Journal of Statistical and Econometric Methods, vol.8, no.2, 2019, 37-49
ISSN: 2241-0384 (print), 2241-0376 (online)
Scienpress Ltd, 2019

Forecasting the Unemployment of Med Counties using Time Series and Neural Network models

Christos Katris[1]

Abstract

In this paper, time series models and neural networks are used for prediction of unemployment in nine Mediterranean countries. Accurate prediction of unemployment is very important for economic policy reasons. FARIMA is a suitable model when long memory exists in a time series and has been applied successfully for predicting unemployment. However, potential data characteristics such as heteroskedasticity, non-normality and non linearity in unemployment time series may reduce the effectiveness of the classical FARIMA model. Here, is made an attempt for the improvement of forecasting accuracy, by applying models which take into account data characteristics, such as FARIMA/GARCH which takes into account heteroskedasticity. Furthermore, non-linearity of the data is better captured by Neural Network models rather than the traditional time series models such as FARIMA and for this reason Artificial Neural Networks with multilayer feed-forward architecture are considered as predictors for unemployment. Non-normality is present at many cases of unemployment data and to further improve the results are considered FARIMA and FARIMA/GARCH models with student t distribution errors. Finally, is proposed a model selection based on data characteristics. We employ monthly seasonally adjusted data (source is Eurostat database) from 2008 M1 to 2016M10 to train the data and we consider 1 step-ahead forecasts for the next 12 months, i.e. until 2017 M10 to compare the performance of the models.

JEL Classification Number: C01, C22, C53
Key words: FARIMA/GARCH, FARIMA, multilayer feedforward neural nets, 1 step-ahead predictions, t-distribution of errors.

---

[1] PhD in Applied Statistics, University of Patras, Greece

Article Info: *Received*: September 23, 2018. *Revised* : December 20, 2018
Published online : April 1, 2019

1 Introduction

Many economic decisions as well as policy designs are based on unemployment prediction. Policy-makers have the difficult task to recognize or better detect early the problem of unemployment and to define effective measures in order to reduce the problem or to avoid its future escalation. Clearly accurate forecasting of unemployment is central to the definition of such effective measures. Many time series models have been employed extensively for the forecasting of macroeconomic variables, including unemployment. ARIMA models have been used in empirical studies such as for Czech Republic (Stoklashová, 2012), for Romania (Dobre and Alexandru, 2008) and for Nigeria (Etuk et al, 2012). Mladenovic et al. (2017) used Seasonal ARIMA model to forecast unemployment at the EU28 level. Many variations of ARMA and GARCH models were compared for forecasting UK unemployment in Floros (2005). GARCH assumes varying heteroskedasticity and offer additional insight in the case of heteroskedastic time series.

The persistent effect of shocks on unemployment rates can be seen as an evidence for hysteresis (Blanchard and Summers, 1986). ARIMA models cannot allow for such persistent effects and fractional ARIMA (FARIMA) models which take into account the long-memory effect seems more suitable for unemployment prediction. These models have been used in studies of (Gil-Alana, 2001) for the forecasting of the UK unemployment rate, of (Kurita, 2010) for forecasting of the Japan's unemployment rate and of Katris (2015) for forecasting Greece's unemployment rate.

Another issue is non-linearity and when is present other approaches are more suitable. In Rothman (1998) six nonlinear models were compared according to their out-of-sample forecasting accuracy, Proietti (2003) examined the forecasting accuracy of several linear and nonlinear forecasting models for the US monthly unemployment rate and Johnes (1999) reports the results of a forecasting competition between linear autoregressive, GARCH, threshold autoregressive and neural network models of the UK monthly unemployment rate series. Neural networks appear promising to model more accurately data which display non-linearity, thus to give better forecasts. The paper of (Aiken, 1996) shows how a neural network may be used to forecast unemployment rates in the United States. More recently, (Olemedo, 2014) used Neural Net techniques for forecasting unemployment in Spain.

This paper is a comparison of time series and neural network models to forecast unemployment of nine Mediterranean countries using monthly seasonally adjusted data for unemployment (Eurostat database). Models which are considered are FARIMA, FARIMA/GARCH and neural networks, while ARIMA and Holt-Winters are used as benchmarks. In section 2 is described the nature of the data and potential problems of heteroskedasticity, non-linearity, non-normality and long-memory, while in section 3 are presented FARIMA, FARIMA/GARCH and Neural Network modeling approaches along with the description of a model selection method based on data characteristics. In section 4 takes place the data analysis and the comparison of models and in section 5 the summary and conclusions of the paper.

2 Unemployment Data Characteristics

Data characteristics play important role for the forecasting accuracy of various models. Models which take into account the specific characteristics of data can offer more accurate predictions. Based on past research on monthly unemployment rates, we observe that models such as FARIMA and GARCH have been used, thus characteristics such as long-memory and heteroskedasticity have been taken into account. Additionally, have been used models such as neural networks, to overcome the problem of non-linearity. One more potential characteristic is the departure from normality in some cases. We are testing for normality and non-linearity characteristics using statistical tests, i.e. Jarque-Bera test for normality (Jarque & Bera, 1980; Cromwell et al, 1994) and White neural network test for non-linearity (Lee et al, 1993). For the detection of long-memory we compute the Hurst exponent (Hurst, 1951) with three methods: the R/S method (Mandelbrot, 1972), the aggregate variance method and the Higuchi method (Taqqu et al, 1995) and we consider the smaller of the values as a conservative approach. A value of the Hurst exponent in $(0.5, 1)$ suggests the existence of long memory and values closer to 1 indicates stronger long memory. Values in $(0, 0.5)$ display antipersistence, i.e. larger values are followed by smaller values and vice versa, while a value equal to 0.5 could be interpreted as independence of the data or exponential decay of their autocorrelation function.

Autocorrelation of the data is checked through the Ljung-Box test in order to ensure that the time series display autocorrelation, thus time series models which model the dependence structure of the data are useful for predictions. Heteroskedasticity of data is checked through the Ljung-Box test on squared residuals of a fitted FARIMA model. This test helps us to decide if a FARIMA model is sufficient, or the volatility of the next period depends on these of past periods. If this is the case, then the use of a GARCH component for the volatility can lead to more accurate modeling of the data.

3 Forecasting Models

In this section we present the considered forecasting approaches for the prediction of unemployment rates. These approaches are compared with ARIMA and Holt-Winters models, which are considered as benchmarks.

3.1 FARIMA Models

The FARIMA forecasting models are extensions of the ARIMA (*p*, *d*, *q*) models where the fractional parameter *d* is allowed to take real, instead of only integers, values. A FARIMA model is given by the equation

$$\Phi_p(L)(1-L)^d(Y_t) = \Phi_q(L)\varepsilon_t,$$

where

L is the lag operator, $\Phi_p(L) = 1 - \varphi_1 L - \cdots - \varphi_p L^p$ and $\Theta_q(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$, $(1-L)^d = \sum_{j=0}^{\infty} \binom{d}{j} (-1)^j L^j$ where $\binom{d}{j}(-1) = \frac{\Gamma(-d+j)}{\Gamma(-d)\Gamma(j+1)}$, and $\varepsilon_t \sim N(0, \sigma^2)$ are the error terms.

Alternatively, there can be considered a model where error terms are following other than the normal distribution. Additionally, in this paper are considered models with the error terms to follow a student-t distribution t(0,σ,ν), where ν>2. The pdf in its location-scale version is

$$f(x;\alpha,\beta,\nu)=\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\beta\nu\pi}\Gamma(\frac{\nu}{2})}\left[1+\frac{(x-\alpha)^2}{\beta\nu}\right]^{\left(-\frac{\nu+1}{2}\right)}$$

with location parameter α, scale parameter β and shape parameter ν. The mean equals to α and here is 0, while the variance is $\frac{\beta\nu}{(\nu-2)}$.

To fit a FARIMA model to a time series the following procedure is applied: At first we ***convert data to a zero-mean series.*** Then we specify the ***order of the model.*** The order (p,q) of the corresponding ARMA model is first determined. For this study, we restrict the autoregressive and moving average orders to be less than or equal to 5 ($0 \le p \le 5,\ 0 \le q \le 5$) and use the lowest Bayes Information Criterion (BIC) for selecting the best combination. The same order (p,q) is used for the FARIMA model. Finally, we ***estimate the parameters of the model.*** After the order of the model has been fixed, the rest of the parameters d, φ_i and θ_j are estimated. The Geweke and Porter-Hudak (GPH) estimator (Geweke and Porter-Hudak, 1983) is our choice for d and is computed using R package fracdiff (Fraley et al, 2012), while a recursive Maximum Likelihood (ML) procedure is used for the estimation of the other parameters. The ML procedure suggested in (Sowell, 1992) goes through nonlinear optimization using the nlminb optimizer or augmented Lagrange method and it is all implemented in the R package rugarch (Ghalanos, 2014).

3.2 FARIMA/GARCH Models

The FARIMA/GARCH models are extensions of FARIMA, in the sense that in addition to all previously discussed assumptions they also assume conditional heteroskedasticity for the errors and then it is possible to give better results than FARIMA. These models were developed and used initially for the prediction of inflation (Baillie et al, 1996) where often the variance appears to be non-stationary. A FARIMA/GARCH model is given by the equation:

$$\Phi_p(L)(1-L)^d(Y_t-\mu)=\Theta_q(L)\varepsilon_t$$

where

$\varepsilon_t|\Omega_{t-1}\sim$ probability distribution $(p_1,\dots,p_k)$ and $\sigma_t^2=\omega+\sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2+\sum_{i=1}^{p}\beta_i\sigma_{t-i}^2$

The new element here, comparing with the description of the FARIMA model, is the fact that given an information set $\Omega_{t-1}$the error terms ε_t follow a probability distribution with parameters $(p_1,\dots,p_k)$. This distribution is usually assumed to be Normal with zero mean and σ_t^2 variance (Bollershev, 1986). Here, we assume both normal and student-t errors as described in 3.2.

The fitting of a FARIMA/GARCH model for this paper will follow the same steps as for the FARIMA, so that the order (p, q) and the fractional parameter d would be the same as in the corresponding plain FARIMA. Additionally, we will consider a GARCH(1,1) model, which in most applications is sufficient for capturing the conditional variance of the errors. The fitting of a FARIMA/GARCH model can also be performed with the use of the R software package rugarch (Ghalanos, 2014).

3.3 Feed-forward Neural Network Models

Since their introduction, neural networks have been successfully applied to many disciplines, including forecasting (Lippmann, 1987; Zhang *et al.*, 1998). They can handle non-linear phenomena more successfully than traditional time series models. ANN forecasting models for time series use sliding windows in the sense that a window with the *k* most recent values is used to predict the next value. More specifically, the forecasting model is expressed as in Eq. (3):

$$x(t+1) = f(\mathbf{x}(t)) \tag{3}$$

where $\mathbf{x}(t) = (x(t), x(t-1\tau), \ldots, x(t-(n-1)\tau)$ is the vector of the lagged variables to be used as input for the forecast. An application of the use of neural networks for time series forecasting can be found at (Frank et al, 2001). The process of the information is the following: the input nodes contain the value of the explanatory variables (in our case past values). Each node connection represents a weight factor and the information reaches a single hidden layer node as the weighted sum of its inputs. Each node of the hidden layer passes the information through a nonlinear activation function and passes it on to the output layer if the calculated value is above a threshold.

There have been a number of different architectures for ANNs and in this paper will be used the multilayer feed-forward design. In order to construct an ANN for time series one-step ahead prediction one needs to decide about the input variables, the number of hidden layers and number of nodes for each layer. Empirical research has shown that one hidden layer is sufficient in most cases; therefore we only have to define the number of input nodes and number of hidden nodes. The neural network used in this paper is a feed-forward ANN comprising of an input layer, one hidden layer and an output node. Each layer is fully connected to the next one and the activation function used in the hidden layer is the sigmoid:

$$S(t) = \frac{1}{1+e^{-t}}.$$

Moreover, a linear function is used in the output layer in order to transform the previous inputs to final outputs. The training of the network has been done with the back-propagation technique (i.e. to find a function that best maps a set of inputs to their correct output, thus determine final neuron weights) (Rumelhart et al, 1986), where the weights of the connections in the neural network are updated using the adaptive gradient descent optimization algorithm (Haykin, 1999).

In this paper, for the construction of an ANN for an unemployment time series, the following steps are performed:

1. ***Determine the resampling rate k:*** We consider k=1, i.e. no resampling is taking place and we apply the neural network model to the full time series.
2. ***Determine the number of input variables:*** We consider all possible models from one to four input nodes.
3. ***Determine the hidden layer nodes and training epochs:*** We consider one hidden layer and decide from 1 to 10 nodes. The training is performed using back-propagation

with the adaptive gradient descent algorithm and for 500 periods of training. The activation function is sigmoid for the hidden layer and linear for the output.

The final topology of the ANN will consist of an input layer with I nodes, one hidden layer with H nodes and an output layer with one node, denoted as $(I, H, 1)$. From the above description, the following applies: $I \in \{1,2,3,4\}$, $H \in \{1, \ldots, 10\}$ and $I, H \in \mathbb{Z}$. The implementation of the models is performed using R package AMORE (Limas et al, 2014).

3.4 Model Selection Method

Except from the above models, a model selection method is considered in order to further improve the accuracy of the FARIMA and FARIMA/GARCH models. The selection is based on the characteristics of normality and of heteroskedasticity of the data. The selection rules are the following:

1. Select FARIMA/GARCH model with student t innovations if non-normality and heteroskedasticity are detected.
2. Select FARIMA model with student t innovations if only non-normality is detected.
3. Select FARIMA/GARCH model with normal innovations if only heteroskedasticity is detected.
4. Select FARIMA with normal innovations elsewhere, i.e. neither non-normality nor heteroskedasticity are detected.

4 Data Analysis

Data are monthly seasonally adjusted unemployment rates of nine Mediterranean countries. Source of the data is the publicly available Eurostat database, the time period is from 2008 M1 to 2016 M10 to train the data and are considered 1 step-ahead predictions for the next 12 months, i.e. until 2017 M10 to compare the performance of the models.

However, there is no global accepted best criterion for comparison of forecasting accuracy of models. The comparison of the forecasting accuracy of the models is performed with the well-known RMSE and MAE criteria. There are considered five different metrics to assess the overall performance of models.

1. Average RMSE.
2. Average MAE.
3. Number of times when a model is the best choice for forecasting.
4. Average position according to RMSE.
5. Average position according to MAE.

To calculate the values of criteria 4 and 5, for every dataset we rank the models according to their performance and we specify their position and then we calculate the average value of the positions for each model. The consideration of different metrics allows a more complete comparison of models.

4.1 Exploratory Analysis of Data

Table 1 displays descriptive statistics of data, i.e. mean, standard deviation (sd), skewness, (excess) kurtosis and coefficient of variation (CV) and Table 2 displays statistical tests for data characteristics, i.e. Ljung-Box test for autocorrelation, Jarque-Bera test for Normality, Ljung-Box test on squared residuals of a FARIMA model (noted on table 3) for heteroskedasticity, White neural network test for non-linearity and the estimation of Hurst exponent, via R/S method, to measure the long memory of the time series.

Table 1: Descriptive statistics of data characteristics

Country	***Mean***	***Sd***	***Skewness***	***Kurtosis***	***CV***
Greece	19.3896	7.3698	-0.4327	-1.4639	0.3801
Spain	20.8755	4.3715	-1.0158	0.6333	0.2094
France	9.5292	0.9087	-1.1743	0.6396	0.0954
Croatia	13.7085	3.1592	-0.4061	-1.2102	0.2305
Italy	9.9849	2.1391	-0.1806	-1.5635	0.2142
Cyprus	10.5302	4.6778	-0.1166	-1.5448	0.4442
Malta	6.1132	0.6592	-0.5561	-0.2729	0.1078
Slovenia	7.9509	1.8241	-0.6650	-0.4843	0.2294
Turkey	10.1783	1.4027	0.7925	0.0026	0.1378

Table 2: Statistical Tests of data characteristics and Hurst exponent estimation

Country	***Auto-correlation (Ljung-Box)***	***Normality (Jarque-Bera)***	***Heteroskedasticity (Ljung-Box on squared.resid.)****	***Non-linearity (White test)***	***Hurst Exponent***
Greece	105.9915 (<0.01)	12.7729 (<0.01)	1.078e-05 (0.9974)	33.1988 (<0.01)	0.8933
Spain	100.8528 (<0.01)	20.0021 (<0.01)	22.97 (<0.01)	26.9608 (<0.01)	0.9479
France	101.5882 (<0.01)	26.1688 (<0.01)	0.003717 (0.9514)	4.8784 (0.087)	0.9344
Croatia	106.3243 (<0.01)	9.3827 (<0.01)	4.206e-05 (0.9948)	5.71 (0.058)	0.9240
Italy	104.4151 (<0.01)	11.3729 (<0.01)	0.01072 (0.9175)	7.7441 (0.021)	0.9041
Cyprus	106.1092 (<0.01)	10.7797 (<0.01)	25.82 (<0.01)	8.5385 (0.014)	0.8652
Malta	95.5515 (<0.01)	5.7923 (0.055)	0.9103 (0.3400)	3.8054 (0.1492)	0.7113
Slovenia	104.6452 (<0.01)	8.8474 (0.012)	1.088 (0.2969)	2.6315 (0.2683)	0.9363
Turkey	104.2006 (<0.01)	11.0957 (<0.01)	23.67 (<0.01)	1.2239 (0.5423)	0.8966

**test in the FARIMA squared residuals*

The time series of unemployment rates are relatively heterogeneous as mean values range from 6.11 to 20.88 and sd's from 0.66 to 7.37. This heterogeneity is reflected to CV values, where ranges from 0.095 for France to 0.444 for Cyprus. Most series are platykurtic (except Spain, France and Turkey) and left skewed (except Turkey).

All datasets display autocorrelation, while at 8 out of 9 time series (exception is Malta) is detected significant deviation from normality (at 5% level). Furthermore, in Greece, Spain, Italy and Cyprus is detected non-linearity (significant at 5% level). Additionally, existence of (strong) long – memory can be detected for all series via Hurst exponent. Finally, heteroskedasticity can be detected through the Ljung-Box test on the squared residuals after the fitting of a FARIMA model (order of the model is displayed at table 3). In Spain, Cyprus and Turkey it can be observed heteroskedasticity and this fact indicates that models such as GARCH which take into account time varying variance can offer additional forecasting accuracy.

4.2 Comparison of Models

In Table 3 are displayed the orders of ARIMA and FARIMA models, the prediction error of ARIMA, FARIMA, FARIMA/GARCH and Holt-Winters models and finally the architecture and the prediction error of ANN model. For each model are calculated the five different performance metrics and the final row displays the number of performance metrics (criteria) for which each model is the best choice.

Table 3: Order and forecasting accuracy of models

			Prediction Error					
Country	***ARIMA order***	***FARIMA order***	***FARIMA***	***ARIMA***	***FARIMA/ GARCH (1,1)***	***Holt-Winters***	***ANN***	
Greece	*(0,2,1)*	*(0, 1.2412, 1)*	*RMSE=0.2598 MAE=0.2098*	*RMSE=0.5533 MAE=0.4560*	***RMSE=0.2068 MAE=0.1582***	*RMSE=0.2422 MAE=0.1912*	*(3,2,1)*	*RMSE=0.2541 MAE=0.1819*
Spain	*(2,2,0)*	*(2, 1.2304, 0)*	*RMSE=0.1208 MAE=0.0954*	*RMSE=0.2275 MAE=0.1947*	***RMSE= 0.1002*** *MAE=0.0809*	*RMSE= 0.1034* ***MAE=0.0792***	*(3,9,1)*	*RMSE=0.1029 MAE=0.0795*
France	*(1,2,0)*	*(1, 1.2103, 0)*	*RMSE= 0.1135* ***MAE=0.0879***	*RMSE=0.1524 MAE=0.1333*	***RMSE=0.1106*** *MAE=0.0900*	*RMSE=0.1164 MAE=0.0954*	*(1,2,1)*	*RMSE=0.1189 MAE=0.0954*
Croatia	*(0,2,1)*	*(0, 1.5974, 1)*	*RMSE=0.1307 MAE=0.1163*	*RMSE=0.3363 MAE=0.3031*	***RMSE=0.1251 MAE=0.1073***	*RMSE=0.1308 MAE=0.1139*	*(2,2,1)*	*RMSE=0.1285 MAE=0.1140*
Italy	*(0,1,0)*	*(0, 1.1258, 0)*	*RMSE= 0.1817 MAE=0.1537*	*RMSE=0.1971 MAE=0.1660*	*RMSE=0.1817 MAE=0.1537*	***RMSE=0.1656 MAE=0.1349***	*(2,2,1)*	*RMSE=0.2189 MAE=0.1828*
Cyprus	*(0,2,2)*	*(0, 1.1126, 2)*	*RMSE= 0.1873 MAE=0.1545*	*RMSE=0.4641 MAE=0.4293*	*RMSE= 0.2006 MAE=0.1669*	***RMSE=0.1805 MAE=0.1505***	*(3,1,1)*	*RMSE=0.2021 MAE=0.1601*
Malta	*(0,1,0)*	*(0, 0.9371, 0)*	*RMSE=0.1196 MAE=0.0843*	***RMSE= 0.0988 MAE=0.0666***	*RMSE=0.1196 MAE=0.0843*	*RMSE=0.1013 MAE=0.0706*	*(4,7,1)*	*RMSE=0.1414 MAE=0.0994*
Slovenia	*(0,2,0)*	*(0, 1.3525, 0)*	***RMSE=0.1402*** *MAE=0.1151*	*RMSE=0.3617 MAE=0.2754*	***RMSE=0.1402*** *MAE=0.1151*	*RMSE= 0.1458* ***MAE=0.1092***	*(1,1,1)*	*RMSE=0.1554 MAE=0.1188*
Turkey	*(2,1,0)*	*(2, 1.1670, 0)*	*RMSE= 0.1620 MAE=0.1431*	*RMSE= 0.1631 MAE=0.1445*	*RMSE= 0.1572 MAE=0.1398*	*RMSE= 0.1647 MAE=0.1450*	*(3,4,1)*	***RMSE=0.1177 MAE=0.0879***
Average RMSE			*0.1573*	*0.2838*	***0.1491***	*0.1501*	*0.1600*	
Average MAE			*0.1289*	*0.2410*	*0.1218*	***0.1211***	*0.1244*	
Best Model			*2*	*1*	***5***	*4*	*1*	
Average Position RMSE			*2.6667*	*4.3333*	***1.6667***	*2.6667*	*3.3333*	
Average Position MAE			*2.7778*	*4.3333*	***2.1111***	***2.1111***	*3.2222*	
number of best criteria			***0***	***0***	***4***	***2***	***0***	

From the above models, FARIMA/GARCH is the best model with four metrics (only on average MAE is outperformed by Holt-Winters). The distribution of errors in ARIMA, FARIMA and FARIMA/GARCH is considered Normal.

4.3 The Effect of Non-Normality and model selection strategy

The non-normality of some datasets suggests that we may improve the forecasting accuracy of models with the use of non-normal errors. To further improve the results, we consider FARIMA and FARIMA/GARCH models with student-t error distributions. Table 4 displays the prediction error of FARIMA with normal and student-t error distributions, the selected model and the prediction error of the model selection strategy which described in section 3.4. Figures 1 and 2 display the comparison of models using average values and average positions respectively, while Figure3 displays the number of times where a model is the best either with RMSE or MAE criterion and the number of criteria with which a model is the best choice.

Table 4: Forecasting accuracy of models

	Prediction Error				***Strategy***	
Country	***FARIMA-N***	***FARIMA-t***	***FARIMA/ GARCH (1,1)-N***	***FARIMA/ GARCH (1,1)-t***	***Model***	***Error***
Greece	*RMSE=0.2598 MAE=0.2098*	*RMSE= 0.2113 MAE=0.1605*	***RMSE=0.2068 MAE=0.1582***	*RMSE= 0.2158 MAE=0.1649*	*FARIMA/ GARCH-t*	*RMSE= 0.2158 MAE=0.1649*
Spain	*RMSE=0.1208 MAE=0.0954*	*RMSE= 0.1004* ***MAE=0.0785***	***RMSE= 0.1002*** *MAE=0.0809*	*RMSE= 0.1013 MAE=0.0788*	*FARIMA/ GARCH-t*	*RMSE= 0.1013 MAE=0.0788*
France	*RMSE= 0.1135* ***MAE=0.0879***	***RMSE= 0.1106*** *MAE=0.0907*	***RMSE=0.1106*** *MAE=0.0900*	*RMSE= 0.1123 MAE=0.0885*	*FARIMA-t*	***RMSE= 0.1106*** *MAE=0.0907*
Croatia	*RMSE=0.1307 MAE=0.1163*	*RMSE= 0.1230* ***MAE=0.1064***	*RMSE=0.1251 MAE=0.1073*	***RMSE= 0.1223 MAE=0.1064***	*FARIMA-t*	*RMSE= 0.1230* ***MAE=0.1064***
Italy	***RMSE= 0.1817 MAE=0.1537***	***RMSE= 0.1817 MAE=0.1537***	***RMSE=0.1817 MAE=0.1537***	***RMSE=0.1817 MAE=0.1537***	*FARIMA/ GARCH-t*	***RMSE=0.1817 MAE=0.1537***
Cyprus	***RMSE= 0.1873 MAE=0.1545***	*RMSE= 0.1908 MAE=0.1578*	*RMSE= 0.2006 MAE=0.1669*	*RMSE= 0.1899 MAE=0.1570*	*FARIMA/ GARCH-t*	*RMSE= 0.1899 MAE=0.1570*
Malta	***RMSE=0.1196 MAE=0.0843***	***RMSE=0.1196 MAE=0.0843***	***RMSE=0.1196 MAE=0.0843***	***RMSE=0.1196 MAE=0.0843***	*FARIMA-N*	***RMSE=0.1196 MAE=0.0843***
Slovenia	***RMSE=0.1402 MAE=0.1151***	***RMSE=0.1402 MAE=0.1151***	***RMSE=0.1402 MAE=0.1151***	***RMSE=0.1402 MAE=0.1151***	*FARIMA-N*	***RMSE=0.1402 MAE=0.1151***
Turkey	*RMSE= 0.1620 MAE=0.1431*	*RMSE= 0.1602 MAE=0.1419*	*RMSE= 0.1572 MAE=0.1398*	***RMSE=0.1540 MAE= 0.1375***	*FARIMA/ GARCH-t*	***RMSE=0.1540 MAE= 0.1375***
Average RMSE	*0.1573*	*0.1486*	*0.1491*	*0.1486*	***0.1485***	
Average MAE	*0.1289*	*0.1210*	*0.1218*	***0.1207***	*0.1209*	
Best Model	*5*	***6***	***6***	*5*	***6***	
Average Position RMSE	*3.2222*	*2.0000*	*2.0000*	*1.8889*	***1.6667***	
Average Position MAE	*2.7778*	*2.1111*	*2.5556*	***1.5556***	*1.7778*	
number of best criteria	***0***	***1***	***1***	***2***	***3***	

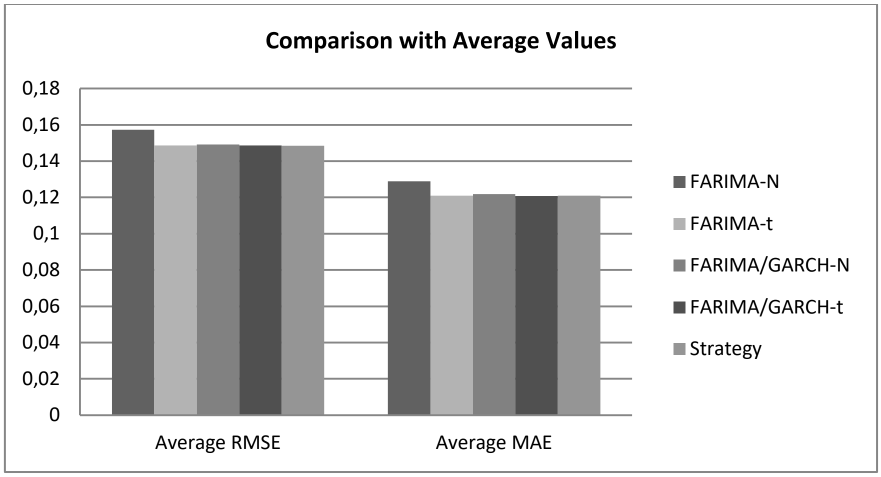

Fig.1: Comparison of models with Average RMSE and MAE

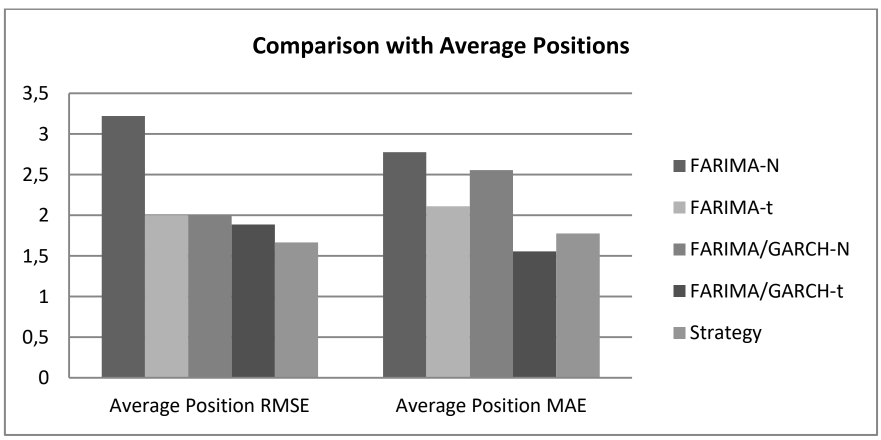

Fig.2: Comparison of models with Average Positions based on RMSE and MAE

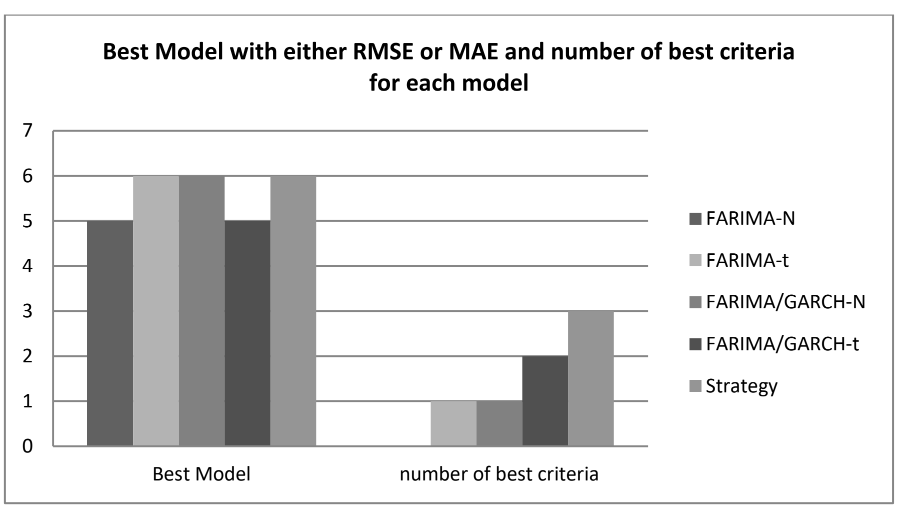

Fig.3: Number of times a model is the best choice with RMSE or MAE and overall

The consideration of student –t errors led to improvement of forecasting accuracy of both FARIMA and FARIMA/GARCH models. FARIMA with student-t errors displayed better results than FARIMA with normal errors in all 5 metrics, while FARIMA/GARCH with student-t errors appeared more accurate than FARIMA/GARCH with normal errors in 4 out of 5 metrics.

The (model selection) strategy is the best approach with 3 metrics (Average RMSE, number of traces where is the best model and average position based on RMSE). The FARIMA/GARCH model with student-t errors is the best individual model approach and is very close to the model selection strategy (better in Average MAE and average position based on MAE).

5 Summary and Conclusions

In this work is presented and applied a number of time series models, multilayer feed-forward neural networks and a model selection method for forecasting of the unemployment time series of nine Mediterranean countries. At the core of time series models is the existence of long-memory and FARIMA model which incorporates this property is the basis of the considered approaches. Data properties (non-linearity, heteroskedasticity and non-normality) are explored and tested and a framework for evaluation is presented.

First, are compared FARIMA and FARIMA/GARCH models with normal error distribution, MLP neural network models with ARIMA and Holt-Winters which are considered as benchmark models. FARIMA/GARCH model found to be the best approach.

To further improve the forecasting accuracy, the effect of non-normality is taken into consideration. FARIMA and FARIMA/GARCH models with student-t errors are considered and led to improved accuracy compared to the same models with normal errors. Finally, a model selection method which is based on data characteristics was even more accurate. The consideration of data characteristics in our model building, offered additional forecasting ability.

References

[1] Aiken, M. (1996). A neural network to predict civilian unemployment rates. Journal of International Information Management, 5(1), 3.

[2] Baillie, R. T., Chung, C. F., & Tieslau, M. A. (1996). Analysing inflation by the fractionally integrated ARFIMA--GARCH model. Journal of applied econometrics, 23-40.

[3] Blanchard, O. J., & Summers, L. H. (1986). Hysteresis and the European unemployment problem. NBER macroeconomics annual, 1, 15-78.

[4] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 31(3), 307-327.

[5] Cromwell, J. B., & Labys, W. C. M. Terraza (1994),"Univariate Tests for Time Series Models".
[6] Dobre, I., & Alexandru, A. A. (2008). Modelling unemployment rate using Box-Jenkins procedure.
[7] Etuk, H., Uchendu, B., & Uyodhu, V. (2012). ARIMA fit to Nigerian unemployment data. Journal of Basic and Applied Scientific Research, 2(6), 5964-5970.
[8] Floros, C. (2005). Forecasting the UK unemployment rate: model comparisons. International Journal of Applied Econometrics and Quantitative Studies, 2(4), 57-72.
[9] Fraley, C., Leisch, F., Maechler, M., Reisen, V., & Lemonte, A. (2012). fracdiff: Fractionally differenced ARIMA aka ARFIMA (p, d, q) models. R package version 1.4-2.
[10] Frank, R. J., Davey, N., & Hunt, S. P. (2001). Time series prediction and neural networks. Journal of intelligent and robotic systems, 31(1-3), 91-103.
[11] Ghalanos, A. (2014). rugarch: Univariate GARCH models, R package version 1.3-3.
[12] Gil-Alana, L. A. (2001). A fractionally integrated exponential model for UK unemployment. Journal of Forecasting, 20(5), 329-340.
[13] Haykin, S. (1999). Neural networks a comprehensive introduction.
[14] Hurst, H. E. (1951). Long term storage capacity of reservoirs.ASCE Transactions, 116(776), 770-808.
[15] Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Economics letters, 6(3), 255-259.
[16] Johnes, G. (1999). Forecasting unemployment. Applied Economics Letters, 6(9), 605-607.
[17] Katris, C. (2015). Dynamics of Greece's unemployment rate: effect of the economic crisis and forecasting models.International Journal of Computational Economics and Econometrics, 5(2), 127-142.
[18] Kurita, T. (2010). A Forecasting Model for Japan"s Unemployment Rate. Eurasian Journal of Business and Economics, 3(5), 127-134.
[19] Lee, T. H., White, H., & Granger, C. W. (1993). Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests. Journal of Econometrics, 56(3), 269-290.
[20] Limas, M. C., Meré, J. B. O., Marcos, A. G., de Pisón Ascacibar, F. J. M., Espinoza, A. V. P., Elías, F. A., & Ramos, J. M. P.. (2014). A MORE flexible neural network package, R package version 0.2-15.
[21] Lippmann, R. (1987). An introduction to computing with neural nets. IEEE Assp magazine, 4(2), 4-22.
[22] Mandelbrot, B. (1972). Statistical methodology for nonperiodic cycles: from the covariance to R/S analysis. In Annals of Economic and Social Measurement, Volume 1, number 3(pp. 259-290). NBER.
[23] Mladenovic, J., Ilic, I., & Kostic, Z. (2017). Modeling The Unemployment Rate At The Eu Level By Using Box-Jenkins Methodology. KnE Social Sciences, 1(2), 1-13.
[24] Olmedo, E. (2014). Forecasting spanish unemployment using near neighbour and neural net techniques.Computational Economics, 43(2), 183-197.

[25] Proietti, T. (2003). Forecasting the US unemployment rate.Computational Statistics & Data Analysis, 42(3), 451-476.
[26] Rothman, P. (1998). Forecasting asymmetric unemployment rates. Review of Economics and Statistics, 80(1), 164-168.
[27] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. nature,323(6088), 533.
[28] Stoklasová, R. (2012). Model of the unemployment rate in the Czech Republic. In Proceedings of 30th International Conference on Mathematical Methods in Economics (pp. 836-841).
[29] Taqqu, M. S., Teverovsky, V., & Willinger, W. (1995). Estimators for long-range dependence: an empirical study. Fractals,3(04), 785-798.
[30] Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. International journal of forecasting, 14(1), 35-62.