Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 52 (2008) 2435–2450
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Neural networks for bandwidth selection in local linear regression of time series

F. Giordano*, M.L. Parrella

Department of Economic and Statistical Sciences, University of Salerno, 84084 Fisciano, SA, Italy

Received 9 October 2006; received in revised form 2 August 2007; accepted 4 August 2007
Available online 30 August 2007

Abstract

The problem of automatic bandwidth selection in nonparametric regression is considered when a local linear estimator is used to derive nonparametrically the unknown regression function. A plug-in method for choosing the smoothing parameter based on the use of the neural networks is presented. The method applies to dependent data generating processes with nonlinear autoregressive time series representation. The consistency of the method is shown in the paper, and a simulation study is carried out to assess the empirical performance of the procedure.

Keywords: Nonparametric regression; Bandwidth selection; Derivative estimation; Dependent data; Local linear estimators; Neural networks

1. Introduction

Let $\{X_t; t = 1, \ldots, n\}$ be a realization of length n from a strictly stationary stochastic process $\{X_t; t \in \mathbb{N}\}$. A general class of models for the conditional mean function is the nonlinear autoregressive model of order 1, that is

$$X_t = m(X_{t-1}) + \varepsilon_t, \tag{1}$$

where $\{\varepsilon_t\}$ are *i.i.d.* random variables with mean 0 and variance $\sigma^2 \in (0, +\infty)$. The unknown function $m(\cdot)$ and its derivatives can be estimated nonparametrically using local polynomial (LP) estimators. These estimators have gained wide interest in recent years for the nonparametric analysis of time series, because they have good theoretical properties and they can be applied to different context, such as identification, estimation and prediction (see, for example, Masry and Fan, 1997).

Beyond their appealing statistical properties, LP estimators are subject to the correct choice of a tuning parameter, the *bandwidth* of the kernel function, which makes critical their practical implementation. In order to reduce the bias of the nonparametric estimate, a relatively small bandwidth should be chosen, but this may cause large standard errors. The best choice of the bandwidth involves, therefore, a trade-off between bias and variance. It is very useful for the analyst to have a data-driven bandwidth selector that estimates the correct amount of smoothing. As a consequence, the problem of automatic choice of smoothing parameters has been widely studied.

* Corresponding author. Tel.: +39 89962213.
E-mail addresses: giordano@unisa.it (F. Giordano), mparrella@unisa.it (M.L. Parrella).

0167-9473/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2007.08.013

The aim of this paper is to propose a plug-in method for the automatic selection of the bandwidth in the LP estimation of the function $m(\cdot)$. We use the Neural Netwrok (NN) technique to estimate the unknown functionals which appear in the asymptotic formula of the "optimal" bandwidth. The method applies to dependent data generating processes with a nonlinear autoregressive time series representation. Therefore, it is also useful for the analysis of econometric and financial time series.

Now we provide some relevant references on the topic, to outline the original contribution of this paper. First of all, the principal bandwidth selection methods studied in the literature can be divided into two broad categories: cross-validation methods and plug-in methods. Both of these methods have been intensively studied, but the bulk of the literature deals with the setting of independent data (for dependent data see, for example, Hart, 1994; Kim, 1997; Sköld, 2000; Hall et al., 1995a,b). Each of these approaches has its own strengths and weaknesses. Nevertheless, the two methods have been compared in the independent data context by several researchers. There are many papers which show evidence of a substantial superiority of plug-in procedures over cross-validation ones (see, Loader, 1999; Ruppert et al., 1995; Park and Turlach, 1992; Hall et al., 1991; Chiu, 1991). Anyway, the main criticism addressed to the plug-in approach is that they are difficult to implement (see, for example, Loader, 1999; Levine, 2006). For dependent observations, Hart (1996) presented a review of the methods for the choice of the smoothing parameter for LP estimators. He also applied the time series cross-validation method to the local linear estimator. Hall et al. (1995a) analysed methods for the choice of the bandwidth based on the block bootstrap and modified cross-validation, when errors have an almost arbitrarily long-range dependence. They applied such methods to kernel estimators as well as to local linear estimators. Opsomer (1997) and Francisco-Fernández et al. (2004) presented a plug-in method, which is a generalization of a method of Ruppert et al. (1995) to the regression model with correlated errors. While in the first paper an entirely non-parametric procedure was proposed, in the paper of Francisco-Fernández et al. (2004) the presence of correlated errors was taken explicitly into account through a parametric correlation function specification. All of the papers mentioned so far dealt with nonparametric regression models with fixed design, where the dependence of the observations arose from the errors. Here we consider a different dependence framework: an autoregressive model, such as the model presented in (1), which is useful to model nonlinear structures commonly encountered in econometric and financial time series (Tjostheim, 1994). This model is also considered the papers of Härdle and Vieu (1992), Kim and Cox (1997) and Masry and Fan (1997). In the first two papers, the cross-validation method was applied to Nadaraya–Watson kernel estimators of the conditional mean function. Masry and Fan (1997) considered LP estimators, which have better asymptotic properties than Nadaraya–Watson's. They identified the theoretical optimal bandwidth for the local estimation, but do not give any suggestion on how to estimate such optimal bandwidth. In this paper we consider the framework of Masry and Fan (1997), and propose a plug-in procedure for the estimation of a global optimal bandwidth for the local linear estimation of the conditional mean function $m(x)$, based on the use of the NNs.The procedure is new and different from those proposed so far, as it will be shown later on in this paper.

The organization of the work is as follows: the next section reports the assumptions on the process and a brief description of the Local Linear estimator and of the NN estimator; we also present in this section two lemmas, which show that these two estimators are consistent for the estimation of the function $m(x)$, under some mild assumptions; in the third section we describe the use of the NN approach in the context of Local Linear estimation, in order to derive a new plug-in bandwidth selector; here we also present the theorems which show theoretically the consistency of the proposed method, while in Section 4 a simulation experiment gives evidence of its finite sample performance; finally, the last section reports some concluding remarks.

2. Assumptions and preliminaries on LP and NN estimators

Given the model (1), we consider the problem of estimating the conditional mean function

$$m(x) = E(X_t | X_{t-1} = x); \tag{2}$$

using the realization $\{X_t; t = 1, \ldots, n\}$. We make the following assumptions on the process:

Assumptions A.

(A1) the process $\{X_t\}$ is strictly stationary, with X_t real valued;

(A2) the function $m(\cdot)$ belongs to $C^2(\mathbb{R})$, where $C^2(\mathbb{R})$ represents the class of functions which have a continuous second order derivative on $\mathbb{R}$.