Available online at www.sciencedirect.com
ScienceDirect
NEUROCOMPUTING
Neurocomputing 70 (2007) 2379–2391
www.elsevier.com/locate/neucom

TVCAR models for forecasting

Federico Palacios-González*

Universidad de Granada, Facultad de CC. EE. y Empresariales, Campus de Cartuja s/n 18011 Granada, Spain

Available online 22 February 2007

Abstract

This paper predicts the 100 missing values in the CATS Benchmark using a Time-Varying Coefficient Autoregressive Model (TVCAR). The TVCAR model is an autoregressive model in which the coefficients vary smoothly with time. The model is fitted to the first differences of the data by minimising the residual sum of squared, subject to certain restrictions that enable the gaps left by the missing observations to be bridged. The path of each time-varying coefficient is initially described by a combination of cosine functions. Later, the method is improved replacing the cosine specifications by piecewise polynomials.

Keywords: Time series; Non-linear models; Time-varying coefficient models; Fourier AR models

1. Introduction

In a constantly changing environment, stable equilibrium systems are not often observed. Nevertheless, smoothly evolving systems, in the short and medium term, can be considered quasi-stable and they allow an approximated stable treatment.

For analysing smoothly evolving systems, it would be appropriate to use non-stationary models capable of absorbing the structural changes that can be observed in such systems. This situation can be observed, for instance, in long-term series, whose generating system has gently and continuously evolved. A time-varying model can be appropriate for fitting such series.

In empirical applications "many series show a non-stationary behaviour (e.g. in economics or sound analysis)" [3]. The topics of Evolutionary Time Series Models and Time-Varying Coefficients Models have already been dealt with in specialised literature. Examples can be found in [3,6,7,9–13,16]. A simple way to build time-varying models is by using, as the starting point, a linear model whose parameters are replaced by time functions with a number of constants. This is a simple idea that is very often employed in non-linear modelling techniques [19]. It is also used in the analysis of time series and in signal processing methods.

Fig. 1 shows that the original series is integrated from another series u_t, whose generating model is unknown. To collect the useful information contained in such a series, a model composed of a locally stationary part [3], plus white noise has been fitted. When integrating, the accumulation of the white noise part makes a random walk appear as an abrupt trend similar to that observed in the data.

This paper recommends the use of an autoregressive model (AR) [1,2,8] for the series u_t, which is obtained by differentiating the given data. In other words, it is a model with the form:

$$u_t = \gamma_0 + \sum_{i=1}^{p} \gamma_j u_{t-j} + \varepsilon_t,$$

with coefficients γ_j that vary smoothly with time. Therefore, the coefficients will be functions of time t that will be specified. The random perturbations ε_t are non-correlated random variables with a mean of zero and constant variance. This type of model is given the name of "Time-Varying Coefficient Autoregressive Model (TVCAR). This non-linear model is fitted by minimising the residual sum of squares (RSS) under four restrictions. The restrictions are required in order to link the predicted values with the adjacent observations. Thus, the predicted values from y_{981}

*Tel.: +34 958 242979.
E-mail address: fpalacio@ugr.es.

0925-2312/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2006.04.010

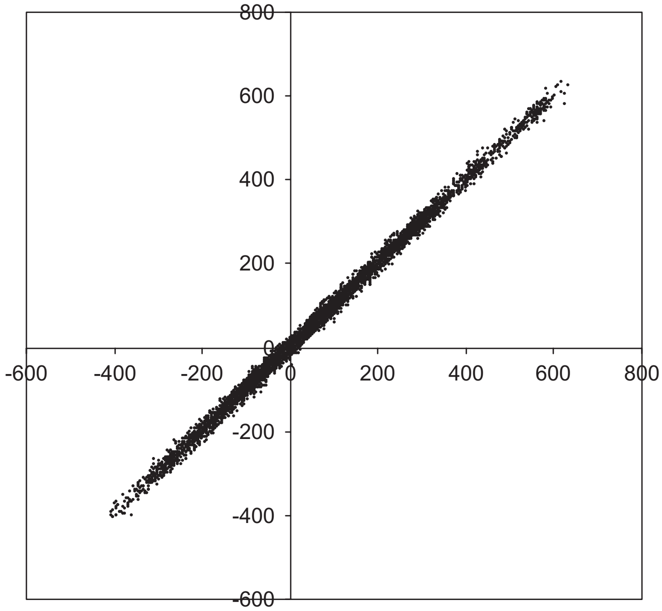

Fig. 1. (y_{t-1}, y_t) pairs.

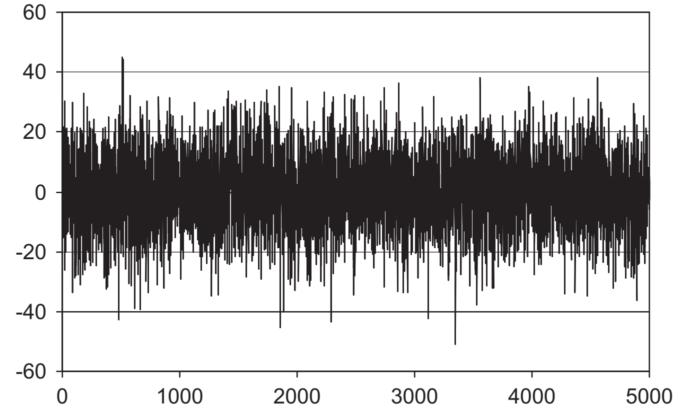

Fig. 2. $(u_t = y_t - y_{t-1})$.

to y_{1000} must link the given values y_{980} and y_{1001}, and likewise for the second, third and fourth blocks.

At the outset, it has been necessary to reduce the data by applying the backwards-difference operator to obtain the series u_t [14].

This paper is organised as follows: Section 2 describes the method used for prediction. Section 3 provides the results obtained on the initial values of the Benchmark. Section 4 examines the results of Section 3, describing the advantages and disadvantages of the current method. Section 5 describes the improved method for prediction. Section 6 applies the method to a new set of data and, finally, Section 7 gives the conclusions.

2. Method for prediction

2.1. Exploratory data analysis

Fig. 1 shows a dispersion graph of the points (y_{t-1}, y_t) which comprise consecutive data values. The model $y_t = y_{t-1}+u_t$ can be taken as an initial working hypothesis; and efforts can be devoted to analysing the series u_t, which is represented in Fig. 2.

The Partial Autocorrelation Function (PACF) [14] can be used in finding the order of an AR process to model the differentiated data. Fig. 3a shows the estimated PACF based on values from u_2 to u_{980}. The estimated PACF using values from u_{1002} to u_{1980}, from u_{2002} to u_{2980}, from u_{3002} to u_{3980} and from u_{4002} to u_{4980}, are shown in Fig. 3b–e, respectively. Graphs from Fig. 3a–e indicate the appropriate order for an AR process describing the data. They suggest that a stationary time-invariant AR model of an order of 16 as follows:

$$u_t = \gamma_0 + \sum_{j=1}^{p} \gamma_j u_{t-j} + \varepsilon_t \tag{1}$$

might be appropriate, where γ_j are real constants and ε_t are non-correlated random variables of mean zero and constant variance.

However, a continuous evolution and fluctuation on the PACF graph bars corresponding to this AR model is also found, and this suggests that the model's coefficients must vary in time. It is known that the bar for the h order lag, in the PACF graph, is the partial correlation coefficient between the variables u_t and u_{t-h}. In turn, the relation between u_t and u_{t-j} is defined in (1) by means of the coefficient γ_j; both amounts are, therefore, related.[1] Moreover, if the PACF estimations on the two consecutive blocks of data for the same series are sufficiently different, then the AR models that are fitted on the said blocks also have to be different. That is, an AR model of constant coefficients for the whole series would, in this case, be inappropriate. The fluctuation observed among graphs 3a–3e leads us to suspect that an AR model of variable coefficients would also be appropriate for u_t.

In order to check whether this evolution over time is real and not merely a random sample fluctuation, an AR(16) model of constant coefficients has been fitted on each block of 979 data points and the results have been ordered in a table of 16 rows and 5 columns. Column j contains the estimated coefficients of the AR(16) in block j, where $j = 1$, 2, 3, 4, 5. This table is subjected to (see Appendix A) a two-way ANOVA test [17]. It is found that the blocks (columns) effect is highly significant (P-value = 5.1E-5). This demonstrates that there are significant differences among the AR(16) models fitted on the five blocks. An

[1] A mathematical relation can be established between both if it is taken into account that the Autocorrelation Function (ACF) verifies the recurrence expression defined by the AR model (from which the Yule–Walker relations are obtained) that allows us to calculate the coefficients γ_j of the AR model from the ACF, and if the definition of the PACF are based on the ACF [14,18].