Forecasting Electricity Spot Price with Generalized Long Memory Modeling: Wavelet and Neural Network

Souhir Ben Amor, Heni Boubaker, Lotfi Belkacem

Abstract—This aims of this paper is to forecast the electricity spot prices. First, we focus on modeling the conditional mean of the series so we adopt a generalized fractional k-factor Gegenbauer process (k-factor GARMA). Secondly, the residual from the k-factor GARMA model has used as a proxy for the conditional variance; these residuals were predicted using two different approaches. In the first approach, a local linear wavelet neural network model (LLWNN) has developed to predict the conditional variance using the Back Propagation learning algorithms. In the second approach, the Gegenbauer generalized autoregressive conditional heteroscedasticity process (G-GARCH) has adopted, and the parameters of the k-factor GARMA-G-GARCH model has estimated using the wavelet methodology based on the discrete wavelet packet transform (DWPT) approach. The empirical results have shown that the k-factor GARMA-G-GARCH model outperform the hybrid k-factor GARMA-LLWNN model, and find it is more appropriate for forecasts.

Keywords—k-factor, GARMA, LLWNN, G-GARCH, electricity price, forecasting.

I. INTRODUCTION

IN power markets, price analysis has become an important topic for all its participants. Background information about the electricity price is crucial for risk management. More precisely, it represents an advantage for a market player facing competition. In fact, forecasting electricity prices at different periods is valuable for all industry stakeholders for cash flow analysis, financial procurement, and capital budgeting, regulatory rule making, and integrated resource planning.

In this vein, both producers and consumers rely on price forecasting information to put forward their corresponding bidding strategies. If a producer has, an accurate price forecast, he can develop a bidding strategy in order to maximize his profit. On the other hand, if an accurate price forecast is available, a consumer can make a plan to minimize his own electricity cost. Hence, the players benefit is greatly affected by the accuracy of price forecast.

However, the behavior of electricity prices differs from other commodity markets. The most obvious of these differences is that electricity is a non-storable merchandize. Therefore, the demand and supply of electricity are highly inelastic and very sensitive to business cycles and weathers,

and so relatively, small changes in load or generation in a matter of hours or minutes can cause huge changes in price.

Moreover, electricity prices present some specific characteristics such as high frequency, multiple seasonality (on annual, weekly, and daily levels), non-stationary behavior, [1]-[3], high volatility, hard nonlinear behavior, long memory, calendar effect, high percentage of unusual prices, mean reversion, price spikes and limited information to the market participants. Hence, these behaviors may affect the prices dramatically. In this respect, there is no other similar market [4]. Thus, we cannot rely on models developed for financial markets or other commodity markets. In this framework, due to the complexity of the electricity market, the electricity price forecasting has been the most challenging task. This has also motivated the researchers to develop intelligent and efficient approaches in order to forecast the prices that all stakeholders in the market can benefit out of it.

In this paper, we are interested in resolving the issues of modeling and forecasting the features of the electricity prices, notably, the existing of the seasonal long memory behavior in the conditional mean and the conditional variance.

In fact, most of the existing studies use models that permit the modelling of one or two features but not more. Specifically they do not model the long memory behavior inside the seasonality in both the conditional mean and the conditional variance. Recall that the long memory models introduced by [5], [6] permit to model an infinite cycle, which is too restrictive for the electricity prices. One of the main characteristics of the high frequencies data sets is the presence of volatility clustering and leptokurtosis, as soon as persistence and cyclical patterns in the conditional mean of the series combined with conditional heteroscedasticity. All these characteristics have been presented inside electricity spot prices. Thus, dynamic modelling of means and variances appears essential for this kind of data sets. In this paper, we propose a new approach, which permit to take into account mainly all these features.

The novelty of our proposed method is its capacity to model the seasonal long memory behavior simultaneously in the conditional mean and in the conditional variance of electricity spot price, using a generalized long memory model called k-factor GARMA-G-GARCH. In addition, we adopt a wavelet estimation approach, which allows us to guarantee a

Souhir Ben Amor is with the Institute of High Commercial Studies of Sousse (IHEC), 40 Route de Ceinture Sahloul III, 4054 Sousse, Tunisia (e-mail: souhirbenamor@live.fr).

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:11, No:9, 2017

parsimonious model with the greater accuracy. More precisely, the methodology of our proposed model consists into two steps; in the first step, the k-factor GARMA model proposed by [7] is adopted to model the conditional mean in the time series. The choice of this model is motivated by the ability to take account simultaneously of long/short-term dependence and seasonal fluctuations at different frequency. The main feature of the k-factor GARMA model is that it allows larger diversity in the covariance structure of a variable witnessed by both the autocorrelation function and the spectral density function, which present k singularities.

Nevertheless, this approach focuses only on modeling the conditional mean, assuming that the residuals are white noises with a constant variance. However, it's well-recognized in the empirical studies that this hypothesis is not verified, and the residuals are rather characterized by a time varying variance. To reproduce these patterns, we include in the second step the G-GARCH model, recently introduced by [8] in order to estimate the volatility, since this model is allows taking into account the long memory phenomenon and seasonality effect in conditional variance. Hence, we obtain the so-called k-factor GARMA-G-GARCH model.

Then, for a comparison purpose, we use a second approach named the local linear wavelet neural network (LLWNN) model, proposed by [9], as a nonlinear non-parametric method to estimate the conditional variance, so we obtain a hybrid k-factor GARMA-LLWNN model. Hence, we can to benefit the power of these two approaches (parametric and non-parametric methods) in forecasting both conditional mean and conditional variance of electricity prices for the Nord Pool market.

This paper provides two contributions. The main one consists in modeling the seasonal long memory behavior of electricity prices simultaneously in the conditional mean and in the conditional variance. For this purpose, we develop a generalized k-factor GARMA-G-GARCH process, which allows for seasonal long-memory behavior associated with k-frequency. In fact, previous studies extend the k-factor GARMA model with the classic GARCH model or the FIGARCH model, and ignore the periodic long memory behavior when modeling the volatility. Concerning the parameters estimation of this model, we adopt an estimation method in the wavelet domain based on the maximal overlap discrete wavelet packet transform (DWPT) proposed by [10].

Our second purpose is to show the performance of the proposed generalized long memory model (the k-factor GARMA-G-GARCH) by comparing it with the hybrid k-factor GARMA-LLWNN model, in order to prove the usefulness of modeling the seasonal long memory in both conditional mean and conditional variance to improve the forecasting accuracy of the electricity price.

The log-return of electricity price for the Nord Pool market is used in this paper in order to show the appropriateness and effectiveness of the proposed model to time series forecasting. The remaining part of the study is organized as follows; Section II presents a brief review of the literature. Section III presents the econometric methodology; which includes the basic concepts of the k-factor GARMA model, the LLWNN model based BP learning algorithm, and illustrate the methodology describing the building of the hybrid k-factor GARMA-LLWNN model and the k-factor GARMA-G-GARCH model with the wavelet- based estimation procedure. Section IV deals with the empirical framework, where the predictive performance of the proposed k-factor GARMA-G-GARCH model is compared with the hybrid k-factor GARMA-LLWNN model, and Section V wrap up the conclusions.

II. LITERATURE REVIEWS

For the electricity market, an appropriate forecast model should consider the features related to the time series of the electricity. In this framework, important methods have been developed and applied for electricity price forecasting from different modeling families. Broadly two methods: the first approach is the statistical or econometric time series models, which considered as parametric tools, and the second approach is the soft computing models, which are considered as non-parametric tools. These two approaches are found to have been applied.

In statistical models, auto regressive integrated moving average ARIMA [11], has been used extensively. However, these models do not allow taking into account the long memory behavior characterizing the electricity prices. To overcome this limitation [5] and [6] introduced the Fractional Autoregressive Moving Average (FARMA) model. Recent works have applied these methods for the electricity prices [2], [12]. In the frequency domain, these models present a peak for very low frequencies near the zero frequency. Hence, it is noteworthy that ARFIMA processes do not allow to take into account the cyclical behavior or persistent periodic in the data. To overcome this insufficiency, [13] introduced a second generation of the long-memory model denoted a generalized long-memory or Gegenbauer Autoregressive Moving Average (GARMA) model that has been developed to model, simultaneously, the presence of persistence and seasonality in the data. Such a model displays a hyperbolic decay of the autocorrelation function at seasonal lags rather than the slow linear decay characterizing the seasonal differencing model (SARIMA). On the other hand, in the frequency domain, this model displays a spectral density that is not necessarily unbounded at the origin, as in the case of the ARFIMA model, but anywhere at a given frequency λ along the interval $[0, \pi]$. This frequency named Gegenbauer frequency or G-frequency. The GARMA model exhibits a long-memory periodic behavior at only one frequency λ, thus implying just one persistent cyclical component. Recently, [7] generalizes the single frequency GARMA model to the so-called k-factor GARMA model that allows the spectral density function to be not necessarily located at one frequency but associated with a

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:11, No:9, 2017

finite number k of frequencies in $[0, \pi]$, known as the Gegenbauer frequencies or G-frequencies. The main feature of this k-factor GARMA model is that it allows for more diversity in the covariance structure of a variable witnessed through both the autocorrelation function and the spectral density function, which presents k singularities.

The k-factor GARMA model applied by several authors to reproduce the seasonal patterns as well as the persistent effects in the stock markets [14]-[16] and Urban transport traffic in Paris area [17]. Despite the compatibility of this model with the characteristics of electricity prices, few applications are oriented in the electricity market [18]-[20].

However, the k-factor GARMA model exhibits two main shortcomings. The first is that it does not handle the nonlinear deterministic trend. [21] proves that omitting the deterministic trend in the fractional integration model leads to a serious bias and high variability in the fractional integration parameter. The second is that it focuses only on modeling the conditional mean, assuming hence that the residuals are white noise with constant variance over time. In practice, this latter assumption is not verified, and the residuals are rather characterized by a time-varying variance.

To reproduce these patterns, two approaches have been considered in the literature: the parametric models such as using GARCH or FIGARCH processes [22], [23], and the non-parametric methods like the neuronal nets for instance.

In the first category, [24] include the GARCH class of models proposed by [25], [26]. Hence, the obtained model called the k-factor GARMA-GARCH process, which allows for long-memory behaviour associated with k-frequency and include a GARCH-type model to describe time varying volatility. In addition, [27] propose the k-factor GARMA-FIGARCH to reproduce the long-range dependence behaviour in the conditional variance of the exchange rate. Recently, [14] proposed a new class of semiparametric generalized long-memory models with FIAPARCH errors that extends the k-factor GARMA model to include nonlinear deterministic trend and allows for time-varying volatility, in some MENA stock markets, using an estimation approach based wavelet theory.

Nevertheless, these models are not fully satisfactory when modelling volatility of intra-daily financial returns series. One important characteristic of such data is the strong evidence of cyclical patterns in the volatility of the series. In fact, the periodic pattern appears as a persistent cyclical behaviour on the autocorrelation functions with oscillations decaying very slowly. Some pronounced peaks at one or more non-zero frequencies in the periodogram are also observed. The empirical evidence so far accumulated emphasises the importance of taking into consideration the periodic dynamics of volatility for a correct modelling. In order to model the empirical evidences of seasonal long memory behaviour in the volatility, [28], [29] proposed new type of GARCH models characterised by periodic long memory behaviour. The suggested category of models introduces generalised periodic long-memory filters, based on Gegenbauer polynomials, into

the equation of standard GARCH model which describe the time-varying volatility. These models, called periodic long-memory GARCH (PLM-GARCH), and generalized long memory GARCH (G-GARCH), this model generalize the FIGARCH and FIEGARCH models, by introducing a reaches dynamics in the conditional variance. The filter used for G-GARCH specfications is the most general and allows the modelling of quite complex seasonal long memory behaviours. In the literature, the generalized long memory GARCH models (or G-GARCH) is used to estimate the financial time series such as the exchange rate by means of Monte Carlo simulations [28], [30]. Few studies apply this approach for the electricity spot price. To exemplify, [31] propose a new approach dealing with the stationary k-factor Gegenbauer process with Asymmetric Power GARCH noise under conditional Student-t distribution. This model called $GGk - APARCH$ model is used to model electricity spot prices coming from some European and American electricity markets. With reference of forecasting criteria, this model shows very good results compared with models using heteroscedastic asymmetric errors.

We can conclude that in the literature of generalized long memory models, the authors use either the k-factor GARMA model, or the G-GARCH model to estimate the conditional mean and the conditional variance of the time series, respectively. However, none of them considers the existing of the long rang periodic behaviour in both the conditional mean and the conditional variance. In this paper, in order to provide robust forecasts for spot electricity prices, we propose a new approach based on dual generalized long memory process, which allows taking into account many stylized facts observed on the electricity spot prices, in particular stochastic volatility, long memory and periodic behaviours.

Concerning the estimation of the parameter's k-frequency GARMA process, [7], [13] considered the time-domain maximum likelihood method. [10] proposed an alternative estimation method in the wavelet domain based on the maximal overlap discrete wavelet packet transform (DWPT). Compared to Fourier analysis, the strength of the wavelet approach lies in its ability to localize simultaneously a process in both time and frequency [32], [33].

In the second category, in order to grapple with the limitations of the parametric models and explain both the nonlinear patterns and time-varying variance that exist in real cases, several nonlinear, non-parametric models have been suggested. In this context, [9] proposed a local linear wavelet neural network (LLWNN). This model replaces the connection weights between the hidden layer units and output units by a local linear model. Thus, it requires only smaller wavelets for a given problem than the case of wavelet neural networks. In addition, this local capacity offers certain advantages such as the efficiency and transparency of the learning structure. The LLWNN model has been used for the electricity price forecasting, [34]-[36].

In fact, both the k-factor GARMA models, as a powerful statistical method, and the LLWNN model, as an advanced AI

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:11, No:9, 2017

method, have achieved successes in their own nonlinear parametric and nonparametric domains respectively. However, none of them is a universal model that is suitable for all circumstances. In other words, a time series is often complex in nature and a single model may not be able to capture the different patterns in the same way, so no method is the greatest for all situations. Thus, using hybrid models or combining several models has become a common practice in order to overcome the limitations of model's components and improve the forecasting accuracy. In general, the statistical methods help in dealing with voluminous datasets and neural network handle the non-linearity.

In the literature, different combination techniques have been proposed in order to overcome the deficiencies of single models. The basic idea of the model combination in forecasting is to use each model's unique feature in order to capture different patterns in the data. The difference between these combination techniques can be described using terminology developed for the classification and neural network literature [37]. Hybrid models can be homogeneous, such as using differently configured neural networks [38], [39]. It can be heterogeneous, with both linear and nonlinear models [40]. In a competitive architecture, the objective is to represent different patterns in the time series by building an appropriate module that including different parts of the time series, and to be able to switch control to the most appropriate [40], [41]. In a cooperative modular combination, the aim is to combine models to build a complete picture from a number of partial solutions [37]. The hybrid techniques that decompose a time series into its linear and nonlinear form are one of the most popular hybrid models categories, which have been shown to be successful for single models [42]-[46].

To sum up, in the literature, different combination technique has been proposed to overcome the limitation of single model by means of hybrid models, which uses the strength of different methods in order to model the different feature existing in the data, and thus enhance the forecasting results. However, the mentioned hybrid methods combine models that are not able to capture some features of the electricity time series, such as the seasonal long memory behavior, non-linearity, etc. In order to grapple with the limitation of such models, we combine in this study the k-factor GARMA model with the LLWNN in order to use each model's unique strength and thus, to capture different patterns in the electricity time series.

III. ECONOMETRIC METHODOLOGIES

A. The k-Factor GARMA Model

The k-frequency GARMA model, proposed by [7], generalizes the ARFIMA [5], [6] model allowing periodic or quasi-periodic movement in the data. The multiple frequency GARMA model is defined as follows;

$$\Phi(L)\prod_{i=1}^{k}\left(I-2\nu_{m,i}L+L^2\right)^{d_{m,i}}(y_t-\mu)=\Theta(L)\varepsilon_t \quad (1)$$

where $\Phi(L)$ and $\Theta(L)$ are the polynomials of the delay operator L such that all the roots of $\Phi(z)$ and $\Theta(z)$ lie outside the unit circle. The parameters $\nu_{m,i}$ provide information about periodic movement in the conditional mean (m), ε_t is a white noise perturbation sequence with variance σ_ε^2, k is a finite integer, $|\nu_{m,i}|<1$, $i=1,2,\ldots k$, $d_{m,i}$ are long memory parameters of the conditional mean indicating how slowly the autocorrelations are damped, μ is the mean of the process, $\lambda_{m,i}=\cos^{-1}(\nu_{m,i})$, $i=1,2,\ldots k$, denote the Gegenbauer frequencies (G-frequencies).

The GARMA model with k-frequency is stationary when $|\nu_{m,i}|<1$, and $d_{m,i}<1/2$ or when $|\nu_{m,i}|=1$, and $d_{m,i}<1/4$, the model exhibits a long memory when $d_{m,i}>0$. The main characteristic of model is given by the presence of the Gegenbauer polynomial:

$$P_m(L)=\prod_{i=0}^{k}\left(I-2\nu_{m,i}L-L^2\right)^{d_{m,j}} \quad (2)$$

This polynomial maybe considered a generalized long-memory filter that models the long-memory periodic behavior at $k+1$ different frequencies. When thinking of the $\lambda_{m,i}$ as the driving frequencies of a cyclical pattern of length S, and $k+1=[S/2]+1$, where $[.]$ stands for the integer part. To highlight the contribution of $P(L)$ at frequencies $\lambda_m=0$ and $\lambda_m=\pi$, equation (2) can be written as:

$$P_m(L)=(I-L)^{d_{m,0}}(I+L)^{d_{m,k}I(E)}\prod_{i=1}^{k+1}\left(I-2\nu_{m,i}L-L^2\right)^{d_{m,j}} \quad (3)$$

where $I(E)=1$ if S is even, and zero otherwise and $k+1=[S/2]+1-I(E)$. For a GARMA model with a single frequency, when $\nu_m=1$, the model is reduced to an ARFIMA (p, d, q) model, and when $\nu_m=1$ and $d=1/2$, the process is an ARIMA model. Finally, when $d_m=0$, we get a stationary ARMA model.

[47] determines the spectral density function and shows that for $d_m>0$ the spectral density function has a pole at $\lambda_m=\cos^{-1}(\nu_m)$, which varies in the interval $[0,\pi]$. It is important to note that when $|\nu_m|<1$, the spectral density function is bounded at the origin for GARMA processes, and thus does not suffer from many problems associated with ARFIMA models.

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:11, No:9, 2017

B. A Local Linear Wavelet Neural Network (LLWNN) Model

[9] propose a local linear wavelet neural network (LLWNN) model for time series forecasting, and they have shown that this model has more accuracy than the traditional WNN. In local linear wavelet neural network (LLWNN) the number of neurons in the hidden layer is equal to the number of inputs and the connection weights between the hidden layer units and output units are replaced by a local linear model. Remind that in the wavelet transform theory the wavelets are in the following form:

$$\varphi = \left\{ \varphi_i = |a_i|^{-1/2} \varphi(\frac{x - b_i}{a_i}) : a_i, b_i \in R^n, i \in Z \right\} \quad (4)$$

$$
\begin{aligned}
x &= (x_1, x_2, \ldots, x_n), \\
a_i &= (a_{i1}, a_{i2}, \ldots a_{in}), \\
b_i &= (b_{i1}, b_{i2}, \ldots b_{in}).
\end{aligned}
\quad (5)
$$

Families of functions that have been generated by a unique function $\varphi(x)$ by the operations of dilation and translation of $\varphi(x)$; which is located in both time space and frequency space, called a mother wavelet and the parameters a_i and b_i are named scale and translation parameters, respectively. The set x represents the inputs to the WNN model. In the standard form of WNN, the output of a WNN is given by:

$$f(x) = \sum_{i=1}^{M} \omega_i \varphi_i(x) = \sum_{i=1}^{M} \omega_i |a_i|^{-1/2} \varphi(\frac{x - b_i}{a_i}) \quad (6)$$

where φ_i, is the activation function of wavelets of the ith unit of the hidden layer and ω_i is the weight connection of the ith unit of the hidden layer to the output layer unit. Note that for the *n*-dimensional input space, the basis function of the multivariate wavelet can be computed by the product of *n* unique wavelet basis functions as follows:

$$\varphi(x) = \prod_{i=1}^{n} \varphi(x_i) \quad (7)$$

Obviously, the location of the ith units of the hidden layer has been determined by the scale parameter a_i and the translation parameter b_i: According to previous research, the two parameters can either be predetermined by the basis of the Wavelet transformation theory or be determined by a learning algorithm. Note that WNN is a kind of neural network in the sense that wavelets represent the basic functions. One limitation of the WNN is that for problems of large dimensions many units of hidden layers are needed. In order to take advantage of the local capacity of the basic wavelet functions

while minimizing the number of hidden units, here we propose another type of WNN.

In the locally linear wavelet neural network (LLWNN) the number of neurons in the hidden layer is equal to the number of inputs and the connection weights between the hidden layer units and the output units are replaced by a local linear model.

In the literature of LLWNN it is known that the local linear model provides a more parsimonious interpolation in a large dimension space and thus provides it with the ability of time series prediction. This local capacity of the LLWNN model offers some advantages such as the efficiency and transparency of the learning structure.

The architecture of the proposed LLWNN is shown in the Fig. 1, and its output from the output layer is given by:

$$
\begin{aligned}
y &= \sum_{i=1}^{M} (\omega_{i0} + \omega_{i1} x_1 + \ldots + \omega_{in} x_n) \varphi_i(x) \\
&= \sum_{i=1}^{M} (\omega_{i0} + \omega_{i1} x_1 + \ldots + \omega_{in} x_n) |a_i|^{-1/2} \varphi(\frac{x - b_i}{a_i})
\end{aligned}
\quad (8)
$$

where $x = [x_1, x_2, \ldots, x_n]$, instead the simple weight ω_i (locally constant model), a linear model has introduced:

$$\upsilon_i = \omega_{i0} + \omega_{i1} x_1 + \ldots + \omega_{in} x_n \quad (9)$$

The linear models $\upsilon_i (i = 1, 2, \ldots, M)$ are determined by the associated locally active wavelet functions $\varphi_i(x)$ and ($i = 1, 2, \ldots, M$), thus υ_i is locally significant. The motivations of introducing local linear models in WNN are as follows: (1) local linear models were studied in some neuro-fuzzy systems and showed good performance, and (2) local linear models could provide a more parsimonious interpolation in large spaces when modeling samples are dispersed. The scale and translation parameters of the locally linear model are randomly initialized at the beginning and are optimized by a learning algorithm. In this paper, we adopt the Back Propagation (BP) algorithm in order to optimize the parameters of the LLWN model.

C. The Back Propagation-Learning Algorithm for Training the LLWNN Model

The backward propagation of errors or back propagation, is classified as a supervised learning algorithm commonly used in training neural networks in general, in combination with an optimization method such as gradient descent.

The backpropagation algorithm minimizes the objective function by adjusting the connection weights employed to develop the models. The gradient of the cost function is calculated according to that particular weight parameter, and then the parameters are updated by means of the negative gradient. The learning rate is a fundamental factor in the backpropagation algorithm. The network learns very slowly if the learning rate is too low, contrariwise, the weights and the

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:11, No:9, 2017

objective function will diverge if this rate is too high. Therefore, an optimum value must be selected to guarantee global convergence, which tends to be a difficult task to achieve. If there are several local and global optima for the objective function, a variable learning rate will do better [48].

Since the backpropagation requires a known and desired output for each value of input in order to determine the loss function gradient, for that reason it is usually considered as a supervised learning method.
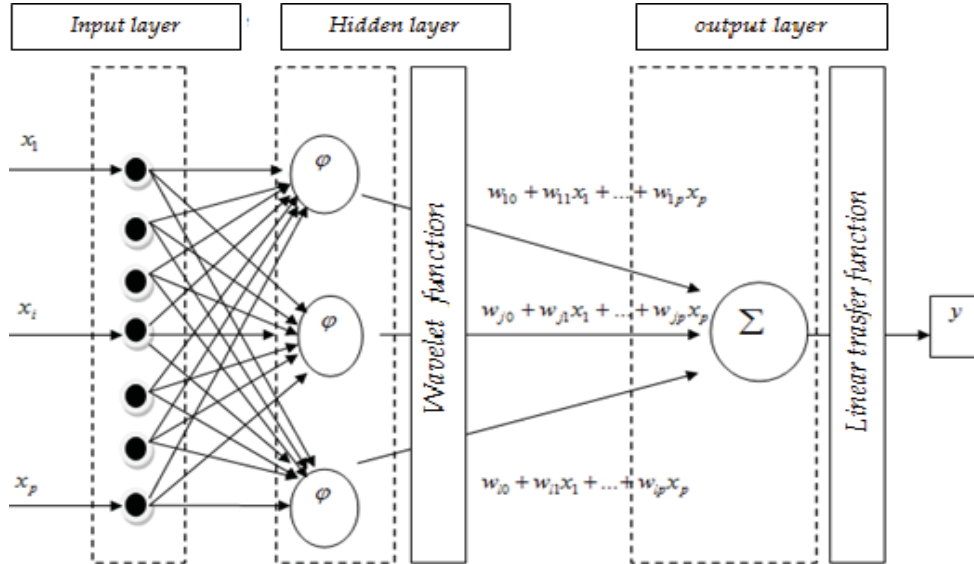

Fig. 1 Local Linear Neural Wavelet Network Architecture

The equations of the backpropagation algorithm are described briefly below. The objective function to minimize is given as:

$$E = \frac{1}{2}[y_t - \omega_{1,0}\varphi_1(x) - \omega_{1,1}p_1\varphi_1(x) - \ldots - \omega_{2,0}\varphi_2(x)\omega_{2,1}p_2\varphi_2(x) - \ldots \ldots \omega_{l,0}\varphi_l(x)\omega_{l,1}p_1\varphi_l(x) - \ldots \omega_{l,p}p_p\varphi_l(x)] \quad (10)$$

where y_t is the desired value, $\varphi(x)$ is the active wavelet function, $\omega_{1,0}$ denote the connection weight, p is the number of input $(i = 1,2,\ldots p)$ and l is the number of the hidden units $(j = 1,2,\ldots l)$. The weight is updated from i^{th} to the $(i+1)^{th}$ iteration, that is from ω_t to ω_{t+1} is given by:

$$\omega_{t+1} = \omega_t + \Delta\omega_t = \omega_t + \left(r\frac{\partial E_t}{\partial \omega_t}\right), \quad (11)$$

Denote that r is the learning rate adopted in the LLWNN model where $\dfrac{\partial E}{\partial \omega}$ for all weights are described by the following equations:

$$\frac{\partial E}{\partial \omega_{i,0}} = \omega_{i,0} + r * e * \left(\frac{1}{2}\right) * \left(x_1^2 + x_2^2 + \ldots + x_p^2\right) \quad (12)$$
$$* \exp(-((x_1 - c_i)^2 + (x_2 - c_i)^2 + \ldots + (x_p - c_i)^2))$$

$$\forall j \neq 0;$$
$$\frac{\partial E}{\partial \omega_{i,j}} = \omega_{i,j} + r * e * \left(\frac{1}{2}\right) * \left(x_1^2 + x_2^2 + \ldots + x_n^2\right) \quad (13)$$
$$* \exp(-((x_1 - c_i)^2 + (x_2 - c_i)^2 + \ldots + (x_n - c_i)^2)) * x_j$$

$$\frac{\partial E}{\partial \omega_{1,0}} = \omega_{1,0} + r * e * \left(\frac{1}{2}\right) * \left(x_1^2 + x_2^2 + \ldots + x_p^2\right) \quad (14)$$
$$* \exp(-((x_1 - c_i)^2 + (x_2 - c_i)^2 + \ldots + (x_p - c_i)^2))$$

$$\frac{\partial E}{\partial \omega_{1,2}} = \omega_{1,2} + r * e * \left(\frac{1}{2}\right) * \left(x_1^2 + x_2^2 + \ldots + x_p^2\right) \quad (15)$$
$$* \exp(-((x_1 - c_i)^2 + (x_2 - c_i)^2 + \ldots + (x_p - c_i)^2)) * x_2$$

$$\frac{\partial E}{\partial \omega_{2,0}} = \omega_{2,0} + r * e * \left(\frac{1}{2}\right) * \left(x_1^2 + x_2^2 + \ldots + x_p^2\right) \quad (16)$$
$$* \exp(-((x_1 - c_i)^2 + (x_2 - c_i)^2 + \ldots + (x_p - c_i)^2))$$

$$\frac{\partial E}{\partial \omega_{2,1}} = \omega_{2,1} + r * e * \left(\frac{1}{2}\right) * \left(x_1^2 + x_2^2 + \ldots + x_p^2\right) \quad (17)$$
$$* \exp(-((x_1 - c_i)^2 + (x_2 - c_i)^2 + \ldots + (x_p - c_i)^2)) * x_1$$

The other weights are also updated in the same way (Fig. 2).

D. Hybrid k-Factor GARMA-LLWNN Model

Both theoretical and empirical results suggested that the combination of different models can be an effective tool to improve the predictive performance of each model, especially when the combined models are very different. In the literature,

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:11, No:9, 2017

several combination techniques have been presented, such as the traditional hybrid model of [42], [43], [49]. These methods use the ARIMA model as a linear model and the multi-layer perceptron neural network to model the non-linear component. Although, these methods have shown, empirically, their effectiveness, insofar as they can be an effective way of improving the accuracy of predictions made by one of the methods used separately, these models are critical.

Indeed, these methods exploited ARIMA modeling to predict the linear trend in the data, but predictions using the $ARIMA(p,d,q)$ model have not always proved to be very effective. The main criticism mainly concerns the modeling of short-term relationships only (short memory), while ignoring the seasonal effect and the presence of long memory that characterizes most financial and economic series.
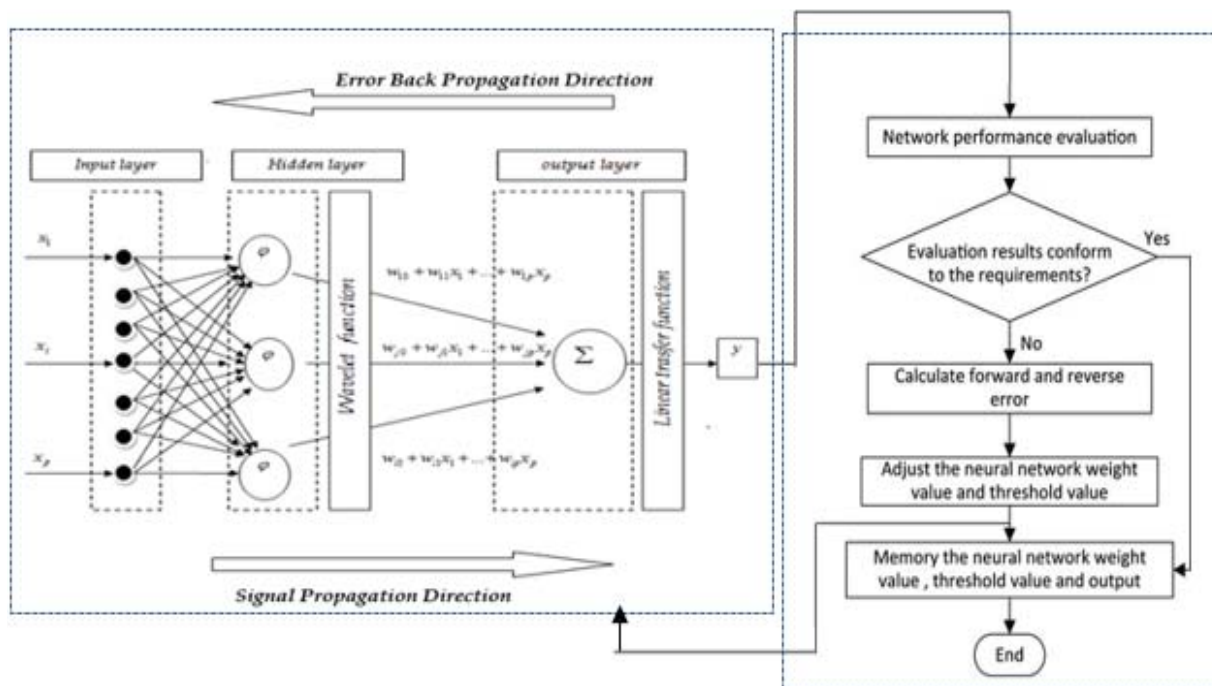

Fig. 2 Local Linear Wavelet Neural Network Model based Back Propagation Algorithm

To overcome this limitation, the k-factor GARMA model offers greater flexibility in modeling simultaneous short and long-term behavior of a seasonal time series.

On the other hand, the hybrid methods existing in the literature neglected the modeling of volatility, a phenomenon that characterizes most financial series. In reality, a good forecast must take into account the time varying variance. Thus, the LLWNN approach has been proposed to take into account the time varying of conditional variance. The choice of LLWNN in our hybrid model is motivated by the wavelet decomposition and its local linear modeling ability.

Furthermore, the previous hybrid models assume that the nonlinear relations are only exist in the residuals and the two components (linear and nonlinear) must be modeled separately, they assume that there are no nonlinear relations in the averages since they are always estimated using a linear model in the first step.

To overcome this limitation, we use the k-factor GARMA model to estimate the nonlinear components in the data, and then we model the residuals using an LLWNN model to predict the volatility. In other words, the first step consists in modeling the conditional mean using a non-linear parametric model (k-factor GARMA). However, residuals are important

in forecasting time series; they may contain some information that is able to improve forecasting performance. Thus, in the second step, the residuals resulting from the first step will be treated according to a local linear wavelet neural network LLWNN.

In our hybrid method, we have combined two models that have different characteristics in order to model the different features existing in the data, thus we adopt a combination of parametric and nonparametric models.

In conclusion, the proposed hybrid model exploits the originality and the strength of the k-factor GARMA model as well as the LLWNN model to detect the different features existing in the data to benefit the complementary characteristics of the models, which compose them. Thus, the proposed hybrid model can be an efficient way giving a more general and accurate model than other hybrid models.

E. The k-Factor GARMA-G-GARCH Model

The k-factor GARMA model is a powerful method in modeling the conditional mean of the time series. Nevertheless, this model assumes that the conditional variance is constant over time. In the empirical studies, it is well recognized that many time series often exhibit volatility

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:11, No:9, 2017

clustering, where time series exhibit both high and low periods of volatility. To reproduce these patterns, we extended the k-factor GARMA described above by inserting a fractional filter in the conditional variance equation. For this reason, we propose the k-factor GARMA-G-GARCH model that is able to capture both seasonality and long memory dependence in the conditional mean and in the conditional variance.

Recently, [28], [29] introduced new GARCH-type models characterized by long memory and seasonality behavior, in order to model the empirical evidences of long memory behavior in the volatility of intra-daily financial returns. These models, named generalized long memory GARCH or Gegenbauer-GARCH (G-GARCH), generalize the FIGARCH and FIEGARCH models by introducing a seasonal long memory in the conditional variance.

In fact, it is also taken into account the periodic long memory patterns in the conditional variances, associated to zero and non-zero frequencies of the power spectrum. Furthermore, to overcome the constraints of non-linear coefficients for the positivity of the variance, the authors propose to model the log-conditional variances. Therefore, G-GARCH nests some traditional specifications of long memory GARCH when adjusted to modeling log-variances.

The fundamental idea of this model is to include the generalized long-memory process into the equation describing the evolution of conditional variance in a GARCH framework. This is why this new class of models is called Gegenbauer-GARCH (G-GARCH). Thus, we consider the following k-factor GARMA process with G-GARCH type innovations to take into account the presence of a time varying conditional variance:

$$y_t = \mu_t + \varepsilon_t = \mu_t + \sigma_t z_t \qquad (18)$$

where μ_t is the conditional mean of y_t modeling using the following k-factor GARMA model:

$$\Phi(L)\prod_{i=1}^{k}(I - 2\nu_{m,i}L + L^2)^{d_{m,i}}(y_t - \mu) = \Theta(L)\varepsilon_t \quad (19)$$

$$\varepsilon_t / I_{t-1} \sim N(0, \sigma_t^2) \qquad (20)$$

where σ_t^2 is the conditional variance, I_{t-1} being the information up to time $t-1$, z_t is an $i.i.d$ random variable with zero mean and unitary variance.

To specify the dynamics of the conditional variance, the starting point is the dynamics of ε_t^2 We assume that ε_t^2 follow a k-factor GARMA model, which describes a cyclical pattern of length S:

$$\left[(I-L)^{d_{v0}}(I+L)^{d_{v,k}I(E)}\prod_{i=1}^{k-1}(I-2\nu_{v,i}L+L^2)^{d_{v,i}}\right]\psi(L)\varepsilon_t^2$$
$$= \gamma + [I - \beta(L)]\vartheta_t \qquad (21)$$

$$P_v(L)\psi(L)\varepsilon_t^2 = \gamma + [I - \beta(L)]\vartheta_t \qquad (22)$$

where $\psi(L) = 1 - \sum_{i=1}^{q}\psi_i L^i$ and $\beta(L) = 1 - \sum_{i=1}^{p}\beta_i L^i$ are suitable polynomials in the lag operator L and $\vartheta_t = \varepsilon_t^2 - \sigma_t^2$ is a martingale difference, $d_{v,0} = d_v / 2$, $I(E) = 1$ if S is even and zero otherwise. With this assumption, the corresponding GARCH-type dynamics for conditional variance is given by;

$$\sigma_t^2 = \gamma + \beta(L)\sigma_t^2 + \{I - \beta(L) - [(I-L)^{d_{v,0}}(I+L)^{d_{v,k}I(E)}$$
$$\prod_{i=1}^{k-1}(I-2\nu_{v,i}L+L^2)^{d_{v,i}}]\psi(L)\}\varepsilon_t^2. \qquad (23)$$

This implies that in the G-GARCH framework of each frequency has been modeled by means of a specific long-memory parameter $d_{v,i}$ (differencing parameter of the conditional variance). When $d_{v,0} = d_{v,1} = \ldots = d_{v,k}$, all the involved frequencies have the same degree of memory.

Model (23) may provide, in particular cases, most of the existing GARCH models. For example, standard GARCH models (included seasonal GARCH [50] can be obtained by putting $d_{v,i} = 0$, $i = 0 \ldots k$. Similarly, the FIGARCH model is equivalent to $S = 1$ and $0 < d_{v,0} < 1$.

It is interesting to mention that generalized long-memory filters, in principle, may be applied to any category of GARCH structure. Nonetheless, due to the constraints needed for conditional variance positivity, G-GARCH models are not always feasible, for this reason, [28] proposed to model the logarithm of the conditional variances. Therefore, a practical computing solution is to apply the filter to a generalized log-GARCH model. This means beginning from the expression:

$$P_v(L)\psi(L)[\ln(\varepsilon_t^2) - \tau] = \gamma + [I - \beta(L)]\vartheta_t \qquad (24)$$

where $P_v(L)$ is the generalized long memory filter introduced into a GARCH structure, $\vartheta_t = \ln(\varepsilon_t^2) - \tau - \ln(\sigma_t^2)$ is a martingale difference and $\tau = E[(\ln(z_t^2))]$. The expected τ value depends on the distribution of the idiosyncratic shock and ensures that ϑ_t is a martingale difference, given that $\ln(\varepsilon_t^2) = \ln(\sigma_t^2) + \ln(z_t^2)$. Under the Gaussian assumption $\tau = -1.27$. The expression for conditional variance implied by (23).

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:11, No:9, 2017

$$\ln(\sigma_t^2) = \gamma + \beta(L)\ln(\sigma_t^2) + \left[I - \beta(L) - P_v(L)\psi(L)\right] \quad (25)$$
$$* \left[\ln(\varepsilon_t^2) - \tau\right]$$

Since we are modeling $\ln(\sigma_t^2)$ instead of σ_t^2, no constraints for variance positivity are necessary. A further approach of bypassing the problem of parameter constraints is to adopt EGARCH versions of our model. The proposed model lies between the FIGARCH and the FIEGARCH representations. For the last model, the Stationarity of the covariance implies a memory parameter $0 \leq d_v < 0.5$. Differently, the covariance Stationarity of FIGARCH model is not obvious and is discussed in [51]-[53], among others. The uncertainty on covariance stationarity existence extends to our model which, despite the dynamic formulation of the log-variance, is closer to the FIGARCH case rather than to the exponential model.

F. Wavelet-Based Estimation Procedure

Concerning the estimation of the k-factor GARMA-G-GARCH model, we adopt an estimation procedure based on the wavelets following the methodology proposed by [10].

The advantage of this approach lies in its ability to simultaneously locate a process in both time and scale. At high frequencies, the wavelet has small centralized low scales, which allow it to focus on short-term phenomena. At low frequencies, the wavelet has a large time supports allowing it to identify long-term behavior. By moving from low to high frequencies, wavelets identify jumps and peaks [32], [33].

The discrete wavelet packet transform (DWPT) [10] is a generalization of the discrete wavelet transform DWT that splits the whole frequency band $[0, 1/2]$ into individual and regularly spaced intervals.

For a given temporal series X of dyadic length $N = 2^J$, the jth level of DWPT is an orthonormal transform giving a vector of dimension N of wavelet packet coefficients $(W_{j,2^{j-1}}, W_{j,2^{j-2}}, \dots W_{j,0})'$ where each $W_{j,n}, n = 0, \dots 2^j - 1$, has

$N_j = N/2^j$ dimension and it is associated with the frequency interval $\kappa_{j,n} = \left[n/2^{j+1}, (n+1)/2^{j+1}\right]$.

Let $\{h_l\}_{l=0}^{L-1}$ and $\{g_l\}_{l=0}^{L-1}$ be the Daubechies wavelet and scaling filters respectively. Starting with the recursion $X = W_{0,0}$ were the tth elements of $W_{j,n}$ have calculated by the following steps of filtering:

$$W_{j,n,t} = \sum u_{n,l} W_{j-1,[n/2],(2t+1-l) \bmod N_{J-10}}, \quad (26)$$
$$t = 0, \dots, N_j - 1$$

$$u_{n,l} = \begin{cases} g_l & \text{if } n \bmod 4 = 0 \text{ ou } 3 \\ h_l & \text{if } n \bmod 4 = 1 \text{ ou } 2 \end{cases} \quad (27)$$

It is notable that the collection of doublets (j, n) (also called nodes) is known as a wavelet packet tree and will be denoted by $T = \{(j,n): \ j = 0 \dots J; n = 0 \dots 2^J - 1\}$. An orthonormal basis $B \subset T$ is obtained when a collection of DWPT coefficient vectors yield a disjoint and there is no overlapping complete covering of the frequency range $[0, 1/2]$ called a "disjoint dyadic decomposition". Hence, in matrix notation, a vector of DWPT coefficients has obtained via $w_B = W_B X$ where W_B an orthonormal $N \times N$ matrix is defining the DWPT through the basis B.

In this paper, to identify the "best base" B, from all possible orthonormal partitions, we use statistical white noise tests named the portmanteau test, following the method of [24]. A Similarly to the MODWPT, the step of the down sampling relative to the DWPT can be also removed by means of a variant of this transform identified as the MODWPT that depend on rescaled versions of the filter $u_{n,l}$ presented above.
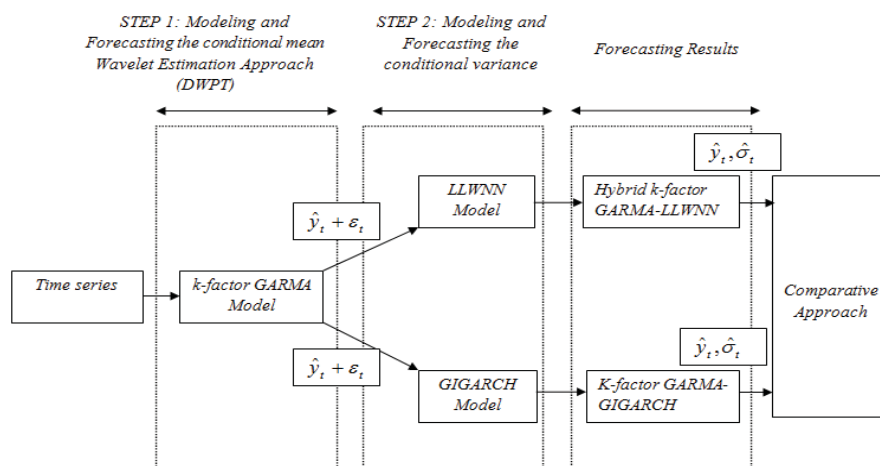

Fig. 3 A schematic representation of the adopted econometric methodology

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:11, No:9, 2017

IV. EMPIRICAL RESULTS

A. The Nord Pool Electricity Market: Data description and Preliminary study

The Nordic electricity market known as Nord Pool market is a stock market affected to the electricity product. This stock market was created in 1992. It includes the three Scandinavian countries: Norway, Sweden and Denmark plus Finland. The spot price is the equilibrium price, and is calculated as the equilibrium point for every 24 hours. It is the unique price throughout the north region, and is determined when the supply and demand curves intersect.

The methodology proposed in this research is tested on hourly spot prices on the Nord Pool electricity market, covering the period between the 1st of January 2015 and 31st of December 2015, in total $N = 8761$ hourly observations, illustrated in Fig. 4. This data was extracted from the official website of Nord Pool market. In this section, we analyze the spot price electricity series on the Nord Pool market, in order to study their statistical and econometric features. In most cases, the econometricians consider the logarithm of their series because the use of the difference logarithm sometimes makes the series stationary and allows modeling returns series. For this reason, we use the series of log-returns spot price (noted RSP).
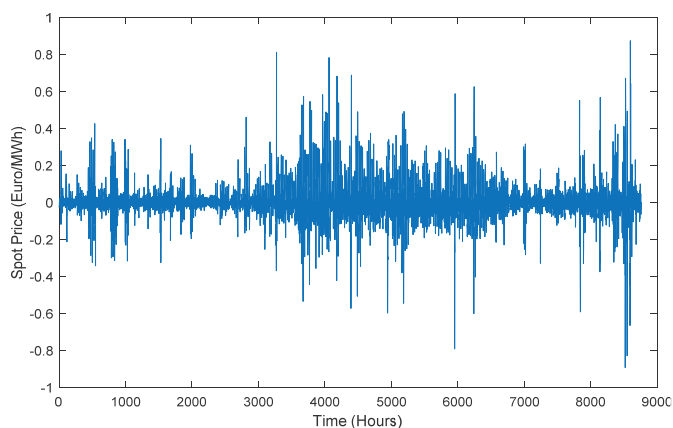

Fig. 4 Hourly spot price for NordPool electricity market

Fig. 4 shows the time evolution of log-return spot price (RSP) indicates that this series seems stationary. This hypothesis can also be supported by the unit root tests (ADF, PP and KPSS). In addition, the series presents a clustering of volatility since periods of low volatility are followed by periods of high volatility. This is a sign of the presence of the ARCH effect in the series.

As shown in Fig. 5, for the return spot price electricity series, the spectral density, traced by the periodogram, shows peaks at equidistant frequencies, which proves the presence of several seasonalities in the conditional mean.

The summary descriptive statistics of the Nord Pool log-returns are reported in Table I. The standard deviation is quite small, while the estimated measure of Skewness indicating a non-symmetric distribution. Moreover, the large value of the

kurtosis statistic suggests that the underlying data are leptokurtic. This significant departure from normality is also confirmed by the large value of the Jarque-Bera (JB) test. Hence, the electricity spot price series (RSP) is not normally distributed.
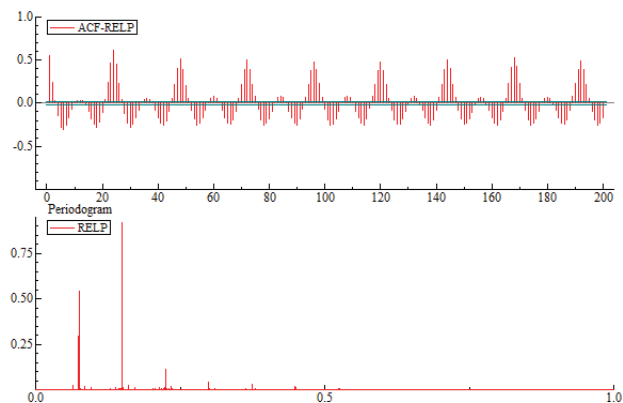

Fig. 5 Nord Pool ACF & Periodogram

TABLE I
DESCRIPTIVE STATISTICS OF THE SPOT PRICES TIME SERIES (LOG-RETURNS)

The Nord Pool log-returns	
Mean	$-6.24 \cdot 10^{-5}$
Standard deviation	0.0835
Skewness	0.8838*** (0.000)
Kurtosis	20.285*** (0.000)
Jarque-Bera	110197 (0.000) ***

Note: levels of significance indicated between squared brackets. *** denotes significance at 1% level.

TABLE II
ADF, PP, KPSS UNIT ROOT TESTING RESULTS FOR NORD POOL LOG-RETURNS

	Model (3) With an intercept and a trend	Model (2) With an intercept	Model (1) Without an intercept
ADF Test	-50.2201*** (0.000)	-50.2230*** (0.000)	-101.3755*** (0.000)
PP Test	-101.8959*** (0.000)	-101.8214*** (0.000)	-
KPSS Test	0.0363	0.04104	-

Note: *** indicate rejection of the null hypothesis at the 1-percent level. ADF and PP: Critical values in the model (3): -3.95 (1%), -3.41 (5%), -3.12 (10%). Critical values in the model (2): -3.43 (1%), -2.86 (5%), -2.56 (10%). Critical values in the model (1): -2.56 (1%), -1.94 (5%), -1.62 (10%). KPSS: Critical values in the model (3): 0.216 (1%), 0.146 (5%), 0.119 (10%). Critical values in the model (2): -0.739 (1%), -0.463 (5%), -0.347 (10%)

We tested for stationarity by performing unit root tests, namely, the augmented Dickey-Fuller (ADF), the Phillips-Perron (PP) and Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) tests, to the Nord Pool log-returns time series. The results reported in Table II indicate that the Nord Pool log-returns time series is significant to reject the hypothesis of non-stationarity. Thus, this series is stationary and suitable for subsequent tests in this study. Using the GPH [54] and LW [55] statistics, we test for the long-range dependence presence

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:11, No:9, 2017

in the conditional mean. Corresponding results shown in Table III indicate evidence of long memory.

TABLE III
RESULTS OF GPH AND LW LONG- RANGE DEPENDENCE TESTS IN THE CONDITIONAL MEAN

	Bandwidth	GPH			LW		
		$\hat{d}$	Standard error	p-value	$\hat{d}$	Standard error	p-value
RSP T=8760	$T^{0.5}=94$	-0.2347	0.0718	0.001	-0.236	0.0515	0.000
	$T^{0.6}=232$	-0.3632	0.0440	0.000	-0.417	0.0328	0.000
	$T^{0.7}=575$	-0.3389	0.0273	0.000	-0.328	0.0208	0.000
	$T^{0.8}=1425$	-0.3810	0.0172	0.000	-0.594	0.0132	0.000

B. The Estimation Results

TABLE IV
ESTIMATION OF THE k-FACTOR GARMA MODEL: A WAVELET BASED APPROACH

Parameters	k-factor GARMA model estimation
$\hat{\Phi}$	0.0357*** (0.000)
$\hat{\Theta}$	-
$\hat{d}_{m,1}$	0.2657*** (0.000)
$\hat{d}_{m,2}$	0.1238*** (0.000)
$\hat{d}_{m,3}$	0.0873*** (0.000)
$\hat{\lambda}_{m,1}$	0.1295*** (0.000)
$\hat{\lambda}_{m,2}$	0.0882 *** (0.000)
$\hat{\lambda}_{m,3}$	0.0479*** (0.000)

Note: *** indicate rejection of the null hypothesis at the 1-percent level.

The seasonality can be easily observed in the frequency domain $\lambda_{m,i}=1/T$; where λ_m is the frequency of the seasonality and T is the period of seasonality. As shown the spectral densities, represented by periodogram (figure 5), are unbounded at equidistant frequencies, which proves the presence of several seasonalities. They show special peaks at frequencies $\hat{\lambda}_{m,1} = 0.1295$ (T= 7.72 ≈ 8 hours $\approx 1/3$ day), $\hat{\lambda}_{m,2} = 0.0882$ (T=11.5 ≈ 12 hours $\approx 1/2$ day), and $\hat{\lambda}_{m,3} = 0.0479$ (T=20.87 hours ≈ 1 day), as shown in Table IV, that corresponding to cycles with third daily, semi-daily and daily periods, respectively.

In the second step, the squared log-returns are used as a proxy of the corresponding volatility. Long memory tests are performed for the resulted time series. As reported in Table V, the results of the GPH and LW indicate the presence of long memory in the conditional variance.

TABLE V
RESULTS OF GPH AND LW LONG RANGE DEPENDENCE TESTS IN THE CONDITIONAL VARIANCE

	Bandwidth	GPH			LW		
		$\hat{\xi}$	Standard error	p-value	$\hat{\xi}$	Standard error	p-value
RSP T=8760	$T^{0.5}=94$	-0.0701	0.0718	0.3293	-0.1164	0.0515	0.0240
	$T^{0.6}=232$	-0.9439	0.0440	0.0000	-0.7501	0.0328	0.0000
	$T^{0.7}=575$	-1.0848	0.02737	0.0000	-0.7834	0.02085	0.0000
	$T^{0.8}=1425$	0.5481	0.01723	0.0000	1.4457	0.01324	0.0000

As shown in Fig. 6, the periodogram of the k-factor GARMA residuals, the spectral density is unbounded at equidistant frequencies that indicates the presence of several seasonalities. Which requires the use of some seasonal long memory G-GARCH method to model such processes.

The residuals from the k-factor GARMA are modelled using two different approaches in order to preserve the best estimation model. In the first approach, we adopt the G-GARCH model in order to estimate the seasonal long memory behaviour in the conditional variance. The estimation results reported in Table VI. The spectral densities, represented by periodogram (Fig. 6), are unbounded at equidistant frequencies, which proves presence of several seasonalities.

They show special peaks at frequencies $\hat{\lambda}_{v,1} = 0.0178$ (T=56h, 17min ≈ 2days) and $\hat{\lambda}_{v,2} = 0.0416$ (T=24h.038min ≈ 1 day), as shown in Table VI, that corresponding to cycles with two day and daily periods, respectively.

In the second approach, residuals resulting from the k-factor GARMA modeling are considered as the input of the LLWNN, and shaped through the network in order to estimate the conditional variance. For the purpose of avoiding the possibility of coupling among different input and to accelerate convergence, all the inputs are normalized within a range of $[0, 1]$ using the most commonly used data smoothing method before applying it to the network.

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:11, No:9, 2017

$$y_{norm} = \frac{y_{org} - y_{min}}{y_{max} - y_{min}} \qquad (28)$$

where y_{norm} is the normalized value, y_{org} is the original value, y_{min} and y_{max} are the minimum and maximum values of the corresponding residuals data.
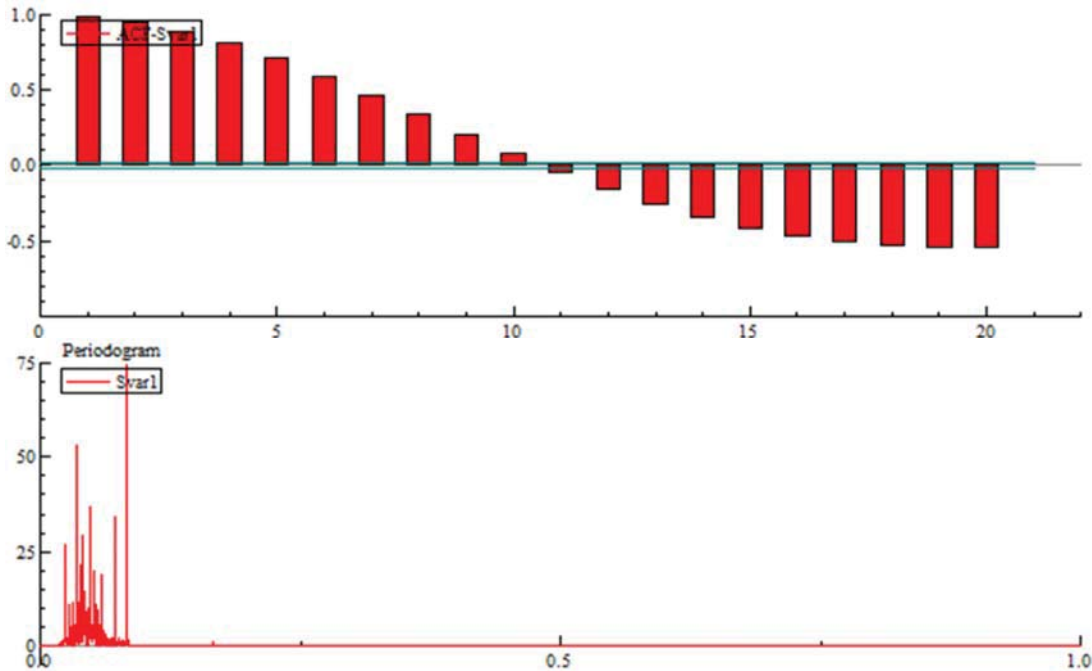

Fig. 6 Residuals ACF & Periodogram

TABLE VI
ESTIMATION RESULTS OF THE G-GARCH MODEL

Parameters	G-GARCH model estimation	
$\hat{\psi}$	0.1341***	
$\hat{\beta}$	0.1652***	
$\hat{\gamma}$	0.0004***	
$\hat{d}_{v,1}$	0.2492***	
$\hat{d}_{v,2}$	0.1295***	
$\hat{\lambda}_{v,1}$	0.0178***	(56h,17min $\approx$ 2days)
$\hat{\lambda}_{v,2}$	0.0416***	(24h.038min $\approx$ 1 day)

Note: *** indicate rejection of the null hypothesis at the 1-percent level.

The datasets is divided into three successive parts as follows: (a) A sample of 200 observations to initialize the network training, (b) a training set and (c) a test set the forecasting experiment is performed over the test set using an iterative forecasting scheme, the model are forecasting for 6, 12, 24, 48 and 72 hours ahead. Details of the datasets are given in the Fig. 7.

In order to find the best neural network architecture, at the beginning the parameters are randomly initialized. Thereafter, using the back-propagation algorithm (BP) these parameters are updated in order to minimize the error between the output values and the real values during the training of the network. Table VII provides the summary of information related to the network architecture.
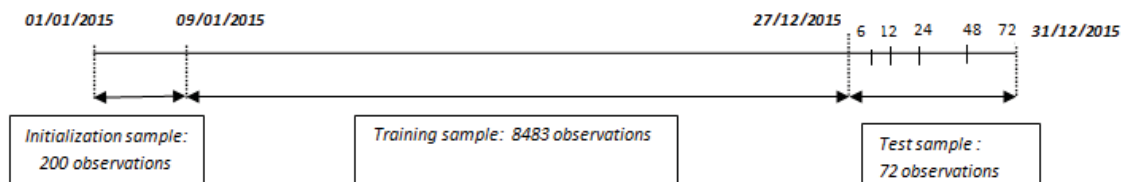

Fig. 7 Details of datasets

C. Forecasting Results: A Comparative Approach

This section is devoted to the evaluation of the estimated models in a multi-step-ahead forecasting task. Since forecasting is basically an out-of-sample problem, we prefer to apply out-of-sample criteria. Accordingly, five different periods (6 hours, 12 hours, one day, two days and tree days, ahead forecasting) were selected in aim to insure the quality

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:11, No:9, 2017

and the robustness of modeling and forecasting results. In order to evaluate the forecasting accuracy, we apply three evaluation criteria, namely, the out-of-sample R^2 of [56], the mean absolute percentage error (MAPE) and the logarithmic loss function (LL). given respectively by:

$$R^2 = 1 - \left(\frac{\sum_{t=t_1}^{N}(y_{t+h} - \hat{y}_{t,t+h})^2}{\sum_{t=t_1}^{N}(y_{t+h} - \overline{y}_{t,t+h})^2} \right), \qquad (29)$$

$$MAPE = \frac{1}{N - t_1} \sum_{t=t_1}^{N} \left| \frac{(y_{t+h} - \hat{y}_{t,t+h})}{y_{t+h}} \right| * 100, \qquad (30)$$

$$LL = \frac{1}{N - t_1} \sum_{t=t_1}^{N} \left(Log\left(\frac{\hat{y}_{t,t+h}}{y_{t+h}} \right) \right)^2. \qquad (31)$$

where N is the number of observations, $N - t_1$ is the number of observations for predictive performance, y_{t+h} is the log-return series through period $t + h$, $\hat{y}_{t,t+h}$ is the predictive log-return series of the predictive horizon h at time t and $\overline{y}_{t,t+h}$ is the historical average forecast.

TABLE VII
LLWNN BASED BP ALGORITHM ARCHITECTURE

Number of hidden layer	10
Learning rate	0.5
Layer Activation function	Wavelet Function
Learning Algorithm	Back Propagation (BP) Learning Algorithm

TABLE VII
OUT OF SAMPLE FORECASTS RESULTS

Models	Criterion	$h = 6$	$h = 12$	$h = 24$	$h = 48$	$h = 72$
k-factor GARMA-LLWNN model	R²	0,9428	0.6040	0.9977	0.9528	0.9889
	MAPE	0.9178%	4.4437%	0.5719%	2.0084%	0.9902%
	LL	1.122×10^{-4}	0.0021	5.734×10^{-5}	5.824×10^{-4}	1.556×10^{-4}
The k-factor GARMA-G-GARCH model	R²	0.9993	0.9998	0.9997	0.9976	0.9981
	MAPE	0.1564%	0.0653%	0.2376%	0.9983%	0.9568%
	LL	4.328×10^{-6}	1.109×10^{-6}	6.567×10^{-6}	1.437×10^{-4}	1.172×10^{-4}

In order to evaluate the prediction performance of the proposed hybrid methodology, this paper has taken into account two approaches: the hybrid k-factor GARMA-LLWNN model trained using the Backpropagation algorithm, and the k-factor GARMA-G-GARCH model. And five-time horizons; 6 hours, 12 hours, one day, 2 days and 3 days ahead forecasting, using the R², the MAPE and the LL out of sample criteria, the forecast evaluation results are reported in Table VII. Figs. 9-13 show that the k-factor GARMA-LLWNN model predictions values for all the five horizons are very close to the real values. The forecasting results (Table VIII) shows that the k-factor GARMA-G-GARCH model outperforms the hybrid model in terms of prediction accuracy. Indeed, k-factor GARMA-G-GARCH model prediction errors are the smallest for all evaluation criteria and for all forecast time horizons.

This can be explained by the fact that the k-factor GARMA-G-GARCH model takes into account the seasonal long-memory in both the conditional mean and the conditional variance, making this model a robust tool that can deal with the features of the electricity prices and thus provide the best forecasting results. On the other hand, despite its capacity as a nonlinear, nonparametric model, and its particularity by having a wavelet activation function and local linearity, the LLWNN model is unable to detect, model and predict the seasonality in the data. Since, when it is compared with the G-GRACH model, this last one provides prediction that is more accurate. This is explained by the ability of the G-GARCH model in modeling the seasonal long memory in the conditional variance. This also proves the importance of taking into account the seasonal long memory behavior to enhance forecasting accuracy [57].
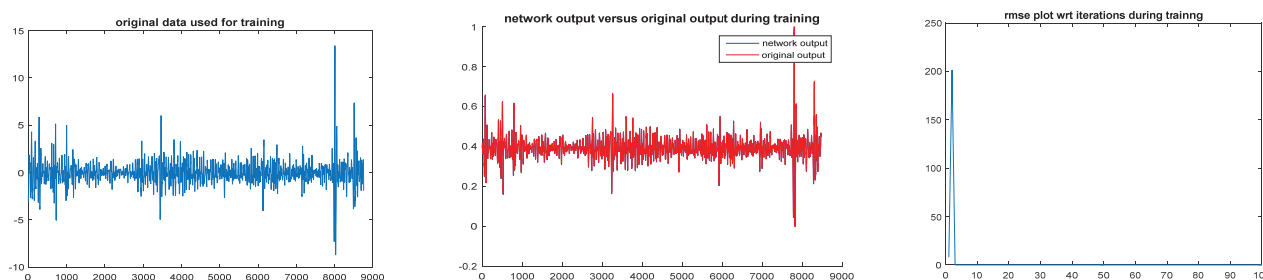

Fig. 8 LLWNN Training with BP algorithm results (residuals of k-factor GARMA modeling)

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
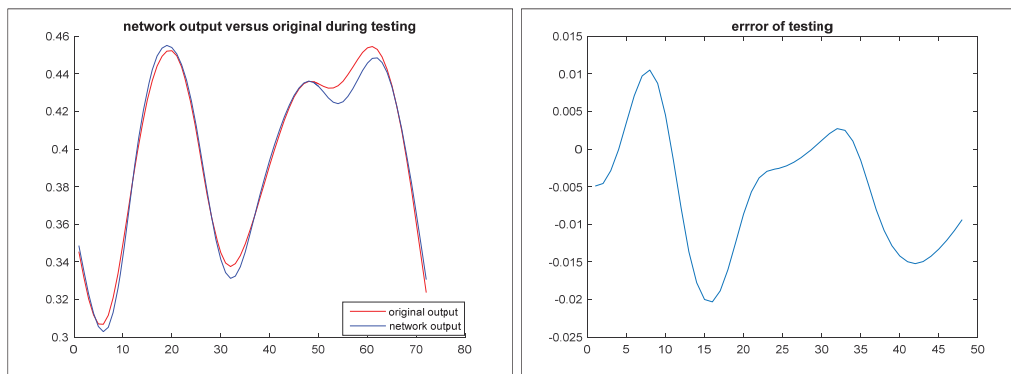Vol:11, No:9, 2017

Fig. 9 Three days (72 hours) ahead prediction during testing (residuals of k -factor GARMA modeling)
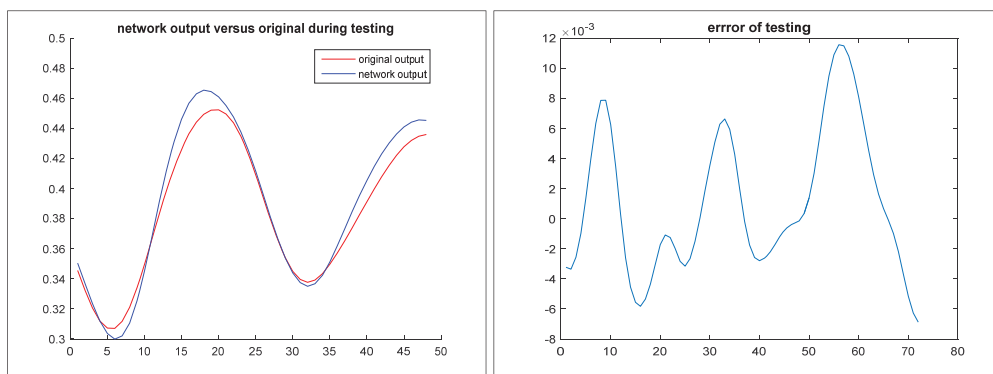

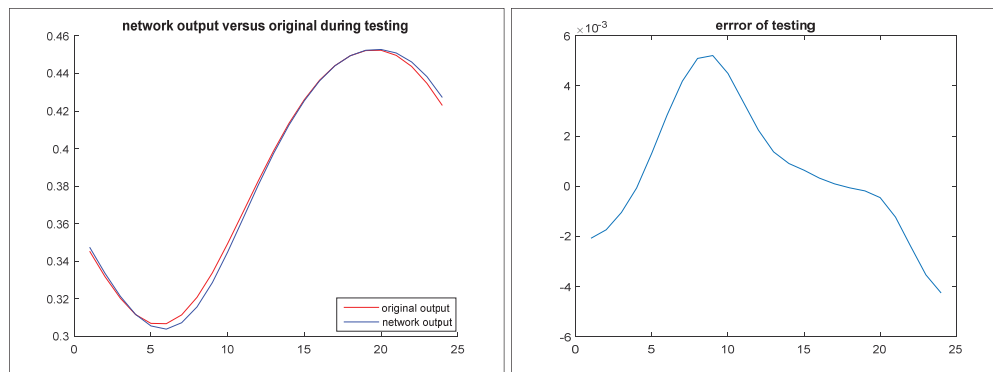Fig. 10 Two days (48 hours) ahead prediction during testing (residuals of k -factor GARMA modeling)

Fig. 11 One day (24 hours) ahead prediction during testing (residuals of k -factor GARMA modeling)
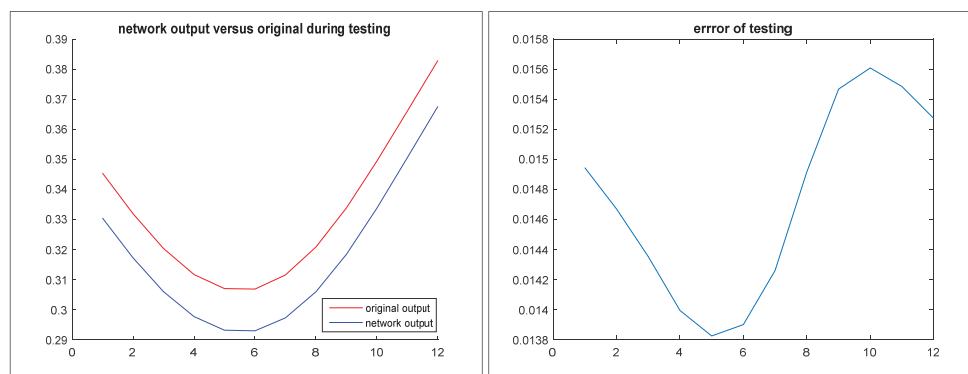

Fig. 12 Semi daily (12 hours) prediction during testing (residuals of k -factor GARMA modeling)
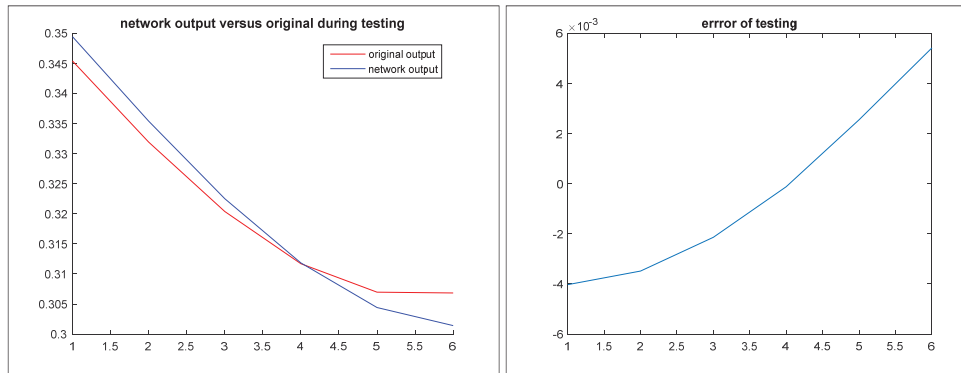
World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:11, No:9, 2017

Fig. 13 6 hours ahead prediction during testing (residuals of k-factor GARMA modeling)

V. CONCLUSIONS

In power markets, price forecasting has become an important topic for all its participants. Background information about the electricity price is crucial for risk management. It may represent an advantage for a market player facing competition.

However, the behavior of electricity prices differs from that of other commodity markets. The most obvious of these differences is that electricity is a non-storable merchandize.

Moreover, electricity prices show some particular characteristics such as high frequency, non-stationary behavior, multiple seasonality, high volatility, hard nonlinear behavior and long memory, which may affect the prices dramatically. Thus, we cannot rely on models developed for financial markets or other commodity markets. In this framework, due to the complexity of the electricity market, the electricity price forecasting has been the most challenging task. In this vein, this paper focuses on resolving the issues of modeling and forecasting the features of the electricity prices, notably the existing of the seasonal long memory behavior in both the conditional mean and the conditional variance.

For this purpose, we focus on the modeling of the conditional mean, we adopt a generalized fractional model with k-factor of Gegenbauer (k-factor GARMA). Thereafter, in order to model and predict the conditional variance we adopt two different approaches; firstly, the local linear wavelet neural network (LLWNN) based BP algorithm is adopted to model and predict the conditional variance (applied to the residual of k-factor GARMA model). Secondly, the G-GARCH model is applied to the residual of the k-factor GARMA, so we estimate a k-factor GARMA-G-GARCH model and we use an estimation approach based on the discrete wavelet transform (DWPT). The R^2, the mean absolute percentage error (MAPE) and the logarithmic loss function (LL) are used as a performance criteria to evaluate the prediction ability of each model. Forecasting results show that the predictive performance of k-factor GARMA-G-GARCH model provides evidence of the power compared to the hybrid parametric and non-parametric k-factor GARMA-LLWNN model. Therefore, this model leads to an improved performance. It can be an effective way in the forecasting task,

especially when higher forecasting accuracy is needed. Obtained results are very interesting in the meaning that it was always difficult to accomplish such precision when forecasting electricity spot prices. This highlights the importance of the k-factor GARMA-G-GARCH methodology as a robust forecasting method.

REFERENCES

[1] Escribano, A., J.I. Peña, and P. Villaplana. 2011. "Modelling electricity prices: international evidence." Oxford bulletin of economics and statistics 73: 622-650.
[2] Koopman, S.J., M. Ooms, and M.A Carnero. 2007. "Periodic seasonal Reg ARFIMA-GARCH models of daily electricity spot prices." Journal of the American Statistical Association 102(477): 16-27.
[3] Knittel, C., and M.R. Roberts. 2005. "An empirical examination of restructured electricity prices." Energy Economics 27(5): 791-817.
[4] Weron, R. 2006. "Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach." The Wiley Finance Series (Book 396), Wiley, Chichester.
[5] Granger, C.W., R. Joyeux. 1981. "An introduction to long-memory time series models and fractional differencing." Journal of time series analysis 1, 15–29.
[6] Hosking, J.R.M. 1981. "Fractional differencing." Biometrika 68 (1): 165-176.
[7] Woodward, W.A., Q.C. Cheng, and H.L. Gray. 1998. "A k-factor GARMA long-memory model." J Time Series Analysis 19: 485-504.
[8] Guégan, D. 2000. "A New Model: The k-factor GIGARCH Process." Journal of Signal Processing 4: 265-271.
[9] Chen, Y., J. Dong, B. Yang, and Y. Zhang, 2004. "A local linear wavelet neural network." Intelligent Control and Automation, IEEE: 1954-1957.
[10] Whitcher, B. 2004. "Wavelet-based estimation for seasonal long-memory processes." Technometrics 46.
[11] Contreras, J., R. Espinola, F.J. Nogales, and A.J. Conejo. 2003. "ARIMA models to predict next day electricity prices." IEEE Transactions on Power systems 18(3): 1014-1020.
[12] Saâdaoui, F., N. Chaâben, and S. Benammou. 2012. "Modelling power spot prices in deregulated European energy markets: a dual long memory approach." Global Business and Economics Review 14: No. 4.
[13] Gray, H.L., N.F. Zhang, and W.A. Woodward. 1989. "On generalized fractional processes". Journal of Time Series Analysis 10 (3): 233-257.
[14] Boubaker, H., and N. Sghaier. 2015. "Semi parametric generalized long-memory modeling of some MENA stock market returns: A wavelet approach." Economic Modeling 50: 254-265.
[15] Caporale, G.M., L.A. Gil-Alana. 2014. "Long-run and cyclical dynamics in the US stock market." J. Forecast. 33 (2): 147-161.
[16] Caporale, G.M., J. Cuñado, and L.A. Gil-Alana. 2012. "Modelling long-run trends and cycles in financial time series data." J. Time Ser. Anal. 34 (2): 405-421.
[17] Ferrara, L., and D. Guegan. 2001. "Forecasting with k-factor Gegenbauer Processes: Theory and Applications." Journal of Forecasting, Wiley 20 (8): 581-601.
[18] Diongue, A.K., D. Guégan, and B. Vignal. 2009. "Forecasting electricity

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:11, No:9, 2017

spot market prices with a k-factor GIGARCH process" Applied Energy 86: 505-510.

[19] Soares, L.J., and L.R. Souza. 2006. "Forecasting electricity demand using generalized long memory." International Journal of Forecasting 22, 17-28.

[20] Diongue, A.K., G. Dominique, and V. Bertrand. 2004. "A k-factor GIGARCH process: estimation and application on electricity market spot prices." Probabilistic Methods Applied to Power Systems: 12-16.

[21] Beran, J. 1999. "SEMIFAR Models-a Semiparametric Framework for Modelling Trends, Long-Range Dependence and Nonstationarity". Center of Finance and Econometrics, University of Konstanz.

[22] Baillie, R.T., T. Bollerslev, and H.O. Mikkelsen. 1996. "Fractionally integrated generalized autoregressive conditional heteroscedasticity". J. Econ 74 (1): 3- 30.

[23] Bollerslev, T., and H.O. Mikkelsen. 19996. "Modeling and pricing long memory in stock market volatility". J. Econometrics 73: 151-184.

[24] Boubaker, H. 2015. "Wavelet Estimation of Gegenbauer Processes: Simulation and Empirical Application". Computational Economics 46: 551-574.

[25] Engle, R.F., 1982. "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation." Econometrica 50: 987-1008.

[26] Bollerslev, T. 1986. "Generalized autoregressive conditional heteroscedasticity". Journal of Econometrics 31: 307-327.

[27] Boubaker; H., and M. Boutahar. 2011. "A wavelet-based approach for modelling exchange rates". Stat Methods Appl 20: 201-220.

[28] Bordignon, S., M. Caporin, and F. Lisi. 2007. "Generalised long-memory GARCH models for intra-daily volatility". Computational Statistics & Data Analysis 51: 5900-5912.

[29] Bordignon, S., M. Caporin, and F. Lisi. 2010. "Periodic long memory GARCH models" Econom Rev 28:60-82.

[30] Caporin, M., and F. Lisi. 2010. "Misspecification tests for periodic long memory GARCH models." Statistical Methods & Applications 19: 47-62.

[31] Diongue, A.K., and Guégan. D. 2008. "Estimation of k-factor GIGARCH process: a Monte Carlo study." ISSN 1955-611X.

[32] Mallat, S., and Zhang. 1993. "Matching pursuits with time-frequency dictionaries." IEEE Transactions on Signal Processing 41: 3397-3415.

[33] Mallat, S. 1999. "A wavelet tour of signal processing". Academic press.

[34] Pany (2011), Short-Term Load Forecasting using PSO Based Local Linear Wavelet Neural Network, International Journal of Instrumentation, Control and Automation (IJICA) ISSN : 2231-1890 Volume-1, Issue-2.

[35] Chakravarty, S., M. Nayak, and M. Bisoi. 2012. "Particle Swarm Optimization Based Local Linear Wavelet Neural Network for Forecasting Electricity Prices, Energy, Automation, and Signal (ICEAS)." Energy, Automation, and Signal (ICEAS), IEEE: 1-6.

[36] Pany, P.K., and S.P. Ghoshal. 2013. "Day-ahead Electricity Price Forecasting Using PSO-Based LLWNN Model." International Journal of Energy Engineering (IJEE) 3: 99-106.

[37] Sharkey, A.J., 2002. Types of multinet system, in: International Workshop on Multiple Classifier Systems. Springer, pp. 108-117.

[38] Armano, G., M. Marchesi, and A. Murru. 2005. "A hybrid genetic-neural architecture for stock indexes forecasting". Information Sciences 170: 3-33.

[39] Yu, L., S. Wang, and K.K. Lai. 2005. A novel nonlinear ensemble-forecasting model incorporating GLAR and ANN for foreign exchange rates." Computers and Operations Research 32: 2523-2541.

[40] Taskaya. T., and M.C. Casey. 2005. "A comparative study of autoregressive neural network hybrids." Neural Networks 18: 781-789.

[41] Tong; H., K.S. Lim. 1980. "Threshold autoregressive, limit cycles and cyclical data." Journal of the Royal Statistical Society Series 42 (3): 245-292.

[42] Zhang, G.P. 2003. "Time series forecasting using a hybrid ARIMA and neural network model." Neurocomputing 50: 159- 175.

[43] Khashei, M. and M. Bijari. 2010. "An artificial neural network model for time series forecasting." Expert Syst. Appl. 37: 479-489.

[44] Valenzuela, O., I. Rojas, F. Rojas, H. Pomares, L. Herrera, A. Guillen, L. Marquez, and M. Pasadas. 2008. "Hybridization of intelligent techniques and ARIMA models for time series prediction." Fuzzy Sets Syst. 159: 821-845.

[45] Tseng, F.M., H.C. Yu, and G.H. Tzeng. 2002. "Combining neural network model with seasonal time series ARIMA model." Technological Forecasting & Social Change 69: 71-87.

[46] Tan. Z., J. Zhang, J. Wang, and J. Xu. 2010. "Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models."Appl Energy 87:3606-3610.

[47] Cheung, Y.W. 1993. "Long memory in foreign-exchanges rates." J Bus Econ Stat 11: 93-101.

[48] Ham F.H. and I. Koslanic. 2001. "Principles of Neurocompuling for Science and Engineering." McGraw-Hill Higher Education.

[49] Khashei, M., and M. Bijari. 2011. "A novel hybridization of artificial neural networks and ARIMA models for time series forecasting." Applied Soft Computing 11: 2664-2675.

[50] Bollerslev, T., and R. Hodrick. 1992. "Financial market efficiency tests". The Handbook of Applied Econometrics I. Macroeconomics, North-Holland, Amsterdam.

[51] Giraitis, L., P. Kokoszka, and R. Leipus. 2000. "Stationary ARCH models: dependence structure and Central Limit Theorem." Econometric Theory 16: 3-22.

[52] Kazakeviéius. V., and R. Leipus. 2002. "On stationarity in the ARCH(∞) model." Econometric Theory 18: 1-16.

[53] Zaffaroni, P., 2004. "Stationarity and memory in ARCH(∞) models." Econometric Theory 20: 147-160.

[54] Gencay, R., M.M. Dacorogna, U.A. Müller, O.V. Pictet, and R.B. Olsen. 2001. "An Introduction to High Frequency Finance." Academic Press, London.

[55] Geweke, J., and S. Porter-Hudak. 1983. "The Estimation And Application Of Long Memory Time Series Models." Journal of Time Series Analysis 4: 221–238.

[56] Campbell, J.Y., and S.B. Thompson. 2008. "Predicting the equity premium out of sample: can anything beat the historical average." Rev. Finance. Stud. 21: 1509-1531.

[57] Nowotarski, J., and R. Weron. 2016. "On the importance of the long-term seasonal component in day-ahead electricity price forecasting." Energy Economics 57: 228-235.