



**HAL**  
open science

## Forecasting electricity spot market prices with a k-factor GIGARCH process

Abdou Kâ Diongue, Dominique Guegan, Bertrand Vignal

► **To cite this version:**

Abdou Kâ Diongue, Dominique Guegan, Bertrand Vignal. Forecasting electricity spot market prices with a k-factor GIGARCH process. *Applied Energy*, 2009, 86 (4), pp.505-510. 10.1016/j.apenergy.2008.07.005 . halshs-00307606v2

**HAL Id: halshs-00307606**

**<https://shs.hal.science/halshs-00307606v2>**

Submitted on 26 Nov 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Forecasting electricity spot market prices with a k-factor GIGARCH process.

Abdou Kâ Diongue\*, Dominique Guégan†, Bertrand Vignal‡

## Abstract

In this article, we investigate conditional mean and conditional variance forecasts using a dynamic model following a  $k$ -factor GIGARCH process. Particularly, we provide the analytical expression of the conditional variance of the prediction error. We apply this method to the German electricity price market for the period August 15, 2000 - December 31, 2002 and we test spot prices forecasts until one month ahead forecast. The forecasting performance of the model is compared with a SARIMA-GARCH benchmark model using the year 2003 as the out-of-sample. The proposed model outperforms clearly the benchmark model. We conclude that the  $k$ -factor GIGARCH process is a suitable tool to forecast spot prices, using the classical RMSE criteria.

*Keywords:* Conditional mean; Conditional variance; Electricity prices; Forecast; GIGARCH process.

---

\*Corresponding author: Université Gaston Berger de Saint-Louis, UFR SAT, BP 234 Saint-Louis Senegal and Research Fellow at Queensland University of Technology, 2 George Street GPO Box 2434, Brisbane QLD 4001 Australia, e-mail: [abdou.diongue@gmail.com](mailto:abdou.diongue@gmail.com)

†Paris School of economics, CES-MSE, Université Paris1 Panthéon-Sorbonne, 106 boulevard de l'Hopital 75647 Paris Cedex 13 France, e-mail: [dguegan@univ-paris1.fr](mailto:dguegan@univ-paris1.fr)

‡Ingénieur EDF R&D, 1 avenue du général de Gaulle 92141 Clamart cedex, e-mail: [bertrand.vignal@edf.fr](mailto:bertrand.vignal@edf.fr)

# 1 Introduction

In finance, commodity price forecasting is a crucial issue, in particular for electricity. Indeed, price forecasts can help evaluating bilateral contracts. For such a commodity, price forecasts can be based on a spot price analysis. Since electricity spot prices have a behavior which presents specific features, the price forecasts problem can be complex. In another hand, electricity price volatility has a marked variability in time. We observe both high and low periods of price reaction. Recently spot price volatility has been studied and modelled using a volatility model depending on time. Using such a model, Benini, Maracini, Pelacchi and Venturini (2002) investigate several markets. In addition, electricity spot prices exhibit long memory behavior combined with a periodic behavior. Recent works have taken into account these last features using related ARFIMA models, see Koopman, Ooms and Carnero (2007) or Diongue and Guégan (2008).

After modelling these electricity prices, the forecasting problem arises. In the literature, two approaches have been considered: parametric models using AR, ARX, AR-GARCH, ARX-GARCH and Regime switching models, Misiorek, Truék and Weron (2006); and non parametric methods like the neuronal nets for instance, Ramsay and Wang (1997) and Conejo, Contreras, Espinola and Plazas (2005).

In this paper, in order to provide robust forecasts for spot electricity prices, we propose a new approach based on the  $k$ -factor GIGARCH process, (Guégan, 2000), which allows taking into account a lot of stylized facts observed on the electricity spot prices, in particular stochastic volatility, long memory and periodic behaviors. The investigation of this model is done in Guégan (2003). Diongue and Guégan (2004) introduced the parameter estimation of the  $k$ -factor GIGARCH process. Here, we provide new developments which

concern the expression of the forecasts using the  $k$ -factor GIGARCH process and we give their properties. We apply these results on the German electricity prices market providing forecasting prices up until a one month-ahead. These results are totally new in the sense that, in most published papers, the previsions concern mostly the one day-ahead horizon and here we are interested by long term prediction. For comparison purpose, we compare our approach with a benchmark model in terms of forecasting, using the RMSE criteria.

This paper is organized as follows. The next Section presents the data and contains the main empirical findings. In Section 3, we specify some notations, define the  $k$ -factor GIGARCH process along with new theoretical results on forecasting in mean and in variance. In the fourth Section, we provide forecasts for the German spot prices data set. The last section is dedicated to conclusions.

## 2 The data set

We consider the hourly series of the EEX spot prices, denoted  $(S_t)_{t=1, \dots, N}$ , from August 15, 2000 to December 31, 2003, which yields  $N = 29616$  hourly observations. For this electricity spot market, the prices are fixed each day, for each of the 24 hours of the next day. This spot market has two particular interests. On the one hand, it allows a buyer (seller) to supply their bid (load) on an hourly basis (physical market) and on the other hand, the spot market could correspond to the price reference for the contracts. In order to make the series  $(S_t)_{t=1, \dots, N}$  stationary, we take its logarithmic form and define  $(X_t)_t$  such that,  $\forall t = 1, \dots, N$ ,  $X_t = \log(S_t)$ . We build two subsets from the original series. The first one, from August 15, 2000 to December 31, 2002 (figure 1) and containing  $T = 20856$  observations, is used for the parameter estimation procedure and the second one, from January 1, 2003

to December 31, 2003, is used for validation. In figure 1, we observe that there are periods of low volatility (Figure 1, Regions 1) following by periods of high volatility (Figure 1, Regions 2). Figure 2 provides a zoom on the evolution of the spot prices, in its logarithmic scale, from October 14, 2000 to October 27, 2000 (two weeks). We observe that the series exhibit two kinds of seasonality: daily and weekly seasonalities.

**Insert Figures 1 and 2**

Table 1 gives the summary statistics of the log electricity spot prices. The mean and the standard deviation are quite small, while the estimated measure of skewness,  $-2.6$ , is significantly negative, indicating that the log electricity spot prices has non-symmetric distribution. Furthermore, the large value of the kurtosis statistic,  $29.67$ , suggests that the underlying data are leptokurtic, or fat-tailed and sharply peaked around the mean when compared with the standard Normal distribution (Figure 3). The results of the non-Normality test agree with previous literature concerning this kind of behavior for these log electricity spot prices data, Misiorek, Trueck and Weron (2003). The Box-Pierce  $Q$ -tests of up to twenty-fourth order serial correlation for the levels and squares of the mean-corrected log electricity spot prices ( $Q(24)$  and  $Q^2(24)$ ) are both significant. In summary, the diagnostics suggest that a GARCH-class model would be appropriate to explain the evolution of the log electricity spot prices, along with an error distribution that allows for greater kurtosis than the Gaussian distribution.

**Insert Table 1 and Figure 3**

The autocorrelation function of the EEX spot prices (Figure 4) exhibits a slow decay pattern at two seasonality lags corresponding to daily and weekly seasonalities. The autocorrelation function does not converge exponentially to zero in covariance sense, (Guégan, 2005, for discussions on this concept),

thus we assume that the series has a long range dependence behavior. Moreover, the spectral density represented by the periodogram (Figure 5), is unbounded at three frequencies corresponding respectively, to the daily, weekly and half daily seasonalities.

### Insert Figures 4 and 5

Thus, this empirical and statistical analysis highly justifies the use of the  $k$ -factor Gegenbauer process with time varying conditional variance errors to model the data set  $(X_t)_t$ .

## 3 The $k$ -factor GIGARCH process

### 3.1 Notations and model

Assume that  $(\xi_t)_{t \in \mathbb{Z}}$  is a white noise process with unit variance and mean zero. Let  $(X_t)_{t \in \mathbb{Z}}$  be a  $k$ -factor GIGARCH process defined by the following equations:

$$\Phi(B) \prod_{i=1}^k (I - 2\nu_i B + B^2)^{d_i} (X_t - \mu) = \Theta(B) \varepsilon_t, \quad (3.1)$$

where

$$\varepsilon_t = \sqrt{h_t} \xi_t \quad \text{with } h_t = a_0 + \sum_{i=1}^r a_i \varepsilon_{t-i}^2 + \sum_{j=1}^s b_j h_{t-j}. \quad (3.2)$$

In the relationship (3.1)-(3.2),  $\mu$  represents the mean of the process  $(X_t)_{t \in \mathbb{Z}}$ ,  $d = (d_1, \dots, d_k)$ , the memory parameters and  $\nu = (\nu_1, \dots, \nu_k)$  the frequency location parameters with,  $0 < d_i < \frac{1}{2}$  if  $|\nu_i| < 1$ , and  $0 < d_i < \frac{1}{4}$  if  $|\nu_i| = 1$  for  $i = 1, \dots, k$ , with  $k$  a nonnegative integer.  $\Phi(B)$  and  $\Theta(B)$  are the autoregressive and the moving average polynomials in  $B$  ( $B^n X_t = X_{t-n}$ ,  $n \geq 0$ ), defined by  $\Phi(B) = I - \sum_{i=1}^p \phi_i B^i$  and  $\Theta(B) = I - \sum_{j=1}^q \theta_j B^j$ , where  $p$  and  $q$  are nonnegative integers. We suppose that all the roots of the polynomials  $\Phi(B)$  and  $\Theta(B)$  are outside the unit circle. We assume that  $a_0 > 0, a_1, \dots, a_r, b_1, \dots, b_s \geq 0$ , and  $\sum_{i=1}^r a_i + \sum_{j=1}^s b_j < 1$ . The  $k$ -factor

GIGARCH process (3.1)-(3.2) under the previous assumptions is well defined. We can use it to forecast electricity spot prices. The interest of such a heteroskedastic model is that it allows emphasizing the optimal predictor for conditional variance. A basic property for the  $k$ -factor GARMA process with time varying innovations following a GARCH( $r, s$ ) model is that the long memory part provides a way to model the conditional mean whereas the heteroscedastic part provides a way to model the conditional variance. From a forecasting point of view, the  $k$ -factor GIGARCH model (Guégan, 2000) and the  $k$ -factor GARMA model (Giraitis and leipus, 1995) will give the same optimal predictor for the conditional mean and the difference between their respective forecasts appears mainly in the building of the confidence interval.

### 3.2 Forecasting equation for the $k$ -factor GIGARCH process

In this section, we provide new results concerning the predictions based on a strong white noise  $(\xi_t)_{t \in \mathbb{Z}}$ .

In a first step, we provide the expression of the forecasts in mean. Observing  $X_1, \dots, X_T$ , using the stationary and invertible  $k$ -factor GIGARCH process defined by equations (3.1)-(3.2), we compute the  $h$ -time-step-ahead least square predictor, denoted  $\hat{X}_t(h)$ ,  $h > 0$ . The predictor is obtained by minimizing the mean squared error between forecasts and observations and our approach follows the work of Ferrara and Guégan (2001). We assume that the parameters of the model (3.1)-(3.2) have been estimated using Whittle method, see Diongue and Guégan (2004). Under the regular conditions specified in section 3.1, this model can be rewritten in its infinite moving average form:

$$X_t = \sum_{j=0}^{\infty} \hat{\beta}_j \left( \hat{d}, \hat{\nu}, \hat{\gamma} \right) \varepsilon_{t-j}, \quad (3.3)$$

where the weights  $\left(\hat{\beta}_j\left(\hat{d}, \hat{\nu}, \hat{\gamma}\right)\right)_{j \in \mathbb{Z}}$  follow the expression:

$$\hat{\beta}_j\left(\hat{d}, \hat{\nu}, \hat{\gamma}\right) = \hat{\psi}_j\left(\hat{d}, \hat{\nu}\right) + \sum_{i=1}^{\min(j,p)} \hat{\phi}_i \hat{\beta}_{j-i}\left(\hat{d}, \hat{\nu}\right) - \sum_{i=1}^{\min(j,q)} \hat{\theta}_i \hat{\psi}_{j-i}\left(\hat{d}, \hat{\nu}, \hat{\gamma}\right), \quad (3.4)$$

and

$$\hat{\psi}_j\left(\hat{d}, \hat{\nu}\right) = \sum_{\substack{0 \leq l_1, \dots, l_k \leq j, \\ l_1 + \dots + l_k = j}} C_{l_1}\left(\hat{d}_1, \hat{\nu}_1\right) \cdots C_{l_k}\left(\hat{d}_k, \hat{\nu}_k\right). \quad (3.5)$$

The weights  $\left(C_{l_i}\left(\hat{d}, \hat{\nu}\right)\right)_{l_i \in \mathbb{Z}}$  are the Gegenbauer polynomials. Denoting the vectorial parameters  $\hat{d} = \left(\hat{d}_1, \dots, \hat{d}_k\right)$ ,  $\hat{\gamma} = \left(\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q\right)$ , and  $\hat{\omega} = \left(\hat{a}_0, \hat{a}_1, \dots, \hat{a}_r, \hat{b}_1, \dots, \hat{b}_s\right)$ , we can derive the  $h$ -time-step-ahead least square predictor,  $\hat{X}_t(h)$ , given by the following expression:

$$\hat{X}_t(h) = \sum_{j \geq h}^{\infty} \hat{\beta}_j\left(\hat{d}, \hat{\nu}, \hat{\gamma}\right) \varepsilon_{t+h-j}, \quad (3.6)$$

which can be rewritten as

$$\hat{X}_t(h) = \sum_{j \geq 0}^{\infty} \hat{\beta}_{j+h}\left(\hat{d}, \hat{\nu}, \hat{\gamma}\right) \varepsilon_{t-j}. \quad (3.7)$$

Unfortunately, in practice, the process  $(\varepsilon_t)_{t \in \mathbb{Z}}$  cannot be observed, thus this latter expression cannot be used to obtain the forecasts in mean from the process. Therefore, we need to investigate an autoregressive form of the  $k$ -factor GIGARCH process. Following Proposition 4 in Ferrara and Guégan (2001), it is straightforward to show that a useful and easy formula to update, for computing the least-squares predictor  $\hat{X}_t(h)$ , is given by

$$\hat{X}_t(h) = - \sum_{j=1}^{h-1} \hat{\alpha}_j\left(\hat{d}, \hat{\nu}, \hat{\gamma}\right) \hat{X}_t(h-j) - \sum_{j \geq 0} \hat{\alpha}_{j+h}\left(\hat{d}, \hat{\nu}, \hat{\gamma}\right) X_{t-j}, \quad (3.8)$$

where the weights  $\left(\hat{\alpha}_j\left(\hat{d}, \hat{\nu}, \hat{\gamma}\right)\right)_{j \in \mathbb{Z}}$  are defined by

$$\hat{\alpha}_j\left(\hat{d}, \hat{\nu}, \hat{\gamma}\right) = \hat{\pi}_j\left(\hat{d}, \hat{\nu}\right) - \sum_{i=1}^{\min(j,p)} \hat{\phi}_i \hat{\pi}_{j-i}\left(\hat{d}, \hat{\nu}\right) + \sum_{i=1}^{\min(j,q)} \hat{\theta}_i \hat{\alpha}_{j-i}\left(\hat{d}, \hat{\nu}, \hat{\gamma}\right), \quad (3.9)$$



and  $\hat{\pi}_j(\hat{d}, \hat{\nu}) = \hat{\psi}_j(-\hat{d}, \hat{\nu})$ . In practice, we use the expression (3.8) to get the forecasts.

Now, we provide the expression for the variance's errors for the forecasts. The  $h$ -time-step-ahead prediction error for the predictor  $\hat{X}_t(h)$  given in (3.8) is denoted  $e_{t,h}$  for  $h > 0$ , and is equal to:

$$\begin{aligned}\hat{e}_{t,h} &= X_{t+h} - \hat{X}_t(h) \\ &= \sum_{j=0}^{h-1} \hat{\beta}_j(\hat{d}, \hat{\nu}, \hat{\gamma}) \varepsilon_{t+h-j}.\end{aligned}\quad (3.10)$$

The conditional mean squared error for this predictor has the following expression:

$$E(\hat{e}_{t,h}^2 | \mathcal{F}_t) = \sum_{j=0}^{h-1} \hat{\beta}_j^2(\hat{d}, \hat{\nu}, \hat{\gamma}) E(\varepsilon_{t+h-j}^2 | \mathcal{F}_t), \quad (3.11)$$

where  $\mathcal{F}_t$  is the  $\sigma$ -algebra generated by  $\varepsilon_0, \dots, \varepsilon_t$ , (the noise which appears in (3.1) - (3.2)). We remark that the conditional mean squared error depends on the future conditional variance of the innovation process  $(\varepsilon_t^2)_{t \in \mathbb{Z}}$ . Its expression, denoted  $\text{var}_t(\hat{e}_{t,h})$ , is given by

$$\begin{aligned}\text{var}_t(\hat{e}_{t,h}) &= \sum_{i=1}^h \hat{\beta}_{h-i}^2(\hat{d}, \hat{\nu}, \hat{\gamma}) \hat{\omega}_i \\ &\quad + \sum_{k=1}^h \hat{\beta}_{h-k}^2(\hat{d}, \hat{\nu}, \hat{\gamma}) \left[ \sum_{i=0}^{s-1} \hat{\delta}_{i,k} h_{t-i} + \sum_{i=0}^{m-1} \hat{\rho}_{i,k} \varepsilon_{t-i}^2 \right],\end{aligned}\quad (3.12)$$

with  $m = \max(r, s)$ ,  $\hat{\omega}_h = e_1'(I + \Gamma + \dots + \Gamma^{h-1}) e_1 \hat{a}_0$ ,  $\hat{\delta}_{j,h} = -e_1' \Gamma^h e_{m+j+1}$  if  $j = 0, \dots, s-1$ ,  $\hat{\rho}_{j,h} = e_1' \Gamma^h (e_{j+1} + e_{m+j+1})$  if  $j = 0, \dots, s-1$ ,  $\hat{\rho}_{j,h} = e_1' \Gamma^h e_{j+1}$  if  $j = s, \dots, m-1$ .

Here  $\{e_j\}_{j=1}^{m+s}$  is the standard base of  $\mathbb{R}^{m+s}$  and the matrix  $\Gamma$  is defined such that its first row is equal to  $(\hat{a}_1 + \hat{b}_1, \dots, \hat{a}_m + \hat{b}_m, -\hat{b}_1, \dots, -\hat{b}_s)$ , and the other elements are all equal to zero, except the first one under diagonal elements which are all equal to one. We use the expression (3.12) to build the confidence interval around  $\hat{X}_t(h)$ .

## 4 A real case study

In this section, we investigate the German Energy EXchange market (EEX) corresponding to the Wholesale spot electricity market and the associated spot prices. The German spot prices have several key features such as: long memory persistence with seasonality and non-constant volatility. These stylized facts strongly suggest to use a GIGARCH process in order to make forecast. Thus, we adjust several  $k$ -factor GIGARCH processes on this data set and compare their forecasting performances with a benchmark model, a SARIMA-GARCH model.

### 4.1 Estimated models

In this subsection, we compare competitively three different modelling approaches in order to forecast this hourly series for different forecast horizons, starting from the first of January 2003. The first model is the SARIMA model of Box and Jenkins (1976) with conditional heteroskedastic noise. We denote it  $M_1$ . The second model is a 3-factor GIGARCH model denoted  $M_2$ . Lastly, the third one is a 1-factor GIGARCH process estimated after removing the weekly seasonality by applying the filter  $1 - B^{168}$  to the data set. We denote it  $M_3$ <sup>1</sup>. Below, the values given in brackets correspond to the standard deviation of the estimated parameters.

#### 4.1.1 The model $M_1$

The autocorrelation function of the EEX spot prices (Figure 4) exhibits an exponential rate at the daily lag. Then, using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), we investigate, in that sense, the best SARIMA( $p, d, q$ )  $\times$  ( $P, D, Q$ )<sub>24</sub>-GARCH(1, 1) model.

---

<sup>1</sup>The results of this model are available from authors upon request.

The selected model  $M_1$  is defined by:

$$(I - B)(I - B^{24})\phi^{(1)}(B)\Phi^{(1)}(B^{24})(X_t - 2.97) = \theta^{(1)}(B)\Theta^{(1)}(B^{24})\varepsilon_t, \quad (4.1)$$

where,

$$\phi^{(1)}(B) = I - \hat{\phi}_1^{(1)}B,$$

with  $\hat{\phi}_1^{(1)} = 0.648(7.425 \times 10^{-4})$ , and

$$\Phi^{(1)}(B^{24}) = I - \hat{\phi}_{24}^{(1)}B^{24} - \hat{\phi}_{48}^{(1)}B^{48},$$

$\hat{\phi}_{24}^{(1)} = 0.1221(5.756 \times 10^{-5})$ ,  $\hat{\phi}_{48}^{(1)} = -0.0579(5.5707 \times 10^{-5})$ . We also get

$$\theta^{(1)}(B) = I - \hat{\theta}_1^{(1)}B,$$

with  $\hat{\theta}_1^{(1)} = 0.7479(5.6393 \times 10^{-4})$ , and

$$\Theta^{(1)}(B^{24}) = I - \hat{\theta}_{24}^{(1)}B^{24} - \hat{\theta}_{168}^{(1)}B^{168},$$

with  $\hat{\theta}_{24}^{(1)} = 0.9184(1.0618 \times 10^{-5})$ ,  $\hat{\theta}_{168}^{(1)} = 0.851, (0.303)$ . Finally

$$\varepsilon_t = \sqrt{h_t^{(1)}}\eta_t, \text{ with } h_t^{(1)} = \hat{a}_0^{(1)} + \hat{a}_1^{(1)}\varepsilon_{t-1}^2 + \hat{b}_1^{(1)}h_{t-1}^{(1)}, \quad (4.2)$$

with  $\hat{a}_0^{(1)} = 0.00872(6.6668 \times 10^{-9})$ ,  $\hat{a}_1^{(1)} = 0.3799(1.8223 \times 10^{-5})$ , and  $\hat{b}_1^{(1)} = 0.4867(1.9028 \times 10^{-5})$ .

#### 4.1.2 Model $M_2$

The periodogram of the EEX spot prices is unbounded at three frequencies, thus we consider that the series  $(X_t)_{t \in \mathbb{Z}}$  can be modelled using a 3-factor GIGARCH process. The observed prices  $(X_t)_t$ , in their logarithmic scale, have the following representation:

$$\Phi^{(2)}(B) \prod_{i=1}^3 \left( I - 2\hat{\nu}_i^{(2)}B + B^2 \right)^{\hat{d}_i^{(2)}} (X_t - 2.97) = \Theta^{(2)}(B)\zeta_t, \quad (4.3)$$

where  $\hat{d}_1^{(2)} = 0.135(0.009)$ ,  $\hat{d}_2^{(2)} = 0.214(0.006)$  and  $\hat{d}_3^{(2)} = 0.141(0.015)$ . The polynomials  $\Phi^{(2)}(B)$  and  $\Theta^{(2)}(B)$  are respectively equal to:

$$\Phi^{(2)}(B) = I - \hat{\phi}_{24}^{(2)} B^{24} - \hat{\phi}_{168}^{(2)} B^{168},$$

where  $\hat{\phi}_{24}^{(2)} = 0.863(0.179)$ ,  $\hat{\phi}_{168}^{(2)} = 0.951(0.221)$  and

$$\Theta^{(2)}(B) = I - \hat{\theta}_{24}^{(2)} B^{24} - \hat{\theta}_{168}^{(2)} B^{168},$$

with  $\hat{\theta}_{24}^{(2)} = 0.720(0.331)$ ,  $\hat{\theta}_{168}^{(2)} = 0.794(0.298)$ . Finally,

$$\zeta_t = \sqrt{h_t^{(2)}} \xi_t, \text{ with } h_t^{(2)} = \hat{a}_0^{(2)} + \hat{a}_1^{(2)} \zeta_{t-1}^2, \quad (4.4)$$

where  $\hat{a}_0^{(2)} = 0.0141(3.39 \cdot 10^{-4})$  and  $\hat{a}_1^{(2)} = 0.756(0.365)$ . Here, we have retained three different G-frequencies for which the spectral density function is unbounded:  $\hat{\nu}_1^{(2)} = \cos\left(\frac{2\pi \times 869}{20856}\right)$ ,  $\hat{\nu}_2^{(2)} = \cos\left(\frac{2\pi \times 124}{20856}\right)$  and  $\hat{\nu}_3^{(2)} = \cos\left(\frac{2\pi \times 1738}{20856}\right)$ .

### 4.1.3 Forecasting results

To evaluate the capability of the models (4.1)-(4.2) and (4.3)-(4.4) to forecast electricity German prices, we use the forecasting method developed in Section 3. The period from January 1, 2003 to December 31, 2003 is considered as the out-of sample forecasts period. We assess the predictive ability of the two models by considering the Root Mean Squared Error (RMSE) of the forecasts defined by

$$RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^M \left( X_{t+i} - \hat{X}_t(i) \right)^2}, \quad (4.5)$$

where  $M$  is the size of the out-of-sample forecast period and  $\hat{X}_t(i)$  is the predicted value of  $X_{t+i}$ . We use this criteria because we are interested in the measure of the variation between the observed data and the forecast data.

For each model, forecasts are computed for  $h = 1, \dots, 720$  hours (from one hour to one month). As we use the prices in a logarithmic scale, we need to compute the exponentials of these log-forecastings. In order to get these forecasts, we assume the Gaussianity of the white noise  $(\xi_t)_{t \in \mathbb{Z}}$  in equation (3.2). Then, the forecasts for the spot prices at time  $T$  can be computed using the following expression

$$\hat{S}_T(h) = \exp \left\{ \hat{X}_T(h) + \frac{1}{2} \text{var} [\hat{e}_{T,h}] \right\}. \quad (4.6)$$

Using two particular weeks, on Figures 6-7, we illustrate the seven step ahead forecasting performance of the  $M_2$  model (3-factor GIGARCH model) and the benchmark  $M_1$  model (SARIMA-GARCH). The first week goes from March 16, 2003 to March 22, 2003 while the second goes from September 8, 2003 to September 14, 2003. We notice that the forecasts from the 3-factor GIGARCH model (model  $M_2$ ) are closer to the observed prices values than those obtained from the benchmark model (model  $M_1$ ), with respect to the RMSE, see Table 2.

### Insert Figures 6 and 7

In Table 2, we provide errors for the thirty days ahead, for models ( $M_1$ ), ( $M_2$ ) and ( $M_3$ ), for some specific hours. We notice that the model ( $M_2$ ) outperforms the benchmark model ( $M_1$ ), even for short-term prediction. In addition, we empirically prove that, with a well specified nonlinear long-memory model, the predictions are more accurate than with a misspecified linear long-memory model. This kind of misspecification and its disastrous consequences for long range forecasting have already been pointed out by Ray (1993) through a simulation study. We also perform forecast-accuracy comparisons in order to evaluate if the accuracy of the SARIMA-GARCH and 3-factor GIGARCH models are equal. Visual inspection of the sample autocorrelation function of the loss differential series in Figure 8 is suggestive of long-range dependence behaviour. Moreover, the Box-Pierce  $Q$ -test of

up to twenty five order,  $Q(25) = 5.6558 \times 10^3$ , is highly significant relative to its asymptotic null distribution of  $\chi_{25}^2$ . According to this latter result, the simple F test, Morgan-Granger-Newbold test and Meese-Rogoof test are not applicable. Therefore, we perform the comparing predictive accuracy test proposed by Diebold and Mariano (1995). For the SARIMA-GARCH and the 3-factor GIGARCH models' forecasts, the estimated  $S_1$  statistic<sup>2</sup>,  $S_1 = -2.3546$  implying a p value of 0.0185, is well above the critical value at the 5% level of significance, thereby rejecting the null hypothesis of equal forecast roots expected square error. However, when considering the 3-factor GIGARCH and 1-factor GIGARCH models, we obtain  $S_1 = -1.6216$ , implying a p value of 0.1049. In this latter case, we do not reject at 5% levels the null hypothesis of equal roots expected square error. Thus, these comparative predictive accuracy tests proposed by Diebold and Mariano (1995) confirm the supremacy of model  $M_3$  for forecasting this electricity data set.

**Insert Table 2**

## 5 Conclusion

In this paper we have investigated forecasting's method using a stochastic model such as a  $k$ -factor GIGARCH process. We derive a least square predictor and its properties and we characterize the conditional variance error of this predictor. The results provided in section 3.2 are new and permit to obtain close form expressions for predictions using a  $k$ -factor GIGARCH process.

---

<sup>2</sup>The estimated statistic  $S_1$  for testing the null hypothesis of equal forecast accuracy is  $\bar{d}/\sqrt{2\pi\hat{f}_d(0)/T}$ , where  $\bar{d}$  is the sample mean loss-differential and  $\hat{f}_d(0)$  is a constant estimate of  $f_d(0)$  given by  $2\pi\hat{f}_d(0) = \sum_{\tau=-1}^{T-1} 1(\tau/S(d))\hat{\gamma}_d(\tau)$ , with  $\hat{\gamma}_d$  is the sample autocovariance of  $d$ ,  $1(\tau/S(d))$  is the uniform lag window,  $S(T)$  is the lag truncation and  $T$  the sample size of the loss differential series.

This forecasting method is applied to the German hourly electricity spot market prices. We adjust three different models on this data set: a SARIMA-GARCH model as a benchmark, a 1-factor GIGARCH model and a 3-factor GIGARCH model. The forecasting results for the year 2003 with the estimated models are highly convincing in the sense of RMSE criteria. The model  $M_2$  (3-factor GIGARCH process) provides better forecasts in the sense of RMSE criteria than model  $M_1$  (SARIMA-GARCH process) and model  $M_3$  (1-factor GIGARCH process), when modelling EEX prices on the period under study. In addition, the comparing predictive accuracy test from Diebold and Mariano (1995) suggest that the expected square error for the models forecasts are not equal, confirming the fact that this new modelling improves the forecasts for this kind of data set.

Thanks to this study, we have detected presence of seasonal long memory and heteroskedascity in electricity spot prices and provide interesting modelling to take them into account. Note that these features appear frequently inside the European electricity market.

**Acknowledgments:** We wish to thank the two referees as well as the editor for helpful comments and suggestions, which have permitted to improve the presentation of this paper.

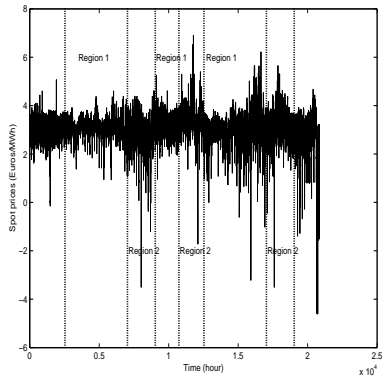


Figure 1: Evolution of the EEX spot market prices in its logarithmic scale from 15/08/2000 to 31/12/2002.

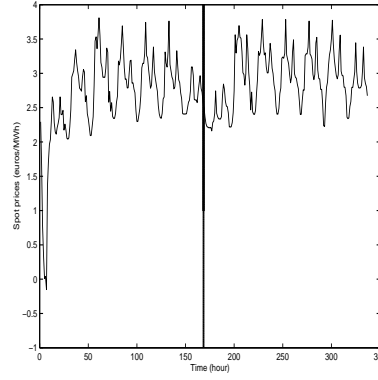


Figure 2: Two weeks zoom concerning the evolution of the EEX spot market prices in its logarithmic scale from October 14, 2000 to October 27, 2000.

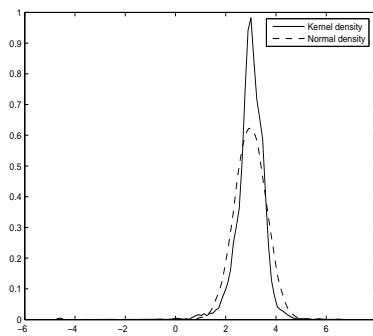


Figure 3: Non-parametric density of the log electricity spot prices (solid line) and probability density function of the Normal distribution (dotted line)



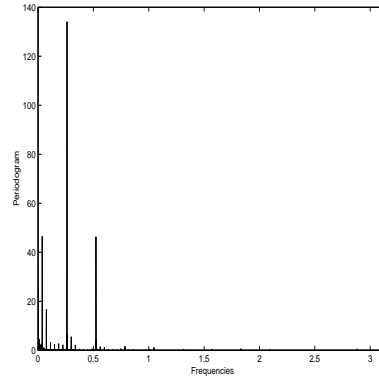
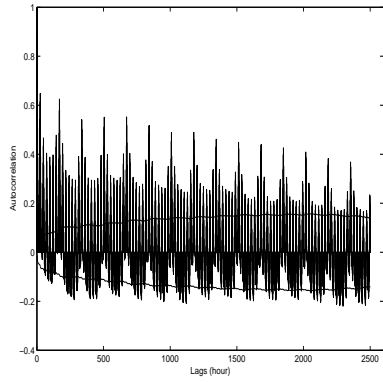


Figure 4: Autocorrelation function for the EEX spot market prices in its logarithmic scale from 15/08/2000 to 31/12/2002. Figure 5: Periodogram of the EEX spot market prices in its logarithmic scale from 15/08/2000 to 31/12/2002.

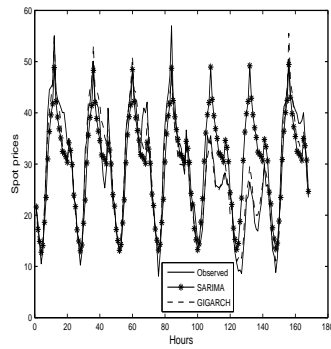
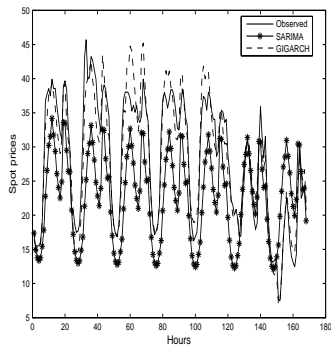


Figure 6: Seven days ahead spot prices forecast for EEX spot market, for both the model  $M_1$  and  $M_2$ , from March 16, 2003 to March 22, 2003. Figure 7: Seven days ahead spot prices forecast for EEX spot market, for both the model  $M_1$  and  $M_2$ , from September 8, 2003 to September 14, 2003.

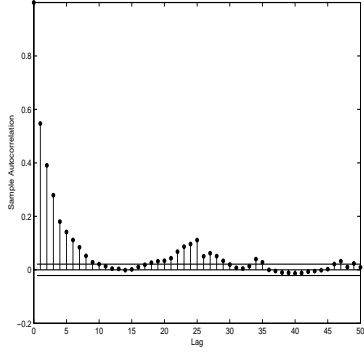


Figure 8: Autocorrelation function for the loss-differential series.

Table 1: Statistics of daily log returns of the S&P 500 stock market index.

Number of observations	20856	Skewness	-2.6
Mean	2.97	Kurtosis	29.67
Standard deviation	0.6241	Jarque-Bera test	$7.3098 \times 10^5$
Minimum	-4.6052	$Q(24)$	$8.4597 \times 10^4$
Maximum	6.5387	$Q^2(24)$	$3.7156 \times 10^4$

The Jarque-Bera test critical value at significance level of 5% is 5.85423.  $Q$  and  $Q^2$  are the Box Pierce statistics for the levels and square of the log electricity spot prices, respectively, using 24 lags. The critical value at level 5% is 36.4150.

Steps ahead	1 h	3 h	5 h	11 h	17 h	22 h	23 h	00 h
1 day	3.94	3.77	3.84	27.30	10.62	6.88	6.58	4.99
	<b>3.63</b>	<b>3.72</b>	<b>3.76</b>	<b>15.29</b>	<b>8.47</b>	<b>5.38</b>	<b>5.63</b>	<b>4.48</b>
	<i>5.39</i>	<i>4.44</i>	<i>4.68</i>	<i>16.17</i>	<i>8.97</i>	<i>5.61</i>	<i>5.45</i>	<i>4.16</i>
10 days	10.21	7.93	7.60	33.69	17.99	14.47	14.18	11.30
	<b>5.37</b>	<b>5.35</b>	<b>5.23</b>	<b>16.98</b>	<b>8.87</b>	<b>6.27</b>	<b>6.20</b>	<b>5.21</b>
	<i>5.55</i>	<i>5.68</i>	<i>5.43</i>	<i>17.59</i>	<i>19.22</i>	<i>6.48</i>	<i>6.33</i>	<i>5.32</i>
20 days	17.68	11.53	10.59	37.72	24.61	21.60	24.01	18.69
	<b>5.63</b>	<b>5.58</b>	<b>5.51</b>	<b>16.83</b>	<b>8.82</b>	<b>6.34</b>	<b>6.24</b>	<b>5.43</b>
	<i>5.85</i>	<i>5.79</i>	<i>5.75</i>	<i>17.55</i>	<i>9.34</i>	<i>6.81</i>	<i>6.60</i>	<i>5.77</i>
30 days	32.56	18.27	16.68	60.34	47.85	43.22	49.67	37.19
	<b>5.99</b>	<b>6.02</b>	<b>6.01</b>	<b>17.94</b>	<b>9.62</b>	<b>6.95</b>	<b>6.82</b>	<b>5.90</b>
	<i>6.26</i>	<i>6.24</i>	<i>6.16</i>	<i>18.73</i>	<i>10.07</i>	<i>7.54</i>	<i>7.25</i>	<i>6.20</i>

Table 2: RMSE of the two models associated to the spot prices at some hours each day from one day-ahead to one month-ahead spot prices forecasts (the bold numbers are RMSE provided by model  $M_2$  while the italic are from model  $M_3$ ).

## References

- [1] Benini M, Marracci M, Pelacchi P, Venturini A. Day-Ahead Market Price Volatility Analysis in Deregulated Electricity Markets. Proc. 2002 IEEE Power Engineering Society Summer Meeting, Chicago, IL, United States 2002; 3: 1354-1359.
- [2] Box GEP, Jenkins GM. *Time series: Theory and Methods*. Springer-Verlag: New York, 1976.
- [3] Conejo AJ, Contreras, J, Espinola R, Plazas, MA. Forecasting electricity prices for a day-ahead pool-based electric energy market. International Journal of Forecasting 2005; 21: 435-462.
- [4] Diebold FX, Mariano, RS. Comparing predictive accuracy. Journal of Business and Economic Statistics 1995; 13(3): 253-263.
- [5] Diongue AK, Guégan D. Estimating parameters of a  $k$ -factor GIGARCH process. C. R. Acad. Sci. Paris 2004; I 339 6: 435-440.
- [6] Diongue AK, Guégan D, Vignal B. A  $k$ -factor GIGARCH process: estimation and application on electricity market spot prices. IEEE Proceedings of the 8<sup>th</sup> International Conference on Probabilistic Methods Applied to Power Systems, Iowa State University, Ames, Iowa 2004; 1-7.
- [7] Diongue AK, Guégan D. The  $k$ -factor Gegenbauer asymmetric Power GARCH approach for modelling electricity spot price dynamics. WP-2008.13, Publications du CES Paris 1 Panthéon-Sorbonne.
- [8] Ferrara L, Guégan D. Forecasting with  $k$ -factor Gegenbauer processes: Theory and Applications. Journal of Forecasting 2001; 20: 1-21.
- [9] Giraitis, L. and Leipus, R. (1995), A generalized fractionally differencing approach in long memory modeling, Lithuanian Mathematical Journal, 35, 65-81.

- [10] Guégan D. A New Model: The  $k$ -factor GIGARCH Process. *Journal of Signal Processing* 2000; 4: 265-271.
- [11] Guégan D. A Prospective Study of the  $k$ -factor Gegenbauer Process with Heteroscedastic Errors and an Application to Inflation Rates. *Finance India* 2003; 17: 165-197.
- [12] Guégan D. How can we define the concept of long memory? An econometric survey. *Econometric Review* 2005; 24: 113-149.
- [13] Koopman, SJ, Ooms, M, Carnero, MA. Periodic seasonal reg-ARFIMA-GARCH models for daily electricity spot prices. *Journal of the American Statistical Association* 2007; 102: 16-27.
- [14] Misiolek, A, Trueck, S, Weron, R. Point and interval forecasting of spot electricity prices: linear vs. non linear time series models. *Studies in Nonlinear Dynamics and Econometrics* 2006; 10: Article 2.
- [15] Ramsay B, Wang AJ. An electricity spot price estimator with particular reference to weekends and public holidays. *Neurocomputing* 1998; 23: 47-57.
- [16] Ray BK. Modelling long-memory processes for optimal long-range prediction. *Journal of Time Series Analysis* 1993; 14: 511-526.