Energy Economics 44 (2014) 143–159

Contents lists available at ScienceDirect

Energy Economics

journal homepage: www.elsevier.com/locate/eneco

Component estimation for electricity prices: Procedures and comparisons

Francesco Lisi [a,*], Fany Nan [b]

[a] Department of Statistical Sciences, University of Padua, Via Cesare Battisti 241, 35121 Padua, Italy
[b] Department of Economics, University of Verona, Via dell'Artigliere 19, 37129 Verona, Italy

ARTICLE INFO

Article history:
Received 16 July 2012
Received in revised form 15 March 2014
Accepted 16 March 2014
Available online 5 April 2014

JEL classification:
C01
C02
C14
C22
C53
Q4

Keywords:
Component estimation
Filtering procedures
Electricity prices
Long-term dynamics
Nonparametric methods

ABSTRACT

Electricity price time series usually exhibit some form of nonstationarity, corresponding to long-term behavior, one or more periodic components as well as dependence on calendar effects. As a result, modeling electricity prices requires accounting for both long-term and periodic components. In the literature, several filtering procedures have been proposed but a standard has not yet been found. Furthermore, since different procedures are applied in contexts that are not homogeneous with respect to data, periods and final goals, a fair comparison is difficult. This work considers several methods for component estimation in a homogeneous framework and compares them according to specific criteria. The final purpose is to find an estimation procedure that performs well, independently of the intended market and that can be proposed as a reference for electricity price time series filtering.

1. Introduction

Modeling and forecasting power prices are important issues for trading and risk management in the liberalized electricity markets and, consequently, many studies in this field have appeared in literature over the last decade (see Aggarwal et al., 2009; Bunn, 2004; Weron, 2006 for reviews).

Electricity price time series usually exhibit some form of nonstationarity, corresponding to long-term behavior, one or more periodic components, as well as dependence on calendar effects and spikes. One way to consider these components is to view them as stochastic processes. Stochastic trends are often modeled by Brownian motion or random walk, assuming the presence of unit roots (Bosco et al., 2007; Bosco et al., 2010) or referring to long-memory (Koopman et al., 2007). Sometimes, also the seasonal component is treated as stochastic (Koopman et al., 2007), allowing joint estimation of the components. Jumps are also often considered as stochastic and treated by using diffusion models with Poisson jump components (for example, see Fanone et al., 2013; Pirino and Renò, 2010), by Markov-Switching models or assuming that the jump size is governed by a normal distribution (Hellström et al., 2012). Only very few works with a focus on prediction, also model spikes (Christensen et al., 2012).

A second way to model electricity prices requires a preliminary estimation of the long-term and periodic behavior to filter out these components in order to achieve stationarity. Also, a good filtering is important because it reduces distorting effects on forecasting and enables a better identification of spikes. Although with different focus, there are a number of works on modeling and prediction electricity prices, which consider this issue. For example, Erlwein et al. (2012), de Jong (2006), Kosater and Mosler (2006), Misiorek et al. (2006), Pilipovic (1998) and Weron et al. (2004) estimate long-term behavior by means of polynomial (usually linear) trends together with to sine or cosine functions. Bosco et al. (2007) use a linear trend and model periodicity with state space models. As a variant, Crespo Cuaresma et al. (2004), Escribano et al. (2011), Lucia and Schwartz (2002) and some of the aforementioned authors, consider monthly dummy variables, sometimes together with a linear trend, to approximate long-run dynamics. Most of these authors describe the weekly periodic component and the daily periodicity through daily and (semi-) hourly dummy variables. Janczura and Weron (2009), Janczura and Weron (2010), Trück et al. (2007) and Weron (2009), use wavelet low-pass filters for the

* Corresponding author. Tel.: +39 049 8274182; fax: +39 049 8274170.
E-mail addresses: francesco.lisi@unipd.it (F. Lisi), fany.nan@univr.it (F. Nan).

http://dx.doi.org/10.1016/j.eneco.2014.03.018
0140-9883/

long-run component and moving average (or median) techniques to estimate the periodic component. Wavelets were also considered by Schlueter (2010) for modeling daily average prices. In the electricity market literature, component estimation based on empirical mode decomposition has been produced by Kurbatsky and Tomin (2010) and Qian et al. (2011). Dordonnat et al. (2010) use cubic splines and sinusoidal functions to approximate the trend and daily means, equivalent to dummy variables, for the weekly component. Spline functions have also been used by Bisaglia et al. (2010), to model the yearly periodic component. Approaches based on local linear regression or local linear trends for long-term and annual components have been considered by Bordignon et al. (2013), Trapero and Pedregal (2009) and Veraart and Veraart (2012). The last two authors use trimmed means to estimate the periodic daily coefficients. It is worth mentioning that in a few cases the long-term component has been implicitly considered by differentiating prices or loads (Sigauke and Chikobvu, 2011; Weron, 2005 and implicitly in Bosco et al., 2010). Gianfreda and Grossi (2012) consider fractionally differentiation. Lastly, authors who consider calendar effects, usually model them by means of dummy variables, accounting for national holidays or other specific calendar conditions.

Which is the best method for filtering components in electricity markets has not yet been assessed in the literature. From a theoretical viewpoint, none of existing methods is strictly preferable. Moreover, filtering procedures have been applied in a wide variety of markets, sample periods and with different final goals. Thus, also from an empirical viewpoint, a fair comparison among them is almost impossible. This work aims to fill this gap and to compare, in a homogeneous framework, several procedures for component estimation with the goal of identifying a procedure that can be used as a standard. We approach component estimation with the topic of prediction in mind; therefore we will look specifically for methods that could lead to good predictive performances. However, a good estimation of components may also be useful for other issues such as spike identification and simulations. Of course, a procedure that is standard does not imply that it is the best in all situations, but only that it can be viewed as a benchmark. Indeed other specific procedures, or filters, may work better for certain issues.

To estimate the long-term component, eleven filtering techniques will be applied. They belong to the following groups: the polynomial–sinusoidal approach, local polynomial regressions, spline functions, wavelets, empirical mode decomposition, singular spectrum analysis and the well-known Kolmogorov–Zurbenko, Hodrick–Prescott and Christiano–Fitzgerald filters. For the periodic component three alternative estimators will be considered. These are based on dummy variables, trimmed means and centered moving medians. Mixing methods for long-term and periodic component estimation leads us to compare 33 different procedures.

Since true components are unobservable, how to compare these filtering methods becomes an issue. In Section 4 we propose three criteria to refer to for procedure evaluation. They are based on three features that are expected of a good component estimation: after filtering there should not be any time-depending pattern; there should not be residual periodicity; the procedure should positively affect the prediction accuracy of original prices.

All methods will be applied and compared using data from three important markets: the British market, the Pennsylvania–New Jersey–Maryland market and the Nord Pool market. These markets have been chosen because they differ substantially, in generation modes, structure organizations and land electricity demand. Indeed, the main fuels used for electricity generation are natural gas, coal and hydro, respectively. Since these factors influence price dynamics in different ways they should guarantee a wide enough scope. Thus, even if there is no guarantee that the results of our study will extend to other markets, we think that findings of our research apply beyond these specific markets and can be considered generic.

The paper is organized as follows. In Section 2 we present some preliminary analyses of our data, suggesting which components to consider and how to model them. This leads us to define the reference model for the components which is based on a deterministic part and a stochastic term. Section 3 is devoted to the description of component estimation. Evaluation criteria for comparing estimation methods are given in Section 4. Section 5 presents empirical results and Section 6 concludes.

2. Preliminary analyses

In our analyses, we consider three main international electricity markets: the British market (APX Power UK, APX-PUK), Pennsylvania–New Jersey–Maryland market (PJM) and Nord Pool market (NP), which operates in Norway, Denmark, Sweden, Finland and Estonia.

The dataset related to the APX-PUK comprises the time series of prices (P_t), national day-ahead demand forecast (D_t) and indicated margin[1] (M_t) for the period 1 April 2005 to 31 December 2010 (100,848 data points, covering $N = 2101$ days). For PJM and NP markets, only the time series of prices and actual demand were available (to us) and these time series were from 1 January 2005 to 31 December 2010 (52,584 data points, covering $N = 2191$ days) for the PJM market and from 1 January 2008 to 31 December 2010 (26,304 data points, covering $N = 1096$ days) for the NP market.

The data have a half-hourly frequency for APX-PUK and an hourly frequency for PJM and NP; therefore each day comprises 48 (for APX-PUK) or 24 (for PJM and NP) load periods with 00:00–00:30 am (00:00–01:00 am) defined as period 1. Spot price is denoted as P_{tj}, where t indicates the day and j indicates the load period ($t = 1,2,\ldots,N$; $j = 1,2,\ldots,24$ or 48). Analogously for D_{tj} and M_{tj}.

In this study, following a widespread practice in literature, each (half-)hourly time series is modeled separately, thereby eliminating the problem of modeling intra-daily periodicity.

Differences in load periods and markets can cause significant variations in price time series. However, a first inspection, based on graphs, spectra and ACFs (see Figs. 1–4) indicates that the series show neither a well-defined long-run behavior nor clear annual dynamics. A common characteristic of price time series is the weekly periodic component (of period 7), suggested by the spectra that show three peaks at the frequencies 1/7, 2/7 and 3/7, and a very persistent autocorrelation function. This indicates that other analyses should be considered to determine how the long-term components should be handled.

To investigate nonstationarity (i.e. deterministic vs. stochastic trends), we employed a robust unit root test based on Lucas' robust pseudo-likelihood ratio (PLR), as described in Bosco et al. (2010).[2] Overall, at the 5% significance level, the null hypothesis of unit root is rejected 91 times over 96 load periods (48 + 24 + 24) and, more specifically, 47 times for APX-UK, 20 times for NP and 24 times for PJM. In the context of electricity prices, which are characterized by outliers, multiple seasonal effects, volatilities and other messy features, such results should be interpreted cautiously and may not be completely reliable. Nevertheless, the use of robust tests leads us to focus on models that assume trend-stationarity and to model the long-term component deterministically. Thus, we assume that the dynamics of log prices can be represented by a nonstationary level component L_{tj}, accounting for level changes and/or long-term or (semi-)periodic behavior as well as for calendar effects, and a residual stationary stochastic component p_{tj}, formally:

$$\log P_{tj} = L_{tj} + p_{tj}. \tag{1}$$

[1] The indicated margin is the available capacity margin and is defined as the difference between the demand forecast and the sum of the maximum export limits nominated by each generator prior to each trading period as its maximum available output capacity.

[2] Lucas' PLR cointegration test, which is based on the Student-t density, can be used to test for a unit root in scalar time series; p-values are calculated through a bootstrap strategy based on Swensen's algorithms (Swensen, 2006).