Journal of Multivariate Analysis 133 (2015) 77–94

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Adaptive estimation of an additive regression function from weakly dependent data

Christophe Chesneau [a,*], Jalal Fadili [b], Bertrand Maillot [a]

[a] LMNO, CNRS-Université de Caen, Campus II, Science 3, 14032, Caen, France

[b] GREYC, CNRS-ENSICAEN-Université de Caen, 6, Bd du Maréchal Juin, 14050, Caen Cedex, France

ARTICLE INFO

Article history:
Received 26 September 2012
Available online 28 September 2014

AMS 2000 subject classifications:
62G07
62G20

Keywords:
Additive regression
Dependent data
Adaptivity
Wavelets
Hard thresholding

ABSTRACT

A d-dimensional nonparametric additive regression model with dependent observations is considered. Using the marginal integration technique and wavelets methodology, we develop a new adaptive estimator for a component of the additive regression function. Its asymptotic properties are investigated via the minimax approach under the $\mathbb{L}_2$ risk over Besov balls. We prove that it attains a sharp rate of convergence which turns to be the one obtained in the i.i.d. case for the standard univariate regression estimation problem.

1. Introduction

1.1. Problem statement

Let d be a positive integer, $(Y_i, \mathbf{X}_i)_{i\in\mathbb{Z}}$ be a $\mathbb{R}\times[0,1]^d$-valued strictly stationary process on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and ρ be a given real measurable function. The unknown regression function associated to $(Y_i, \mathbf{X}_i)_{i\in\mathbb{Z}}$ and ρ is defined by

$$g(\mathbf{x}) = \mathbb{E}(\rho(Y)|\mathbf{X}=\mathbf{x}), \quad \mathbf{x}=(x_1,\ldots,x_d)\in[0,1]^d.$$

In the additive regression model, the function g is considered to have an additive structure, i.e. there exist d unknown real measurable functions $g_1,\ldots,g_d$ and an unknown real number μ such that

$$g(\mathbf{x}) = \mu + \sum_{\ell=1}^{d} g_\ell(x_\ell). \tag{1.1}$$

For any $\ell\in\{1,\ldots,d\}$, our goal is to estimate g_ℓ from n observations $(Y_1,\mathbf{X}_1),\ldots,(Y_n,\mathbf{X}_n)$ of $(Y_i,\mathbf{X}_i)_{i\in\mathbb{Z}}$.

1.2. Overview of previous work

When $(Y_i,\mathbf{X}_i)_{i\in\mathbb{Z}}$ is a i.i.d. process, this additive regression model becomes the standard one. In such a case, Stone in a series of papers [34–36] proved that g can be estimated with the same rate of estimation error as in the one-dimensional case.

* Corresponding author.
E-mail address: christophe.chesneau@gmail.com (C. Chesneau).

http://dx.doi.org/10.1016/j.jmva.2014.09.005

The estimation of the component g_ℓ has been investigated in several papers via various methods (kernel, splines, wavelets, etc.). See, e.g., [4,21,23,29,30,1,2,33,40,32,17].

In some applications, as dynamic economic systems and financial times series, the i.i.d. assumption on the observations is too stringent (see, e.g., [19,38]). For this reason, some authors have explored the estimation of g_ℓ in the dependent case. When $(Y_i, \mathbf{X}_i)_{i\in\mathbb{Z}}$ is a strongly mixing process, this problem has been addressed by [5,11], and results for continuous time processes under a strong mixing condition have been obtained by [12,13]. In particular, they have developed non-adaptive kernel estimators for g_ℓ and studied its asymptotic properties.

1.3. Contributions

To the best of our knowledge, adaptive estimation of g_ℓ for dependent processes has been addressed only by [18]. The lack of results for adaptive estimation in this context motivates this work. To reach our goal, as in [40], we combine the marginal integration technique introduced by [28] with wavelet methods. We capitalize on wavelets to construct an adaptive thresholding estimator and show that it attains sharp rates of convergence under mild assumptions on the smoothness of the unknown function. By adaptive, it is meant that the parameters of the estimator do not depend on the parameter(s) of the dependent process nor on those of the smoothness class of the function. In particular, this leads to a simple estimator.

More precisely, our wavelet estimator is based on term-by-term hard thresholding. The idea of this estimator is simple: (i) we estimate the unknown wavelet coefficients of g_ℓ based on the observations; (ii) then we select the greatest ones and ignore the others; (iii) and finally we reconstruct the function estimate from the chosen wavelet coefficients on the considered wavelet basis. Adopting the minimax point of view under the $\mathbb{L}_2$ risk, we prove that our adaptive estimator attains a sharp rate of convergence over Besov balls which capture a variety of smoothness features in a function including spatially inhomogeneous behavior. The attained rate corresponds to the optimal one in the i.i.d. case for the univariate regression estimation problem (up to an extra logarithmic term).

1.4. Paper organization

The rest of the paper is organized as follows. Section 2 presents our assumptions on the model. In Section 3, we describe wavelet bases on [0, 1], Besov balls and tensor product wavelet bases on $[0, 1]^d$. Our wavelet hard thresholding estimator is detailed in Section 4. Its rate of convergence under the $\mathbb{L}_2$ risk over Besov balls is established in Section 5. A comprehensive simulation study is reported and discussed in Section 6. The proofs are detailed in Section 7.

2. Notations and assumptions

In this work, we assume the following on our model:

Assumptions on the variables.

- For any $i \in \{1, \ldots, n\}$, we set $\mathbf{X}_i = (X_{1,i}, \ldots, X_{d,i})$. We suppose that
  - for any $i \in \{1, \ldots, n\}$, $X_{1,i}, \ldots, X_{d,i}$ are identically distributed with the common distribution $\mathcal{U}([0, 1])$,
  - $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are identically distributed with the common known density f.
- We suppose that the following identifiability condition is satisfied: for any $\ell \in \{1, \ldots, d\}$ and $i \in \{1, \ldots, n\}$, we have

$$\mathbb{E}(g_\ell(X_{\ell,i})) = 0. \tag{2.1}$$

Strongly mixing assumption. Throughout this work, we use the strong mixing dependence structure on $(Y_i, \mathbf{X}_i)_{i\in\mathbb{Z}}$. For any $m \in \mathbb{Z}$, we define the mth strongly mixing coefficient of $(Y_i, \mathbf{X}_i)_{i\in\mathbb{Z}}$ by

$$\alpha_m = \sup_{(A,B)\in\mathcal{F}^{(Y,\mathbf{X})}_{-\infty,0}\times\mathcal{F}^{(Y,\mathbf{X})}_{m,\infty}} \left|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)\right|, \tag{2.2}$$

where $\mathcal{F}^{(Y,\mathbf{X})}_{-\infty,0}$ is the σ-algebra generated by $\ldots, (Y_{-1}, \mathbf{X}_{-1}), (Y_0, \mathbf{X}_0)$ and $\mathcal{F}^{(Y,\mathbf{X})}_{m,\infty}$ is the σ-algebra generated by $(Y_m, \mathbf{X}_m), (Y_{m+1}, \mathbf{X}_{m+1}), \ldots$.

We suppose that there exist two constants $\gamma > 0$ and $\upsilon > 0$ such that, for any integer $m \geq 1$,

$$\alpha_m \leq \gamma \exp(-\upsilon m). \tag{2.3}$$

This kind of dependence is reasonably weak. Further details on strongly mixing dependence can be found in [3,39,16,27,6].

Boundedness assumptions.

- We suppose that $\rho \in \mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_\infty(\mathbb{R})$, i.e. there exist constants $C_1 > 0$ and $C_2 > 0$ (supposed known) such that

$$\int_{-\infty}^{\infty} |\rho(y)|dy \leq C_1, \tag{2.4}$$

$$\text{and} \quad \sup_{y\in\mathbb{R}} |\rho(y)| \leq C_2. \tag{2.5}$$