Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Statistical Planning and Inference 136 (2006) 730–769

journal of statistical planning and inference

www.elsevier.com/locate/jspi

On semiparametric M-estimation in single-index regression

Michel Delecroix, Marian Hristache*, Valentin Patilea[1]

CREST-ENSAI, Campus de Ker Lann, Rue Blaise Pascal, BP 37203, 35172 Bruz Cedex, France

Received 17 November 2003; accepted 2 September 2004
Available online 31 October 2004

Abstract

In this paper we analyze a large class of semiparametric M-estimators for single-index models, including semiparametric quasi-likelihood and semiparametric maximum likelihood estimators. Some possible applications to robustness are also mentioned. The definition of these estimators involves a kernel regression estimator for which a bandwidth rule is necessary. Given the semiparametric M-estimation problem, we propose a natural bandwidth choice by joint maximization of the M-estimation criterion with respect to the parameter of interest and the bandwidth. In this way we extend a methodology first introduced by Härdle et al. (Ann. Statist. 21 (1993) 157) for semiparametric least-squares. We prove asymptotic normality for our semiparametric estimator. We derive the asymptotic equivalence between our bandwidth and the optimal bandwidth obtained through weighted cross-validation. Empirical evidence obtained from simulations suggests that our bandwidth improves the higher order asymptotics of the semiparametric M-estimator when it replaces the usual bandwidth chosen by cross-validation.

MSC: 62G05; 62G08; 62G20; 62J12

Keywords: Semiparametric M-estimator; Single-index model; Bandwidth selection; Cross-validation; U-processes; Semiparametric quasi-likelihood; Robustness

* Corresponding author.
E-mail addresses: delecroi@ensai.fr (M. Delecroix), hristach@ensai.fr (M. Hristache), valentin.patilea@ensai.fr (V. Patilea).

[1] Part of the work for this paper was accomplished while this author was at LEO, Université d'Orléans.

0378-3758/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2004.09.006

1. Introduction

Consider the problem of estimating a regression function $m(x) = E(Y|X = x)$ from independent copies $(Y_1, X_1^{\mathrm{T}})^{\mathrm{T}}, \ldots, (Y_n, X_n^{\mathrm{T}})^{\mathrm{T}}$ of a random vector $(Y, X^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{d+1}$. In GLM (generalized linear models; e.g., McCullagh and Nelder, 1989) it is assumed that $m(x) = r_0(x\theta_0)$ with r_0 known. Hereafter, $x\theta$ is a notation for $x^{\mathrm{T}}\theta$ when $x, \theta \in \mathbb{R}^d$. The function r_0 is the inverse of the so-called link function. Moreover, the conditional density $f_{Y|X=x}$ of Y given $X = x$ belongs to the linear exponential family, that is

$$f_{Y|X=x}(y) = \exp\left[B(r_0(x\theta_0)) + C(r_0(x\theta_0))y + D(y)\right],$$

where B, C and D are known functions.

A natural extension of GLM is provided by the semiparametric single-index models (SIM), where one only assumes the existence of some $\theta_0 \in \mathbb{R}^d$ (unique up to a scale normalization factor) such that

$$E(Y \mid X) = E(Y \mid X\theta_0), \tag{1.1}$$

that is $m(x)=r_0(x\theta_0)$, with unknown r_0. Since the regression $r_0(t)=E(Y \mid X\theta_0=t)$ depends on θ_0, hereafter, we shall write r_{θ_0} instead of r_0. In SIM framework, both θ_0 and r_{θ_0} are to be estimated. Numerous semiparametric approaches for root-n consistent estimation of θ_0 have been proposed: M-estimation (e.g., Ichimura, 1993; Sherman, 1994b; Delecroix and Hristache, 1999; Xia and Li, 1999; Xia et al., 1999), direct (average derivative based) estimation (e.g., Powell et al., 1989; Härdle and Stoker, 1989; Hristache et al., 2001a, b), iterative methods (e.g., Weisberg and Welsh, 1994; Chiou and Müller, 1998; Bonneu and Gba, 1998; Xia and Härdle, 2002).

Typically, the semiparametric M-estimators mentioned above can be written as

$$\widehat{\theta} = \arg\max_{\theta} \frac{1}{n} \sum_{i=1}^{n} \psi\left(Y_i, \widehat{r}_{\theta,h}^{\,i}(X_i\theta)\right) \tau_n(X_i), \tag{1.2}$$

where $\widehat{r}_{\theta,h}^{\,i}(t)$ is, for instance, the leave-one-out Nadaraya–Watson estimator (with bandwidth h) of $r_\theta(t)=E(Y \mid X\theta=t)$, $-\psi$ is a contrast function and $\tau_n(\cdot)$ is a so-called trimming function introduced to guard against small values for the denominators appearing in $\widehat{r}_{\theta,h}^{\,i}(t)$. Finally, the regression function $m(x)$ is estimated by $\widehat{r}_{\widehat{\theta},h}(x\widehat{\theta})$. Other smoothers, such as local polynomials and splines, can replace the Nadaraya–Watson estimator.

In order to estimate θ_0 and $r_{\theta_0}(\cdot\theta_0)$, two smoothing parameters seem to be necessary. First, after choosing a primary bandwidth h, the estimator $\widehat{\theta}$ is computed as in (1.2). Afterwards, $r_{\theta_0}(x\theta_0)$ is estimated by $\widehat{r}_{\widehat{\theta},h^*}(x\widehat{\theta})$, a kernel estimator, with bandwidth h^*, of the expectation of Y given $x\widehat{\theta}$. The rates of decay for the two bandwidths should verify some conditions. When $\psi(y,r) = -(y-r)^2$, Härdle et al. (1993) defined more directly

$$\left(\widehat{\theta}, \widehat{h}\right) = \arg\max_{\theta,h} \frac{1}{n} \sum_{i=1}^{n} \psi\left(Y_i, \widehat{r}_{\theta,h}^{\,i}(X_i\theta)\right) I_A(X_i). \tag{1.3}$$