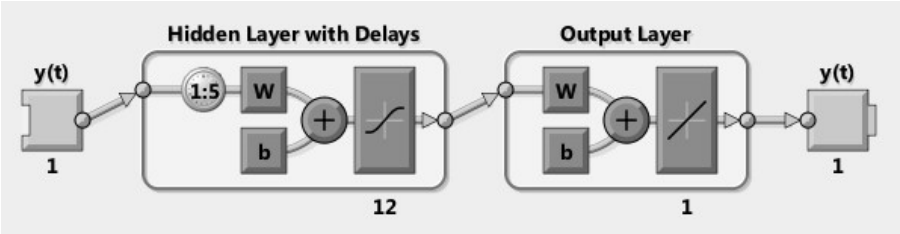
Dependent variable	**Unemployment rate in Harghita County**			
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.844177	0.661996	10.33869	0.0000
AR(1)	0.931595	0.024561	37.93010	0.0000
MA(1)	0.428266	0.065383	6.550110	0.0000
R-squared	0.941462			
Adj. R-squared	0.940874			
AIC	1.214645			

Source: author's own calculations based on INSSE (2018) data

with 20 observation values. In the case of the logarithmed unemployment rate in Brașov County, the network architecture was set to 1 input, 12 hidden neurons and d = 3 number of delay, while in the case of the logarithmed unemployment rate in Harghita County it was set to 1 input, 12 hidden neurons and d = 5 number of delay (Figure 4). With this neuron network architecture and lag values, the errors are autocorrelated.



a.



b.

Source: author's own design

Figure 4. NAR neuron network construction of monthly unemployment rates in Brașov (a) and Harghita (b) counties

The time steps were divided into three groups: the training group (85%), the validation group (5%), and the testing group (10%). In both Artificial Neural Networks, the Levenberg-Marquardt training algorithm was used. The prediction errors became uncorrelated, after a retraining process (Figure 5), and this way the final form of the ANN models was validated for the second forecasting.

The two neuron network models presented above were used for a medium-term forecasting of unemployment rates in Brașov and Harghita counties.

Both the ARIMA model and the NAR model forecasted higher unemployment rates for Harghita than for Brașov County (Table 4), as that was the most common

Table 4. Results of the Box-Jenkins and the Artificial Neural Network forecasting of monthly unemployment rates in Brașov and Harghita counties

Month	ARMA models		NAR models	
	Brașov	Harghita	Brașov	Harghita
	%	%	%	%
2016M12	3.740316	5.736696	3.8851	5.4722
2017M01	4.143789	4.946636	3.7602	5.7280
2017M02	4.038858	5.556964	4.0287	5.3310
2017M03	4.066147	5.703539	3.8241	5.8240
2017M04	3.814162	5.776752	4.1868	5.0325
2017M05	3.879695	5.473426	3.8578	6.0922

Source: author's own calculations based on INSSE (2018) data

In the end, I compared the forecasted values to the officially registered unemployment rates from the same period (Table 5).

Table 5. Comparison of forecasted values with the officially registered unemployment rates

Month	Registered value of the unemployment rate		Difference to ARMA model forecasting		Difference to NAR model forecasting	
	Brașov	Harghita	Brașov	Harghita	Brașov	Harghita
	%	%	%	%	%	%
2016M12	3.60	5.80	-0.14	0.06	-0.29	0.33
2017M01	3.60	5.80	-0.54	0.85	-0.16	0.07
2017M02	3.60	5.90	-0.44	0.34	-0.43	0.57
2017M03	3.50	5.20	-0.57	-0.50	-0.32	-0.62
2017M04	3.20	4.90	-0.61	-0.88	-0.99	-0.13
2017M05	3.20	4.80	-0.68	-0.67	-0.66	-1.29

Source: author's own calculations based on INSSE (2018) data

We can observe that the officially registered unemployment rates follow the same trend as in the two estimated models. By the end of the forecasting period, the differences between the real and predicted values were higher for the NAR model-based forecasting than for the ARMA model-based forecasting, while at the beginning they were almost the same.

Conclusions

In this paper, two forecasting models were developed for predicting monthly unemployment rates in Brașov and Harghita counties, using the time series for the period January 2000 – November 2016. Based on the Box-Jenkins methodology,

Longhi, S.–Nijkamp, P.–Reggianni, A.–Maierhofer, E. 2005. Neural Network Modeling as a Tool for Forecasting Regional Employment Patterns. *International Regional Science Review* 28(3), 330–346.

Madaras, Sz. 2009. A munkanélküliségi helyzet elemzése a Központi Régió megyéiben. *Közgazdász Fórum* 12(5), 45–58.

Madaras, Sz. 2014. A gazdasági válság hatása a munkanélküliség alakulására országos és megyei szinten Romániában. *Közgazdász Fórum* 17(1–2), 136–149.

Mayor, M.–López, A. J.–Pérez, R. 2007. Forecasting Regional Employment with Shift–Share and ARIMA Modelling. *Regional Studies* 41(4), 543–551.

Morariu, N.–Iancu, E.–Vlad, S. 2009. A Neural Network Model for Time-Series Forecasting. *Romanian Journal of Economic Forecasting* 2009(4), 213–223.

Pecican, Ș. E. 2006. *Econometrie*. București: Editura Beck.

Schanne, N.–Wapler, R.–Weyh, A. 2008. Regional unemployment forecasts with spatial interdependencies. *IAB-Discussion Paper* 28/2008, http://nbn-resolving.de/urn:nbn:de:0168-ssoar-307640, downloaded: 18.02.2018.

Tsay, R. S. 2015. *Analysis of Financial Time Series*. Hoboken, NJ: John Wiley & Sons.

Warner, B.–Misra, M. 1996. Understanding Artificial Neural Networks as Statistical Tools. *The American Statistician* 50(4), 284–293.