Energy Policy 38 (2010) 715–725

Contents lists available at ScienceDirect

Energy Policy

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

Electric power demand forecasting using interval time series: A comparison between VAR and iMLP

Carolina García-Ascanio, Carlos Maté *

ICAI, School of Industrial Engineering, Pontificia Comillas University, Alberto Aguilera 25, 28015 Madrid, Spain

A R T I C L E I N F O

Article history:
Received 21 April 2009
Accepted 3 October 2009
Available online 22 November 2009

Keywords:
Interval data
Interval multi-layer perceptron
Vector autoregressive model

A B S T R A C T

Electric power demand forecasts play an essential role in the electric industry, as they provide the basis for making decisions in power system planning and operation. A great variety of mathematical methods have been used for demand forecasting. The development and improvement of appropriate mathematical tools will lead to more accurate demand forecasting techniques.

In order to forecast the monthly electric power demand per hour in Spain for 2 years, this paper presents a comparison between a new forecasting approach considering vector autoregressive (VAR) forecasting models applied to interval time series (ITS) and the iMLP, the multi-layer perceptron model adapted to interval data.

In the proposed comparison, for the VAR approach two models are fitted per every hour, one composed of the centre (mid-point) and radius (half-range), and another one of the lower and upper bounds according to the interval representation assumed by the ITS in the learning set. In the case of the iMLP, only the model composed of the centre and radius is fitted. The other interval representation composed of the lower and upper bounds is obtained from the linear combination of the two.

This novel approach, obtaining two bivariate models each hour, makes possible to establish, for different periods in the day, which interval representation is more accurate. Furthermore, the comparison between two different techniques adapted to interval time series allows us to determine the efficiency of these models in forecasting electric power demand. It is important to note that the iMLP technique has been selected for the comparison, as it has shown its accuracy in forecasting daily electricity price intervals.

This work shows the ITS forecasting methods as a potential tool that will lead to a reduction in risk when making power system planning and operational decisions.

1. Introduction

Accurate models for electric power demand and price forecasting are necessary for the operation and planning of power systems and its consequences are critical in energy efficiency and sustainability issues. Nowadays, the role of forecasting in deregulated energy industries is essential in key decision making, such as purchasing and generating electric power, load switching, and infrastructure development. In addition, sustainability analysis (see, for example, Linares et al. (2008) for the Spanish electricity sector) relies on accurate quantitative predictions.

Depending on the time horizon selected, demand forecasting can be classified as: short-term from 1 h to 1 week, medium-term from a week to a year and long-term for more than a year (see Hahn et al. (2009) for details and implications in decision making).

Most forecasting methods based on classic data (single valued) use statistical techniques or artificial intelligence tools. The initial studies were based on statistical models, using for instance integrated moving average models (ARIMA) (Abdel-Aal and Al-Garni, 1997; Saab et al., 2001) or models based on regression (Mohgram and Rahman, 1989; Papalexopoulos and Hesterberg, 1990). Their low accuracy in time series with non-linear characteristics prompted the application of artificial intelligence techniques, such as neural networks (Papalexopoulos et al., 1994; Chowand and Leung, 1996), hybrids methods (Srinivasan et al., 1999; Padmakumari et al., 1999) or genetic algorithms (Tzafestas and Tzafestas, 2001). Taylor and McSharry (2007) evaluate different short-term load forecasting methods: (i) ARIMA model; (ii) Periodic AR model; (iii) an extension for double seasonality of Holt–Winters exponential smoothing method; (iv) an alternative exponential smoothing method; (v) a method based on the principal component analysis (PCA) of the daily demand profiles, concluding from the results obtained that the double seasonal Holt–Winters exponential smoothing method is the best of these

* Corresponding author. Tel.: +34 915422800 2430; fax: +34 915411132.
E-mail addresses: carolina.gda@icai.es (C. García-Ascanio), cmate@upcomillas.es (C. Maté).

0301-4215/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.enpol.2009.10.007

methods. Franco et al. (2006) apply an extension of the VAR models, the Vector error correction models (VECM) to forecast the electricity demand of the Venezuelan electric system for the period 2004–2024. Erdogdu (2007) analyzed electricity demand using cointegration and ARIMA modelling. Arroyo and Maté (2009) adapt the k-Nearest Neighbours (k-NN) algorithm to forecast histogram time series (HTS).

As it can be seen, a large variety of mathematical methods and ideas have been used for demand forecasting (see Hahn et al. (2009) for a recent survey). Further development and improvement of mathematical tools will surely lead to more accurate demand forecasting techniques, such as indicated by Gonzalez-Romera et al. (2006).

In the field of interval time-series (ITS) forecasting, different techniques have been developed in recent years. Arroyo et al. (2007) develop three exponential smoothing methods for ITS. Maia et al. (2006) apply autoregressive moving average models (ARMA) to ITS in a hybrid model with neural nets and Maia et al. (2008) present approaches to interval-valued time-series forecasting based on AR, ARIMA and Artificial Neural Networks (ANN) models. Cheung (2007) proposes an empirical model for daily highs and lows for three US stock indexes using the VECM. Han et al. (2008) propose an interval linear model to investigate the dynamic relationships between interval processes.

Moreover, Neural Nets applied to Interval data (INN) have been developed by several researchers (Ishibuchi et al., 1993; Simoff, 1996; Beheshti et al., 1998; Rossi and Conan-Guez, 2002; Patiño-Escarcina et al., 2004; Muñoz San Roque et al., 2007). The model proposed by Muñoz San Roque et al. (2007) is the iMLP used here for comparison.

In addition, Zhao et al. (2008) propose a statistical approach for interval forecasting of the electricity price, but the method is based on classic time series (single valued) not in interval time series. The prediction interval is obtained by forecasting the price value and its variance. The Support Vector Machine (SVM) is employed to forecast the value of the price, and a statistical model – obtained by introducing a heteroskedastic variance equation to the SVM – to forecast the price variance.

Two interval time-series forecasting models are proposed in this paper in order to show their accuracy in demand forecasting. In the work reported here the VAR technique is adapted to ITS as it is done in Maia et al. (2006) for ARMA models. The iMLP is applied in accordance with Muñoz San Roque et al. (2007), whose model presents accuracy results in electricity price forecasting, in order to observe its behaviour in electricity demand forecasting. Depending on the representation of the interval assumed by the ITS, the interval time series are split up into two time series: the time series of the lower bound and the time series of the upper bound, or the time series of the centre and the time series of the radius. Once the two time series are obtained, the application of the VAR and the iMLP methods to obtain the forecasts of the electric power demand is straightforward.

The paper is divided into five sections: Section 2 introduces the interval analysis; Section 3 introduces the vector auto-regression models; Section 4 shows the iMLP; Section 5 shows the application of the forecasting technique to the interval time series under study, compares both methods and analyzes the results. Finally, Section 6 concludes.

2. Interval analysis

2.1. Introduction

Under the assumption that observations and estimations in the real world are incomplete to represent real data exactly, Ramon Moore in 1959 proposed the Interval Analysis as a tool for automatic control of the errors in a computed result that arise from input error, from rounding errors during computation, and from truncation errors when using a numerical approximation to the mathematical problem. Hence, if precision is needed, data must be represented by intervals. Since 1960, Interval Analysis has been an active focus on research.

Moore (1966) establishes the basis for Interval Analysis and Moore and Bierbaum (1979) deal with an important set of techniques providing a mathematically rigorous and complete error analysis for computational results. They show that interval analysis provides a powerful set of tools with direct applicability to important problems in scientific computing. In addition, Chavent and Saracco (2008) have obtained basic descriptive statistics such as central tendency and dispersion measures for interval data. More recently, Moore et al. (2009) present an updated introduction to Interval Analysis.

2.2. Interval data

Interval data is a particular case of symbolic data. See Billard and Diday (2003), Billard and Diday (2006) and Diday and Noirhomme (2008) for key references in the field. The main difference between classic and symbolic data is that a classic data point takes as its value a single point in p-dimensional space, whereas a symbolic one takes as its value a hypercube (or hyperrectangle) in p-dimensional space, or it is the Cartesian product of p distributions in p-dimensional space or a mixture of both. In short, symbolic data have internal variation and structure. It is important to mention that interval data may be in many instances the result of an aggregation procedure, spatial or temporal, over information collected at a very disaggregated level.

An interval $[x]$ over the base set $(E, \leq)$ is an ordered pair $[x]=[x_L; x_U]$ where $x_L, x_U \in E$ are the endpoints or bounds of the interval such that $x_L \leq x_U$.

Table 1 shows the interval-valued variables in every month of the hourly spot electricity price and the electric energy demand in Spain in 2007.

An interval can be represented by its lower and upper bounds $[x]=[x_L, x_U]/-\infty < x_L \leq x_U < \infty$ or by its centre (mid-point) and radius (half-range) as $[x]=\langle x_C, x_R \rangle$ where $x_C=(x_L+x_U)/2$ and $x_R=(x_U-x_L)/2$. In Fig. 1 the structure of an interval is presented.

2.3. Interval time series

An interval time series (ITS) is a chronological sequence of interval-valued variables and it is denoted by $\{[x_t]\} = \{[x_{Lt}, x_{Ut}]\} = \{\langle x_{Ct}, x_{Rt} \rangle\}$ for $t=1,2,\ldots,n$. The interval time series of the monthly electric power demand in 2000 for the hour 1, H1, is shown in Fig. 2.

2.4. Interval time-series forecasting techniques

Several interval time-series forecasting techniques have been developed in recent years (see Arroyo et al. (2010) for a survey).

Table 1
Interval-valued variables (data available at http://www.ree.es and http://www.omel.es).

Year 2007	Final electricity price (€/MWh)	Electricity demand (MWh)
January	[10, 84.79]	[18,644, 43,201]
February	[5, 77]	[20,082, 40,745]
March	[7, 61.32]	[19,907, 39,593]
…	…	…
…	…	…