Available online at www.sciencedirect.com
SCIENCE DIRECT®
Environmental Modelling & Software 20 (2005) 1389–1400
Environmental Modelling & Software
www.elsevier.com/locate/envsoft

# Modelling the load curve of aggregate electricity consumption using principal components

Matteo Manera[a,*], Angelo Marzullo[b]

[a] *Department of Statistics, University of Milano-Bicocca and Fondazione Eni Enrico Mattei, Milano, Italy*
[b] *Enifin, Eni S.p.A., Milano, Italy*

Received 31 December 2003; received in revised form 10 August 2004; accepted 24 September 2004

**Abstract**

Since oil is a non-renewable resource with a high environmental impact, and its most common use is to produce combustibles for electricity, reliable methods for modelling electricity consumption can contribute to a more rational employment of this hydrocarbon fuel. In this paper we apply the Principal Components (PC) method to modelling the load curves of Italy, France and Greece on hourly data of aggregate electricity consumption. The empirical results obtained with the PC approach are compared with those produced by the Fourier and Constrained Smoothing Spline estimators. The PC method represents a much simpler and attractive alternative to modelling electricity consumption since it is extremely easy to compute, significantly reduces the number of variables to be considered, and generally increases the accuracy of electricity consumption forecasts. As an additional advantage, the PC method is able to accommodate relevant exogenous variables such as daily temperature and environmental factors, and is extremely versatile in computing out-of-sample forecasts.


*JEL classification:* C51; C53; Q30; Q40

*Keywords:* Electricity; Load curves; Principal components; Fourier estimator; Constrained smoothing spline estimator; Temperature; Non-renewable resources; Hydrocarbon fuels; Environment

## 1. Introduction

In a Europe characterized by strong incentives towards the liberalization of national electricity markets, researchers and market operators are increasingly interested in obtaining reliable estimates and forecasts of the short-run demand for electricity.

The energy sector, which is intimately related with the oil and gas industry, is also crucial for its environmental implications. According to the European Energy Agency (2002), about 90% of the greenhouse effect is directly or indirectly attributable to the use of hydrocarbon fossils and deforestation. The International Energy Agency (2000) defines an important energy indicator, the so-called ‘‘total primary energy supply’’ (TPES), that is the total amount of energy produced by all existing sources. In 2002, the world TPES was about 10,000 megatons oil equivalent (MTOE), which are equivalent to $1.2 \times 10^8$ Giga-Watt per hour. It is important to notice that crude oil only represents 35% of the TPES, and that the sum of all fossil combustibles (i.e. oil, gas and coal) amounts to 76.2% of the TPES. Another crucial energy indicator is the ‘‘total final consumption’’ (TFC). Since no energy plant is 100% efficient, TFC is less than TPES, and has been estimated around 7000 MTOE. Out of 7000 MTOE, 75% is given by fossil combustibles, which is equivalent to $6 \times 10^4$ Tera-Watt per hour.

* Corresponding author. Department of Statistics, University of Milano-Bicocca, Via Bicocca degli Arcimboldi, 8 Building U7 20126 Milano, Italy. Tel.: +39 02 64487319; fax: +39 02 6473312.
*E-mail address:* matteo.manera@unimib.it (M. Manera).

1364-8152/$ - see front matter 
doi:10.1016/j.envsoft.2004.09.019

Since oil is a non-renewable resource with a high environmental impact, academics, research institutions and public opinion are engaged in a fast-growing debate on how to reduce the dependence of national economic systems on oil. Besides the development of alternative and more efficient ways to exploit renewable resources, the simplest method to control this type of dependence is to reduce oil consumption. As the most common use of oil is to produce combustibles for electricity and transportation, more reliable methods to model, estimate and forecast electricity consumption can contribute to a more rational employment of this fundamental hydrocarbon fuel.

Early studies on the analysis of the load curve (e.g. Cargill and Meyer, 1971) generally concentrate only on the long-run features of electricity consumption. Alternative traditional approaches which explicitly take into account the short-run movements in the load curve are spline and Fourier models (see Hendricks et al., 1979; Mouchart and Roche, 1987), while a more recent methodology which embeds both splines and Fourier is the constrained smoothing splines estimator (see Rodriguez-Poo, 2000).

The Principal Components (PC) method, combined with traditional regression analysis, represents a much simpler and attractive alternative to modelling and forecasting electricity consumption as it is extremely easy to compute, significantly reduces the number of variables to be considered, and generally contributes to more accurate electricity consumption forecasts (see Nogales et al., 2002; Ramanathan et al., 1997; Granger et al., 1979; Rodriguez-Poo, 1992). As an additional advantage, the PC method is able to accommodate relevant exogenous variables such as daily temperature (see Vogelsang and Franses, in press for a study of the time trends in monthly temperatures of different European countries), and is versatile in computing out-of-sample forecasts.

In this paper we apply the PC method to model and forecast the load curves of three European countries, namely Italy, France and Greece, using hourly aggregate electricity consumption data. The empirical results obtained with the PC approach are compared with those produced by Fourier and constrained smoothing spline estimators.

The paper is organized as follows. Section 2 provides a description of cubic splines, Fourier and constrained smoothing estimators for modelling the loading curve. An illustration of the PC method is given in Section 3. The data are presented in Section 4. In Section 5 the empirical results obtained with the PC method are presented and compared with the Fourier and constrained smoothing spline estimators. In Section 6 the PC method is used to obtain out-of-sample forecasts of electricity consumption for the French market. Section 7 provides some concluding comments.

## 2. Classical models of the load curve

Early research on the daily load curve was generally based on a two-stage estimation approach (see, among others, Cargill and Meyer, 1971). In the first stage, a simple ARMA time series model is fitted to consumption data for each consumption unit (e.g. household, firm, or industrial plant). In the second stage, the estimated coefficients of each ARMA model are regressed on a set of residential, demographical and socio-economic variables. One limitation of this method is that it concentrates only on long-run features of the data, since the selected explanatory variables do not change within a single day. On the other hand, load curves, which are subject to physical as well as atmospherical conditions, are characterized by intra-day marked variations.

Alternative approaches which concentrate on short-run movements in the load curve are given by spline and Fourier models. In both cases, the general problem can be described as follows. Indicate with $y_i$, $i = 1,\ldots,n$, the electricity consumption between time $t_i - 1$ and time $t_i$ for a given sample of data. The index $i = 1,\ldots,n$ denotes the data frequency. We are interested in the statistical model:

$$y_i = m(t_i) + \varepsilon_i, \tag{1}$$

where $\varepsilon_i$, are independent and identically distributed error terms with zero mean and constant variance, $t_i = i/n$ is a time index, $n$ indicates the total number of observations, and the function $m(t_i)$ is to be specified.

### 2.1. Cubic splines

A spline estimator to fit the hourly load curve has been used by Hendricks et al. (1979) and Mouchart and Roche (1987). The function $m(t_i)$ is specified as a cubic spline, that is, a polynomial series with continuous first and second derivatives, and a step-wise third derivative. The polynomials are thus linked by a series of nodes which correspond to flex points. The number of flex points $q$ determines estimation accuracy. If $q = n$, a smoothing spline is obtained, while $q < n$ gives a parametric spline.

Assuming $q = n$, a cubic spline function can be interpreted as a non-parametric regression estimator which arises from the solution to the following problem:

$$\max_{m \in W_2^{(2)}[0,1]} L_n(m), \tag{2}$$

where

$$L_n(m) = n^{-1} \sum_{i=1}^{n} w_{in} [y_i - m(t_i)]^2 + \lambda \int_0^1 \left[ \frac{\mathrm{d}^2 m(t)}{\mathrm{d}t^2} \right]^2 \mathrm{d}t. \tag{3}$$

In this context, $W_2^{(2)}[0,1]$ indicates the class of all twice periodic differentiable functions, whereas the loss