Journal of Statistical Planning and Inference 140 (2010) 3734–3741

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Semiparametric distribution forecasting

Benjamin Kedem *, Richard E. Gagnon

Department of Mathematics, University of Maryland, College Park, MD 20742, USA

ARTICLE INFO

Available online 21 May 2010

Keywords:
Reference distribution
Exponential tilt
Empirical likelihood
Predictive distribution
Residual
Regression models

ABSTRACT

Given m time series regression models, linear or not, with additive noise components, it is shown how to estimate semiparametrically the predictive probability distribution of one of the time series conditional on past random covariate data. This is done by assuming that the distributions of the residual components associated with the regression models are tilted versions of a reference distribution.

1. Introduction

The problem of time series prediction is revisited in this paper, given m time series that are regressed linearly or nonlinearly on covariate time records. The m time series need not be of the same length, they may be stationary or nonstationary, dependent or independent, and they may be relatively short. The approach depends on well behave residuals in a system of m regression equations, and certain "tilt" relationships between their probability density functions. That is, relationships between probability densities $g_1,\ldots,g_q$ and a reference or baseline density $g\equiv g_m$ of the form

$$g_j(x)=\exp\{\alpha_j+\boldsymbol{\beta}_j'\boldsymbol{h}(x)\}g(x),\quad j=1,\ldots,q, \tag{1}$$

with scalars α_j, $p\times 1$ vectors $\boldsymbol{\beta}_j$, and a known $p\times 1$ vector of real-valued functions $\boldsymbol{h}(x)$.

Using the combined residual data from several time series regression models it is shown how to estimate the probability distribution of a "reference" time series and use it in conditional prediction. In essence, it is a certain extension of a semiparametric method to time series discussed in the context of random samples in Fokianos et al. (2001), Gilbert et al. (1999), Qin (1993), Qin and Lawless (1994), Qin and Zhang (1997), and Zhang (2000a, b).

The present paper should be viewed as a review of some recent developments akin to the cited references and related work. It presents a novel concept in time series prediction and some supporting empirical evidence in terms of real data. Our basic idea of estimating predictive distributions is an alternative to Bayesian methods described, for example, in Geisser (1993) (general), De Oliveira et al. (1997) (spatial interpolation), and Kedem and Fokianos (2002) (time series).

2. A density ratio model for time series

Consider now the following system of $m= q+1$ time series regression models:

$$x_{1t}=f_1(\mathbf{z}_{1,t-1})+\varepsilon_{1t},\quad t=1,\ldots,n_1,$$

$$\vdots$$

* Corresponding author. Fax: +1 301 314 0827.
E-mail addresses: bnk@math.umd.edu (B. Kedem), RichardEGagnon@aol.com (R.E. Gagnon).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2010.04.038

$$x_{qt} = f_q(\mathbf{z}_{q,t-1}) + \varepsilon_{qt}, \quad t = 1, \ldots, n_q,$$

$$x_{mt} = f_m(\mathbf{z}_{m,t-1}) + \varepsilon_{mt}, \quad t = 1, \ldots, n_m, \tag{2}$$

where the vectors $\mathbf{z}_{k,t-1}$ contain past values of covariate time series possibly including even past values of $x_{1\ t}, \ldots, x_{qt}, x_{mt}$, and where the ε_k, k=1,..., q, m, are independent noise components. A special case of (2) is multivariate first order autoregression. *We approach time series prediction through the distribution of the noise components.*

To motivate the basic idea of the paper it is helpful to first treat rigorously the scenario where each noise sequence $\{\varepsilon_{kt}\}$, corresponding to the kth regression model, consists of independent and identically distributed (iid) random variables. In applications the $\{\varepsilon_{kt}\}$ are replaced by the corresponding residuals $\{\hat{\varepsilon}_{kt}\}$, as is done in two examples below.

Suppose that for each t, ε_{jt} is distributed according to an unknown probability density $g_j(x)$, j=1,...,q, m. Designate $g(x) \equiv g_m(x)$ as the *reference* density. Then, emboldened by the results in the aforementioned references, we shall assume that each g_j is a tilt or distortion of the reference g as in (1). Zhang (2000a) advocates the use of $h(x)= x$ or $\boldsymbol{h}(x) = (x,x^2)$ in logistic discriminant analysis. The choice of $\boldsymbol{h}(x) = (x,x^2)$ also has been found useful in cluster detection which requires repeated multiple testing (Kedem and Wen, 2007). In an application to radar meteorology, $h(x)$=log x was used in conjunction with reflectivity data (Kedem et al., 2004). Notice that the $\boldsymbol{\beta}_j$ are scalars whenever $h(x)$ is a real valued function.

The objective is to estimate all the $\alpha_j, \boldsymbol{\beta}_j$, the reference density g, and the corresponding cdf G, for the purpose of predicting the future reference value $x_{m,t+1}$. This is done using the combined noise "data" from all the m "samples"

$$\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n) \equiv \{(\varepsilon_1, \ldots, \varepsilon_{1n_1}), \ldots, (\varepsilon_{q1}, \ldots, \varepsilon_{qn_q}), (\varepsilon_{m1}, \ldots, \varepsilon_{mn_m})\} \tag{3}$$

of length $n = n_1 + \cdots + n_q + n_m$. Then, using the estimator $\hat{G}$ of G we can estimate future probabilities of events formulated in terms of the "reference" $x_{m,\ t+1}$ conditional on $\mathbf{z}_{m,t}$.

2.1. Comparison distributions

Interestingly, tilting may be viewed as a variation of *comparison densities*, a concept advanced by Emanuel Parzen in many of his papers, for example Parzen (2004). Accordingly, when F and G are both continuous with probability densities f and g, respectively, and $F \ll G$, define the *comparison distribution*

$$D(u; G,F) = F(G^{-1}(u)), \quad 0 < u < 1.$$

The corresponding *comparison density* is defined as

$$d(u; G,F) = f(G^{-1}(u))/g(G^{-1}(u))$$

or

$$f(G^{-1}(u)) = d(u; G,F)g(G^{-1}(u)) \tag{4}$$

which formally is similar to (1). That is, (4) is a form of tilting of a reference g. An analogous relationship holds in the discrete case.

3. Semiparametric estimation

A maximum likelihood estimator of $G(x)$ can be obtained by maximizing the likelihood over the class of step cdf's with jumps at the values $\tau_1, \ldots, \tau_n$ (Fokianos et al., 2001; Gilbert et al., 1999; Qin and Lawless, 1994; Qin and Zhang, 1997).

Let $w_j(\tau) = \exp\{\alpha_j + \boldsymbol{\beta}_j' \boldsymbol{h}(\tau)\}$, j=1,..., q, and $p_i = dG(\tau_i)$, i=1,..., n. Then the *empirical likelihood* becomes (Fokianos et al., 2001; Owen, 2001; Qin and Zhang, 1997)

$$\begin{aligned}\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_q, G) &= \prod_{i=1}^{n} p_i \prod_{j=1}^{n_1} \exp(\alpha_1 + \boldsymbol{\beta}_1' \boldsymbol{h}(\varepsilon_{1j})) \cdots \prod_{j=1}^{n_q} \exp(\alpha_q + \boldsymbol{\beta}_q' \boldsymbol{h}(\varepsilon_{qj})) \\ &= \prod_{i=1}^{n} p_i \prod_{j=1}^{n_1} w_1(\varepsilon_{1j}) \cdots \prod_{j=1}^{n_q} w_q(\varepsilon_{qj}).\end{aligned} \tag{5}$$

Fix $\boldsymbol{\alpha}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_q$. Then maximizing (5) with respect to the p_i subject to the constraints

$$\sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i[w_1(\tau_i) - 1] = 0, \ldots, \sum_{i=1}^{n} p_i[w_q(\tau_i) - 1] = 0$$

we obtain (Fokianos et al., 2001; Kedem and Wen, 2007),

$$p_i = \frac{1}{n_m} \cdot \frac{1}{1 + \rho_1 w_1(\tau_i) + \cdots + \rho_q w_q(\tau_i)}, \tag{6}$$

where $\rho_j = n_j / n_m$, j=1,..., q, are the relative series sizes. Substituting the p_i in (6) back into (5) gives the log-likelihood as a