THE USV ANNALS OF ECONOMICS AND PUBLIC ADMINISTRATION | VOLUME 16, SPECIAL ISSUE, 2016

MODELLING CHAOTIC TIME SERIES USING HYBRID APPROACHES

Lecturer PhD **Sorin VLAD**
"Ştefan cel Mare" University of Suceava, Romania
sorinv@seap.usv.ro

Lecturer PhD **Mariana VLAD**
"Ştefan cel Mare" University of Suceava, Romania
marianav@seap.usv.ro

Abstract:
Chaotic time series have as main feature, the possibility of short time prediction of their future evolution. The time series of the natural systems are seldom generated by purely linear or nonlinear systems. They are usually the result of the evolution in time of nonlinear systems. Therefore, the time series will encapsulate both linear and nonlinear components. This hybrid character of such time series makes the prediction process very complex and not accurate using a single technique. Combining different techniques and models increases the chances of representing and modeling complex relationships between data and to improve the quality of time series prediction. The paper presents the results obtained in the case of the electricity spot price time series modeled using ARIMA technique for the linear part and a neural network for the nonlinear component.

Key words: chaotic dynamics, time series prediction, ARIMA modeling, neural networks

JEL classification: C02

1. INTRODUCTION

The literature referring to hybrid approaches in modeling, offers many examples of papers describing the use of ARIMA technique for modeling the linear part of the time series and neural networks for the nonlinear component. [7], [4]

A model of a univariate time series tries to explain the evolution of a variable based on its previous values. Usually, it is considered that the variable is the result of a stochastic process, its values are evolving according to probabilistic laws.

Since in the time series analyzed by this paper elements indicating the presence of chaotic dynamics have been identified, we will apply the technique of phase space reconstruction. According to Takens's theorem, since each variable contains information about the dynamics of the whole system, one can reconstruct the phase space by using a univariate time series of the observed variable. Reconstructed phase space (also called embedding space) will be equivalent to the original topological space phase and, therefore, equivalent to the original attractor. The reconstructed phase space is parameterized by the embedding dimension m and the delay time τ. M dimensional vectors formed applying the equation 1 are describing an object topologically equivalent to the attractor of the original time series.

$$Y_i = \{y_i, y_{i-\tau}, y_{i-2\tau}, \ldots, y_{i-(m-1)*\tau}\} \quad (1)$$

2. ARIMA MODELING TECHNIQUE

The ARIMA modeling procedure analyzes univariate time series using Autoregressive Integrated Moving Average (ARIMA). The ARIMA model (also called Box Jenkins model) allows forecasting the future values of a time series as linear combinations between past values of the series and previous errors (called shocks or innovations). An ARIMA model also uses correlations between time series values.

The process of finding the appropriate ARIMA model comprises two steps:

- Identifying the model – determines the model best fitting the time series behavior. If the time series is affected by the seasonal component, the model to be determined is called SARIMA (Seasonal ARIMA) and it's represented as ARIMA (p, d, q) (P, D, Q)s, where d and D represents seasonal and non-seasonal ordinal differences, p and P are seasonal and non-seasonal orders of the autoregressive terms, and q and Q represents ordinal moving average terms. S signifies seasonal difference. Finding the model parameters is achieved by studying the plots of autocorrelation function (ACF) and the partial correlation (PACF).
- Model estimation - performance evaluation to describe the behavior of the time series. If the performance is not satisfactory one will return to the previous step, modifying the structure of the model until it is appropriate to capture the dynamics of time-series.

The time series used for developing a model are assumed to be stationary. Stationarity is a mathematical concept introduced to simplify theoretically and practically the modeling of the stochastic processes. In practice, the chances that time series to be nonstationary are increasing with the number of the values contained by that time series. A time series is non-stationary if its properties do not change regardless of when it is observed. Trend and seasonal component causes the series to be nonstationary. Time series stationarizing is usually done through differentiation (repeated as many times as needed) and involves calculating differences between the elements of the time series.

The mathematical expression of the general ARIMA model is:

$$y_t = c + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} - \theta_1 e_{t-1} - ... - \theta_p e_{t-p} \quad (2)$$

The equation above corresponds to an ARIMA(p,d,q) model, where p is the order of the autoregression part (the number of previous values to be taken into account when calculating the current value) d the number of differences , and q is the order of the moving average part (specify how the deviation from the mean of previous values is taken into calculation of the current value).

Obtaining an optimum And RIMA model implies to find its optimal order, the residuals resulting from the application of the model must be independent and identically distributed with zero mean (i.e. to be uncorrelated with each other) white noise type.

The initial time series is considered to contain a linear autocorrelated component and a nonlinear component, such that we can decompose the original time series like in the equation 3:

$$y_t = L_t + N_t \quad (3)$$

After ARIMA modeling process is completed, the time series can be expressed as the sum of linear component and the residuals:

$$y_t = L_t^m + e_t \quad (4)$$

The residuals will contain the nonlinear relationships that can't be captured by the ARIMA model.

The nonlinear part of the time series will be modeled using a multilayer perceptron neural network. The values at the output of the network are given by:

$$y_t = b + \sum_{j=1}^{q} \alpha_j g\left(b_0 + \sum_{i=1}^{p} w_{ij} y_{t-i}\right) + \varepsilon_t \quad (5)$$

where α_i, w_{ij} are the network weights, p and q are the numbers of neurons on the input layer, hidden layer respectively, and g is the activation function of the hidden neurons (usually hyperbolic tangent).

The neural network determines the output values as a combination of the previous values, like the nonlinear regression model.

$$y_t = f(y_{t-1}, y_{t-2}, ... y_{t-p}, w) + \varepsilon_t \quad (6)$$

where w is the vector of all network parameters and f is the function determined by the neural network.

The method further applied aims to combine the abilities and performances of the two methods in time series modeling. ARIMA is used for the prediction of the linear component of time series and the neural network for the residual modeling (i.e. the nonlinear part of the time series). Due to its universal approximator property, the multilayer perceptron is extensively used in time series modeling and prediction. In the majority of this type of problems, one hidden layer of neurons is sufficient.

3. ELECTRICITY SPOT MARKET MODELING

In the spot market, or the day-ahead market, the participants are trading contracts to supply electricity for the next day. Electricity prices are set daily, for one-hour intervals and covers the whole 24 hours range of the next day. The deadline for collecting all purchase orders and sales information is stored in the form of aggregated supply and demand curves for each hour. The prices determined using this procedure are called steady spot prices or system prices.

The models for spot price are the foundations of decision making and strategic planning processes. The need of spot price prediction, even for short-time periods is obvious and very important for every player of the trading market.

Most models for spot price market are using at least two risk factors: short-term dynamics characterized by mean reversion and a very strong volatility and the second one represents the long-term behavior of the price [6], [5], [2].

The mean reversion is modeled by a Ornstein-Uhlenbeck process given by:

$$dS_t = -\lambda(S_t - a)dt + \varphi dW_t \tag{7}$$

where S_r is the spot price, W_t is a standard Brownian motion, φ is the process volatility and λ is the speed of the process reverts to the mean a.

Another classical model consists of two terms:

$$dS_t = -\lambda(S_t - Y_t)dt + \varphi dW_t \tag{8}$$

where Y_t is a Brownian motion type process.

An alternative model is given by:

$$S_t = \exp(X_t + Y_t) \tag{9}$$

where X_t is the Ornstein-Uhlenbech process describing short term variation and Y_t is the Brownian motion describing long term dynamics.

There are two completely opposite opinions regarding power consumption: the first one considers that is chaotic due to the influence of multiple factors (temperature, electricity price, network distribution operating conditions, economic environment, season, etc.) and another one that states that power consumption is random.

We will work with the electricity spot price time series for one month, which we identified as exhibiting chaotic dynamics.

There are different approaches in analyzing the time series as univariate or multivariate [1]. The trading value of the electricity for each hour of the current day being set up in the previous one, the time series is seen by some researchers as multivariate. Other papers are working with 24 distinct values per day and the time series is univariate. In our analysis we have considered the time series as univariate.

The electricity spot price time series has some specific features:

1. Cyclical patterns represented by different types of seasonally behavior: annual seasonality (the price depends on the current season, in the cold season the price is higher), weekly season (the price depends on the day of the week) and daily (the price is different from hour to hour depending on the load).
2. Mean reverting. Referring to the short-time behavior, significant variations of the price can suddenly occur, and then the price is coming back to the values prior the increase. These

outliers can occur because storing large quantities of electricity is not possible, it has to be consumed when produced.

3. Another specific property of the electrical energy trading market is the fact that the price can be negative or zero, this happening mostly over night [3]. If there is an energy surplus, and the consumption low, the price is negative because uncoupling the generators or the nuclear plants would be very costly. The EEX German trading market was the first European market to allow negative prices.

4. RESULTS

The time series was differentiated until the mean stationarity is obtained. Generally, we can use logarithmic values of the time series for variance stabilization. We have chosen to work with non logarithmic values

The outliers were not removed in order not to destroy the dynamics of the time series, even if these values are the result of the exceptional events or can be produced by errors.

Finding the values for p, d and q parameters implies the analysis of ACF (autocorrelation function) and PACF (partial autocorrelation function) plots. ACF measures the strength of the relationship between the current value y_t and k previous values of the y_{t-k} series for different k values. PACF measures the extent of which y_t and y_{t-k} are correlated, but after removing the intermediate values effect (correlation for k time step unexplained by lower order correlations).

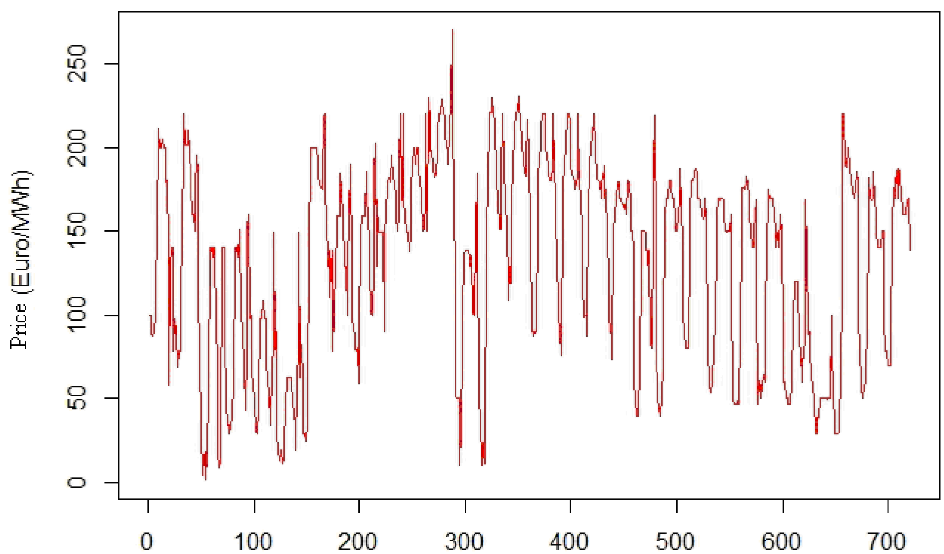
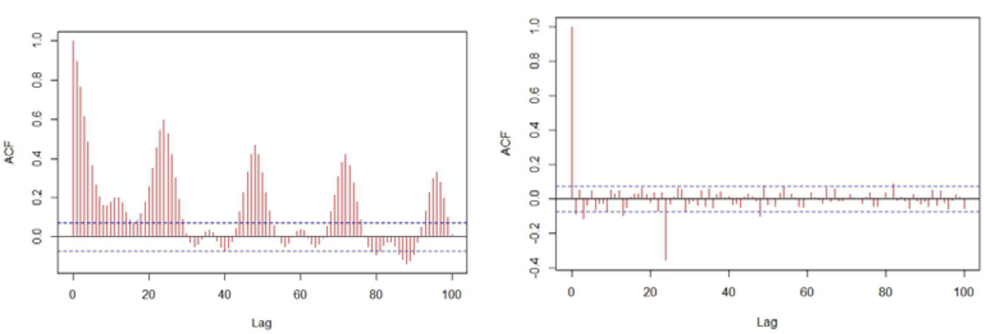


Figure no. 1. The spot price time series seasonally adjusted for trend removal.



a) b)

Figure no. 2. The plot of the ACF function for the spot price time series. a) before differentiation and b) after both simple and seasonal differentiation

A seasonal differentiation was carried out because the ACF plot obtained after removing the trend shows peaks outside the significance range for intervals 24, 48, 72 etc, as can be observed in Figure no. 2.

The optimum adjusted model is SARIMA(1,0,5)$(1,0,1)_{24}$ with AIC=6111.14. This index, the Akaike Information Criterion (AIC) is a relative measure of the quality of statistical models. The optimum model is the one with the lowest AIC. Table 1 presents the values of AIC index for different ARIMA models.

Table no. 1. AIC index for different possible models..

Model	AIC
ARIMA(0,0,0)	6348.15
ARIMA(0,0,0)(1,0,0)	6253.717
ARIMA(0,0,0)(2,0,0)	6192.9
ARIMA(0,0,1)(1,0,0)	6247.509
ARIMA(0,0,3)(2,0,0)	6182.547
ARIMA(0,0,4)	6337.326
ARIMA(1,0,0)(1,0,0)	6247.276
ARIMA(1,0,5)(2,0,0)	6188.617

The resulted model passes the Box – Ljung test. The null hypothesis (i.e. the model is not adequate) is rejected because the value of the parameter p (the level of uncorrelated residuals) is 0.7204, greater than 0.05 (the reference value). We can say that the results are not correlated (uncorrelated with no other predictor), as one can observe from the plot of ACF and PACF for the residuals time series. All the value are in the significance range.

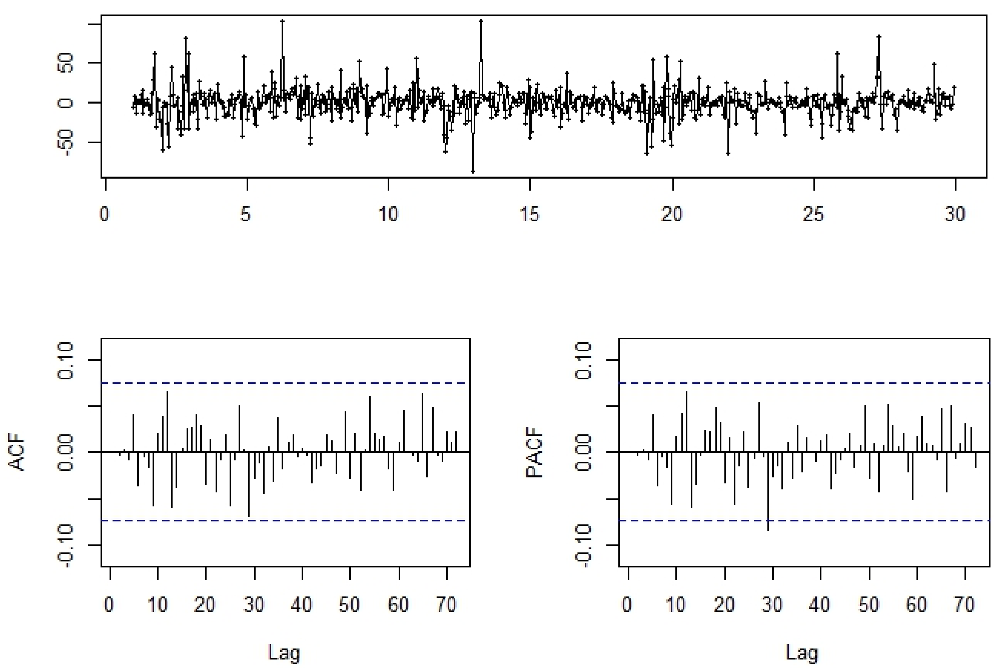


Figure no. 3. Autocorrelation and partial autocorrelation functions for the residuals of SARIMA model.

The result of time series modeled with optimal SARIMA model is presented in Figure no. 4.

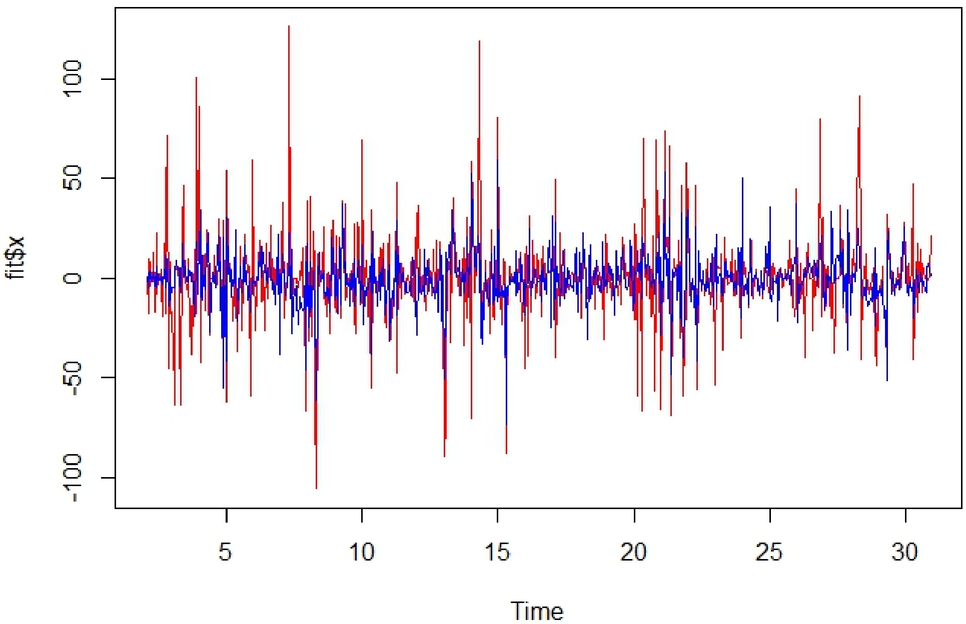


Figure no. 4. ARIMA modeled time series (blue) vs. original time series.

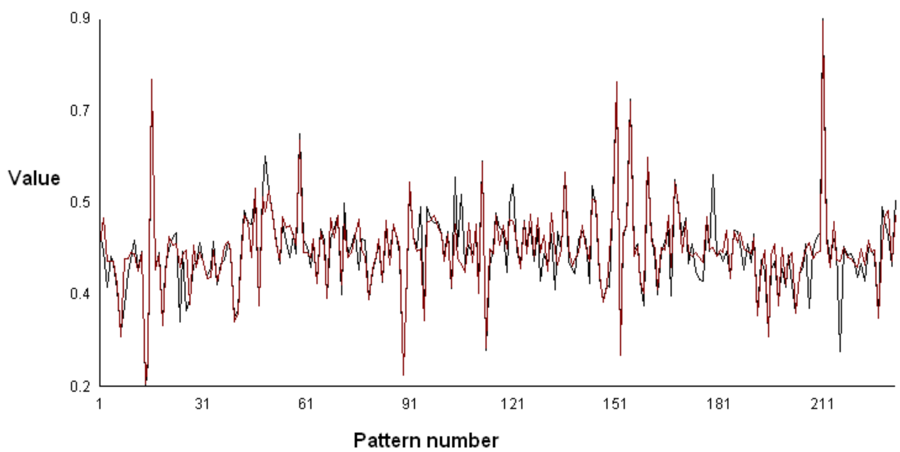


Figure no. 5. Learning results for the ARIMA residuals.

The mean squared error obtained as a result of model fitting is 0.00050813. The neural network with 2-24-1 architecture applied on the learning set leads to an error of 0.003.

5. CONCLUSIONS

Time series representing the evolution in time of a variable produced by a nonlinear system are difficult to predict. If the time series exhibits chaotic dynamics, the analysis and modeling is even more difficult. The paper present the results obtained in modeling of a time series corresponding to the electricity spot price values collected for one month. The nonlinear part of the time series was modeled using the ARIMA technique and the nonlinear part which is formed by the residuals from the ARIMA model, was modeled by a multilayer perceptron neural network. The results are proving that hybrid techniques are able to provide better models than each procedure separately.

BIBLIOGRAPHY

1. Jakasa T., Androcec I., Sprcic P., *Electricity price forecasting – ARIMA model approach*, 2011 8th International Conference on the European Energy Market, 25-27 May 2011, Zagreb
2. Liebl D., *Modeling and forecasting electricity spot prices: a functional data perspective*, The Annals of Applied Statistics, vol. 7, no. 3, p. 1562-1592, 2013
3. Schneider S., *Power spot price models with negative prices*, 2010, http://mpra.ub-uni-muenchen.de/29958/
4. Unsihuay Vila C., Zambroni de Souza A. C., Maragon J. W., *LimaElectricity demand and spot price forecasting using evolutionary computation combined with chaotic nonlinear dynamic model*, Electrical Power and Energy Systems 32(2010)
5. Weron R., *Electricity price forecasting: a review of the state-of-the-art with a look into the future*, International Journal of Forecasting, vol. 30, Issue 4, 2014
6. Weron R., *Modeling and forecasting electricity loads and prices – a statistical approach*, Wiley Finance 2006, ISBN 978-0-470-05753-7
7. Zhang G. P., *Time series forecasting using hybrid ARIMA and neural network model*, Neurocomputing 50, p. 159-175, 2003