Journal of Multivariate Analysis 140 (2015) 283–290

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Minimax rate of convergence for an estimator of the functional component in a semiparametric multivariate partially linear model

Michael Levine

Purdue University, United States

ARTICLE INFO

Article history:
Received 10 May 2014
Available online 21 May 2015

AMS subject classifications:
62G08
62G10

Keywords:
Multivariate semiparametric partial linear model
Minimax rate of convergence
Functional component
Lower bound
Fano's lemma
Varshamov–Gilbert bound

ABSTRACT

A multivariate semiparametric partial linear model for both fixed and random design cases is considered. Earlier, in Brown et al. (2014), the model has been analyzed using a difference sequence approach. In particular, the functional component has been estimated using a multivariate Nadaraya–Watson kernel smoother of the residuals of the linear fit. Moreover, this functional component estimator has been shown to be rate optimal if the Lipschitz smoothness index exceeds half the dimensionality of the functional component domain. In the current manuscript, we take this research further and show that, for both fixed and random designs, the rate achieved is the minimax rate under both risk at a point and the L_2 risk. The result is achieved by proving lower bounds on both pointwise risk and the L_2 risk of possible estimators of the functional component.

1. Introduction

Semiparametric models have a long history in statistics and have received considerable attention in the last several decades. The main reason they are of considerable interest is that, quite often, the relationships between the response and predictors are very heterogeneous in the same model. Some of the relationships are clearly linear whereas the detailed information about others is hard to come by. In many situations, a small subset of variables is presumed to have an unknown relationship with the response that is modeled nonparametrically while the rest are assumed to have a linear relationship with it. As an example, Engle, Granger, Rice, and Weiss [3] studied the nonlinear relationship between temperature and electricity usage where other related factors, such as income and price, are parameterized linearly.

The model we consider in this paper is a semiparametric partial linear multivariate model

$$Y_{\mathbf{i}} = a + X_{\mathbf{i}}'\beta + f(U_{\mathbf{i}}) + \varepsilon_{\mathbf{i}} \tag{1.1}$$

where $X_{\mathbf{i}} \in \mathbb{R}^p$ and $U_{\mathbf{i}} \in \mathbb{R}^q$, β is an unknown $p \times 1$ vector of parameters, a is an unknown intercept term, $f(\cdot)$ is an unknown function and $\varepsilon_{\mathbf{i}}$ are independent and identically distributed random variables with mean 0 and constant variance σ^2. We consider two cases with respect to U: a random design case whereby $U_{\mathbf{i}}$ is a q-dimensional random variable and a fixed design case with $U_{\mathbf{i}}$ being a q-dimensional vector where each coordinate is defined on an equispaced grid on [0, 1]. In the fixed design case the errors are independent of $X_{\mathbf{i}}$ while in the random design case they are independent of $(X_{\mathbf{i}}', U_{\mathbf{i}})$. To obtain meaningful results, the function f is assumed to belong in the Lipschitz ball $\Lambda_\alpha(M)$ where α is the Lipschitz exponent. Of

E-mail address: mlevins@purdue.edu.

http://dx.doi.org/10.1016/j.jmva.2015.05.010

particular interest is the fact that, to be coherent, in the fixed design case when $q > 1$ the model (1.1) must have multivariate indices. The version with $q = 1$ was earlier considered in [11] while [1] considered the case of $q > 1$ in detail. The latter defined two conceptually similar difference based estimators of the parametric component for fixed and random design cases, respectively, and showed $\sqrt{n}$ asymptotic normality of both of these estimators. Moreover, it was also established that, in order for the estimator of the parametric component to be efficient, the order of the difference sequence must go to infinity. Brown, Levine and Wang [1] also obtained a uniform over a Lipschitz ball $\Lambda^\alpha(M)$ convergence result for an estimator of the functional component, establishing the rate of convergence $n^{-2\alpha/(2\alpha+q)}$.

To the best of our knowledge, the optimal in the minimax sense rate of convergence for estimators of the nonparametric component of multivariate partial linear model (1.1) has not been established. Since results of [1] amount to establishing the upper bound of that rate, the remaining task is to establish the lower bound. In this manuscript, we are doing just that for both fixed and random designs as well as for the two different functional distances. The first distance considered is the difference at a given fixed point and the second is that $L_2[0, 1]^q$ distance. A number of different techniques are used to obtain these results.

Before proceeding, it is probably useful to recap quickly how the functional component estimator is constructed. The detailed discussion is available in [1]. We only describe what happens in the fixed design case. We begin with (normalized) "diagonal" differences of observations $Y_{\mathbf{i}}$. As in Cai, Levine and Wang [2] and Munk, Bissantz, Wagner and Freitag [9], we select first a set of q-dimensional indices $J = \{(0, \ldots, 0), (1, \ldots, 1), \ldots, (\gamma, \ldots, \gamma)\}$. Some specialized notation is needed first to describe resulting differences. For any vector $u \in \mathbb{R}^q$, a real number v and a set $A \subset \mathbb{R}^q$, we define the affine transformation of the set A is the set $B = u + vA = \{y \in \mathbb{R}^q : y = u + va, a \in A \subset \mathbb{R}^q\}$; then, we introduce a set R that consists of all indices $\mathbf{i} = (i_1, \ldots, i_q)$ such that $R + J \equiv \{(\mathbf{i} + j)|\mathbf{i} \in R, j \in J\} \subset \{1, \ldots, m\}^q$. Let a subset of $R + J$ corresponding to a specific $\mathbf{i} \in R$ be $\mathbf{i}+J$. In order to define a difference of observations of order γ, we define first a sequence of real numbers $\{d_j\}$ such that $\sum_{j=0}^{\gamma} d_j = 0$, and $\sum_{j=0}^{\gamma} d_j^2 = 1$ and $\sum_{j=0}^{\gamma} d_j j^k = 0$ for any power $k = 1, \ldots, \gamma$. Moreover, denote $c_k = \sum_{i=0}^{\gamma-k} d_i d_{i+k}$. Then the difference of order γ "centered" around the point $Y_{\mathbf{i}}$, $\mathbf{i} \in R$ is defined as

$$D_{\mathbf{i}} = \sum_{j\in J} d_j Y_{\mathbf{i}+J}. \tag{1.2}$$

Note that this particular choice of the set J makes numbering of difference coefficients d_j very convenient; since each q-dimensional index j consists of only identical scalars, that particular scalar can be thought of as a scalar index of d; thus, $\sum_{j\in J} d_j$ is the same as $\sum_{j=0}^{\gamma} d_j$ whenever needed. Now, let $Z_{\mathbf{i}} = \sum_{j\in J} d_j X_{\mathbf{i}+J}$, $\delta_{\mathbf{i}} = \sum_{j\in J} d_j f(U_{\mathbf{i}+J})$, and $\omega_{\mathbf{i}} = \sum_{j\in J} d_j \varepsilon_{\mathbf{i}+J}$, for any $\mathbf{i} \in R$. Then, by differencing the original model (2.1), one obtains $D_{\mathbf{i}} = Z_{\mathbf{i}}'\beta + \delta_{\mathbf{i}} + \omega_{\mathbf{i}}$ for all $\mathbf{i} \in R$. The ordinary least squares solution for β can then be written as

$$\hat{\beta} = argmin \sum_{\mathbf{i}\in R} (D_{\mathbf{i}} - Z_{\mathbf{i}}'\beta)^2.$$

In [1], the estimator of the nonparametric component f has been constructed in several stages. First, the vector coefficient β has been estimated as described above. Next, the intercept a has been estimated using the natural estimator $\hat{a} = \frac{1}{n}\sum_{\mathbf{i}\le n}(Y_{\mathbf{i}} - X_{\mathbf{i}}'\hat{\beta})$. Finally, the multivariate Nadaraya–Watson kernel smoother has been applied to the residuals $r_{\mathbf{i}} = Y_{\mathbf{i}} - \hat{a} - X_{\mathbf{i}}'\hat{\beta}$ from that fit to estimate the unknown function f. To construct the kernel smoother, one can, for example, select a univariate kernel function $K(U^l)$ for a specific coordinate U^l, $l = 1, \ldots, q$ such that $\int K(U^l)\,dU^l = 1$ and that has $\lfloor \alpha \rfloor$ vanishing moments. Next, one would usually chose an asymptotically optimal bandwidth $h = n^{-1/(2\alpha+q)}$ (see, for example, [4]), and define the univariate rescaled kernel as $K_h(U^l) = h^{-1}K(h^{-1}U^l)$. The q-dimensional product kernel is, then $K_h(U) = h^{-q}\prod_{l=1}^{q} K(h^{-1}U^l)$. Armed with this framework, the Nadaraya–Watson kernel weights can be defined as $W_{\mathbf{i},h}(U - U_{\mathbf{i}}) = \frac{K_h(U-U_{\mathbf{i}})}{\sum_{\mathbf{i}\le n} K_h(U-U_{\mathbf{i}})}$ Finally, the resulting kernel estimator of the function $f(U)$ can then be defined as

$$\hat{f}(U) = \sum_{\mathbf{i}\le n} W_{\mathbf{i},h}(U - U_{\mathbf{i}}) r_{\mathbf{i}}.$$

Note that in the univariate case, Wang, Brown and Cai [11] used the Gasser–Müller kernel to obtain the estimator of the functional component; for the multivariate case, Nadaraya–Watson estimator seems to be a better choice because it can be generalized easier to the multivariate case.

The next two sections present detailed results for the fixed and random design cases, respectively.

2. Optimal rates of convergence for the deterministic design case

We consider the following semiparametric model

$$Y_{\mathbf{i}} = a + X_{\mathbf{i}}'\beta + f(U_{\mathbf{i}}) + \varepsilon_{\mathbf{i}} \tag{2.1}$$

where $X_{\mathbf{i}} \in \mathbb{R}^p$, $U_{\mathbf{i}} \in S = [0, 1]^q \subset \mathbb{R}^q$, $\varepsilon_{\mathbf{i}}$ are iid zero mean random variables with variance σ^2 and finite absolute moment of the order $\delta+2$ for some small $\delta > 0$, that is, $E\,|\varepsilon_{\mathbf{i}}|^{\delta+2} < \infty$. As noticed earlier in Brown, Levine and Wang [1], the model (2.1) must have multidimensional indices $\mathbf{i} = (i_1, \ldots, i_q)'$ to be coherent. Throughout this work, we will use bold font $\mathbf{i}$ to refer to multivariate indices and regular font to refer to coordinates of a multivariate index. For some positive integer m, we can take