Ann Inst Stat Math (2010) 62:249–275
DOI 10.1007/s10463-008-0185-1

Oracle inequality for conditional density estimation and an actuarial example

Sam Efromovich

Received: 22 January 2007 / Revised: 29 February 2008 / Published online: 16 July 2008

Abstract Conditional density estimation in a parametric regression setting, where the problem is to estimate a parametric density of the response given the predictor, is a classical and prominent topic in regression analysis. This article explores this problem in a nonparametric setting where no assumption about shape of an underlying conditional density is made. For the first time in the literature, it is proved that there exists a nonparametric data-driven estimator that matches performance of an oracle which: (i) knows the underlying conditional density, (ii) adapts to an unknown design of predictors, (iii) performs a dimension reduction if the response does not depend on the predictor, (iv) is minimax over a vast set of anisotropic bivariate function classes. All these results are established via an oracle inequality which is on par with ones known in the univariate density estimation literature. Further, the asymptotically optimal estimator is tested on an interesting actuarial example which explores a relationship between credit scoring and premium for basic auto-insurance for 54 undergraduate college students.

Keywords Dimension reduction · Fixed and random design · MISE · Nonparametric regression

---

Supported in part by NSF Grant DMS–0604558, NSA Grant 0710075, and Grant TAF/CAS-07.

S. Efromovich (✉)
Department of Mathematical Sciences, The University of Texas at Dallas,
Richardson, TX 75083-0688, USA
e-mail: efrom@utdallas.edu

1 Introduction

Suppose that observations are n independent pairs (Y_l, X_l), $l = 1, \ldots, n$, and the problem is to estimate the conditional density (c.d.) of the response Y given the predictor X. Two classical regression designs will be considered simultaneously. The former design is random where pairs of observations are independent samples from a pair of two random variables Y and X. Under this design, suppose that the joint density $f(y, x)$ exists and the marginal density $p(x) := \int_{-\infty}^{\infty} f(y, x)\,\mathrm{d}y$ of the predictor is positive on its support, then we are estimating the conditional density $f(y|x) := \frac{f(y,x)}{p(x)}$. In what follows we shall assume that the support of $p(x)$ is [0, 1]. The latter design is where predictors are a permutation of deterministic $X_{(1)}, \ldots, X_{(n)}$ with $X_{(0)} = 0$, $X_{(n+1)} = 1$, $\int_{X_{(l)}}^{X_{(l+1)}} p(x)\,\mathrm{d}x = (n+1)^{-1}$, $l = 0, 1, \ldots, n$ and $p(x)$ being a positive probability density supported on [0, 1]. From now on, $p(x)$ will be referred to as the design density regardless of an underlying regression design. A discussion of these two designs can be found in Neter et al. (1996) and Eubank (1999); an interesting probabilistic point of view is presented in Arnold et al. (1999).

The considered statistical problem is to estimate the c.d. $f(y|x)$, as a bivariate function, under the Mean Integrated Squared Error (MISE) criterion. Neither the type of design, nor the design density, nor the shape/smoothness of an underlying conditional density are supposed to be known, and thus a suggested estimator should adapt to an underlying design and smoothness of the c.d.; see an interesting discussion about design and smoothness adaptive estimation in Fan (1992) and Efromovich (1999). Another desirable feature of a conditional density estimator is to perform as well as a univariate marginal density estimator under a traditional null hypothesis "the response does not depend on the predictor." This hypothesis implies that $f(y|x) \equiv f(y)$, and then the familiar curse of multidimensionality can be overcome. Note that the latter issue can be also referred to as a dimension–reduction property.

To satisfy the last property, the following expansion of the c.d. can be suggested:

$$f(y|x) = f(y) + \psi(y, x), \tag{1}$$

where $\psi(y, x)$ vanishes when the response does not depend on the predictor. Expansion (1) also explains how to choose an oracle which can be a benchmark for any c.d. estimator: this oracle should know both $f(y)$ and $\psi(y, x)$.

The literature on nonparametric c.d. estimation is not vast and it is primarily devoted to developing ad hoc estimators with the main theoretical emphasis on the bias–variance analysis. The interested reader can find a discussion and nice examples in the books by Prakasa Rao (1983), Fan and Gijbels (1996), Efromovich (1999) and Fan and Yao (2003), some possible extensions in Abramovich and Sapatinis (1999) and Koul and Sakhanenko (2005), as well as in Efromovich (2007) where a set of plausible conjectures about c.d. estimation is outlined. This article proves one of those conjectures: it is possible to develop a data-driven estimator that matches performance of an oracle which has all of the above–outlined wished statistical properties.

The content of this article is as follows. Section 2 defines an oracle to match and describes its statistical properties. A data–driven c.d. estimator, which mimics the oracle, is presented in Sect. 3. This section also contains an oracle inequality which shows how well the estimator matches performance of the oracle under the MISE criterion. Further, the suggested asymptotically optimal estimator is tested on an actuarial example with just 54 observations. Proofs can be found in Sect. 4.

It will benefit the reader if we finish the Introduction by presenting notation used in all sections of this article. Namely: i always denotes the complex unit, that is, $i^2 = -1$; $\mathrm{Re}\{\cdot\}$ is the real part; $o(1)$'s are generic sequences in n such that $o(1) \to 0$ as $n \to \infty$; C's are generic positive constants; $(x)_+ := \max(0, x)$; $\lfloor x \rfloor$ is the integer part of x; $I(\cdot)$ is the indicator; the cosine basis on $[0, 1]$ is denoted as $\varphi_0(x) := 1, \varphi_j := 2^{1/2}\cos(\pi j x),\ j = 1, 2, \ldots$ Two given arrays of nonnegative numbers $\{0 = b_1 < b_2 < \ldots\}$ and integers $\{b'_1 = 0, b'_2 = 1 + \lfloor \ln^{3/4}(n + 3) \rfloor, b'_{s+1} = b'_s + \lfloor b'_2(1 + 1/\ln\ln(n + 3))^{s-2} \rfloor; s = 2, 3, \ldots\}$ will be used to define blocks, and two given arrays of positive numbers $\{t_1, t_2, \ldots\}$ and $\{t_{k\tau} := 1/2 \ln\ln((k + 3)(\tau + 3)); k, \tau = 1, 2, \ldots\}$ will denote thresholds. Two different arrays of blocks are used for estimation of the univariate $f(y)$ and bivariate $\psi(y, x)$ components of the c.d. $f(y|x)$; remember (1). The former is $B_k := [b_k, b_{k+1})$, and the latter is $B_{k\tau} := \{(u, r) : u \in [b'_k, b'_{k+1}), r \in \{b'_\tau + 1, \ldots, b'_{\tau+1}\}$. The corresponding lengths/cardinality of these blocks are $L_k := b_{k+1} - b_k$ and $L_{k\tau} := (b'_{k+1} - b'_k)(b'_{\tau+1} - b'_\tau)$. We shall also use adjusted lengths

$$L^*_{k\tau} := L_{k\tau} \Bigg/ \left[\sum_{r=1}^{\infty} \int_0^\infty I((u, r) \in B_{k\tau}) \left[\left| \int_0^1 p^{-1}(x)\varphi_{2r}(x)\,\mathrm{d}x \right| + \int_0^1 \left| \int_{-\infty}^{\infty} e^{iuy} f(y|x)\,\mathrm{d}y \right|^2 \mathrm{d}x \right] \mathrm{d}u \right]. \quad (2)$$

2 Oracle for conditional density estimation

The primary aim of this section is to suggest an oracle, based on both data and underlying conditional and design densities (note that these densities are unknown to the statistician), which has three wished properties: (i) It can be considered as a benchmark for any estimator due its good statistical properties; (ii) It should be relatively simple for mimicking by a data-driven estimator; (iii) If the response is independent of the predictor and $f(y|x) \equiv f(y)$, then the bivariate oracle becomes a univariate one (performs a dimension reduction).

To incorporate the third property, expansion (1) will be used where the terms are defined via corresponding characteristic functions. Namely, write

$$f(y) = (2\pi)^{-1} \int_{-\infty}^{\infty} h_0(u)\mathrm{e}^{-iuy}\,\mathrm{d}y, \quad (3)$$

and the second term as

$$\psi(y, x) = \sum_{r=1}^{\infty} (2\pi)^{-1} \int_{-\infty}^{\infty} h_r(u)\mathrm{e}^{-iuy}\,\mathrm{d}u\varphi_r(x). \quad (4)$$

Here, given a particular $u \in (-\infty, \infty)$,

$$h_r(u) := \int_0^1 h(u|x)\varphi_r(x)\,\mathrm{d}x, \quad r = 0, 1, \ldots \tag{5}$$

is the rth Fourier coefficient of the conditional characteristic function

$$h(u|x) := \int_{-\infty}^{\infty} f(y|x)\mathrm{e}^{iuy}\,\mathrm{d}y. \tag{6}$$

Note that (3) and (4) are valid for both random and fixed regression designs.

The underlying idea of using representations (3) and (4) is that the original problem of c.d. estimation is converted into the problem of characteristic function estimation. Then it is possible to employ Efromovich–Pinsker (EP) oracle methodology motivated by the Wiener filter; the interested reader can find a comprehensive discussion of this oracle methodology in Efromovich (1999, 2000, 2005).

Let us define EP oracles $\tilde{f}^*(y)$ and $\tilde{\psi}^*(y, x)$ for estimation of the components $f(y)$ and $\psi(y, x)$, respectively. Write

$$\tilde{f}^*(y) := \pi^{-1} \int_0^{\infty} \mathrm{Re}\{\tilde{h}_0(u)\mathrm{e}^{-iuy}\}\,\mathrm{d}u, \tag{7}$$

where

$$\tilde{h}_0(u) := \sum_{k=1}^{K} \mu_k \hat{h}_0(u) I(u \in B_k), \quad u \geq 0. \tag{8}$$

Here and in what follows $\hat{h}_r(u)$ is a statistic

$$\hat{h}_r(u) := n^{-1} \sum_{l=1}^{n} e^{iuY_l}\varphi_r(X_l)\hat{p}^{-1}(X_l), \quad r = 0, 1, \ldots. \tag{9}$$

At the same time, the oracle employs shrinkage coefficients μ_k depending on unknown to the statistician functions $f(y)$ and $p(x)$:

$$\mu_k := \frac{\Theta_k}{\Theta_k + d(p)n^{-1}}, \tag{10}$$

where

$$d(p) := \int_0^1 p^{-1}(x)\,\mathrm{d}x \tag{11}$$

and Θ_k is a Sobolev functional

$$\Theta_k := L_k^{-1} \int_{B_k} |h_0(u)|^2\,\mathrm{d}u. \tag{12}$$

The cutoff K, used in (8), is the minimal integer such that $b_{K+1} > \max(b_3, n^{1/3} \ln \ln(n+3))$. It is chosen from a minimax consideration explained below.

Finally we need to define the design density estimator $\hat{p}(x)$ used in (9). (Let us note that the oracle knows p but uses it only in the shrinkage procedure; this approach simplifies mimicking of the oracle.) Set

$$\hat{p}(x) := \max(1/\ln \ln(n+3), \tilde{p}(x)), \tag{13}$$

where the pivotal design density estimator $\tilde{p}(x)$ is an orthogonal series one:

$$\tilde{p}(x) := 1 + n^{-1} \sum_{r=1}^{n^{1/3}} \sum_{l=1}^{n} \varphi_r(X_l)\varphi_r(x). \tag{14}$$

Note that this estimator is universal for the fixed and random regression designs.

Now, we are in a position to describe an oracle for the bivariate component $\psi(y, x)$. Here, we are following the methodology of Efromovich (2000), developed for the setting of multivariate densities, and set:

$$\tilde{\psi}^*(y, x) := \pi^{-1} \sum_{k,\tau=1}^{T} \mu_{k\tau} \sum_{r=1}^{\infty} \int_0^{\infty} I((u, r) \in B_{k\tau}) \mathrm{Re}\{\hat{h}_r(u) \mathrm{e}^{-iuy}\} \, \mathrm{d}u \varphi_r(x), \tag{15}$$

where $\hat{h}_r$ is defined in (9),

$$\mu_{k\tau} := \frac{\Theta_{k\tau}}{\Theta_{k\tau} + d(p)n^{-1}}, \tag{16}$$

$$\Theta_{k\tau} := L_{k\tau}^{-1} \sum_{r=1}^{\infty} \int_0^{\infty} I((u, r) \in B_{k\tau}) |h_r(u)|^2 \, \mathrm{d}u, \tag{17}$$

and T is the minimal integer such that $b'_{T+1} > \max(b'_3, (n^{1/4} \ln \ln(n+3))$.

Then EP oracle is defined as

$$\tilde{f}^*(y|x) := \tilde{f}^*(y) + \tilde{\psi}^*(y, x). \tag{18}$$

We shall see shortly that it is easy to evaluate the oracle's MISE. But first we need to introduce two assumptions.

Assumption 1 An estimated conditional density $f(y|x)$ belongs to a Sobolev class of differentiable bivariate functions:

$$\mathcal{S}(1,1,Q) := \left\{ f(y|x): \quad f(y|x) = \sum_{r=0}^{\infty} (2\pi)^{-1} \int_{-\infty}^{\infty} h_r(u) \mathrm{e}^{-iuy} \, \mathrm{d}u \varphi_r(x), \right.$$
$$f(y|x) \geq 0, \int_{-\infty}^{\infty} f(y|x) \, \mathrm{d}y \equiv 1, (y,x) \in (-\infty, \infty) \times [0,1],$$
$$\left. \sum_{r=0}^{\infty} \pi^{-1} \int_{0}^{\infty} [u^2 + (\pi r)^2] |h_r(u)|^2 \, \mathrm{d}u \leq Q < \infty \right\}. \tag{19}$$

Assumption 2 The design density $p(x)$, $x \in [0,1]$ is supported and bounded below from zero on $[0,1]$, and its first derivative $p^{(1)}(x)$ exists and is bounded on $[0,1]$.

Theorem 1 *The cases of random and fixed designs of predictors are considered simultaneously. Consider a particular sample size n and the pair $(f(y|x), p(x))$ of conditional and design densities satisfying Assumptions 1 and 2. Then the MISE of the oracle (18) satisfies*

$$E_{(f(y|x),p(x))} \int_0^1 \int_{-\infty}^{\infty} (\tilde{f}^*(y|x) - f(y|x))^2 \, \mathrm{d}y \, \mathrm{d}x$$
$$= \pi^{-1} n^{-1} d(p) \left[\sum_{k=1}^{K} L_k \mu_k + \sum_{k,\tau=1}^{T} L_{k\tau} \mu_{k\tau} \right]$$
$$+ \pi^{-1} \left[\sum_{k>K} L_k \Theta_k + \sum_{k,\tau=1}^{\infty} I((k,\tau) \notin [1,T]^2) L_{k\tau} \Theta_{k\tau} \right] + \delta_n^*, \tag{20}$$

where for any two arrays $\{\nu_k \in (0,1); k = 1,2,\ldots\}$ and $\{\nu_{k\tau} \in (0,1); k,\tau = 1,2,\ldots\}$

$$|\delta_n^*| \leq \pi^{-1} d(p) n^{-1} \sum_{k=1}^{K} L_k \mu_k \left[\nu_k + C \nu_k^{-1} \mu_k \left(L_k^{-1} + n^{-1/4} \right) \right]$$
$$+ \pi^{-1} d(p) n^{-1} \sum_{k,\tau=1}^{T} L_{k\tau} \mu_{k\tau} \left[\nu_{k\tau} + C \nu_{k\tau}^{-1} \mu_{k\tau} (L_{k\tau}^*)^{-1} \right]. \tag{21}$$

Note that in (20) the notation $E_{(f(y|x),p(x))}(\cdot)$ emphasizes the fact that the expectation is taken given the design, the conditional density and the design density.

The following technical result will help us to analyze the oracle's MISE.

Lemma 1 *The adjusted lengths satisfy the inequality*

$$(L_{k\tau}^*)^{-2} \leq C[\ln^{-9/4}(n+3) + \ln^{-3/2}(n+3)(I(\tau=1) + I(k=1))], \tag{22}$$

and the cutoff T satisfies the inequality $T \le C \ln(n+3) \ln\ln(n+3)$.

Theorem 1 is the main tool in establishing minimax properties of the oracle (recall that this is one of the main criteria in choosing a benchmark). Because smoothness of $f(y|x)$ in y and x may be different, it is reasonable to consider anisotropic bivariate conditional density classes; here classes introduced in Efromovich (2007) will be studied.

We begin with a Sobolev anisotropic class of conditional densities

$$\mathcal{S}(m_Y, m_X, Q) := \Bigg\{ f(y|x) : \ f(y|x) = \sum_{r=0}^{\infty} (2\pi)^{-1} \int_{-\infty}^{\infty} h_r(u) e^{-iuy} \, du \varphi_r(x),$$
$$f(y|x) \ge 0, \int_{-\infty}^{\infty} f(y|x) \, dy \equiv 1, (y, x) \in (-\infty, \infty) \times [0, 1],$$
$$\sum_{r=0}^{\infty} \pi^{-1} \int_0^{\infty} [u^{2m_Y} + (\pi r)^{2m_X}] |h_r(u)|^2 \, du \le Q < \infty, \ m_Y \ge 1, m_X \ge 1 \Bigg\}. \quad (23)$$

Another anisotropic class to consider is an analytic–Sobolev one,

$$\mathcal{AS}(\gamma, m_X, Q) := \Bigg\{ f(y|x) : \ f(y|x) = \sum_{r=0}^{\infty} (2\pi)^{-1} \int_{-\infty}^{\infty} h_r(u) e^{-iuy} \, du \varphi_r(x),$$
$$f(y|x) \ge 0, \int_{-\infty}^{\infty} f(y|x) \, dy \equiv 1, \quad (y, x) \in (-\infty, \infty) \times [0, 1],$$
$$\sum_{r=0}^{\infty} \pi^{-1} \int_0^{\infty} [e^{\gamma u} + (\pi r)^{2m_X}] |h_r(u)|^2 \, du \le Q < \infty, \ m_X \ge 1, \gamma > 0 \Bigg\}. \quad (24)$$

Note that this class includes, among others, classical normal, Student and Cauchy conditional densities as well as their mixtures and one-to-one transformations which are typical in the additive regression; see a discussion in Efromovich (1999, 2007).

Finally, we are considering an anisotropic analytic class

$$\mathcal{A}(\gamma_1, \gamma_2, Q) := \Bigg\{ f(y|x) : \ f(y|x) = \sum_{r=0}^{\infty} (2\pi)^{-1} \int_{-\infty}^{\infty} h_r(u) e^{-iuy} \, du \varphi_r(x),$$
$$f(y|x) \ge 0, \int_{-\infty}^{\infty} f(y|x) \, dy \equiv 1, \quad (y, x) \in (-\infty, \infty) \times [0, 1],$$
$$\sum_{r=0}^{\infty} \pi^{-1} \int_0^{\infty} [e^{\gamma_1 u} + e^{\gamma_2 r}] |h_r(u)|^2 \, du \le Q < \infty, \ \gamma_1 > 0, \gamma_2 > 0 \Bigg\}. \quad (25)$$

A direct calculation implies the following technical result.

Lemma 2 *Suppose that $L_k \to \infty$ and $L_{k+1}/L_k \to 1$ as $k \to \infty$. Then for $\mathcal{F}$ being one of the above–specified anisotropic bivariate c.d. classes,*

$$\sup_{f(y|x)\in\mathcal{F}} \pi^{-1}n^{-1}d(p)\left[\sum_{k=1}^{K} L_k\mu_k + \sum_{k,\tau=1}^{T} L_{k\tau}\mu_{k\tau}\right] = R_n(p,\mathcal{F})(1+o(1)), \quad (26)$$

and cutoffs K and T are such that

$$\sup_{f(y|x)\in\mathcal{F}} \left[\sum_{k>K} L_k\Theta_k + \sum_{k,\tau=1}^{\infty} I((k,\tau)\notin[1,T]^2)L_{k\tau}\Theta_{k\tau}\right] = o(1)R_n(p,\mathcal{F}). \quad (27)$$

Here:

(a) *For an anisotropic Sobolev class $\mathcal{F} = \mathcal{S}(m_Y, m_X, Q)$*

$$R_n(p,\mathcal{S}) = [P(\alpha,\beta)Q^{1/(2\tau+1)}][d(p)n^{-1}]^{2\tau/(2\tau+1)}, \quad (28)$$

where $\alpha = m_Y$, $\beta = m_X$, $1/(2\tau) := 1/(2\alpha) + 1/(2\beta)$,

$$P(\alpha,\beta) := \pi^{-4\tau/(2\tau+1)}[J_1(\alpha,\beta)]^{-1/(2\tau+1)}J_2(\alpha,\beta), \quad (29)$$

and

$$J_1(\alpha,\beta) := \int_{\{(u,v):\, u^{2\alpha}+v^{2\beta}\le 1;\, u,v\ge 0\}} \left([u^{2\alpha}+v^{2\beta}]^{1/2} - [u^{2\alpha}+v^{2\beta}]\right) dv\, du, \quad (30)$$

$$J_2(\alpha,\beta) := \int_{\{(u,v):\, u^{2\alpha}+v^{2\beta}\le 1;\; u,v\ge 0\}} (1 - [u^{2\alpha}+v^{2\beta}]^{1/2})\, dv\, du. \quad (31)$$

(b) *For an analytic–Sobolev class $\mathcal{F} = \mathcal{AS}(\gamma, m_X, Q)$*

$$R_n(p,\mathcal{AS}) = P(m_X)Q^{1/(2m_X+1)}(d(p)/n)^{2m_X/(2m_X+1)} \times [2m_X \ln(n)/((2m_X+1)\pi\gamma)]^{2m_X/(2m_X+1)}, \quad (32)$$

where

$$P(m) = (2m+1)^{1/(2m+1)}[m/(\pi(m+1))]^{2m/(2m+1)}. \quad (33)$$

(c) *For an anisotropic analytic class $\mathcal{F} = \mathcal{A}(\gamma_1, \gamma_2, Q)$*

$$R_n(p,\mathcal{A}) = (\pi\gamma_1\gamma_2)^{-1}d(p)n^{-1}\ln^2(n). \quad (34)$$

It is proved in Efromovich (2007) that, for the considered anisotropic bivariate c.d. classes, $R(p,\mathcal{F})$ is an asymptotically sharp lower bound for the minimax MISE where the maximum is taken over all possible c.d. estimators. Further, it is plain to see that if $L_k \to \infty$ as $k \to \infty$ then, according to (21) and Lemma 1, we can always find ν_k and $\nu_{k\tau}$ such that

$$\sup_{f(y|x)\in\mathcal{F}} |\delta_n^*| = o(1)R_n(p, \mathcal{F}). \tag{35}$$

These two remarks imply the following proposition.

Corollary 1 *Let Assumptions 1 and 2 hold, $L_k \to \infty$ and $L_{k+1}/L_k \to 1$ as $k \to \infty$. Then the EP oracle (18) is simultaneously sharp minimax over anisotropic Sobolev, analytic-Sobolev and analytic bivariate function classes defined in (23)–(25), respectively.*

Now let us consider the case where $f(y|x) \equiv f(y)$, the case of the response being independent of the predictor. In this case $\Theta_{k\tau} = \mu_{k\tau} \equiv 0$ for all $k, \tau \geq 1$. Then the right side of (20) simplifies into a familiar expression for the MISE of the univariate EP oracle; see Efromovich (1985, 2005). Further, it is well known that the univariate EP oracle is sharp minimax over a vast set of density classes. We get the following result.

Corollary 2 *Let the assumption of Corollary 1 hold and $f(y|x) \equiv f(y)$, that is, the response is independent of the predictor. Then the bivariate EP oracle (18) is rate minimax over classical univariate Sobolev and analytic function classes introduced in Efromovich (2005). Moreover, if the design is uniform ($p(x) = I(0 \leq x \leq 1)$) then the oracle is simultaneously sharp minimax over these two univariate density classes.*

This corollary implies that the bivariate EP oracle performs the wished dimension reduction whenever the response is independent of the predictor.

We have established all the desired statistical properties of EP oracle. As a result, it can be considered as a benchmark for any c.d. estimator. The next section shows that performance of the oracle can be matched by a data-driven estimator which mimics the oracle.

3 Estimator and its oracle inequality

The oracle's knowledge of conditional and design densities is used "only" for optimal shrinkage of empirical characteristic functions; recall (10) and (16). This fact dramatically simplifies mimicking of the oracle: the idea is to use a plug-in shrinkage procedure. Namely, in place of μ_k we use

$$\tilde{\mu}_k := \frac{\tilde{\Theta}_k}{\tilde{\Theta}_k + \tilde{d}n^{-1}} I(\tilde{\Theta}_k > t_k\tilde{d}n^{-1}), \tag{36}$$

where

$$\tilde{\Theta}_k := L_k^{-1} \int_{B_k} |\hat{h}_0(u)|^2 \, \mathrm{d}u - \tilde{d}n^{-1} \tag{37}$$

and

$$\tilde{d} := \int_0^1 \hat{p}^{-1}(x)\,\mathrm{d}x. \tag{38}$$

In place of $\mu_{k\tau}$ we use

$$\tilde{\mu}_{k\tau} := \frac{\tilde{\Theta}_{k\tau}}{\tilde{\Theta}_{k\tau} + \tilde{d}n^{-1}} I(\tilde{\Theta}_{k\tau} > t_{k\tau}\tilde{d}n^{-1}), \tag{39}$$

where

$$\tilde{\Theta}_{k\tau} := L_{k\tau}^{-1} \sum_{r=1}^{\infty} \int_0^{\infty} I((u,r) \in B_{k\tau})|\hat{h}_r(u)|^2\,\mathrm{d}u - \tilde{d}n^{-1}. \tag{40}$$

Then

$$\tilde{f}(y) := \pi^{-1} \sum_{k=1}^{K} \tilde{\mu}_k \int_{B_k} \mathrm{Re}\{\hat{h}_0(u)\mathrm{e}^{-iuy}\}\,\mathrm{d}u \tag{41}$$

estimates $f(y)$ while

$$\tilde{\psi}(y,x) := \pi^{-1} \sum_{k,\tau=1}^{T} \tilde{\mu}_{k\tau} \sum_{r=1}^{\infty} \int_0^{\infty} I((u,r) \in B_{k\tau}) \mathrm{Re}\left\{\hat{h}_r(u)\mathrm{e}^{-iuy}\right\} \mathrm{d}u\varphi_r(x) \tag{42}$$

estimates $\psi(y,x)$.

Further, the suggested data–driven c.d. estimator is

$$\tilde{f}(y|x) := \tilde{f}(y) + \tilde{\psi}(y,x). \tag{43}$$

Note that the estimator does not depend on the type of an underlying design. Further, all integrals and sums are taken over finite sets and easily calculated, for instance the estimate (42) can be written as

$$\tilde{\psi}(y,x) = n^{-1}\pi^{-1} \sum_{k,\tau=1}^{T} \tilde{\mu}_{k\tau} \sum_{l=1}^{n} \left[\sum_{r=b'_\tau+1}^{b'_{\tau+1}} \varphi_r(X_l)\varphi_r(x) \right] \times \frac{\sin((Y_l - y)b'_{k+1}) - \sin((Y_l - y)b'_k)}{\hat{p}^{-1}(X_l)(Y_l - y)}.$$

Now, we are in a position to show that the MISE of the estimator matches the MISE of the oracle that knows the conditional and design densities.

Theorem 2 (Oracle inequality) *The cases of random and fixed designs of predictors are considered simultaneously. Consider a particular sample size n as well as a particular pair* $(f(y|x), p(x))$ *of conditional and design densities satisfying Assumptions* 1 *and* 2. *Then the following oracle inequality holds:*

$$E_{(f(y|x),p(x))} \int_0^1 \int_{-\infty}^{\infty} (\tilde{f}(y|x) - f(y|x))^2 \, dy \, dx$$
$$\leq E_{(f(y|x),p(x))} \int_0^1 \int_{-\infty}^{\infty} (\tilde{f}^*(y|x) - f(y|x))^2 \, dy \, dx + \delta_n, \tag{44}$$

where

$$\delta_n \leq Cn^{-1} \left[\sum_{k=1}^{K} \left[L_k \mu_k \left(t_k^{1/2} + L_k^{-1/2} t_k^{-3/2} \right) + L_k^{-2} t_k^{-5} \right] \right.$$
$$\left. + \sum_{k,\tau=1}^{T} \left[L_{k\tau} \mu_{k\tau} \left(t_{k\tau}^{1/2} + (L_{k\tau}^*)^{-1/2} t_{k\tau}^{-3/2} \right) + (L_{k\tau}^*)^{-2} t_{k\tau}^{-5} \right] \right]. \tag{45}$$

Theorem 2 shows how well the estimator matches its benchmark, and δ_n indicates the difference. Let us stress that the oracle inequality is not asymptotic and it is valid for any c.d. and any design satisfying Assumptions 1 and 2. At the same time, the constant C depends on the design density; to relax this dependence a class of design densities uniformly satisfying Assumption 2 may be considered.

This oracle inequality is a remarkable technical tool for the analysis of the suggested data-driven estimator. Suppose that

$$L_{k+1}/L_k \to 1 \quad \text{and} \quad t_k \to 0 \quad \text{as} \quad k \to \infty, \qquad \sum_{k=1}^{\infty} L_k^{-2} t_k^{-5} < \infty. \tag{46}$$

Then the oracle inequality, together with Theorem 1, Lemma 1 and Corollary 1, immediately establishes sharp minimaxity of the estimator over the bivariate c.d. classes (23)–(25).

Now let us show that the estimator does perform the wished dimension reduction. Indeed, if $f(y|x) \equiv f(y)$, $x \in [0, 1]$, then

$$E_{(f(y|x),p(x))} \int_0^1 \int_{-\infty}^{\infty} (\tilde{f}^*(y|x) - f(y|x))^2 \, dy \, dx$$
$$= \pi^{-1} n^{-1} d(p) \left[\sum_{k=1}^{K} L_k \mu_k + \sum_{k>K} L_k \Theta_k \right] + \delta_n^*, \tag{47}$$

where for any array $\{\nu_k \in (0, 1); k = 1, 2, \dots\}$

$$|\delta_n^*| \le d(p)n^{-1} \sum_{k=1}^{K} L_k \mu_k \left[\nu_k + C\nu_k^{-1}\mu_k \left(L_k^{-1} + n^{-1/4}\right)\right]. \tag{48}$$

Further, if additionally $\sum_{k=1}^{K} L_k \mu_k \to \infty$ as $n \to \infty$ meaning that the density $f(y)$ is nonparametric in the sense of Efromovich (1985), then (45) yields

$$\delta_n \le o(1) E_{(f(y|x),p(x))} \int_0^1 \int_{-\infty}^{\infty} \left(\tilde{f}^*(y|x) - f(y|x)\right)^2 \mathrm{d}y\, \mathrm{d}x + Cn^{-1}. \tag{49}$$

These results match known ones for the univariate density case, discussed in Efromovich (1985, 1999, 2005), up to the factor $d(p) = \int_0^1 p^{-1}(x)\,\mathrm{d}x$ in the MISE convergence. This allows us to conclude that the bivariate c.d. estimator asymptotically attains the sharp minimax convergence within the factor-penalty $d(p)$. At the same time, if the design is uniform then $d(p) = 1$ and the estimator is sharp minimax not only over the bivariate c.d. classes but also over univariate density classes whenever the classical null hypothesis "the response is independent of the predictor" holds.

Let us combine all these results into a proposition.

Corollary 3 *Let Assumption 2 and (46) hold. Then:*

(a) *The data-driven estimator (43) is simultaneously sharp minimax over anisotropic bivariate Sobolev, analytic-Sobolev and analytic c.d. classes defined in (23)–(25), respectively.*
(b) *Suppose that $f(y|x) \equiv f(y)$ (the response is independent of the predictor). Then the bivariate data-driven c.d. estimator (43) is rate minimax over univariate Sobolev and analytic density classes introduced in Efromovich (2005). Moreover, the MISE of the estimator is within the factor $d(p)$ of the sharp minimax MISE. In particular, if the design density is uniform, that is $p(x) = I(0 \le x \le 1)$, then $d(p) = 1$ and the bivariate data-driven estimator (43) is sharp minimax over those univariate density classes.*

Using these results we can conclude that the suggested estimator matches performance of the oracle in all wished categories: its MISE is close to the oracle's MISE, it is sharp minimax over anisotropic bivariate c.d. classes, and its MISE is within factor $d(p)$ of the sharp univariate minimax MISE whenever the response is independent of the predictor (in other words, the estimator performs the wished dimension reduction).

At the same time, it is fair to say that on first glance the estimator (43) looks formidable. Thus, it is absolutely natural to raise the following question. Can the suggested asymptotic estimator, as well as its underlying Fourier-approximation idea, be feasible for small sample sizes? Let us check this using an actuarial example which explores a relationship between credit score and premium for basic auto-insurance paid by 54 undergraduate college students taking an introductory statistical course. The nice feature of this example is that a visualization of the data, as well as its well-understood nature, will allow the reader to be the judge of the estimator (43). Of

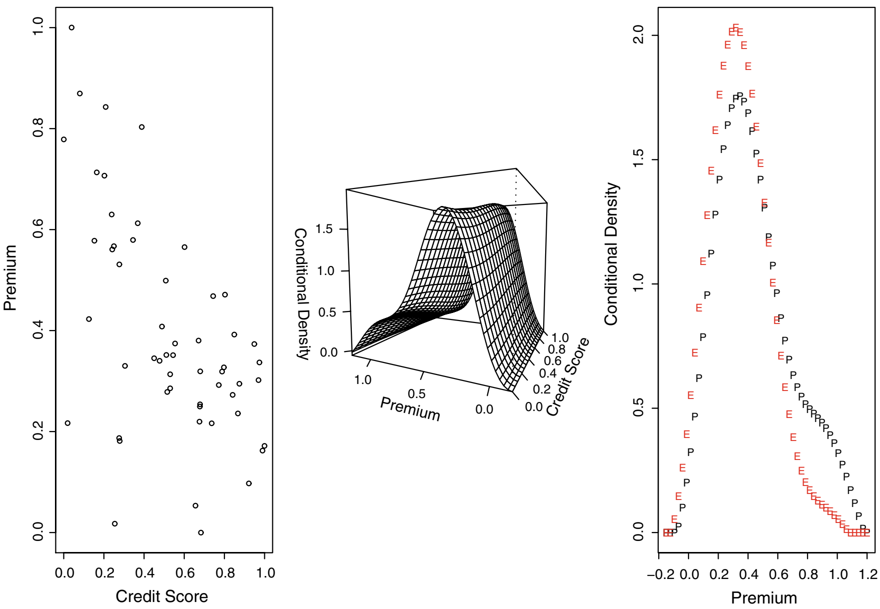

Fig. 1 Study of relationship between credit score and premium paid for basic auto-insurance. The left diagram shows the scattergram for 54 observations; data is linearly rescaled onto $[0, 1]^2$. The middle diagram shows the bivariate conditional density estimate (43). The right diagram shows two univariate "slices" of the estimate for the lowest and largest credit scores shown by letters "P" and "E", respectively

course, the sample size $n = 54$ is considered small even for univariate nonparametric estimates according to Efromovich (1999). Thus the reader should take into account two issues: the challenging nature of a small data-set for nonparametric estimation and the fact that (43) is purely asymptotic in its nature with the parameters chosen to simplify proofs rather than to be good for small sample sizes.

The data, collected by the author for his work on a project sponsored by the Grant TAF/CAS-07, is linearly rescaled onto $[0, 1]^2$ and shown in the left diagram in Fig. 1. The middle diagram shows the bivariate estimate (43) with $b_k := b'_k$ and $t_k := t_{k,1}$. Let us note that in this example $K = T = 3$ and $b_1 = 0, b_2 = 3, b_3 = 6$, and $b_4 = 11$. Finally, the right diagram shows the univariate conditional densities ("slices" of the surface) for the minimal credit score (which is "poor" according to the industry standard and thus denoted by letter "P") and the largest credit score (which is "excellent" according to the industry standard and thus denoted by letter "E"). In the univariate "slices" the projection of Efromovich (1999) on the class of densities was used. The bivariate conditional density is probably oversmoothed by the large blocks. Nonetheless the estimate does show the main underlying characteristics of the data. First of all, we clearly see that the excellent credit score implies the smaller mean payment (which should be the case because some insurers use credit score to calculate premiums). Second, the estimate indicates that the poor credit score implies a larger variability in the paid premiums. This is explained by the fact that an insurer may or may not take credit score into account, and on the top of this the penalty for a poor credit scoring varies dramatically from one insurer to another. Finally, the estimate correctly shows that premiums can be smaller and larger than the ones paid by the 54 surveyed students. This is an extra and "free" benefit of the Fourier approach which does not require the statistician to choose an underlying support of the estimator (43).

Let us stress that the main purpose of this example is to explain that (43) is a feasible data-driven estimator. A special paper (or a book similar to Efromovich 1999) will be devoted to creating a good modification of (43) for the practically important case of small sample sizes.

4 Proofs

Proof of Theorem 1 For the both considered designs we can write

$$\int_0^1 \int_{-\infty}^{\infty} (\tilde{f}^*(y|x) - f(y|x))^2 \, dy \, dx = \int_{-\infty}^{\infty} (\tilde{f}^*(y) - f(y))^2 \, dy + \int_0^1 \int_{-\infty}^{\infty} (\tilde{\psi}^*(y,x) - \psi(y,x))^2 \, dy \, dx. \quad (50)$$

Using the Plancherel identity we get for the first integral in the right side of (50),

$$\begin{aligned}\int_{-\infty}^{\infty} (\tilde{f}^*(y) - f(y))^2 \, dy &= \pi^{-1} \int_0^{\infty} |\tilde{h}_0(u) - h_0(u)|^2 \, du \\ &= \pi^{-1} \sum_{k=1}^{K} \int_{B_k} |\mu_k \hat{h}_0(u) - h_0(u)|^2 \, du \\ &\quad + \pi^{-1} \int_{b_{K+1}}^{\infty} |h_0(u)|^2 \, du. \end{aligned} \quad (51)$$

Similarly, the second integral in the right side of (50) can be written as

$$\begin{aligned}&\int_0^1 \left[\int_{-\infty}^{\infty} (\tilde{\psi}^*(y,x) - \psi(y,x))^2 \, dy \right] dx \\ &= \pi^{-1} \sum_{k,\tau=1}^{T} \sum_{r=1}^{\infty} \int_0^{\infty} I((u,r) \in B_{k\tau}) |\mu_{k\tau} \hat{h}_r(u) - h_r(u)|^2 \, du \\ &\quad + \pi^{-1} \sum_{k,\tau=1}^{\infty} I((k,\tau) \notin \{1,\dots,T\}^2) \sum_{r=1}^{\infty} \int_0^{\infty} I((u,r) \in B_{k\tau}) |h_r(u)|^2 \, du. \end{aligned} \quad (52)$$

To evaluate integrals involving the statistic $\hat{h}_r(u)$ we need a technical result describing its properties. It is convenient to rewrite $\hat{h}_r(u)$ as a sum of two parts where the former one does not involve the estimate $\hat{p}(x)$ of the design density:

$$\begin{aligned}\hat{h}_r(u) &= n^{-1}\sum_{l=1}^{n} e^{iuY_l}\varphi_r(X_l)\hat{p}^{-1}(X_l)\\ &= n^{-1}\sum_{l=1}^{n} e^{iuY_l}\varphi_r(X_l)p^{-1}(X_l) + n^{-1}\sum_{l=1}^{n} e^{iuY_l}\varphi_r(X_l)(\hat{p}^{-1}(X_l) - p^{-1}(X_l))\\ &=: \check{h}_r(u) + \delta_{rn}(u). \qquad (53)\end{aligned}$$

From now on we need to consider random and fixed designs in turn. Let us begin with the random design.

Lemma 3 *Let Assumptions 1 and 2 hold, and consider a random design regression. Then $\check{h}_r(u)$ and $\delta_{rn}(u)$, defined in (53), satisfy*

$$E\{\check{h}_r(u)\} = h_r(u), \qquad (54)$$

$$E|\check{h}_r(u) - h_r(u)|^2 = n^{-1}[d - |h_r(u)|^2 + 2^{-1/2}\pi'_{2r}I(r > 0)], \qquad (55)$$

where $d := \int_0^1 p^{-1}(x)\,dx$ and $\pi'_r := \int_0^1 \varphi_r(x)p^{-1}(x)\,dx$, and for $r < (1-\nu_1)n^{1/3}$, $\nu_1 \in (0,1)$

$$E|\delta_{rn}(u)|^2 \le Cn^{-4/3}\left[\ln^6(n) + \int_0^1 |\partial h(u|x)/\partial x|^2\,dx\right] + Cn^{-1}\int_0^1 |h(u|x)|^2\,dx. \qquad (56)$$

Proof of Lemma 3 Write

$$\begin{aligned}E\{\check{h}_r(u)\} = E\{e^{iuY}\varphi_r(X)p^{-1}(X)\} &= \int_0^1 \left[\varphi_r(x)\int_{-\infty}^{\infty} f(y|x)e^{iuy}\,dy\right]dx\\ &= \int_0^1 h(u|x)\varphi_r(x)\,dx = h_r(u).\end{aligned}$$

This verifies (54). Using an elementary trigonometric formula $\varphi_r^2(x) = 1 + 2^{-1/2}\varphi_{2r}(x)I(r > 0)$, we get

$$\begin{aligned}E|\check{h}_r(u) - h_r(u)|^2 &= E\left\{|n^{-1}\sum_{l=1}^{n} e^{iuY_l}\varphi_r(X_l)p^{-1}(X_l) - h_r(u)|^2\right\}\\ &= n^{-1}E|e^{iuY_l}\varphi_r(X_l)p^{-1}(X_l) - h_r(u)|^2\\ &= n^{-1}\left[\int_0^1 \varphi_r^2(x)p^{-1}(x)\,dx - |h_r(u)|^2\right]\\ &= n^{-1}\left[d - |h_r(u)|^2 + 2^{-1/2}\pi'_{2r}I(r > 0)\right].\end{aligned}$$

This verifies (55). To check (56) we write,

$$\begin{aligned}\delta_{rn}(u) &= n^{-1}\sum_{l=1}^{n} e^{iuY_l}\varphi_r(X_l)(p(X_l)-\hat{p}(X_l))\hat{p}^{-1}(X_l)p^{-1}(X_l)\\ &= n^{-1}\sum_{l=1}^{n} e^{iuY_l}\varphi_r(X_l)(p(X_l)-\tilde{p}(X_l))p^{-2}(X_l)\\ &\quad + n^{-1}\sum_{l=1}^{n} e^{iuY_l}\varphi_r(X_l)(\tilde{p}(X_l)-\hat{p}(X_l))p^{-2}(X_l)\\ &\quad + n^{-1}\sum_{l=1}^{n} e^{iuY_l}\varphi_r(X_l)(\hat{p}(X_l)-p(X_l))^2\hat{p}^{-1}(X_l)p^{-2}(X_l)\\ &=: A_1(u)+A_2(u)+A_3(u). \end{aligned} \tag{57}$$

Plainly

$$\begin{aligned}E|A_1(u)|^2 &= n^{-2}E\left\{\sum_{l,m=1}^{n} e^{iu(Y_l-Y_m)}\varphi_r(X_l)\varphi_r(X_m)(\tilde{p}(X_l)\right.\\ &\quad \left. -\, p(X_l))(\tilde{p}(X_m)-p(X_m))p^{-2}(X_l)p^{-2}(X_m)\right\}\\ &= n^{-1}E\left\{(\tilde{p}(X_n)-p(X_n))^2p^{-4}(X_n)\varphi_r^2(X_n)\right\}\\ &\quad + n^{-2}n(n-1)E\left\{h(u|X_1)\bar{h}(u|X_2)\varphi_r(X_1)\varphi_r(X_2)\right.\\ &\quad \left.\times\ (\tilde{p}(X_1)-p(X_1))(\tilde{p}(X_2)-p(X_2))p^{-2}(X_1)p^{-2}(X_2)\right\}\\ &=: A_{11}(u)+A_{12}(u).\end{aligned}$$

To evaluate $A_{11}(u)$ we first recall a familiar inequality which holds for any natural k (see, for instance, Efromovich 2005)

$$\begin{aligned}\max_{x\in[0,1]}\ \max\left(E|p(x)-\tilde{p}(x)|^{2k}, E|p(x)-\hat{p}(x)|^{2k}\right)\\ \le C_k \ln^{2k+1}(n)n^{-2k/3},\ C_k<\infty.\end{aligned} \tag{58}$$

Second, let $\tilde{p}_n(x)$ be another notation, apart of $\tilde{p}(x)$, for the density estimator (14) which stresses the fact that the estimator is based on n observations $(X_1,\dots,X_n)$. Then we can write

$$\begin{aligned}\tilde{p}_n(x) &= n^{-1}\sum_{j=1}^{n^{1/3}}\varphi_j(X_n)\varphi_j(x) + \left[1+n^{-1}\sum_{j=1}^{n^{1/3}}\sum_{l=1}^{n-1}\varphi_j(X_l)\varphi_j(x)\right]\\ &= n^{-1}\sum_{j=1}^{n^{1/3}}[\varphi_j(X_n)-(n-1)^{-1}\sum_{l=1}^{n-1}\varphi_j(X_l)]\varphi_j(x)+\tilde{p}_{n-1}(x).\end{aligned}$$

This immediately yields that

$$|\tilde{p}_n(x) - \tilde{p}_{n-1}(x)| < 4n^{-2/3}.$$

Combining these two results we conclude that $A_{11}(u) \leq Cn^{-1}n^{-1/2}$. To evaluate $A_{12}(u)$ we need some preliminary calculations. Write

$$\tilde{p}(x) = 1 + n^{-1}\sum_{j=1}^{n^{1/3}}[\varphi_j(X_1) + \varphi_j(X_2)]\varphi_j(x) + \sum_{j=1}^{n^{1/3}} \check{\pi}_j \varphi_j(x)$$

where $\check{\pi}_j := n^{-1}\sum_{l=3}^{n} \varphi_j(X_l)$. Note that the $\check{\pi}_j$'s do not depend on (X_1, X_2), and we can write,

$$\begin{aligned} A_{12}(u) = n^{-1}(n-1)E\Bigg\{&h(u|X_1)\bar{h}(u|X_2)p^{-2}(X_1)p^{-2}(X_2)\varphi_r(X_1)\varphi_r(X_2)\\ &\times \prod_{l=1}^{2}\Bigg[n^{-1}\sum_{j=1}^{n^{1/3}}(\varphi_j(X_1)+\varphi_j(X_2))\varphi_j(X_l)\\ &\quad + \sum_{j=1}^{n^{1/3}}(\check{\pi}_j - \pi_j)\varphi_j(X_l) - \sum_{j>n^{1/3}} \pi_j\varphi_j(X_l)\Bigg]\Bigg\}. \end{aligned}$$

Plainly $|n^{-1}\sum_{j=1}^{n^{1/3}}(\varphi_j(X_1) + \varphi_j(X_2))\varphi_j(X_l)| \leq 4n^{-2/3}$, and also $\varphi_r(x)\varphi_j(x) = 2^{-1/2}[\varphi_{r-j}(x) + \varphi_{r+j}(x)]$ for $j, r \geq 1$. Denote $\kappa_r(u) := \int_0^1 h(u|x)p^{-1}(x)\varphi_r(x)\,\mathrm{d}x$, and let us make several preliminary remarks. First, $E\{\check{\pi}_j\} = (n-2)n^{-1}E\{\varphi_j(X)\} = \pi_j - 2n^{-1}\pi_j$. Second, for any natural j and t

$$\begin{aligned} E(\check{\pi}_j - \pi_j)(\check{\pi}_t - \pi_t) &= E(\check{\pi}_j - (n-2)n^{-1}\pi_j - 2n^{-1}\pi_j)(\check{\pi}_t - (n-2)n^{-1}\pi_t - 2n^{-1}\pi_t)\\ &= E(\check{\pi}_j - (n-2)n^{-1}\pi_j)(\check{\pi}_t - (n-2)n^{-1}\pi_t) + 4n^{-2}\pi_j\pi_t. \end{aligned}$$

Third,

$$\begin{aligned} &E(\check{\pi}_j - (n-2)n^{-1}\pi_j)(\check{\pi}_t - (n-2)n^{-1}\pi_t)\\ &= n^{-2}E\sum_{l,m=3}^{n}(\varphi_j(X_l) - \pi_j)(\varphi_t(X_m) - \pi_t)\\ &= n^{-2}(n-2)E(\varphi_j(X) - \pi_j)(\varphi_t(X) - \pi_t)\\ &= n^{-2}(n-2)[2^{-1/2}(\pi_{j-t} + \pi_{j+t}) - \pi_j\pi_t]. \end{aligned}$$

Fourth, we note that $\int_0^1 h(u|x)p^{-1}(x)\varphi_j(x)\varphi_r(x)\,\mathrm{d}x = 2^{-1/2}[\kappa_{j-r}(u) + \kappa_{j+r}(u)]$. Our fifth remark is that for the considered $r < (1-\nu_1)n^{1/3}$ and under Assumption 2 we can write

$$\left[\sum_{j>n^{1/3}} \pi_j(\kappa_{j-r}(u)+\kappa_{j+r}(u))\right]^2 \le 2\sum_{j>n^{1/3}} \pi_j^2 \sum_{j>n^{1/3}} \left[\kappa_{j-r}^2(u)+\kappa_{j+r}^2(u)\right]$$
$$\le Cn^{-4/3}\int_0^1 \left[p^{(1)}(x)\right]^2 \mathrm{d}x$$
$$\times \int_0^1 |\partial(h(u|x)p^{-1}(x))/\partial x|^2 \,\mathrm{d}x$$
$$\le Cn^{-4/3}\int_0^1 (|\partial h(u|x)/\partial x|^2 + |h(u|x)|^2)\,\mathrm{d}x.$$

Also, using the first, second and third remarks we get

$$\left| E\sum_{j,t=1}^{n^{1/3}} (\check{\pi}_j-\pi_j)(\check{\pi}_t-\pi_t)2^{-1}(\kappa_{j-r}(u)+\kappa_{j+r}(u))(\kappa_{t-r}(u)+\kappa_{t+r}(u))\right|$$
$$\le C\left|\sum_{j,t=1}^{n^{1/3}} \left\{n^{-1}\left[2^{-1/2}(\pi_{j-t}+\pi_{j+t})-\pi_j\pi_t\right]+4n^{-2}\pi_j\pi_t\right\}(\kappa_{j-r}(u)\right.$$
$$\left. +\ \kappa_{j+r}(u))(\kappa_{t-r}(u)+\kappa_{t+r}(u))\right|$$
$$\le Cn^{-1}\sum_{j=0}^{\infty}\kappa_j^2(u) \le Cn^{-1}\int_0^1 |h(u|x)|^2\,\mathrm{d}x.$$

In the last two inequalities we used the Cauchy inequality, the Parseval identity and $\sum_{j=0}^{\infty}|\pi_j| \le C$. We conclude that

$$A_{12}(u) \le Cn^{-4/3}\left[1+\int_0^1 |\partial h(u|x)/\partial x|^2\,\mathrm{d}x\right] + Cn^{-1}\int_0^1 |h(u|x)|^2\,\mathrm{d}x.$$

Combining the obtained results verifies (56) with δ_{rn} being replaced by $A_1(u)$ [that is, by its first component on the right side of (57)]. Evaluation of $A_2(u)$ is based on a direct application of (58) and a remark that $0 \le \hat{p}(x)-\tilde{p}(x) \le 1/\ln\ln(n+3)$ and that for all sufficiently large n the event $\hat{p}(x) \ne \tilde{p}(x)$ implies $|\tilde{p}(x)-p(x)| > c$ for some positive constant c. This together with the Chebyshev inequality yields (4.7) with δ_{rn} being replaced by $A_2(u)$. Finally, using (58) together with $\hat{p}(x) \ge 1/\ln\ln(n+3)$ and $p(x) > C > 0$ implies $E|A_3(u)|^2 \le C\ln^6(n)n^{-4/3}$. Lemma 3 is proved.

Now we have a tool to evaluate integrals in (51) and (52). To make the evaluation shorter, from now on we may skip subscripts whenever no confusion occurs. Consider an interval $D \subset [0,\infty)$ and a nonnegative integer r. Using Lemma 3 we can write for $\check{h}_r(u)$ used in place of $\hat{h}_r(u)$,

$$E\int_D |\mu\check{h}_r(u) - h_r(u)|^2\,\mathrm{d}u = E\int_D \mu^2|\check{h}_r(u) - h_r(u)|^2\,\mathrm{d}u + (\mu-1)^2\int_D |h_r(u)|^2\,\mathrm{d}u$$
$$= \mu^2 n^{-1}\int_D \left[d - |h_r(u)|^2 + 2^{1/2}\pi'_{2r}I(r>0)\right]\mathrm{d}u$$
$$+(\mu-1)^2\int_D |h_r(u)|^2\,\mathrm{d}u.$$

This allows us to write,

$$E\sum_{r=0}^{\infty}\int_0^{\infty} I((u,r)\in B)|\mu\check{h}_r(u) - h_r(u)|^2\,\mathrm{d}u$$
$$= \left[\mu^2 n^{-1}Ld + \frac{(dn^{-1})^2 L\Theta}{(\Theta + dn^{-1})^2}\right]$$
$$+\mu^2 n^{-1}\sum_{r=0}^{\infty}\int_0^{\infty} I((u,r)\in B)\left[-|h_r(u)|^2 + 2^{1/2}\pi'_{2r}I(r>0)\right]\mathrm{d}u$$
$$= L\mu n^{-1}d + \mu n^{-1}\left[-\mu\sum_{r=0}^{\infty}\int_0^{\infty} I((u,r)\in B)|h_r(u)|^2\,\mathrm{d}u\right.$$
$$\left.+\mu\sum_{r=0}^{\infty}\int_0^{\infty} I((u,r)\in B)2^{1/2}\pi'_{2r}I(r>0)\,\mathrm{d}u\right].$$

Also, Assumption 1 together with (56) yields

$$\mu^2\sum_{r=0}^{\infty}\int_0^{\infty} I((u,r)\in B)E|\delta_{rn}(u)|^2\,\mathrm{d}u$$
$$\leq C\mu Ln^{-1}\left[\mu n^{-1/3}\ln^6(n) + \mu\sum_{r=0}^{\infty}\int_0^{\infty} I((u,r)\in B)\int_0^1 |h(u|x)|^2\,\mathrm{d}x\,\mathrm{d}uL^{-1}\right].$$

Combining obtained results we conclude that

$$E\left\{\sum_{r=0}^{\infty}\int_0^{\infty} I((u,r)\in B)|\mu\hat{h}_r(u) - h_r(u)|^2\,\mathrm{d}u\right\} = L\mu\mathrm{d}n^{-1} + \delta_n^*(B), \quad (59)$$

and

$$|\delta_n^*(B)| \leq L\mu\mathrm{d}n^{-1}[\nu + C\nu^{-1}\mu(n^{-1/3}\ln^6(n) + 1/L^*)]. \quad (60)$$

Here L^* denotes either $L^*_{k\tau}$ or L_k for the double-index or single-index blocks, respectively.

Then (51), (52), (59) and (60), together with a simple calculation, prove Theorem 1 for the case of random design. Now let us consider the case of fixed design. The main difference here is in the analysis of $\hat{h}_r$. The next lemma, which is an analog of Lemma 3, presents necessary technical results for the fixed–design case.

Lemma 4 *Let Assumptions 1 and 2 hold and consider a fixed design regression. Then* $\check{h}_r(u)$ *and* $\delta_{rn}(u)$, $r = 0, 1, \ldots, \lfloor n^{1/4} \ln\ln(n) \rfloor$ *defined in (53), satisfy*

$$E\{\check{h}_r(u)\} = h_r(u) + \gamma_r(u) \tag{61}$$

where

$$|\gamma_r(u)| \le Cn^{-1}\left[1 + r\int_0^1 |h(u|x)|\,\mathrm{d}x + \int_0^1 |\partial h(u|x)/\partial x|\,\mathrm{d}x\right], \tag{62}$$

$$E|\check{h}_r(u) - h_r(u)|^2 = n^{-1}[d - |h_r(u)|^2 + 2^{1/2}\pi'_{2r} I(r > 0)] + n^{-2}\gamma_{1r}(u), \tag{63}$$

where $d := \int_0^1 p^{-1}(x)\,\mathrm{d}x$, $\pi'_r := \int_0^1 \varphi_r(x) p^{-1}(x)\,\mathrm{d}x$,

$$|\gamma_{1r}(u)| \le C\left(1 + r + \int_0^1 |\partial h(u|x)/\partial x|\,\mathrm{d}x\right), \tag{64}$$

and

$$E|\delta_{rn}(u)|^2 \le Cn^{-1}\left[\int_0^1 |h(u|x)|^2\,\mathrm{d}x + n^{-1/3}\left(1 + \int_0^1 |\partial h(u|x)/\partial x|^2\,\mathrm{d}x\right)\right] \tag{65}$$

Proof of Lemma 4 Without any loss of generality it is possible to assume that $X_1 < X_2 < \cdots < X_n < X_{n+1} := 1$. Recall that $\int_{X_l}^{X_{l+1}} p(x)\,\mathrm{d}x = (n+1)^{-1}$ and write,

$$\begin{aligned}
E\{\check{h}_r(u)\} &= n^{-1}\sum_{l=1}^n h(u|X_l)\varphi_r(X_l)p^{-1}(X_l)\\
&= n^{-1}(n+1)\sum_{l=1}^n h(u|X_l)\varphi_r(X_l)p^{-1}(X_l)\int_{X_l}^{X_{l+1}} p(x)\,\mathrm{d}x\\
&= n^{-1}(n+1)\sum_{l=1}^n \int_{X_l}^{X_{l+1}} h(u|x)\varphi_r(x)\,\mathrm{d}x + n^{-1}(n+1)\\
&\quad\times \sum_{l=1}^n \int_{X_l}^{X_{l+1}} \left[h(u|X_l)\varphi_r(X_l)p^{-1}(X_l) - h(u|x)\varphi_r(x)p^{-1}(x)\right] p(x)\,\mathrm{d}x\\
&=: \int_0^1 h(u|x)\varphi_r(x)\,\mathrm{d}x + \gamma_r(u) = h_r(u) + \gamma_r(u).
\end{aligned}$$

Using $|h(u|x)\varphi_r(x)p^{-1}(x)| \le C$ together with the fact that for a differentiable $\psi(x)$

$$\begin{aligned}\left|\int_{X_l}^{X_{l+1}}[\psi(X_l)-\psi(x)]p(x)\,\mathrm{d}x\right| &= \left|\int_{X_l}^{X_{l+1}}\left[\int_{X_l}^{x}\psi^{(1)}(z)\,\mathrm{d}z\right]p(x)\,\mathrm{d}x\right| \\ &= \left|\int_{X_l}^{X_{l+1}}\psi^{(1)}(z)\left[\int_{z}^{X_{l+1}}p(x)\,\mathrm{d}x\right]\mathrm{d}z\right| \\ &\le n^{-1}\int_{X_l}^{X_{l+1}}|\psi^{(1)}(x)|\,\mathrm{d}x, \qquad (66)\end{aligned}$$

we verify (61) and (62). Now let us recall that $\bar{h}$ denotes the conjugate of h, and $\varphi_r^2(x) = 1 + 2^{1/2}\varphi_{2r}(x)I(r>0)$. Write,

$$\begin{aligned}E|\check{h}_r(u)-h_r(u)|^2 &= E\left\{\left|n^{-1}\sum_{l=1}^{n}[\mathrm{e}^{iuY_l}\varphi_r(X_l)p^{-1}(X_l)-h_r(u)]\right|^2\right\} \\ &= n^{-2}\sum_{l=1}^{n}E|\mathrm{e}^{iuY_l}\varphi_r(X_l)p^{-1}(X_l)-h_r(u)|^2 \\ &\quad + n^{-2}\sum_{l,m=1}^{n}I(l\ne m)[h(u|X_l)\varphi_r(X_l)p^{-1}(X_l) \\ &\quad -h_r(u)][\bar{h}(u|X_m)\varphi_r(X_m)p^{-1}(X_m)-\bar{h}_r(u)] \\ &= n^{-2}\sum_{l=1}^{n}[\varphi_r^2(X_l)p^{-2}(X_l)-2\mathrm{Re}\{h(u|X_l) \\ &\quad \times\varphi_r(X_l)\bar{h}_r(u)p^{-1}(X_l)\}+|h_r(u)|^2] \\ &= n^{-2}\sum_{l=1}^{n}|h(u|X_l)\varphi_r(X_l)p^{-1}(X_l)-h_r(u)|^2 \\ &\quad +n^{-2}\left|\sum_{l=1}^{n}\left[h(u|X_l)\varphi_r(X_l)p^{-1}(X_l)-h_r(u)\right]\right|^2 \\ &= n^{-2}(n+1)\sum_{l=1}^{n}[\varphi_r^2(X_l)p^{-2}(X_l)-2\mathrm{Re}\{h(u|X_l)\bar{h}_r(u)\} \\ &\quad \times\varphi_r(X_l)p^{-1}(X_l)]\int_{X_l}^{X_{l+1}}p(x)\,\mathrm{d}x+n^{-1}|h_r(u)|^2-n^{-2}(n+1) \\ &\quad \times\sum_{l=1}^{n}|h(u|X_l)\varphi_r(X_l)p^{-1}(X_l)-h_r(u)|^2\int_{X_l}^{X_{l+1}}p(x)\mathrm{d}x+n^{-2}(n+1) \\ &\quad \times\left|\sum_{l=1}^{n}[h(u|X_l)\varphi_r(X_l)p^{-1}(X_l)-h_r(u)]\int_{X_l}^{X_{l+1}}p(x)\,\mathrm{d}x\right|^2\end{aligned}$$

$$
\begin{aligned}
&=: n^{-1}\left[\int_0^1 (1+2^{1/2}\varphi_{2r}(x)I(r>0))p^{-1}(x)\,\mathrm{d}x\right.\\
&\quad\left.-\ 2\int_0^1 \mathrm{Re}\{h(u|x)\bar{h}_r(u)\}\varphi_r(x)\,\mathrm{d}x\right]+n^{-1}|h_r(u)|^2\\
&\quad-n^{-1}[\int_0^1[|h(u|x)|^2\varphi_r^2(x)p^{-1}(x)\,\mathrm{d}x-|h_r(u)|^2]\,\mathrm{d}x+n^{-2}\gamma_{1r}(u)\\
&= n^{-1}[d-|h_r(u)|^2+2^{1/2}\pi'_{2r}I(r>0)]+n^{-2}\gamma_{1r}(u),
\end{aligned}
$$

where $\gamma_{1r}(u)$ satisfies (64). Now let us check (65). Using (57), $\delta_{rn}(u)$ can be evaluated via estimation of 3 terms $A_1(u)$, $A_2(u)$ and $A_3(u)$ in turn with the following simplification. For the considered fixed design case the estimate $\tilde{p}(x)$ is no longer a random function, and it will be verified shortly that

$$
\max_{x\in[0,1]}|\tilde{p}(x)-p(x)|\le Cn^{-1/3}. \tag{67}
$$

The last inequality immediately implies that $\hat{p}(x)\equiv\tilde{p}(x), x\in[0,1]$ for all sufficiently large n. To verify (67) we recall that $\tilde{p}(x):=1+\sum_{j=1}^{n^{1/3}}\tilde{\pi}_j\varphi_j(x)$,

$$
\tilde{\pi}_j := n^{-1}\sum_{l=1}^{n}\varphi_j(X_l)=n^{-1}(n+1)\sum_{l=1}^{n}\varphi_j(X_l)\int_{X_l}^{X_{l+1}}p(x)\,\mathrm{d}x =: \pi_j+\gamma_j \tag{68}
$$

where $\pi_j := \int_0^1\varphi_j(x)p(x)\,\mathrm{d}x$ and

$$
|\gamma_j|\le Cn^{-1}(1+j)\le Cn^{-1}n^{1/3}. \tag{69}
$$

Then using $|\sum_{j>n^{1/3}}\pi_j|\le Cn^{-1/3}$ verifies (67). Now we can evaluate $A_1(u)$. Write,

$$
\begin{aligned}
E|A_1(u)|^2 &= n^{-2}E\left\{\sum_{l,m=1}^{n}e^{iu(Y_l-Y_m)}\varphi_r(X_l)\varphi_r(X_m)(\tilde{p}(X_l)-p(X_l))(\tilde{p}(X_m)\right.\\
&\qquad\left.-\ p(X_m))p^{-2}(X_l)p^{-2}(X_m)\right\}\\
&= n^{-2}\sum_{l=1}^{n}(\tilde{p}(X_l)-p(X_l))^2p^{-4}(X_l)\varphi_r^2(X_l)+n^{-2}\\
&\quad\times\sum_{l,m=1}^{n}I(l\ne m)h(u|X_l)\bar{h}(u|X_m)\varphi_r(X_l)\varphi_r(X_m)(\tilde{p}(X_l)\\
&\quad-p(X_l))(\tilde{p}(X_m)-p(X_m))p^{-2}(X_l)p^{-2}(X_m)\}\\
&=: A_{11}(u)+A_{12}(u).
\end{aligned}
$$

The term $A_{11}(u)$ is at most $Cn^{-5/3}$, and also

$$\left| A_{12}(u) - n^{-2} \left| \sum_{l=1}^{n} h(u|X_l)\varphi_r(X_l)(\tilde{p}(X_l) - p(X_l))p^{-1}(X_l) \right|^2 \right| \le Cn^{-5/3}.$$

Using (66) and a straightforward calculation we get

$$\begin{aligned} n^{-2} &| \sum_{l=1}^{n} h(u|X_l)\varphi_r(X_l)(\tilde{p}(X_l) - p(X_l))p^{-1}(X_l)|^2 \\ &\le 2 \left| \int_0^1 h(u|x)\varphi_r(x)(\tilde{p}(x) - p(x))p^{-1}(x)\,\mathrm{d}x \right|^2 \\ &\quad + Cn^{-2} \left[1 + \int_0^1 |\partial h(u|x)/\partial x|\,\mathrm{d}x \right]^2. \end{aligned}$$

The first integral is evaluated using (67) and (68), namely we can write

$$\begin{aligned} &\left| \int_0^1 h(u|x)\varphi_r(x)(\tilde{p}(x) - p(x))p^{-1}(x)\,\mathrm{d}x \right|^2 \\ &= \left| 2^{-1/2} \int_0^1 h(u|x) \left[\sum_{j=1}^{n^{1/2}} (\tilde{\pi}_j - \pi_j)(\varphi_{j-r}(x) + \varphi_{j+r}(x)) \right.\right. \\ &\qquad \left.\left. + \sum_{j>n^{1/3}} \pi_j(\varphi_{j-r}(x) + \varphi_{j+r}(x)) \right] p^{-1}(x)\,\mathrm{d}x \right|^2 \\ &= 2 \left| \sum_{j=1}^{n^{1/3}} (\tilde{\pi}_j - \pi_j)(\kappa_{j-r}(u) + \kappa_{j+r}(u)) + \sum_{j>n^{1/3}} \pi_j(\kappa_{j-r}(u) + \kappa_{j+r}(u)) \right|^2 \\ &\le Cn^{-1} \sum_{j=1}^{n^{1/3}} |\kappa_{j-r}(u) + \kappa_{j+r}(u)|^2 + \sum_{j>n^{1/3}} \pi_j^2 \sum_{j>n^{1/3}} |\kappa_{j-r}(u) + \kappa_{j+r}(u)|^2 \\ &\le C \left[n^{-1} \int_0^1 |h(u|x)|^2\,\mathrm{d}x + n^{-4/3} \int_0^1 |\partial h(u|x)/\partial x|^2\,\mathrm{d}x \right], \end{aligned}$$

where we used notation $\kappa_j(u) := \int_0^1 h(u|x)p^{-1}(x)\varphi_j(x)\,\mathrm{d}x$. We conclude that

$$A_{12}(u) \le Cn^{-4/3} \left[n^{-1/3} + \int_0^1 |\partial h(u|x)/\partial x|^2\,\mathrm{d}x \right],$$

and then we note that the same upper bound holds for $E|A_1(u)|^2$. The term $A_2(u) \equiv 0$ for all sufficiently large n. Using (67) we establish that $|A_3(u)|^2 \le Cn^{-4/3}$, and this finishes the proof of Lemma 4.

The rest of the proof of Theorem 1 for the fixed design case is identical to the above-presented proof for the random design case. Theorem 1 is proved.

Proof of Lemma 1 Set $\pi'_{2r} := \int_0^1 p^{-1}(x)\varphi_{2r}(x)\,\mathrm{d}x$ and note that, due to Assumption 2, $\sum_{r=1}^{\infty} r^2(\pi'_{2r})^2 < C$. Using this and Cauchy–Schwarz inequality we get

$$\sum_{r=1}^{\infty}\int_0^{\infty} I((u,r)\in B_{k\tau})\left|\int_0^1 p^{-1}(x)\varphi_{2r}(x)\,\mathrm{d}x\right|\,\mathrm{d}u$$
$$\leq C(b'_{k+1}-b'_k)\left[\sum_{r=b'_\tau+1}^{b'_{\tau+1}} r^{-2}\right]^{1/2} \leq C(b'_{k+1}-b'_k)\min((b'_{\tau+1}-b'_\tau)^{1/2}(b'_\tau+1)^{-1},1).$$

Similarly, using Assumption 1, we write,

$$\sum_{r=1}^{\infty}\int_0^{\infty} I((u,r)\in B_{k\tau})\int_0^1\left|\int_{-\infty}^{\infty} e^{iuy}f(y|x)\,\mathrm{d}y\right|^2\,\mathrm{d}x\,\mathrm{d}u$$
$$=\sum_{r=1}^{\infty}\int_0^{\infty} I((u,r)\in B_{k\tau})\sum_{s=0}^{\infty}\pi^{-1}|h_s(u)|^2\,\mathrm{d}u \leq C(b'_{\tau+1}-b'_\tau)(b'_k+1)^{-2}.$$

These inequalities, together with an elementary calculation, prove Lemma 1.

Proof of Lemma 2 The assertion is established by a direct calculation similarly to Efromovich (1985, 2005); it is straightforward and thus skipped here.

Proof of Theorem 2 We begin with the random design case. Write,

$$\int_0^1\int_{-\infty}^{\infty}(\tilde f(y|x)-f(y|x))^2\,\mathrm{d}y\,\mathrm{d}x$$
$$=\int_{-\infty}^{\infty}(\tilde f(y)-f(y))^2\,\mathrm{d}y+\int_0^1\int_{-\infty}^{\infty}(\tilde\psi(y,x)-\psi(y,x))^2\,\mathrm{d}y\,\mathrm{d}x. \tag{70}$$

Using the Plancherel identity we get

$$\int_{-\infty}^{\infty}(\tilde f(y)-f(y))^2\,\mathrm{d}y=\pi^{-1}\int_0^{\infty}|\tilde h_0(u)-h_0(u)|^2\,\mathrm{d}u$$
$$=\pi^{-1}\sum_{k=1}^{K}\int_{B_k}|\tilde\mu_k\hat h_0(u)-h_0(u)|^2\,\mathrm{d}u$$
$$+\pi^{-1}\int_{b_{K+1}}^{\infty}|h_0(u)|^2\,\mathrm{d}u, \tag{71}$$

and

$$\int_0^1 \left[\int_{-\infty}^{\infty} (\tilde{\psi}(y,x) - \psi(y,x))^2 \, dy\right] dx$$
$$= \sum_{r=1}^{\infty} \pi^{-1} \int_0^{\infty} |\tilde{h}_r(u) - h_r(u)|^2 \, du$$
$$= \pi^{-1} \sum_{k,\tau=1}^{T} \sum_{r=1}^{\infty} \int_0^{\infty} I((u,r) \in B_{k\tau}) |\tilde{\mu}_{k\tau} \hat{h}_r(u) - h_r(u)|^2 \, du$$
$$+ \pi^{-1} \sum_{k,\tau=1}^{\infty} I((k,\tau) \notin \{1,\dots,T\}^2) \sum_{r=1}^{\infty} \int_0^{\infty} I((u,r) \in B_{k\tau}) |h_r(u)|^2 \, du. \quad (72)$$

From now on we may skip subscripts whenever no confusion occurs. Consider a particular $\nu \in (0,1)$, an interval $D \subset [0,\infty)$, and a nonnegative integer r. Using the inequality

$$\int_D |\tilde{\mu}\hat{h}_r(u) - h_r(u)|^2 \, du \le (1+\nu) \int_D |\mu \hat{h}_r(u) - h_r(u)|^2 \, du$$
$$+ (1+\nu^{-1}) \int_D |(\tilde{\mu} - \mu)\hat{h}_r(u)|^2 \, du, \quad (73)$$

we can evaluate integrals in (71) and (72). The first integral in the right side of (73) was estimated in (59). For the second integral we can write,

$$\sum_{r=1}^{\infty} \int_0^{\infty} I((u,r) \in B) |(\tilde{\mu} - \mu)\hat{h}_r(u)|^2 \, du$$
$$= (\tilde{\mu} - \mu)^2 L(\tilde{\Theta} + \tilde{d}n^{-1})[I(\tilde{\Theta} > t\tilde{d}n^{-1}) + I(\tilde{\Theta} \le t\tilde{d}n^{-1})] =: \rho_1 + \rho_2. \quad (74)$$

The terms ρ_1 and ρ_2 will be evaluated in turn. For ρ_1 we can write,

$$\rho_1 = \left[\frac{\tilde{\Theta}}{\tilde{\Theta} + \tilde{d}n^{-1}} - \frac{\Theta}{\Theta + dn^{-1}}\right]^2 L(\tilde{\Theta} + \tilde{d}n^{-1}) I(\tilde{\Theta} > t\tilde{d}n^{-1})$$
$$= \frac{n^{-2}(d\tilde{\Theta} - \tilde{d}\Theta)^2 L I(\tilde{\Theta} > t\tilde{d}n^{-1})}{(\tilde{\Theta} + \tilde{d}n^{-1})(\Theta + dn^{-1})^2}.$$

Note that $d\tilde{\Theta} - \tilde{d}\Theta = d(\tilde{\Theta} - \Theta) + (d - \tilde{d})\Theta$, and then

$$\rho_1 \le \frac{2n^{-2}d^2(\tilde{\Theta} - \Theta)^2 L I(\tilde{\Theta} > t\tilde{d}n^{-1})}{(\tilde{\Theta} + \tilde{d}n^{-1})(\Theta + dn^{-1})^2}$$
$$+ \frac{2n^{-2}\Theta^2(d - \tilde{d})^2 L I(\tilde{\Theta} > t\tilde{d}n^{-1})}{(\tilde{\Theta} + \tilde{d}n^{-1})(\Theta + dn^{-1})^2} =: \rho_{11} + \rho_{12}. \quad (75)$$

We need a technical result.

Lemma 5 *Let Assumptions 1 and 2 hold and $L \le n^{1/2}\ln\ln(n)$. Then*

$$E(\tilde{d}-d)^8 \le C\ln^{10}(n)n^{-8/3}, \tag{76}$$
$$E(\tilde{\Theta}-\Theta)^6 \le C(L^*)^{-3}n^{-3}(\Theta+n^{-1})^3. \tag{77}$$

Proof of Lemma 5 The Hölder inequality implies that

$$|\tilde{d}-d|^8 = \left|\int_0^1 (\hat{p}^{-1}(x)-p^{-1}(x))\,\mathrm{d}x\right|^8$$
$$= \left|\int_0^1 (p(x)-\hat{p}(x))[\hat{p}(x)p(x)]^{-1}\,\mathrm{d}x\right|^8 \le C(\ln\ln(n))^8\int_0^1 \left|p(x)-\hat{p}(x)\right|^8\mathrm{d}x.$$

Then using (58) verifies (76). To check (77) we note that

$$\tilde{\Theta} = L^{-1}\sum_{r=1}^{\infty}\int_0^1 I((u,r)\in B)|\hat{h}_r(u)|^2\,\mathrm{d}u - \tilde{d}n^{-1}$$
$$= \left[L^{-1}\int_B |\hat{h}_r(u)|^2\,\mathrm{d}u - \mathrm{d}n^{-1}\right] + (d-\tilde{d})n^{-1}.$$

Then (76) implies that the term $(\tilde{d}-d)n^{-1}$ is sufficiently small, and this together with a direct calculation, similar to the proof of Lemma 3 in Efromovich (1985), verifies (77). Lemma 5 is proved.

Now we can continue evaluation of terms ρ_{11} and ρ_{12} defined in (75). Write,

$$\rho_{11} = \rho_{11}I(|\tilde{d}-d| > d/2) + \rho_{11}I(|\tilde{d}-d| \le d/2) =: \rho_{111}+\rho_{112}.$$

Lemma 5 and the Cauchy–Schwarz inequality imply

$$E\{\rho_{111}\} \le Cn^{-2}(L^*)^{-1}n^{-1}(\Theta+dn^{-1})L\ln^5(n)n^{-4/3}[n^{-1}(\Theta+dn^{-1})^2]^{-1}$$
$$\le Cn^{-2}n^{-1/4}(L/L^*).$$

Using line (5.10) from Efromovich (1985) together with Lemma 5 yields

$$(1+t^{-1/2})E\{\rho_{112}\} \le Cn^{-1}[L\mu((L^*)^{-1}t^{-3/2}) + (L^*)^{-2}t^{-5}].$$

Also, a plain calculation based on (76) implies $E\{\rho_{12}\} \le CL\mu n^{-1}n^{-2/3}\ln^2(n)$. Combining the results and using $L \le n^{1/2}\ln\ln(n)$ we get

$$(1+t^{-1/2})E\{\rho_1\} \le Cn^{-1}[L\mu((L^*)^{-1}t^{-3/2}) + (L^*)^{-2}t^{-5}]. \tag{78}$$

Now we are considering ρ_2 defined in (74). Write,

$$\rho_2 = \rho_2I(|\tilde{d}-d| > d/2) + \rho_2I(|\tilde{d}-d| \le d/2) =: \rho_{21}+\rho_{22}.$$

Lemma 2 together with plain $\tilde{d} \le \ln\ln(n)$ allows us to write

$$E\{\rho_{21}\} = \mu^2 L E\{(\tilde{\Theta} + \tilde{d}n^{-1}) I(\tilde{\Theta} \le t\tilde{d}n^{-1}) I(|\tilde{d} - d| > d/2)\} \le \mu L n^{-1}(Cn^{-2}).$$

Using line (5.11) from Efromovich (1985) together with Lemma 5 implies

$$(1 + t^{-1/2})E\{\rho_{22}\} \le Cn^{-1}L\mu[t^{1/2} + (L^*)^{-1/2}t^{-3/2}],$$

and thus for the considered blocks

$$(1 + t^{-1/2})E\{\rho_2\} \le Cn^{-1}\mu L[t^{1/2} + (L^*)^{-1/2}t^{-3/2}].$$

Combining the obtained results in (73), choosing a particular $\nu = t^{1/2}$, and then using (70)–(72) proves Theorem 2 for the considered case of random design. The fixed-design case is considered absolutely similarly with the help of Lemma 4. Theorem 2 is proved.

References

Abramovich, F., Sapatinis, T. (1999). *Bayesian approach to wavelet decomposition and shrinkage*. New York: Springer.

Arnold, B. C., Castillo, E., Sarabia, J. M. (1999). *Conditional specification of statistical models*. New York: Springer.

Efromovich, S. (1985). Nonparametric estimation of a density with unknown smoothness. *Theory of Probability and its Applications*, *30*, 557–568.

Efromovich, S. (1999). *Nonparametric curve estimation: methods, theory and applications*. New York: Springer.

Efromovich, S. (2000). On sharp adaptive estimation of multivariate curves. *Mathematical Methods of Statistics*, *9*, 117–139.

Efromovich, S. (2005). Estimation of the density of regression errors. *The Annals of Statistics*, *33*, 2194–2227.

Efromovich, S. (2007). Conditional density estimation in a regression setting. *The Annals of Statistics*, *33*, 2504–2535.

Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. New York: Marcel Dekker.

Fan, J. (1992). Design-adaptive nonparametric regression. *The Journal of American Statistical Association*, *87*, 998–1004.

Fan, J., Gijbels, I., (1996). *Local polynomial modeling and its applications*. New York: Chapman and Hall.

Koul, H., Sakhanenko, L. (2005). Goodness-of-fit testing in regression: A finite sample comparison of bootstrap methodology and Khmaladze transformation. *Statistics and Probability Letters*, *74*, 290–302.

Neter, J., Kutner, M., Nachtsheim, C., Wasserman, W. (1996). *Applied linear models*, 4th ed. Boston: McGraw-Hill.

Prakasa Rao, B. L. S. (1983). *Nonparametric functional estimation*. New York: Academic Press.

Samarov, A. M. (1992). Lower bound for the integral risk of density function estimates. *Topics in Nonparametric Estimation*, *8*, 1–7.