Unobserved Component Model for Forecasting Electricity Prices and Their Volatilities

Carolina García-Martos, Julio Rodríguez and María Jesús Sánchez*

Abstract

The liberalization of electricity markets about ten years ago in the vast majority of developed countries has introduced the need of modelling and forecasting electricity prices. Moreover, it is interesting to compute accurate forecasts of electricity prices both for the short-run and for the long term.

Thus, there is a need of providing methodology that is able to deal with the most important features of electricity price series, that are well known for presenting not only structure in conditional mean but also their conditional variance evolves over time.

In this work we propose a new model, that allows to extract conditionally heteroskedastic common factors from the vector of electricity prices. These common factors, their relationship with the original vector of series, as well as the dynamics affecting both their conditional mean and variance are jointly estimated. Considering that ARCH and GARCH effects can be handled under the state-space formulation, the estimation of the model is carried out in this way.

*Carolina García-Martos is Graduate Student and Teaching Assistant, Escuela Técnica Superior Ingenieros Industriales, Universidad Politécnica de Madrid, Madrid, Spain (email: garcia.martos@upm.es), Julio Rodríguez is Associate Professor, Facultad de Ciencias Económicas y Empresariales, Universidad Autónoma de Madrid, Madrid, Spain (email: jr.puerta@uam.es) and María Jesús Sánchez is Associate Professor, Escuela Técnica Superior Ingenieros Industriales, Universidad Politécnica de Madrid, Madrid, Spain (email: mjsan@etsii.upm.es).

The new model proposed is applied to extract seasonal common dynamic factors as well as common volatility factors for electricity prices. Then, the estimation results are used to forecast electricity prices and their volatilities in the Spanish Market.

1 Introduction

Electricity markets are deregulated in most developed countries for more than a decade. This implies that, every day, producers and users submit their hourly bids for purchase and sale, respectively, to the market operator. The marginal price for each hour is defined as the bid submitted by the last generation unit that is needed to be producing for the whole demand being satisfied. Thus, there is a marginal price for each of the 24 hours of the day, and a 24-dimensional vector of prices is generated for each day. This procedure is illustrated in Figure 1.

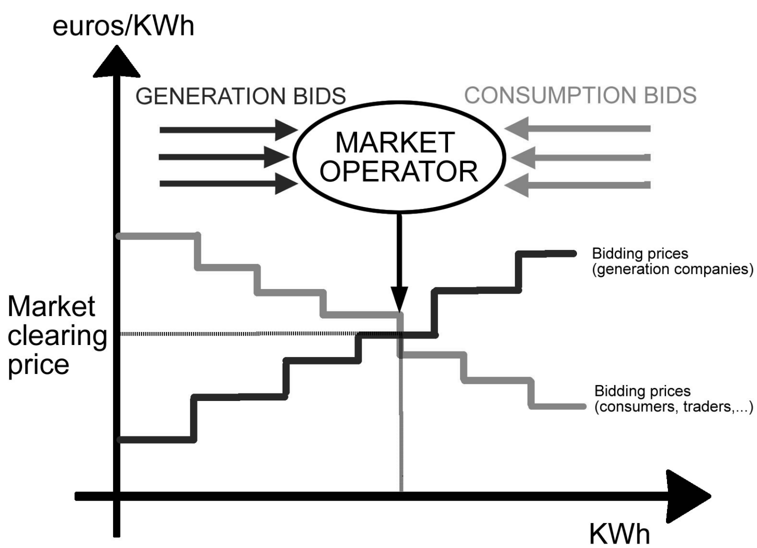

Figure 1: Market clearing price

Therefore, demand has long ago ceased to be the only variable of interest, and thus the purpose of many researchers and electrical companies is now price modelling and forecast, so a wide field of research is now opened.

The special features that electricity presents (not being able to be stored, and demand to be satisfied instantaneously) are responsible for the largely unpredictable behavior of its price. This has created the need of developing specific models that deal with this problem, since it has a great importance for a strategic sector in the economy of any country.

Besides, it must be considered that the high-dimensional vector of series of electricity prices present structure both in the mean and in the variance. So, although similarities with financial markets exist with respect to its operations, electricity price dynamics are more complex, and the mean process of electricity log prices cannot be described simply by a random walk because of its specific characteristics (Escribano et al., 2002 and Bunn and Karakatsani, 2003).

There are many authors that have developed methodology for modelling and forecasting electricity prices, and most of them deal with one-day-ahead forecasting (useful for scheduling of the power generation units). But not many of them face the problem of reducing the risk that every bilateral contract imply, this can be done by forecasting electricity prices with an horizon that covers, at least, the length of the contract, usually one year, which means long-term forecasting.

Some authors whose well known works are related to short-term forecasting are Nogales et al. (2002), Contreras et al. (2003), Conejo et al. (2005) which used time series models to produce one-step-ahead forecasts for electricity prices in some weeks, both in the Spanish and the Californian Market. More recently, García-Martos, Rodríguez and Sánchez (2007) provided a computational experiment to obtain the combination of univariate time series models with the best global performance in the period under study (1998-2003). Weron and Misiorek (2008) made a comparison among several parametric and non-parametric time series methodologies.

On the other hand, Vehvilainen and Pyykkonen (2005) provided medium-term forecasts for monthly electricity prices in the Nord Pool. They included exogenous variables that affect the prices, such as temperature. Conejo et al. (2009) used the prices of the futures market products as explanatory variables in the long term and provided the uncertainty associated to pool prices in the medium and long term in the Leipzig Market (EEX). Alonso et al. (2008) computed point forecasts and bootstrap forecast intervals for electricity prices in the Spanish market using a Seasonal Dynamic Factor Model.

Higgs (2009) focused on modelling volatility inter-relationships in the Australian National Electricity Market, after removing the structure in the conditional mean by autoregressive processes and fitting a multivariate GARCH model that is able to asses the volatility spillovers between the 4 australasian markets (New South Wales, Queensland, Victoria and South Australia).

Concerning the joint modelling of conditional mean and variance, Smith and Cottet (2006) developed a bayesian approach for the estimation of a multivariate stochastic volatility model for the 48 half-hourly time series of prices from the New South Wales market in Australia. Koopman, Ooms and Carnero (2007) provided novel periodic extensions of dynamic long-memory regression models with autoregressive conditional heteroskedastic errors for the analysis of daily electricity spot prices in several european deregulated markets.

Moreover, it is important to produce not only daily but intradaily forecasts of the prices, since some power generators are focused only on submitting their bids to cover peak prices that occur at certain hours (Panagiotelis and Smith, 2008). They provided methodology for building density forecasts of intradaily prices using multivariate skewed t distribution, and carried out the estimation using Markov Chain Monte Carlo (MCMC) techniques.

Focusing on factor models: on the one hand, Stock and Watson (2002) explored dimensionality reduction in panel data used to explain one variable, on the other hand, Peña and Box (1987) proposed a simplifying structure for a vector of time series valid only for the stationary case. Lee and Carter (1992) extended Principal Component Analysis to the case in which the variables are time series, and computed long-run forecasts of mortality and fertility rates by means of extracting a single common factor. Most recently, Peña and Poncela (2004, 2006) extended the Peña-Box model to the Non-Stationary case, and Alonso et al. (2008) provided the Seasonal Dynamic Factor Analysis. When dealing with financial time series, we can quote the factor GARCH model of Engle (1987). More recently, Connor et al. (2006) developed a dynamic approximate factor for extracting common and specific components of dynamic volatility, and García-Ferrer, González-Prieto and Peña (2008) developed a multivariate generalized independent factor GARCH model.

In this work we allow unobserved common factors extracted from the 24 hourly time series of prices to be conditionally heteroskedastic, following a seasonal VARIMA plus ARCH or GARCH processes, so not only the common structure in mean is extracted, but also the common volatility factors. The Seasonal Dynamic Fator Analysis proposed by Alonso et al. (2008), and the works by Diebold and Nerlove (1989) and Harvey, Ruiz and Sentana (1992) for dealing with conditionally heteroskedastic disturbances in state space models are applied and extended.

The rest of the paper is organized as follows. In Section 2 a descriptive analyisis of the dataset of prices in the period 1998-2007 in the Spanish Market is provided to justify the methodology presented in this paper. In Section 3 the formulation of the Conditionally Heteroskedastic Seasonal Dynamic Factor Analysis (GARCH-SeaDFA) is introduced, as well as its estimation procedure by Quasi-Maximum Likelihood. Its application for modelling the electricity prices in the Spanish Market for the period 1998-2007, as well as the computation of hourly (one-day-ahead and year-ahead) forecasts for the prices for the whole year 2008 is provided in Section 4, as well as volatility forecasts and prediction intervals that incorporate conditional heteroskedasticity. Finally, in Section 5 concludes.

2 Descriptive analysis of electricity price data

In this paper the electricity prices coming from the Spanish Market are considered in the period 1998-2008. The data coming for this Market are known for being more difficult to forecast than others (Contreras et al., 2003), so checking the performance of the methodology proposed in this paper for this market will be a challenging application.

Firstly, we provide a descriptive analysis of the data in the period 1998-2007, since the last year would be used in Section 4 to validate the performance of the GARCH-SeaDFA in terms of prediction accuracy.

Figures 2 and 3 show the boxplot of the hourly prices in the period under study (1998-2007), and only differ between them on the scale. Variations affecting both the level and variability of the prices depend on the hour of the day considered. These changes are related to the load pattern provided in Figure 4.

According to Figures 2 and 3, we have selected hours 4, 9, 12 and 21 as representative ones, just to plot the evolution over time of these hourly prices instead of all of them, that would give a less clear figure.

Figure 5 shows the four dimensional vector of these series in the period of interest (1998-2008). The last year is the one we are going to forecast (year-ahead forecasting and also one-day-ahead reestimating the model to update the data daily) using the data of the previous ones. A common pattern can be detected not only in the conditional mean but also in the conditional variance. The common patterns can also be appreciated when plotting all the series, the 24 ones (see Figure 6). The scale of the figures allow to detect that

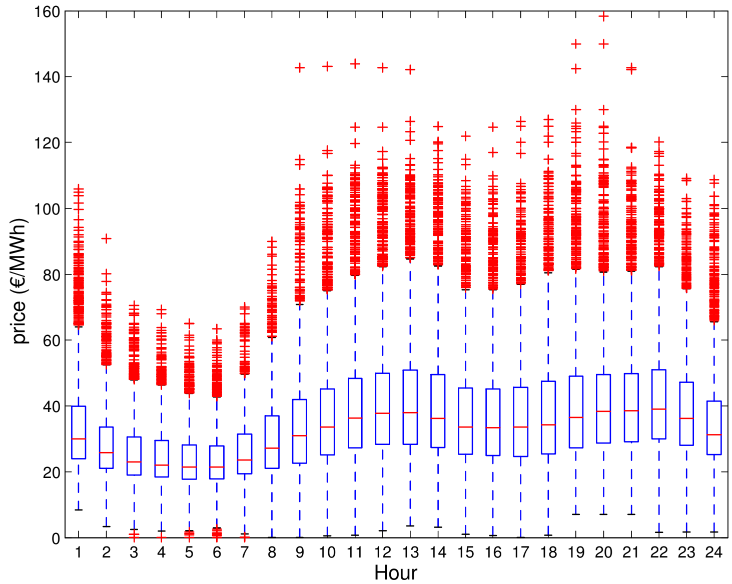

Figure 2: Boxplot of hourly prices (1998-2007).

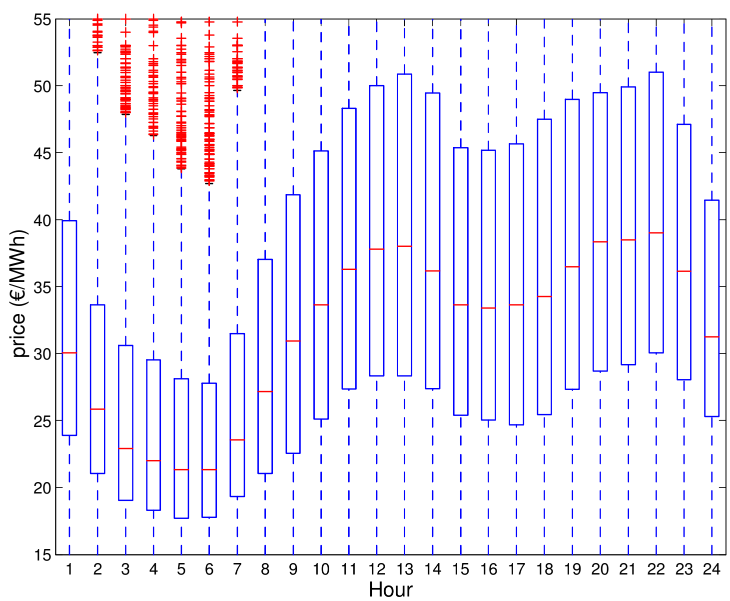

Figure 3: Boxplot of hourly prices (1998-2007). Y-Scale modified.

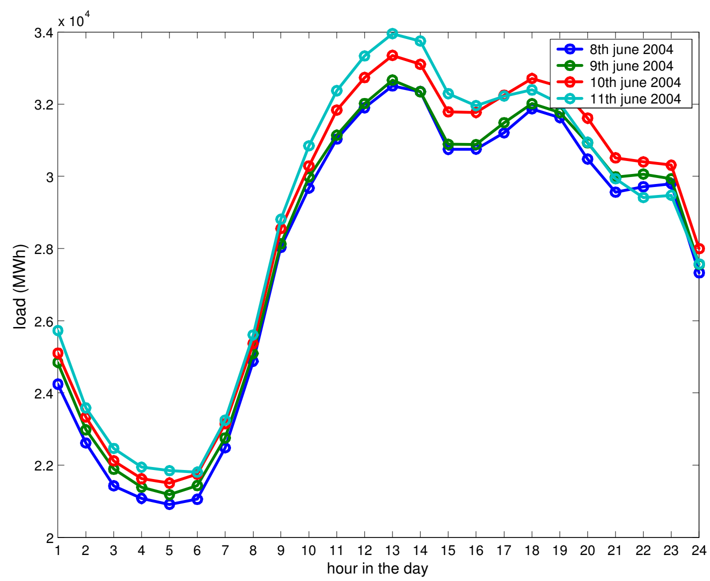

Figure 4: Examples of the hourly load pattern, 8th-11th june 2004.

the periods in which there are high prices and volatilities for the 24 hourly time series of electricity prices, so due to this common pattern not only in the dynamics of the conditional mean but also in the dynamics of the conditional variance, this evidence suggests to try and extract unobserved common factors in mean as well as common volatility factors.

Moreover, the hourly prices evolve over the time and the series present yearly patterns and weekly seasonality.

Figure 7 shows the first differences of the log-prices for 4 of the 24 hourly time series considered for the Spanish Market in the period 1998-2007. Clear patterns of volatility clustering are shown. Apart from the typical features of daily prices from other financial markets, it can be emphasized that the dynamics of electricity prices is more intrincate, since not only the conditional variance is evolving over time, but also the conditional mean.

Besides, a common pattern in the volatility clusters can be observed in Figure 7, since high and low volatilities correspond to same dates for all the hours presented. So, the extraction of common volatility factors, as proposed in this paper is justified bearing in mind those figures.

Once a common pattern has been detected not only in the dynamics of the conditional mean but also in the conditional variance of the electricity prices in the Spanish Market, the Conditionally Heteroskedastic Dynamic Factor Analysis is presented in the next section. It will be used to model the 24-dimensional vector of prices in the period 1998-2007 and to

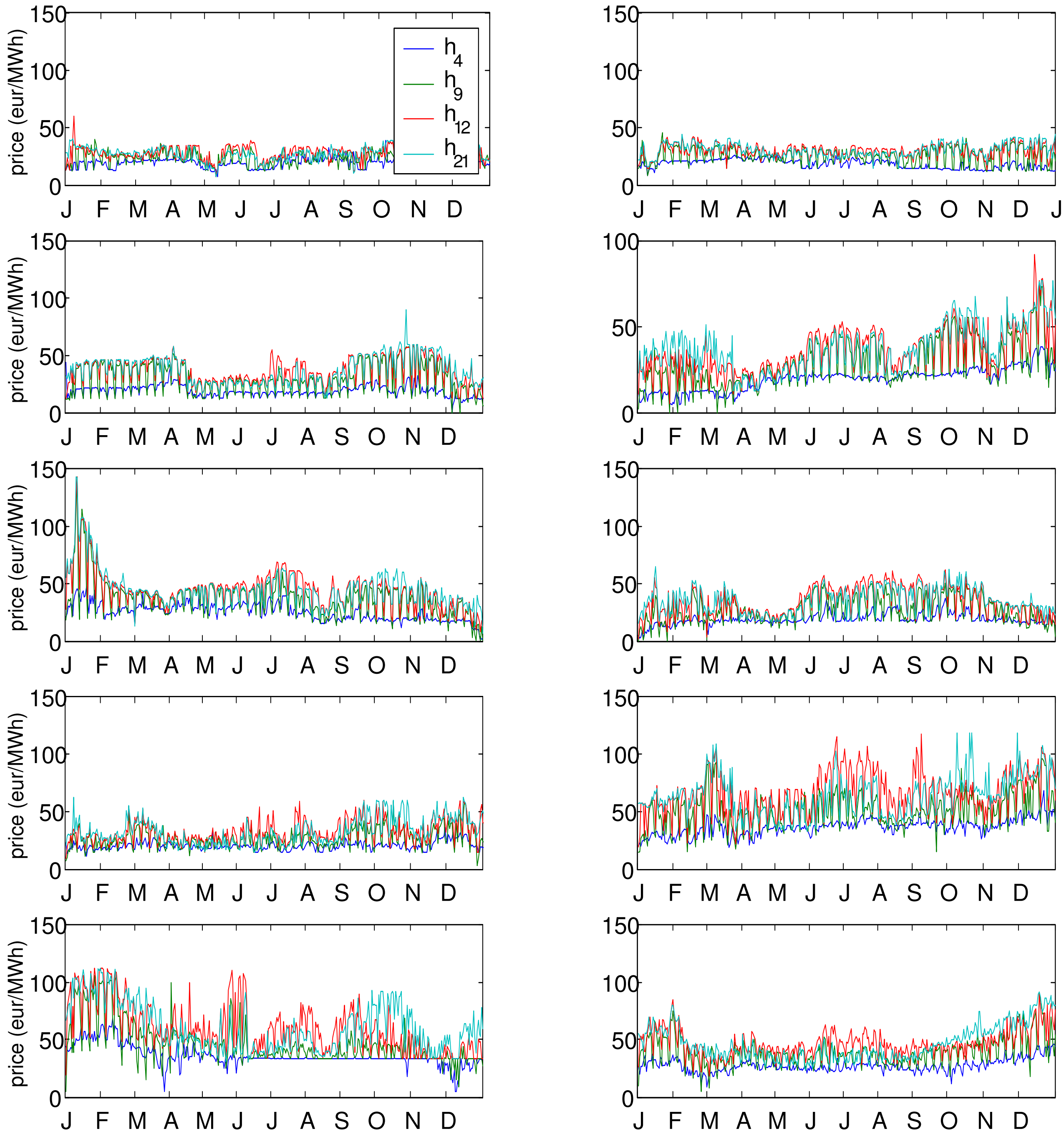

Figure 5: Left to right and top to bottom, prices in the years 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006 and 2007.

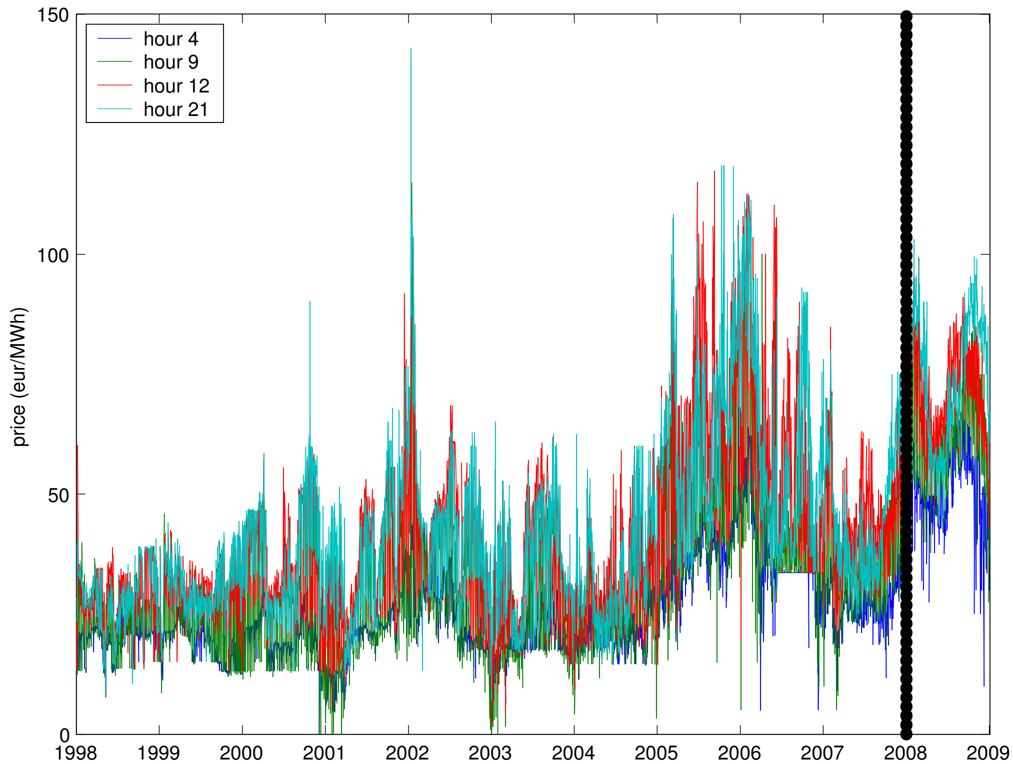

Figure 6: Prices for hours 4, 9, 12 and 21 in the period 1998-2008. Common pattern in conditional mean as well as in the conditional pattern.

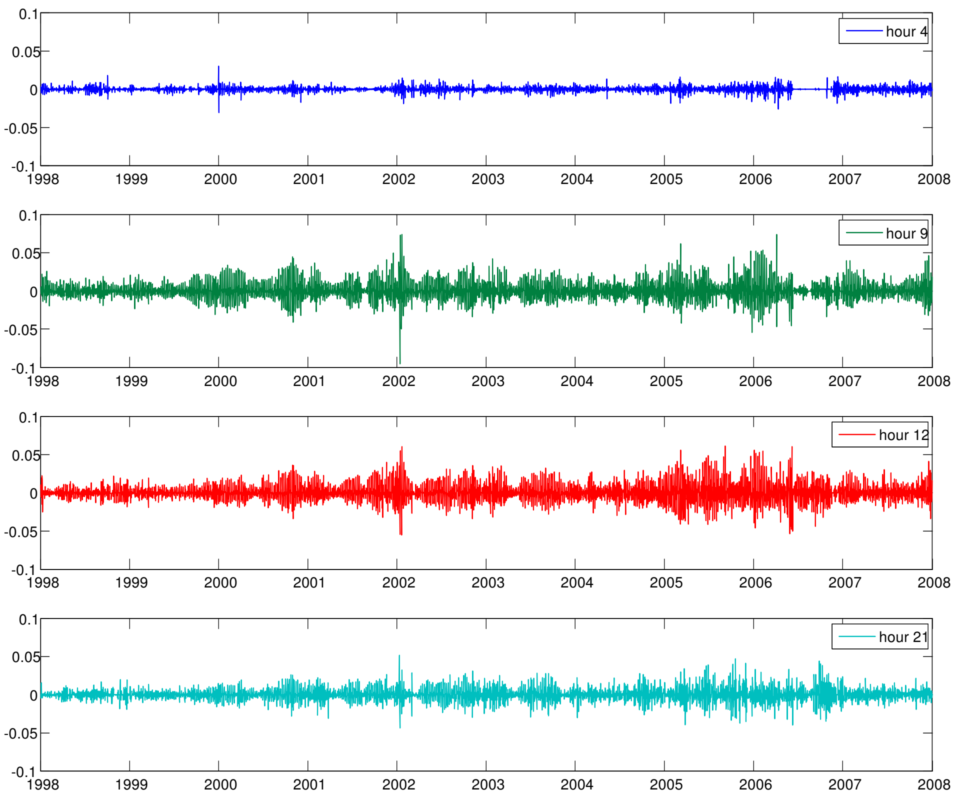

Figure 7: First differences of log-prices. Hours 4, 9, 12 and 21.

produce forecasts for the whole year 2008.

3 Formulation of the Model

3.1 Seasonal Dynamic Factor Analysis with Homoskedastic disturbances

Alonso et al. (2008) developed the Seasonal Dynamic Factor Analysis (SeaDFA), which is able to extract a r-dimensional vector of unobserved seasonal common factors from a m-dimensional observed vector of time series (where $r < m$). They assume that vector $\mathbf{y}_t$ can be written as a linear combination of the unobserved common factors, $\mathbf{f}_t$, plus $\boldsymbol{\varepsilon}_t$, to which we will refer from now on as specific components or specific factors:

$$\mathbf{y}_t = \Omega_1 \mathbf{f}_t + \boldsymbol{\varepsilon}_t,\ \ \boldsymbol{\varepsilon}_t \sim NID(0, S) \tag{1}$$

Besides, these common factors are allowed to follow a multiplicative seasonal Vector AutoRegressive Integrated Moving Average, VARIMA$(p, d, q) \times (P, D, Q)_s$ model:

$$(\mathbf{I} - B)^d (\mathbf{I} - B^s)^D \boldsymbol{\phi}(B) \boldsymbol{\Phi}(B^s) \mathbf{f}_t = \mathbf{c}_1 + \boldsymbol{\theta}(B) \boldsymbol{\Theta}(B^s) \mathbf{w}_t^1, \tag{2}$$

where $\mathbf{c}_1$ is the r-dimensional constant of the model of the common factors, $\boldsymbol{\phi}(B) = (\mathbf{I} - \boldsymbol{\phi}_1 B - \boldsymbol{\phi}_2 B^2 - \ldots - \boldsymbol{\phi}_p B^p)$, $\boldsymbol{\Phi}(B^s) = (\mathbf{I} - \boldsymbol{\Phi}_1 B^s - \boldsymbol{\Phi}_2 B^{2s} - \ldots - \boldsymbol{\Phi}_p B^{Ps})$, $\boldsymbol{\theta}(B) = (\mathbf{I} - \boldsymbol{\theta}_1 B - \boldsymbol{\theta}_2 B^2 - \ldots - \boldsymbol{\theta}_q B^q)$ and $\boldsymbol{\Theta}(B^s) = (\mathbf{I} - \boldsymbol{\Theta}_1 B - \boldsymbol{\Theta}_2 B^{2s} - \ldots - \boldsymbol{\Theta}_Q B^{Qs})$ are polynomial matrices $r \times r$, B is the backshift operator such that $B\mathbf{y}_t = \mathbf{y}_{t-1}$, the roots of $|\boldsymbol{\phi}(B)| = 0$ and $|\boldsymbol{\Phi}(B^s)| = 0$ are on or outside the unit circle, the roots of $|\boldsymbol{\theta}(B)| = 0$ and $|\boldsymbol{\Theta}(B^s)| = 0$ are outside the unit circle and $\mathbf{w}_t^1 \sim \mathbf{N}_r(\mathbf{0}, \mathbf{Q}_1)$ is serially uncorrelated for all leads and lags.

The homoskedastic SeaDFA is given by equations (1) and (2), and can be rewritten under the state-space formulation, just reformulating them as an observation and transition equation, (3) and (4), respectively:

$$\mathbf{y}_t = \Omega \mathbf{F}_t + \boldsymbol{\varepsilon}_t,\ \ \boldsymbol{\varepsilon}_t \sim NID(0, S), \tag{3}$$

$$\mathbf{F}_t = \mathbf{c} + \boldsymbol{\Psi} \mathbf{F}_{t-1} + \mathbf{w}_t, \mathbf{w}_t \sim NID(0, \mathbf{Q}). \tag{4}$$

Concerning the evolution of the common factors over time, in general, taking into account Ansley and Kohn (1986), any multiplicative seasonal VARIMA $(p, d, q) \times (P, D, Q)_s$ model as given by equation (2), can be easily rewritten as a transition equation like (4). For example, in the case of an unique seasonal ARI$(p,d)\times(P,D)_{12}$ common factor extracted from monthly data ($s=12$), $\mathbf{\Psi}$ has dimensions $b\times b$, where $b=r\cdot(s\cdot(D+P)+d+p)$.

$$\mathbf{Q}=\begin{pmatrix}\mathbf{Q}_1 & \mathbf{0}_{r\times(b-r)}\\ \mathbf{0}_{(b-r)\times r} & \mathbf{0}_{(b-r)\times(b-r)}\end{pmatrix}_{b\times b},\ \mathbf{c}=\begin{pmatrix}\mathbf{c}_1\\ \mathbf{0}_{(b-r)\times 1}\end{pmatrix}_{b\times 1 b\times 1} \text{ and } \mathbf{w}_t=\begin{pmatrix}\mathbf{w}_t^1\\ \mathbf{0}_{(b-r)\times 1}\end{pmatrix}.$$

Focusing on equation (3), $\mathbf{\Omega}=(\mathbf{\Omega}_1\ \mathbf{0}_{r\times(b-1)})$. For equations (3) and (4) $\mathbf{F}_t$ contains $\mathbf{f}_t$ and its $(b-1)$ lags.

We also assume that the disturbances $\boldsymbol{\varepsilon}_t$ and $\mathbf{w}_t$ are uncorrelated for all lags, i.e. $E(\mathbf{w}_t\boldsymbol{\varepsilon}'_{t-h})=0,\forall h$.

Moreover, the model is unidentified under rotations, this problem is solved imposing restrictions like $\mathbf{Q}_1=\mathbf{I}$ or $\mathbf{\Omega}_1'\mathbf{\Omega}_1=\mathbf{I}$, as well as $\omega_{ij}=0$, for $j>i$, where the ω_{ij}'s are the elements in $\mathbf{\Omega}_1$. This is not restrictive at all since a rotation can be applied for interpretation purposes when needed. For further details on this identification issue see Geweke and Singleton (1981).

3.2 Conditionally Heteroskedastic Seasonal Dynamic Factor Analysis

In this paper we introduce the possibility of the unobserved common factors having structure both in mean and variance, since we allow for seasonal VARIMA+ARCH/GARCH (Generalized Autorregressive Conditionally Heteroskedastic, Engle, 1982 and Bollerslev, 1986) unobserved common factors. For this purpose the conditionally heteroskedastic disturbances $\mathbf{w}_t^*$ are added in the transition equation:

$$\mathbf{y}_t = \Omega\mathbf{F}_t+\boldsymbol{\varepsilon}_t,\ \boldsymbol{\varepsilon}_t\sim NID(0,S) \tag{5}$$

$$\mathbf{F}_t = \mathbf{c}+\mathbf{\Psi}\mathbf{F}_{t-1}+\mathbf{w}_t^*, \tag{6}$$

Disturbances $\boldsymbol{\varepsilon}_t,\mathbf{w}_t^*$, which appear in equations (5) and (6) are mutually independent, and $\mathbf{w}_t^*=\begin{pmatrix}\mathbf{w}_t^{*1}\\ \mathbf{0}_{(b-r)\times 1}\end{pmatrix}_{b\times 1}$. In the simplest case, we allow the disturbances of the transition equation to follow univariate ARCH(1) models:

$$\mathbf{w}_t^{*1}|I_{t-1} \sim N\left(\mathbf{0}_{r\times 1},\ diag(\sigma_{1t}^2, \sigma_{2t}^2, ..., \sigma_{rt}^2)\right) = N(\mathbf{0}_{r\times 1}, \mathbf{Q}_{1t}),$$
$$\mathbf{w}_t^{*}|I_{t-1} \sim N\left(\mathbf{0}_{b\times 1},\ diag(\sigma_{1t}^2, \sigma_{2t}^2, ..., \sigma_{rt}^2, \mathbf{0}_{(b-1)})\right) = N(\mathbf{0}_{b\times 1}, \mathbf{Q}_{t})$$
$$w_{j,t}^* = \sigma_{jt} a_{jt},\ a_{jt} \sim NID(0,1),\ \sigma_{jt}^2 = \alpha_{0j} + \alpha_j w_{j,t-1}^{*2}, \text{for } j = 1, ..., r \tag{7}$$

where I_{t-1} refers to all the information available at time $t-1$.

It cannot be forgotten that this is a Dynamic Factor Analysis, and that the restrictions provided in the homoskedastic case remain here, since otherwise the model would be unidentified under rotations.

3.2.1 Quasi-Maximum Likelihood Estimation and Augmented Kalman filter

For estimating the parameters involved in this model, we must maximize the log-likelihood function, and it is well known that this function is calculated for models expressed under state space formulation using the expression:

$$\log L = -\frac{1}{2}\sum_{t=1}^{T} \log((2\pi)^n |\Sigma_t|) - \frac{1}{2}\sum_{t=1}^{T} \upsilon_t \Sigma_t^{-1} \upsilon_t',$$

where υ_t are the innovations and Σ_t its variance-covariance matrix. For calculating these quantities, υ_t and Σ_t, the Kalman Filter must be run, and some difficulties arise when conditional heteroskedasticity is present in the disturbances. The equations of the Kalman Filter (see Durbin and Koopman, 2001 for details about Kalman filtering and likelihood evaluation) are:

$$\mathbf{F}_{t|t-1} = \mathbf{c} + \mathbf{\Psi}\mathbf{F}_{t-1|t-1} \tag{8}$$
$$P_{t|t-1} = \mathbf{\Psi} P_{t-1|t-1}' \mathbf{\Psi}' + \mathbf{Q}_t = \mathbf{\Psi} P_{t-1|t-1}' \mathbf{\Psi}' + diag(\sigma_{1t}^2, \sigma_{2t}^2, ..., \sigma_{rt}^2, \mathbf{0}_{(b-r)}) \tag{9}$$
$$\upsilon_t = \mathbf{y}_t - \Omega \mathbf{F}_{t|t-1} \tag{10}$$
$$\Sigma_t = \Omega P_{t|t-1}' \Omega + S \tag{11}$$
$$\mathbf{F}_{t|t} = \mathbf{F}_{t|t-1} + P_{t|t-1}' \Omega \Sigma_t^{-1} \upsilon_t \tag{12}$$
$$P_{t|t} = P_{t|t-1} - P_{t|t-1}' \Omega \Sigma_t^{-1} \Omega P_{t|t-1} \tag{13}$$

When processing equation (9) the computation of the matrix whose diagonal contains $(\sigma_{1t}^2, \sigma_{2t}^2, ..., \sigma_{rt}^2)$ is needed. Bearing in mind (7), we realize that for obtaining $(\sigma_{1t}^2, \sigma_{2t}^2, ..., \sigma_{rt}^2)$, the terms $w^{*2}_{j,t-1}$, for $j = 1...,r$, must be calculated. Since they are unobservable, the equations of the Kalman Filter (8) to (13) cannot be operated directly. For solving this problem Harvey, Ruiz and Sentana (1992) proposed substituting $w^{*2}_{j,t-1}$ in (7) by their condicional expectations:

$$\sigma^2_{jt} = \alpha_{0j} + \alpha_{1j} w^{*2}_{j,t-1} \rightarrow \sigma^2_{jt} = \alpha_{0j} + \alpha_{1j} E(w^{*2}_{j,t-1}|I_{t-1}) \text{ for } j = 1, ..., r. \tag{14}$$

Here, we propose extending the Homoskedastic Seasonal Dynamic Factor Analysis to the Conditionally Heteroskedastic case, by means of the idea introduced by Harvey, Ruiz and Sentana (1992), which consists of including the conditionally heteroskedastic shocks of the common factors into the "original" state vector. This is needed because the expectation of $w^{*2}_{j,t-1}$ conditional on the information available at time $t-1$, i.e. $E(w^{*2}_{j,t-1}|I_{t-1})$, must be calculated. This quantity is obtained as an output from the Kalman Filter if $w^{*}_{j,t-1}$ is a latent or state variable. For the transition equation, incorporating $\mathbf{w}^*_t$ in the "original" state vector gives:

$$\begin{pmatrix} \mathbf{F}_t \\ \mathbf{w}^{1*}_t \end{pmatrix} = \begin{pmatrix} \mathbf{c} \\ \mathbf{0}_{r\times 1} \end{pmatrix} + \begin{pmatrix} \mathbf{\Psi}_{b\times b} & \mathbf{0}_{b\times r} \\ \mathbf{0}_{r\times b} & \mathbf{0}_{r\times r} \end{pmatrix} \begin{pmatrix} \mathbf{F}_{t-1} \\ \mathbf{w}^{*}_{t-1} \end{pmatrix} + \begin{pmatrix} \mathbf{w}^{*}_t \\ \mathbf{w}^{1*}_t \end{pmatrix}$$

$$\begin{pmatrix} \mathbf{F}_t \\ \mathbf{w}^{1*}_t \end{pmatrix} = \begin{pmatrix} \mathbf{c} \\ \mathbf{0}_{r\times 1} \end{pmatrix} + \begin{pmatrix} \mathbf{\Psi}_{b\times b} & \mathbf{0}_{b\times r} \\ \mathbf{0}_{r\times b} & \mathbf{0}_{r\times r} \end{pmatrix} \begin{pmatrix} \mathbf{F}_{t-1} \\ \mathbf{w}^{*}_{t-1} \end{pmatrix} + \begin{pmatrix} I_r \\ 0_{(b-r)\times r} \\ I_r \end{pmatrix} \mathbf{w}^{1*}_t, \tag{15}$$

and its matrix form:

$$\mathbf{F}^A_t = \mathbf{c}^A + \mathbf{\Psi}^A \mathbf{F}^A_{t-1} + \mathbf{G}^A \mathbf{v}^A_t, \tag{16}$$

where $\mathbf{F}^A_t = \begin{pmatrix} \mathbf{F}_t \\ \mathbf{w}^{1*}_t \end{pmatrix}$, $\mathbf{c}^A = \begin{pmatrix} \mathbf{c} \\ \mathbf{0}_{r\times 1} \end{pmatrix}$, $\mathbf{G}^A = \begin{pmatrix} I_r \\ 0_{(b-r)\times r} \\ I_r \end{pmatrix}$ and $\mathbf{\Psi}^A = \begin{pmatrix} \mathbf{\Psi} & \mathbf{0}_{b\times r} \\ \mathbf{0}_{r\times b} & \mathbf{0}_{r\times r} \end{pmatrix}$. Be aware also that in equation (15), the r-dimensional, $\mathbf{w}^{1*}_t$ are playing both roles of state-vector and disturbances, $\mathbf{v}^A_t$ denotes $\mathbf{w}^{1*}_t$ when it is a disturbance. The conditional expectation $E(\mathbf{v}^A_t \mathbf{v}^{A\prime}_t|I_{t-1}) = Q^A_t = diag(\sigma^2_{1t}, ..., \sigma^2_{rt})$.

Moreover, the observation equation must be replaced by:

$$\mathbf{y}_t = \begin{pmatrix} \mathbf{\Omega} & \mathbf{0}_{m\times r} \end{pmatrix} \begin{pmatrix} \mathbf{F}_t \\ \mathbf{w}^{1*}_t \end{pmatrix} + \boldsymbol{\varepsilon}_t, \tag{17}$$

or in its matrix formulation:

$$\mathbf{y}_t = \Omega^A \mathbf{F}_t^A + \boldsymbol{\varepsilon}_t, \text{ and } E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = S, \tag{18}$$

where $\Omega^A = \begin{pmatrix} \boldsymbol{\Omega} & \mathbf{0}_{m\times r} \end{pmatrix}$.

Once we have the formulation of the Conditionally Heteroskedastic SeaDFA given by equations (18) and (16), the Kalman filter can be run for this "augmented" formulation. Kalman filter recursions for the augmented formulation are given by:

$$\begin{aligned}
\mathbf{F}^A_{t|t-1} &= \mathbf{c}^A + \Phi^A \mathbf{F}^A_{t-1|t-1} \\
P^A_{t|t-1} &= \boldsymbol{\Psi}^A P^A_{t-1|t-1} \boldsymbol{\Psi}^{A\prime} + \mathbf{G}^A Q^A_t \mathbf{G}^{A\prime} \\
\upsilon^A_t &= y_t - \Omega^A \mathbf{F}^A_{t|t-1} \\
\Sigma^A_t &= \Omega^A P^A_{t|t-1} \Omega^{A\prime} + S \\
\mathbf{F}^A_{t|t} &= \mathbf{F}^A_{t|t-1} + P^A_{t|t-1} \Omega^{A\prime} (\Sigma^A_t)^{-1} \upsilon^A_t \\
P^A_{t|t} &= P^A_{t|t-1} - P^A_{t|t-1} \Omega^A (\Sigma^A_t)^{-1} \Omega^A P^A_{t|t-1}
\end{aligned} \tag{19}$$

For the calculation of Q^A_t in equation (19), $\sigma^2_{1t}, ..., \sigma^2_{rt}$ must be computed. For this purpose the approximation given in equation (14) is used, and $E(w^*_{j,t-1}|I_{t-1})$ and $E(w^{*2}_{j,t-1}|I_{t-1})$ for $j = 1, ..., r$ must be calculated. But now, in the "augmented" formulation, this is easy since $\mathbf{w}^*_{t-1}$ can be expressed in terms of :

$$\mathbf{w}^*_{t-1} = E(\mathbf{w}^*_{t-1}|I_{t-1}) - (E(\mathbf{w}^*_{t-1}|I_{t-1}) - \mathbf{w}^*_{t-1}), \tag{20}$$

where $E(\mathbf{w}^*_{t-1}|I_{t-1})$ are given by the last r elements in $\mathbf{f}^A_{t-1|t-1}$. Besides, concerning the variances, they can be expressed as:

$$E(\left(\mathbf{w}^*_{t-1}\right)^2 |I_{t-1}) = \left(E(\mathbf{w}^*_{t-1}|I_{t-1})\right)^2 + E\left[(\mathbf{w}^*_{t-1} - E(\mathbf{w}^*_{t-1}|I_{t-1}))^2\right], \tag{21}$$

where $E\left[(\mathbf{w}^*_{t-1} - E(\mathbf{w}^*_{t-1}|I_{t-1}))^2\right]$ are given by the elements $(b+1)$ to $(b+r)$ in the diagonal of $P^A_{t-1|t-1}$. So the elements in the diagonal of Q^A_t, i.e., $\sigma^2_{1t}, ..., \sigma^2_{rt}$ are calculated using (14), (20) and (21), and all the involved quantities are given by the Augmented Kalman Filter:

$$\begin{aligned}
\sigma^2_{jt} &= \alpha_{0j} + \alpha_{1j} E(w^{*2}_{j,t-1}|I_{t-1}) = \\
&= \alpha_{0j} + \alpha_{1j} \left\{ \left(E(w^{*2}_{j,t-1}|I_{t-1})\right)^2 + E\left[(w^*_{j,t-1} - E(w^*_{j,t-1}|I_{t-1}))^2\right] \right\}.
\end{aligned} \tag{22}$$

Finally, the parameters of the model are estimated maximizing the expression for the log-likelihood in the "augmented" formulation:

$$\log L^A = -\frac{1}{2}\sum_{t=1}^{T}\log((2\pi)^n|\Sigma_t^A|) - \frac{1}{2}\sum_{t=1}^{T}\upsilon_t^A\left(\Sigma_t^A\right)^{-1}\upsilon_t^{A\prime}. \tag{23}$$

To let the model being more general, it should manage not only transitory disturbances following ARCH processes, but also GARCH ones, since in practice the conditional variance of most of the data can be sufficiently described by a GARCH(1,1) model (for a revision of these models see Tsay, 2005). So, if equation (7) is modified in such a way that the disturbances are allowed to follow univariate GARCH(1,1) processes, then the evolution over time of the transitory disturbances is given by:

$$\mathbf{w}_t^*|I_{t-1} \sim N\left(\mathbf{0}_{r\times 1},\ diag(\sigma_{1t}^2, \sigma_{2t}^2, ..., \sigma_{rt}^2, \mathbf{0}_{(b-r)})\right),$$
$$w_{j,t}^* = \sigma_{jt}a_{jt},\ a_{jt} \sim NID(0,1),\ \sigma_{jt}^2 = \alpha_{0j} + \alpha_j w_{j,t-1}^{*2} + \beta_j\sigma_{j,t-1}^2, \text{for } j = 1, ..., r \tag{24}$$

When the values σ_{jt}^2 are needed in the Augmented Filter to calculate Q_t^A in equation (19), the slight modification introduced consists of considering an additional term $\beta_{1j}\sigma_{j,t-1}^2$, since $\sigma_{jt}^2 = \alpha_{0j} + \alpha_{1j}w_{j,t-1}^{*2} + \beta_{1j}\sigma_{j,t-1}^2$. Thus, the following approximation is applied to calculate Q_t^A:

$$\sigma_{jt}^2 = \alpha_{0j}+\alpha_{1j}w_{j,t-1}^{*2}+\beta_{1j}\sigma_{j,t-1}^2 \rightarrow \sigma_{jt}^2 = \alpha_{0j}+\alpha_{1j}E(w_{j,t-1}^{*2}|I_{t-1})+\beta_{1j}E(\sigma_{j,t-1}^2|I_{t-2}),\ j = 1, ..., r.$$

where $E(w_{j,t-1}^{*2}|I_{t-1})$ is obtained in (22) and $E(\sigma_{j,t-1}^2|I_{t-2}) = \widehat{\sigma}_{j,t-1}^2$, for $j = 1, ..., r$, and it is available from $Q_{t-1}^A = diag(\sigma_{1,t-1}^2, ..., \sigma_{r,t-1}^2)$.

4 Application to electricity price data

There are two interesting problems to be solved concerning electricity price forecasting. On the one hand, power generation units must be scheduled for the forthcoming day (24-hour ahead) trying to maximize their profits. For this purpose the disposal of accurate one-step-ahead forecasts is crucial.

On the other hand, the risk that bilateral contracts imply should be reduced. By means of a bilateral contract consumers agree to purchase some amount of energy at a certain price to a seller for a long period of time, usually a year. But everyday the sellers must go to the *pool* and submit its own bid to buy the energy needed to cover the needs of all their customers. So, there is also a need of computing accurate year-ahead forecasts.

4.1 The model for electricity price data in the Spanish Market (1998-2007).

In this subsection, the results obtained when estimating the Conditionally Heteroskedastic Seasonal Dynamic Factor Analysis for the prices in the Spanish Market are provided. In Section 2, some of the most important features of prices were remarked, illustrating them by means of a descriptive analysis. A common pattern (not only in conditional mean but also in conditional variance) arose. So, conditionally heteroskedastic common factors could be extracted. The model proposed and explained in the previous section, is able obtain from a vector of time series, the common factors for the common structure in conditional mean. Since these factors can present ARCH or GARCH disturbances, also common volatility factors are extracted.

First of all, and using the test proposed by Peña and Poncela (2004) the number of common factors is selected. Concerning the data under, the $m = 24$ hourly time series of electricity prices in the period 1998-2007, $r = 2$ common factors are extracted. Moreover, electricity prices exhibit a weekly seasonal pattern since there is an instantaneous relationship between load and price and the consumption heavily depends on the day of the week (see García-Martos et al., 2007), so seasonality of order $s = 7$ is present. The model chosen for the common factors is a VARIMA(1,0,0)$\times$(1,1,0)$_7$ with univariate GARCH(1,1) disturbances. So, the dimensions of the transition matrix $\mathbf{\Psi}$ is $b \times b$, with $b = 30$. The celebrated GARCH(1,1), Bollerslev (1986) is well-known for being able to capture the structure in the vast majority of series whose conditional variance evolves over time.

Thus, the GARCH-SeaDFA model given by (17) and (15) that is estimated for the the particular case of Spanish Market in the period 1998-2007 is given by the following equations:

$$
\begin{aligned}
(\mathbf{y}_t)_{24\times 1} &= \begin{pmatrix}(\mathbf{\Omega}_1)_{24\times 2} & \mathbf{0}_{24\times 28} & \mathbf{0}_{24\times 2}\end{pmatrix} \begin{pmatrix}(\mathbf{F}_t)_{28\times 1} \\ (\mathbf{w}_t^{1*})_{2\times 1}\end{pmatrix} + (\boldsymbol{\varepsilon}_t)_{24\times 1}, \\
\begin{pmatrix}\mathbf{F}_t \\ \mathbf{w}_t^{1*}\end{pmatrix} &= \begin{pmatrix}\mathbf{c}_{30\times 1} \\ \mathbf{0}_{2\times 1}\end{pmatrix} + \begin{pmatrix}\mathbf{\Psi}_{30\times 30} & \mathbf{0}_{30\times 2} \\ \mathbf{0}_{2\times 30} & \mathbf{0}_{2\times 2}\end{pmatrix} \begin{pmatrix}\mathbf{F}_{t-1} \\ \mathbf{w}_{t-1}^{*}\end{pmatrix} + \begin{pmatrix} I_2 \\ \mathbf{0}_{(28)\times 2} \\ I_2 \end{pmatrix} \mathbf{w}_t^{1*},
\end{aligned}
$$

where $\omega_{12} = 0$, $\mathbf{w}_t^{1*}$ has dimension 2 and contains the GARCH disturbances, and $\mathbf{c} = (c_1, c_2, \mathbf{0}_{1\times 28})$. Notice that from the 30×30 matrix $\mathbf{\Psi}$, only 8 parameters must be estimated (those involved . This parameters, or linear and nonlinear relationships between them, appear in the first $r = 2$ rows. All the elements in rows from $r + 1$ to b, are ones or zeros (see Ansley and Kohn, 1986 for details and Alonso et al., 2008, for an example).

Figure 8 shows the 24- dimensional vector of hourly prices as well as the conditionally heteroskedastic common factors estimated.

In Figure 9 the loading matrix that relates observed series of prices and unobserved common factors is provided. Loads of hourly series in the first factor are all positive and higher in those hours in which the variability is higher according to Figure 3. Besides, the loads relating observed series and second common factors are positive for these hours in the nigth and positive. There is a clear realtionship between the loads estimated and the pattern of load given in Figure 4.

Figure 10 shows the common factors, their estimated GARCH(1,1) disturbances, $\mathbf{w}_t^*$, and volatilities σ_{1t}^2 and σ_{2t}^2, where $\mathbf{w}_t^*|I_{t-1} \sim N\left(\mathbf{0}_{2\times 1}, diag(\sigma_{1t}^2, \sigma_{2t}^2)\right)$.

The results here provided, are used to compute long and short-term forecasts of electricity prices in the Spanish Market. These forecasts are provided in the next two subsections.

4.2 Long-run forecasting

The conditionally heteroskedastic dynamic factor model estimated for the data in the period 1998-2007 is used to compute forecasts for the whole year 2008, which are very useful for the negotiation of bilateral contracts as explained at the beginning of this section. No matter which day in 2008 the hourly forecasts are being computed for, the last data used corresponds to the 31^{st} December 2007.

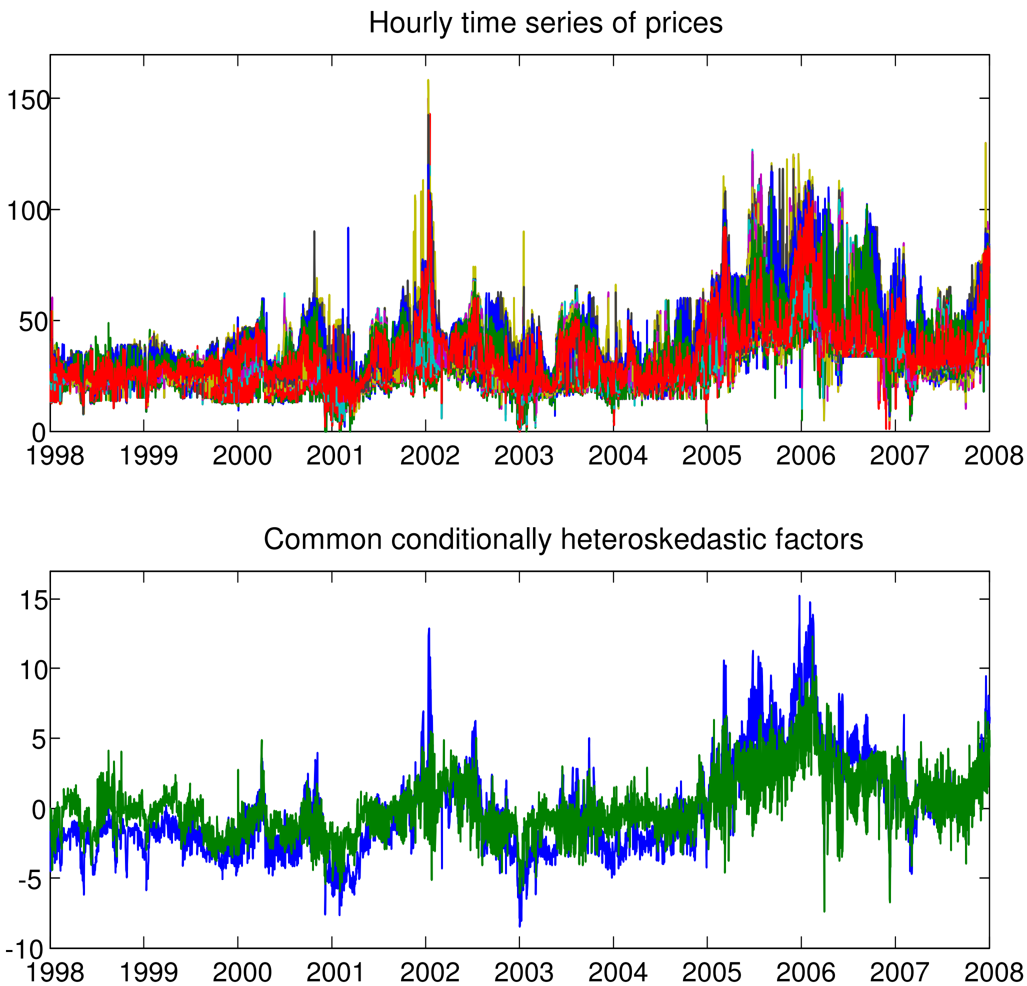

Figure 8: Vector of hourly prices and unobserved conditionally heteroskedastic common factors (1998-2007).

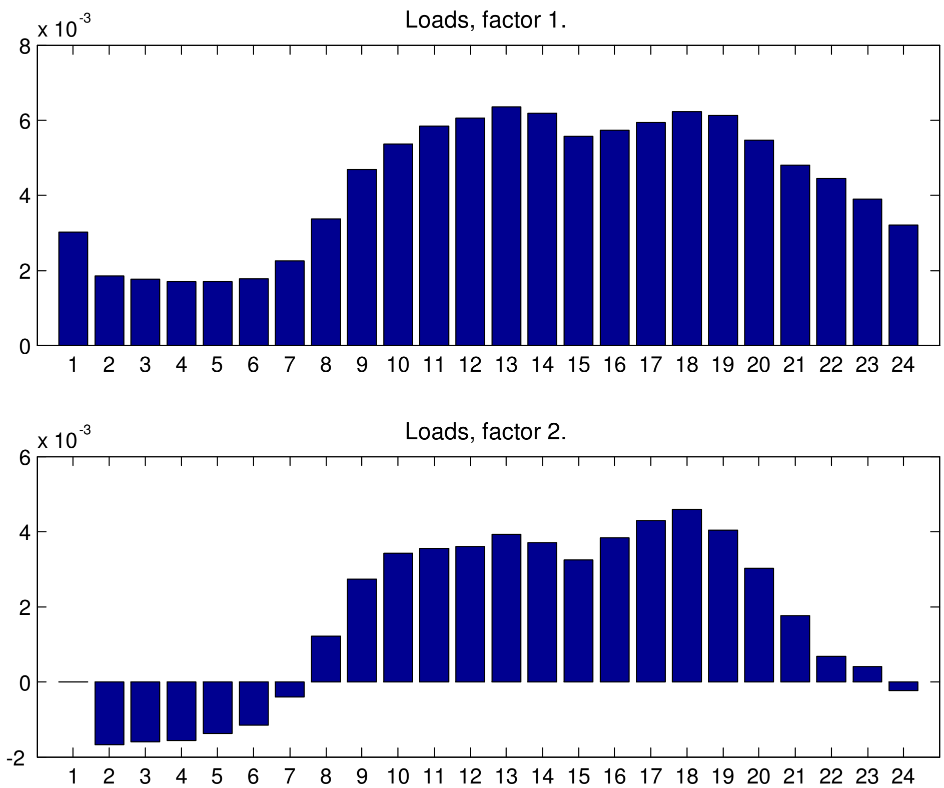

Figure 9: Estimated loading matrix, $\widehat{\Omega}_A$.

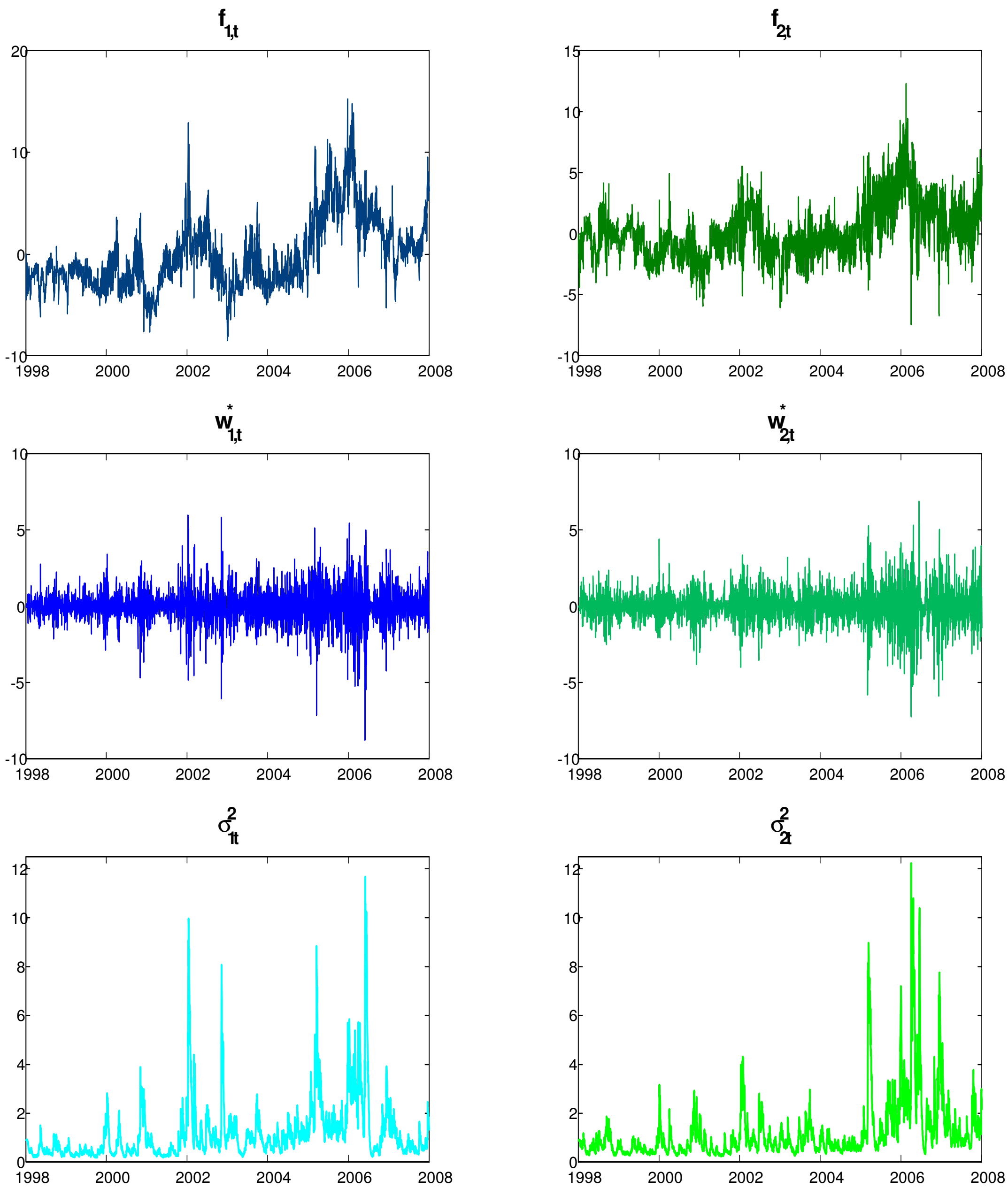

Figure 10: Common factors, $f_{1,t}$ and $f_{2,t}$, conditionally heteroskedastic disturbances w_{1t} and $w_{2,t}$, and its σ^2_{1t} and σ^2_{2t}.

The accuracy metrics used to check the performance of the model proposed in this paper, are those usually encountered in previous works (Conejo et al., 2005). These accuracy metrics are the Mean Average Percentage Error (MAPE) and the MAPE2, defined respectively as:

$$
\begin{aligned}
MAPE &= \frac{1}{D}\sum_{d=1}^{D} emean_d, \\
MAPE2 &= \frac{1}{D}\sum_{d=1}^{D} emedian_d,
\end{aligned}
$$

where $emean_d = mean(e_{1,d}, ...e_{24,d})$ and $emedian_d = median(e_{1,d}, ...e_{24,d})$. The percentage error at hour h, day d, $e_{h,d}$ is given by the expression:

$$
e_{h,d} = \frac{|p_{h,d} - \widehat{p}_{h,d}|}{p_{h,d}},
$$

where $p_{h,d}$ is the real price at hour hour h of day d, and $\widehat{p}_{h,d}$ is its forecast computed using the conditionally heteroskedastic dynamic factor analysis.

In Table 1 the monthly MAPE and MAPE2 for the whole year 2008, using the model estimated with the data from 1998 up to 2007 are shown.

It should be pointed out, that the models usualy handled to compute one-day-ahead forecasts are no longer valid for year ahead forecasting (long-term). For example, using the Mixed Model provided in García-Martos et al. (2007) the errors for the predicitions in 2008 are above 35%, much bigger than the 16.15% we get using the Conditionally Heteroskedastic Model developed in this paper. This Mixed Model is used as benchmarking since is the one with the smallest errors when dealing with short-term forecasts. Moreover, it should be emphasized that the Mixed Model is doing intervention analysis for the outliers, and the GARCH-SeaDFA is not, so although in this aspect the comparison is not fair, GARCH-SeaDFA gets much better results.

4.3 One-day-ahead forecasting

The GARCH-SeaDFA model not only produce accurate forecasts in the long-run, but also in the short-term. If the model is reestimated everyday, and one-day-ahead forecasts are computed for all the days in 2008.

Table 1: Year-ahead forecasting errors.

Month	MAPE (%)	MAPE2 (%)
January 2008	20.25	19.87
February 2008	18.00	17.91
March 2008	13.06	12.20
April 2008	13.31	12.40
May 2008	12.82	12.20
June 2008	11.10	10.38
July 2008	16.63	15.94
August 2008	18.30	17.73
September 2008	20.66	18.93
October 2008	17.31	15.60
November 2008	14.98	14.17
December 2008	17.39	15.27
TOTAL	16.15	15.22

Usually, in previous works dealing with electricity price forecasting, the last week or the one before last week, depending on which one is a complete week, are used to evaluate the results obtained. (Conejo et al., 2005).

In Table 2, the MAPE for these weeks is provided, as well as those obtained using the Mixed Model by García-Martos et al. (2007). The Mixed Model was specifically designed for one-day-ahead forecasting, and intervention analysis is performed to deal with outliers. In the GARCH-SeaDFA model presente d in this paper, we are not incorporating intervention analysis to the QML estimation procedure described in Section 3. This could be done as further research, since it exceeds the goals of this paper, thus, the forecasting errors could be improved.

Anyway, the results for the short-term are slightly better using the GARCH-SeaDFA provided here, compared to Mixed Model. Nevertheless for long-run forecasting the forecasts produced with Mixed Model give much bigger errors as pointed out in the previous Subsection.

Table 2: Day-ahead forecasting errors for the third week of each month in 2008.

Week	MAPE (GARCH-SeaDFA)	MAPE Mixed Model (García-Martos et al., 2007)
21-27 January 2008	6.59	8.30
18-24 February 2008	9.13	7.59
24-30 March 2008	6.60	9.86
21-27 April 2008	6.72	5.14
19-25 May 2008	6.72	6.22
23-29 June 2008	5.40	5.18
21-27 July 2008	5.76	5.31
25-31 August 2008	6.77	5.52
22-28 September 2008	6.41	4.33
20-26 October 2008	6.65	9.98
17-23 November 2008	7.35	7.80
22-28 December 2008	5.66	15.61
AVERAGE	**6.65 %**	**7.57 %**

Finally, and just to illustrate some of the results given in Table 2, Figure 11 shows the forecasts obtained using GARCH-SeaDFA and those coming from the Mixed Model for the week 22^{nd}–28^{th} December 2008. Usually this is a week difficult to forecast, since peak demand due to Christmas affects the prices. MAPE using GARCH-SeaDFA is 5.66% and using the Mixed Model is 15.61%.

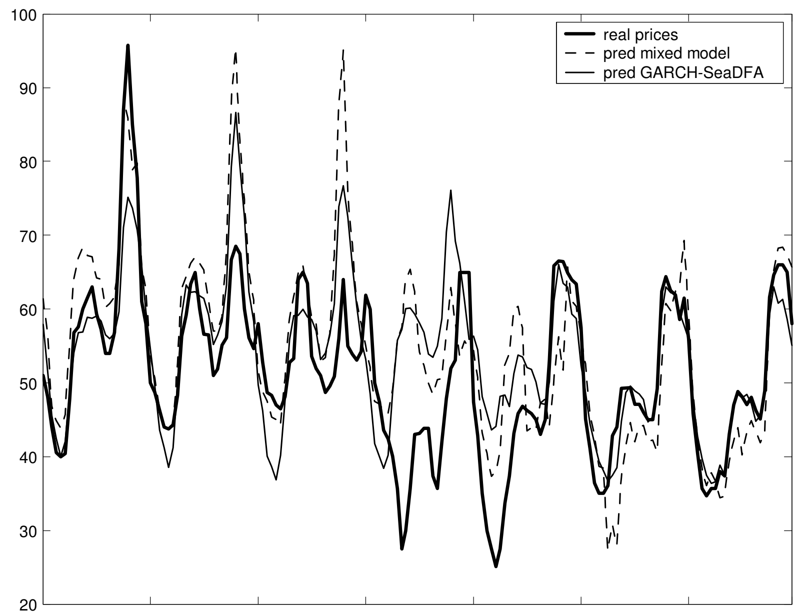

Figure 11: One-day-ahead forecasts and real prices, 22^{nd}–28^{th} December 2008.

4.3.1 Volatility forecasting and prediction intervals for electricity prices

Once the model has been estimated, and price forecasts computed, it is also possible to compute one-step-ahead forecasts for the common volatility factors, $\mathbf{w}_t^*$. Thus, those volatility forecasts can be used to compute prediction intervals for the prices that take into account the fact that electricity price data present structure both in conditional mean and variance.

Given equations (5) and (6) where $\mathbf{w}_t^*|I_{t-1} \sim N(0, Q_t)$, and assuming that the future prediction errors are normally distributed, prediction intervals for electricity prices are given by (see Rodríguez and Ruiz (2009) for details):

$$(\widehat{\mathbf{y}}_{T+k|T} - z_{1-\alpha/2}\sqrt{\Sigma_{T+k|T}}, \widehat{\mathbf{y}}_{T+k|T} + z_{1-\alpha/2}\sqrt{\Sigma_{T+k|T}}),$$

where the point prediction $\widehat{\mathbf{y}}_{T+k|T}$, and its MSE, $\Sigma_{T+k|T}$ are obtained assuming known parameters. In practice we replace them by consistent estimates ontained by the QML procedure described in subsection 3.2. QML estimators have nice asymptotic properties (Durbin and Koopman, 2001).

The results shown here to illustrate the procedure described, are one-step-ahead forecasts for the week 17^{th}-23^{rd} November 2008. In general the third week of each month and more specifically the third one in November are used by different authors to check the performance of different forecasting methodologies applied to electricity markets. (Nogales et al., 2002, Conejo et al., 2005).

The results shown here illustrate the procedure described, are one-step-ahead forecasts for the week $17^{th} - 23^{rd}$ November 2008. In general, the third week of each month and more specifically the third one in November are used by different authors to check the performance of different forecasting methodologiesapplied to electricity markets (Nogales et al., 2002 and Conejo et al., 2005).

Figure 12 shows real prices, point forecasts and prediciton intervals for the prices in the third week in November 2008. These intervals incorporate conditionally heteroskedasticity present in common factors.

5 Conclusions

This paper faces the important problem of forecasting electricity prices and their volatilities in deregulated markets. Moreover, the problem of computing short-term forecasts and specially that of long-run forecasting is still an open problem, and the methodology here introduced is motivated by these important issues.

Electricity prices present structure both in the conditional mean and conditional variance, so it is necessary to develop methodology that is able to capture jointly both dynamics. Although electricity prices share some features with those data coming from financial markets, also the structure in the conditional mean must be modelled. The high dimension of the problem, 24 hourly time series, justifies the applicaion of dimensionality reduction techniques.

In this paper we develop the Conditionally Heteroskedastic Seasonal Dynamic Factor Analysis, allowing for unobservable common factors that exhibit conditional heteroskedasticity.

The methodology developed by Harvey, Ruiz and Sentana (1992) is applied to extend

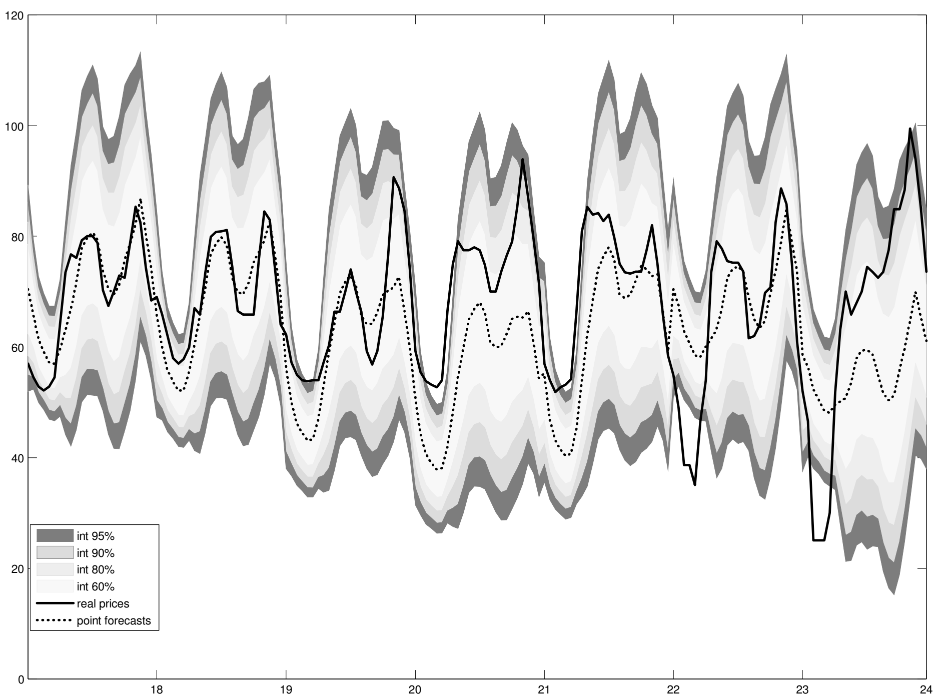

Figure 12: Point forecasts and forecasting intervals in the week $17^{th}-23^{rd}$ November 2008.

the Homoskedastic Seasonal Dynamic Factor Analysis by Alonso et al. (2008), to the case in which the common factors present conditional heteroskedasticity. This new model proposed is able to extract conditionally heteroskedastic common factors from any high-dimensional vector of time series.

Detailed numerical results are provided for the Spanish Market, in which the prices and volatilities have been modelled for the period 1998-2007 using the new GARCH-SeaDFA model. Moreover, concerning forecasting, the results for all the hourly prices in the year 2008 are calculated and validated, both for the short and long term, comparing these results with other methodologies that had been published recently and that are used here as benchmarking models. In terms of prediction accuracy a very good performance of the model here presented is shown both for the short and long-term. Moreover, results concerning volatility forecasting are also included, as well as forecasting intervals that incorporate the evolution of volatility over time.

The methods here proposed, in which the common trend of a vector of series is captured both for the conditional mean and variance, could be of application in the field of macroeconomic data.

6 Acknowledgements

Authors would like to thank financial support by ISF-SAS Grant 2007.

References

[1] Alonso, A.M., García-Martos, C., Rodríguez, J. and Sánchez, M.J. (2008). "Seasonal Dynamic Factor Analysis and Bootstrap Inference: application to electricity market forecasting", Working Paper 08-14, Statistics and Econometrics Series, Universidad Carlos III de Madrid.

[2] Ansley, C.F. and Kohn, R. (1986), "Estimation Prediction and Interpolation for ARIMA models with missing data," Journal of the American Satistical Association.

[3] Bollerslev, T. (1986). "Generalized autoregressive conditional heteroskedasticity," Journal of Econometrics.

[4] Bunn, D.W., Karakatsani, N. (2003), "Forecasting electricity prices", London Business School, London.

[5] Conejo A. J., Contreras J., Espínola R., Plazas M. A. (2005). "Forecasting electricity prices for a day-ahead pool-based electric energy market". International Journal of Forecasting.

[6] Conejo, A.J., Carrión, M., Morales, J.M. and Nogales, J. (2009). "Electricity Pool Prices: Long-Term Uncertainty Characterization for Futures-Market Trading and Risk Management," Journal of the Operational Research Society.

[7] Conejo, A.J., Plazas, M.A., Espinola, R. and Molina, A.B. (2005), "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models", IEEE Transactions on Power Systems.

[8] Connor, G., Korajczyk, R. and Linton, O.B. (2006). "The common and specific components of dynamic volatility," Journal of Econometrics.

[9] Contreras, J., Espínola, R. Nogales, F.J., Conejo, A.J. (2003), "ARIMA Models to Predict Next-Day Electricity Prices". IEEE Transactions on Power Systems.

[10] Diebold, F.X. and Nerlove, M. (1989). "The dynamics of exchange rate volatility: a multivariate latent factor ARCH model". Journal of Applied Econometrics.

[11] Durbin and Koopman, (2001). "Time series analysis by satate space methods", Oxford University Press.

[12] Escribano, A., Peña, J.I. and Villaplana, P. (2002), "Modelling Electricity Prices: International Evidence", Working Paper 02-27, Economic Series, Universidad Carlos III de Madrid.

[13] Engle, R.F. (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", Econometrica.

[14] Engle, R.F. (1987). "Multivariate ARCH with factor structures -cointegration in variance," University of California Press.

[15] García-Ferrer, A., González-Prieto, E. and Peña, D. (2008). "A multivariate generalized independent factor GARCH model with an application to financial stock returns," , Working Paper 08-75, Statistics and Econometrics Series, Universidad Carlos III de Madrid.

[16] García-Martos, C., Rodríguez, J. and Sánchez, M.J. (2007), "Mixed models for short-run forecasting of electricity prices: application for the Spanish market," IEEE Transactions on Power Systems.

[17] Geweke, J. and Singleton, K.J. (1981), "Maximum likelihood 'confirmatory' factor analysis of economic time series," International Economic Review.

[18] Harvey, A., Ruiz, E. and Sentana, E. (1992). "Unobserved Component Time Series Models with ARCH Disturbances," Journal of Econometrics.

[19] Higgs, H. (2009). "Modelling price and volatility inter-relationships in the Australian wholesale spot electricity markets," Energy Economics.

[20] Koopman, S.J., Ooms, M. and Carnero, M.A. (2007), "Periodic Seasonal Reg-ARFIMA-GARCH Models for Daily Electricity Spot Prices," Journal of the American Satistical Association.

[21] Lee R. D., Carter L. R. (1992), "Modeling and Forecasting U. S. Mortality," Journal of the American Statistical Association.

[22] Nogales, F.J., Contreras, J., Conejo, A.J., Espínola, R. (2002),"Forecasting Next-Day Electricity Prices by Time Series Models". IEEE Transactions on Power Systems.

[23] Nogales F. J. and Conejo A. J. (2006), "Electricity Price Forecasting through Transfer Function Models". Journal of the Operational Research Society.

[24] Panagiotelis, A. and Smith, M. (2008). "Bayesian density forecasting of intraday electricity prices using multivariate skew t distributions," International Journal of Forecasting.

[25] Peña, D. and Box, G.E.P. (1987), "Identifying a simplifying structure in time series," Journal of the American Statistical Association.

[26] Peña, D. and Poncela, P. (2004), "Forecasting with Nonstationary Dynamic Factor Models," The Journal of Econometrics.

[27] Peña, D. and Poncela, P. (2006), "Nonstationary Dynamic Factor Analysis," Journal of Statistical Planning and Inference.

[28] Smith, M. and Cottet, R. (2006), "Estimation of a longitudinal multivariate stochastic volatility model for the analysis of intra-day electricity prices," Working Paper coded ECMT2006-1, Sydney University, School of Economics and Political Science.

[29] Rodríguez, A.F. and Ruiz, E. (2009). "Bootstrap Prediction Intervals in State Space Models," Journal of Time Series Analysis.

[30] Stock, J. H. and M. Watson (2002), "Forecasting using principal components from a large number of predictors," Journal of the American Statistical Association, 97, 1167–79.

[31] Tsay, R. (2005). "Analysis of Financial Time Series". Wiley Press.

[32] Vehvilainen, I. and Pyykkonen, T. (2005), "Stochastic factor model for electricity spot price: the case of the Nordic market", Energy Economics.

[33] Weron, R and Misiorek, A. (2008). "Forecasting Spot Electricity Prices: A Comparison of Parametric and Semiparametric Time Series Models ", International Journal of Forecasting.