Journal of Multivariate Analysis 165 (2018) 262–269

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Minimax linear estimation at a boundary point

Wayne Yuan Gao

Department of Economics, Yale University, 28 Hillhouse Avenue, New Haven, CT 06511, USA

ARTICLE INFO

Article history:
Received 15 June 2017
Available online 4 February 2018

AMS subject classifications:
62G05
62G08

Keywords:
Boundary point
Minimax linear estimation
Modulus problem
Regression discontinuity

ABSTRACT

This paper characterizes the minimax linear estimator of the value of an unknown function at a boundary point of its domain in a Gaussian white noise model under the restriction that the first-order derivative of the unknown function is Lipschitz continuous. The result is then applied to construct the minimax optimal estimator for the regression discontinuity design model, where the parameter of interest involves function values at boundary points.

1. Introduction

This paper studies the problem of finding the minimax linear estimator of the value of an unknown function f at a boundary point when f is known to lie in a particular smoothness class $\mathscr{F}$ of functions. Following Zhao [20], this paper takes $\mathscr{F}$ to be the space of continuously differentiable functions whose first-order derivatives are Lipschitz continuous with constant C, which is also known as the Hölder class of the second order. The following Gaussian regression model is also considered, in which a stochastic process $\{Y(t) : t \in \mathbb{R}_+ \equiv [0, \infty)\}$ is observed and assumed to be generated according to

$$dY(t) = f(t)\, dt + \sigma\, dW(t)\,, \tag{1}$$

where W represents standard Brownian motion. The parameter of interest considered here is $Lf = f(0)$, i.e., the value of f at the boundary point of its domain, and we seek to obtain a linear estimator $\hat{L}$ for $f(0)$ that minimizes the worst-case mean squared error, i.e., $\hat{L}$ solves

$$\inf_{\hat{L}\in\mathscr{L}} \sup_{f\in\mathscr{F}} \mathrm{E}_f\{(\hat{L} - Lf)^2\}.$$

As shown by Donoho [5], the problem of minimax linear estimation is essentially equivalent to solving the modulus problem, i.e., finding the least favorable function f^*. Following Zhao [20], it can be shown that this is also equivalent to minimizing $\|f\|^2$, where $\|\cdot\|$ denotes the L^2 norm, subject to an "initial condition" $f(0) = b$ in $\mathscr{F}$.

In Zhao [20], the domain of the unknown function f is taken to be $\mathbb{R}$ and the parameter of interest $f(0)$ is the value of f at an interior point. As a result, by the strict convexity of $\|\cdot\|^2$, the solution f^* must be symmetric about 0, i.e., f^* must be an even function. This implies that $(f^*)'(0) = 0$, imposing an initial condition on $(f^*)'$ in addition to $f^*(0) = b$. Zhao [20] then provides a characterization of the solution f^*.

The key difference between this paper and Zhao's lies in the fact that here, $f(0)$ is the value of f at a boundary point of the domain of f. As a result, the convexity of $\|\cdot\|^2$ no longer imposes any initial condition on the first order derivative

E-mail address: wayne.gao@yale.edu.

https://doi.org/10.1016/j.jmva.2018.01.001
0047-259X/

of f, and the least favorable function f^* requires the optimal specification of $(f^*)'(0)$. This paper provides a complete characterization of the solution f^*, which has two main features. First, the Lipschitz continuity constraint on $(f^*)'$ is binding on $\mathbb{R}_+$, i.e., $|(f^*)''| = C$ on $\mathbb{R}_+$ (almost everywhere). Second, assuming $f^*(0) > 0$ without loss of generality, the optimal initial condition on the first-order derivative satisfies $(f^*)'(0) < 0$.

This paper complements Zhao [20] and other research on the solution to the minimax estimation problem under various settings, such as [4,6,7,10,11,13–17,19]. See Armstrong and Kolesár [1, 2] for a review and the construction of honest confidence intervals for nonparametric regression models based on minimax optimal kernels.

A particular application of the present problem is the estimation of the regression-discontinuity parameter in the regression-discontinuity (RD) design model which, originally introduced by Thistlethwaite and Campbell [18], arises frequently in economic scenarios. Imbens and Lemieux [9] provide a review of various settings of the RD design model and a practical guide for estimation based on the work of Cheng et al. [4], who show the minimax optimality of their method in the Taylor class. As noted in [15,20], often times the Hölder class is a more natural restriction to make than the Taylor class because the former imposes a uniform bound on a derivative of a certain order globally, while latter only bounds the derivative locally at a point. The global restriction allowed by the Hölder class, when empirically reasonable, helps improve the minimax optimality of estimation.

Consider the Gaussian nonparametric regression model

$$\dot{Y}_t = f(t) + \sigma \dot{W}_t, \quad W_t \sim BM \; (1)$$

where f is a function defined on $\mathbb{R}$. Suppose that there is a regression discontinuity at 0, i.e.,

$$Lf = \lim_{x \searrow 0} f(x) - \lim_{x \nearrow 0} f(x) \neq 0.$$

We might be interested in estimating the RD parameter Lf: if a "treatment" is assigned whenever $x \geq 0$, then the parameter of interest Lf captures the "treatment effect" at $x = 0$. Define $f_+, f_- : \mathbb{R}_+ \to \mathbb{R}$ by

$$f_+(x) = \begin{cases} f(x) & \text{if } x > 0, \\ \lim_{y \searrow 0} f(y) & \text{if } x = 0, \end{cases} \qquad f_-(x) = \begin{cases} f(-x) & \text{if } x > 0, \\ \lim_{y \searrow 0} f(-y) & \text{if } x = 0. \end{cases}$$

The RD parameter $Lf = f_+(0) - f_-(0)$ is then the difference between f_+ and f_- at the boundary point of their domain $\mathbb{R}_+$. In Section 3, we show how the main result of this paper can be applied to construct the minimax linear estimator for the RD parameter.

The paper is organized as follows. Section 2 presents the model and the main result. Section 3 considers in more detail the application to the RD design model. Section 4 concludes the paper. The Appendix contains all proofs.

2. Model and main result

Following Donoho [5], consider the general model

$$y = Kf + \epsilon, \tag{2}$$

where $y, \epsilon \in \mathscr{Y}$, a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$, and ϵ is standard Gaussian with respect to $\langle \cdot, \cdot \rangle$, i.e., for all $g \in \mathscr{Y}$, $\langle g, \epsilon \rangle \sim \mathcal{N}(0, \|g\|^2 \sigma^2)$ with $\sigma^2 > 0$ known. Here, $f \in \mathscr{F}$ is an unknown real-valued function, where $\mathscr{F}$ is the set of admissible functions to which we restrict our attention, and $K : \mathscr{F} \to \mathscr{Y}$ is a known linear operator on $\mathscr{F}$. Suppose $L : \mathscr{F} \to \mathbb{R}$ is a given functional and that we are interested in estimating the parameter Lf.

Following Zhao [20], this paper considers the special case of the Gaussian regression model, in which we observe $\{Y(t) : t \in [0, \infty)\}$ as defined in Eq. (1) with σ a known parameter, and f an unknown function in $\mathscr{F}$. Equivalently, we specialize the general model (2) by taking $y = \dot{Y}$, $Kf = f$, $\epsilon = \sigma \dot{W}$ and $\mathscr{Y} = L_2(\mathbb{R}_+)$ with the standard inner product defined, for all $g, h \in L_2(\mathbb{R}_+)$, by

$$\langle g, h \rangle = \int_0^\infty g(t) \, h(t) \, dt.$$

By the standard Gaussianity of ϵ, we have, for all $g \in \mathscr{Y}$,

$$\langle g, \epsilon \rangle = \sigma \int_0^\infty g(t) \, dW(t) \sim \mathcal{N}(0, \sigma^2 \|g\|^2).$$

Moreover, we restrict f to be continuously differentiable and f' to be Lipschitz continuous with parameter C, i.e.,

$$\mathscr{F} = \mathscr{F}_H(2, C) = \{g \in L_2(\mathbb{R}_+) \cap C^1(\mathbb{R}_+) : \forall_{t,s \in \mathbb{R}_+} \; |g'(t) - g'(s)| \leq C\,|t - s|\},$$

where $\mathscr{F}_H(\gamma, C)$, as defined in [12], denotes the Hölder class with order $\gamma > 0$ and constant $C > 0$, viz.

$$\mathscr{F}_H(\gamma, C) = \{g \in L_2(\mathbb{R}_+) \cap C^{\lfloor \gamma \rfloor}(\mathbb{R}_+) : \forall_{t,s \in \mathbb{R}_+} \; |g^{(\lfloor \gamma \rfloor)}(t) - g^{(\lfloor \gamma \rfloor)}(s)| \leq C|t - s|^{\gamma - \lfloor \gamma \rfloor}\}$$

with $\lfloor \gamma \rfloor := \max\{k \in \mathbb{N} : k < \gamma\}$.