AISM (2007) 59: 617–654
DOI 10.1007/s10463-006-0067-3

Sam Efromovich

Optimal nonparametric estimation of the density of regression errors with finite support

Received: 30 March 2005 / Revised: 26 December 2005 /
Published online: 25 July 2006

Abstract Knowledge of the probability distribution of error in a regression problem plays an important role in verification of an assumed regression model, making inference about predictions, finding optimal regression estimates, suggesting confidence bands and goodness of fit tests as well as in many other issues of the regression analysis. This article is devoted to an optimal estimation of the error probability density in a general heteroscedastic regression model with possibly dependent predictors and regression errors. Neither the design density nor regression function nor scale function is assumed to be known, but they are suppose to be differentiable and an estimated error density is suppose to have a finite support and to be at least twice differentiable. Under this assumption the article proves, for the first time in the literature, that it is possible to estimate the regression error density with the accuracy of an oracle that knows "true" underlying regression errors. Real and simulated examples illustrate importance of the error density estimation as well as the suggested oracle methodology and the method of estimation.

Keywords and phrases Adaptation · Error depending on predictor · Heteroscedasticity · Minimax · Pinsker oracle

1 Introduction

Let n identical and independently distributed (iid) observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ of the pair of random variables (X, Y) be available. The classical (and simplest) regression problem is to find and then infer about the regression function

S. Efromovich
Department of Mathematical Sciences,
The University of Texas at Dallas,
P.O. Box 830688, EC35, Richardson,
TX 75083-0688, USA
E-mail: efrom@utdallas.edu

$m(x) := E(Y|X = x)$ in the model

$$Y = m(X) + \xi, \tag{1}$$

where X is uniformly distributed on [0, 1] predictor and ξ is independent of the predictor regression error. See a discussion of this homoscedastic nonparametric regression model in Fan and Gijbels (1996), Hart (1997), Eubank (1999) and Efromovich (1999).

Suppose that the statistician uses a regression estimator $\tilde{m}(x)$, then $R_l := Y_l - \tilde{m}(X_l), l = 1, \ldots, n$ are called residuals, and they are traditionally used as proxies for corresponding unobserved regression errors. The latter is the foundation of the classical residual analysis used for model validation, hypothesis testing and prediction (Neter et al., 1996; Fan and Gijbels, 1996).

Interestingly, despite the fact that the residual analysis is widely used in applied nonparametric statistic, so far no mathematically rigorous result about a possibility to use residuals for an optimal (in any sense) error density estimation is known. The available literature is practically next to none with just several known articles devoted to consistent estimation of the regression error density (Cheng, 2002, 2004). At the same time, there is a vast literature devoted to the estimation and application of functionals of the error density (Akritas and Van Keilegom, 2001; Müller et al., 2004).

It has been conjectured (with outlined ideas of a possible proof) in Efromovich (2005) that under a mild assumption on smoothness of the regression function and the error density, residuals can proxy underlying regression errors in the following optimal sense. Consider an oracle, which is a traditional density estimator with best known properties under the mean integrated squared error (MISE) criteria, that knows underlying regression errors. Then, it is possible to suggest a data-driven regression estimator whose residuals can be plugged in the oracle in place of regression errors and the obtained data-driven error density estimator will have the same MISE convergence as the oracle based on "true" regression errors. In other words, the conjecture is that error density estimation based on either residuals or underlying regression errors imply the same asymptotic MISE. It has been recommended to use a Pinsker oracle, which is a classical blockwise-shrinkage orthogonal series density estimator for the case of known errors with a finite support. This choice is justified by the fact that this series estimator is sharp-minimax over a vast class of densities including both Sobolev and analytic ones, superefficient and also an excellent plug-in estimate for many traditionally studied functionals (Brown et al., 1997; Efromovich, 1998, 1999, 2001; Bickel and Ritov, 2003; Wasserman, 2005). Let us note that while regression errors with a finite support dominate applied settings, theoretical residual analysis is also interested in errors with infinite support (like normal errors); this important setting will be considered in a separate publication because proofs are too lengthy for a single paper.

The main aim of this article is to prove the above-described conjecture for the case of regression errors with a finite support. Also, we shall consider a more general heteroscedastic regression model

$$Y = m(X) + \sigma(X)\xi, \tag{2}$$

where X is the predictor with the design density $p(x)$ supported on $[0, 1], \sigma(x) > 0$ is the scale function, and ξ is a regression error which may depend on the predictor such that $E\{\xi|X\} = 0$ and $\mathrm{Var}\{\xi|\mathrm{X}\} = 1$. Neither the regression function $m(x)$ nor the scale function $\sigma(x)$ nor the design density $p(x)$ is suppose to be known. Let us also note that the information about dependence or independence of the predictor and error is unavailable to the statistician, thus in general a corresponding marginal error density is the estimand. Of course, in the case of the dependency the condition density of the error may be of interest; this problem will be considered in a separate paper.

The context of the article is as follows. Section 2 formally describes the problem and its solution. It also presents the analysis of a real dataset and a simulated example which shed light on the problem and the proposed method of estimation. Sections 3–5 contain proofs.

2 Estimation of the error density

We begin with describing the considered regression model, then describe Pinsker oracle, define plugged-in residuals, present a proposition about optimal error density estimation based on the residuals and finish with examples.

2.1 Nonparametric regression model

We are considering a general heteroscedastic regression model (2) where observations are n iid realizations $(X_1, Y_1), \dots, (X_n, Y_n)$ from the pair (X, Y) of the predictor and the response. The predictor X is distributed according to an unknown design density $p(x)$ supported on $[0, 1]$. Neither the regression function $m(x)$ nor the scale function $\sigma(x)$ nor the design density $p(x)$ is assumed to be known. The regression error ξ satisfies $E\{\xi|X\} = 0$ and $\mathrm{Var}(\xi|\mathrm{X}) = 1$, it does not take values beyond a known finite interval $[a, a + b]$, and it may depend on the predictor according to an unknown conditional density $b^{-1}\psi([v - a]/b|x)$, $v \in [a, a + b]$.

The problem of interest is to estimate the (in general marginal) probability density of the regression error ξ and to show that, under a mild assumption, appropriately calculated residuals can proxy underlying regression errors unavailable to the statistician. Without any loss of generality, from now on we shall consider a transformed error $\epsilon := (\xi - a)/b$ as the object of interest, refer to ϵ as the error and be interested in the estimation of its (marginal) density $f(u) = \int_0^1 \psi(u|x)p(x)\mathrm{d}x$, $u \in [0, 1]$. (Let us note that due to the zero mean of ξ, we cannot assume that it is supported on $[0, 1]$.)

2.2 Pinsker oracle

The statistician needs to estimate the error density f based solely on n pairs of observations (X_l, Y_l), $l = 1, \dots, n$. If we look one more time at (2), then it becomes clear that the problem is indirect and it involves nuisance functions. In such a complicated indirect setting, it is reasonable to employ a popular in the nonparametric literature oracle approach where an estimator is compared with an oracle (guru, pseudo-estimator) that knows underlying regression errors. Note that, formally the latter means that an oracle knows the regression and scale functions. Then the oracle becomes a natural benchmark for any data-driven estimator of the error density.

Let us define Pinsker oracle used in this article. Assume that $Z_1^n := (Z_1, \ldots, Z_n)$ is the vector of n iid observations distributed according to an estimated density $f(z)$ supported on [0, 1]. Then Pinsker oracle is a data-driven (adaptive) blockwise-shrinkage orthogonal series density estimate defined as follows. Consider a classical cosine basis $\{1, \varphi_j(z) := 2^{1/2}\cos(\pi jz), j = 1, 2, \ldots\}$ on [0, 1], this is the place where the unit support becomes handy. Introduce an increasing to infinity sequence of positive integers $1 = q_1 < q_2 < \ldots$, which divides frequencies of the basis into blocks $B_k := \{q_k, q_k + 1, \ldots, q_{k+1} - 1\}$ having lengths $L_k := q_{k+1} - q_k, k = 1, 2, \ldots$ Also a sequence of corresponding positive and finite thresholds t_k is introduced. To be specific, set $L_k = k^2$ and $t_k = \ln^{-2}(2 + k)$. Then Pinsker oracle is

$$\hat{f}_P(z, Z_1^n) := 1 + \sum_{k=1}^{K} \bar{\mu}_k \sum_{j \in B_k} \bar{\theta}_j \varphi_j(z), \quad z \in [0, 1], \tag{3}$$

where K is a minimal integer such that $\sum_{k=1}^{K} L_k \geq n^{1/5} b_n$, $b_n := 4 + \ln\ln(n + 20)$, $\{\bar{\theta}_j\}$ are empirical Fourier coefficients [estimates of Fourier coefficients $\theta_j := \int_0^1 f(z)\varphi_j(z)\mathrm{d}z$]

$$\bar{\theta}_j := n^{-1} \sum_{l=1}^{n} \varphi_j(Z_l), \tag{4}$$

and the shrinkage coefficients are

$$\bar{\mu}_k := \frac{L_k^{-1} \sum_{j \in B_k} \bar{\theta}_j^2 - n^{-1}}{L_k^{-1} \sum_{j \in B_k} \bar{\theta}_j^2} I\Big(L_k^{-1} \sum_{j \in B_k} \bar{\theta}_j^2 > (1 + t_k) n^{-1}\Big). \tag{5}$$

This oracle, as a data-driven (adaptive) estimator based on n direct observations Z_1^n from the density f, is a sharp-minimax for Sobolev and analytic densities, superefficient and has a whole bouquet of other excellent statistical properties (Efromovich, 1985, 1999, 2005; Donoho and Johnstone, 1995; Brown et al., 1997; Zhang, 2005).

There are two ways to explain why the oracle has such nice properties. The former is to note that (5) mimics a familiar blockwise Wiener filter, which employs optimal shrinkage coefficients $\mu_k^* := \Theta_k/(\Theta_k + n^{-1})$, $\Theta_k := L_k^{-1} \sum_{j \in B_k} \theta_j^2$; because Wiener filter, is based on Fourier coefficients of the estimated (and unknown to the statistician) density of errors, it is the "ultimate" oracle. The latter is to realize that if in (5) we replace the used hard thresholding by a soft thresholding, then (3) is transformed into a classical Stein shrinkage procedure. This point of view was first expressed in Donoho and Johnstone (1995), and the discussion of Stein shrinkage in nonparametric curve estimation can be found, for instance, in Efromovich (1999) and Wasserman (2005).

If we set $Z_1^n = \epsilon_1^n$, then (3) can be referred to as Pinsker oracle for the considered error density estimation problem. Then Pinsker oracle, which knows "true" underlying regression errors, becomes a natural benchmark for any data-driven error density estimator based on observations $(X_1, Y_1), \dots, (X_n, Y_n)$. (Please note the following hierarchy among the oracles: Pinsker oracle is less powerful than Wiener oracle because the latter knows the underlying error density and "true" regression errors while the former knows only "true" regression errors. At the same time, under the MISE criteria, these oracles have the same asymptotic minimax properties, and this fact justifies the choice of Pinsker oracle as the benchmark for a data-driven error density estimate.)

2.3 Assumptions

We need one assumption about the regression model and another about smoothness of the conditional density $\psi(u|x)$ which together with the design density $p(x)$ defines the density of interest $f(u) = \int_0^1 \psi(u|x)p(x)dx, u \in [0, 1]$.

Assumption A Model (2) is considered where the regression error ξ may depend on the predictor X, $E\{\xi|X\} = 0$, $\mathrm{Var}(\xi|\mathrm{X}) = 1$ and $P(\xi \in [a, a+b]) = 1$, where $a < b$ are two given real numbers. Pairs of observations $(X_1, Y_1), \dots, (X_n, Y_n)$ are iid. The regression function $m(x)$, the design density $p(x)$ and the scale function $\sigma(x)$ are differentiable and their derivatives are bounded and integrable on [0, 1]. Also, $\min_{x\in[0,1]} \min(\sigma(x), p(x)) > 0$ and $\int_0^1 p(x)dx = 1$.

Assumption B The conditional density $\psi(u|x)$ is such that $(\partial/\partial x)(\partial^2/\partial u^2)\psi(u|x)$ exists and is bounded and integrable on $[0, 1]^2$, and $\psi(u|x) = 0$ for $u \notin (0, 1)$, $x \in [0, 1]$.

2.4 Residuals

Let us explain how to find residuals that can proxy underlying regression errors. Recall notation $b_n = 4 + \ln\ln(n + 20)$ and define several more sequences in n: $n_2 := n - 3n_1$; n_1 is the smallest integer larger than n/b_n; $S := S_n$ is the smallest integer larger than $n^{1/3}$; $o(1) \to 0$ as $n \to \infty$. In what follows, we always consider sufficiently large n such that $\min(n_1, n_2) > 4$, and integrals are taken over [0, 1].

Now we can define the procedure. The first n_1 observations are used to estimate the design density $p(x)$, the next n_1 observations are used to estimate the regression function $m(x)$, the next n_1 observations are used to estimate the scale function $\sigma(x)$ and the last n_2 observations are used to estimate the error density of interest $f(u)$. Note that $n_2 \geq [1 - 3(b_n^{-1} + n^{-1})]n$ and thus using either n_2 or n observations implies the same MISE convergence. The three nuisance functions are estimated using a truncated cosine series estimator. The design density estimator is

$$\hat{p}(x) = \max\left(b_n^{-1},\ n_1^{-1}\sum_{s=0}^{S}\sum_{l=1}^{n_1}\varphi_s(X_l)\varphi_s(x)\right). \tag{6}$$

The regression estimator is

$$\hat{m}(x) = \sum_{s=0}^{S}\hat{\kappa}_s\varphi_s(x), \tag{7}$$

where

$$\hat{\kappa}_s = n_1^{-1} \sum_{l=n_1+1}^{2n_1} Y_l \hat{p}^{-1}(X_l)\varphi_s(X_l). \tag{8}$$

The scale estimator is

$$\hat{\sigma}(x) = \left[\min(\max(\tilde{\sigma}^2(x), b_n^{-2}), b_n^2)\right]^{1/2}, \tag{9}$$

where $\tilde{\sigma}^2(x)$ is a regression estimator defined identically to (7) and (8) only with pairs $\{(X_l, Y_l), l = n_1 + 1, \dots, 2n_1\}$ being replaced by $\{(X_l, [Y_l - \hat{m}(X_l)]^2), l = 2n_1 + 1, \dots, 3n_1\}$.

Then transformed onto [0,1] residuals are defined as

$$\hat{\epsilon}_l := \frac{Y_l - \hat{m}(X_l)}{b\hat{\sigma}(X_l)} - \frac{a}{b}, \qquad l = n - n_2 + 1, \dots, n. \tag{10}$$

2.5 Assertion

Denote by $\hat{\mathbf{Z}}$ a vector $(\hat{\epsilon}_{n-n_2+1}, \dots, \hat{\epsilon}_n)$ of residuals and by $\mathbf{Z}$ a vector of "true" errors $(\epsilon_1, \dots, \epsilon_n)$; recall that the errors are known to Pinsker oracle. It is possible to show that, under the made assumption, the MISE of plugged-in Pinsker oracle $\hat{f}_P(u, \hat{\mathbf{Z}})$ asymptotically matches the MISE of Pinsker oracle $\hat{f}_P(u, \mathbf{Z})$.

Theorem 1 *Suppose that Assumptions A and B hold. Then, for all sufficiently large samples such that* $\min(n_1, n_2) > 4$, *there exists a finite constant* P^* *such that the MISE of plugged-in Pinsker oracle* $\hat{f}_P(u, \hat{\mathbf{Z}})$ *satisfies the following oracle inequality:*

$$E\int(\hat{f}_P(u, \hat{\mathbf{Z}}) - f(u))^2 du$$
$$\le (1 + P^* \ln^{-1}(b_n))E\int(\hat{f}_P(u, \mathbf{Z}) - f(u))^2 du + P^* b_n^3 n^{-1}. \tag{11}$$

Let us make several comments about the result. First of all, the oracle inequality (11) is not asymptotic. Second, plainly $P^* b_n^3 n^{-1} < C[\ln\ln(n)]^3 n^{-1}$, $C < \infty$. Also recall that the fastest nonparametric rate of the MISE convergence, traditionally considered in the minimax nonparametric density estimation literature, is $\ln(n)n^{-1}$ for analytic densities, and for a Sobolev class of order α the rate is $n^{-2\alpha/(2\alpha+1)}$. Thus we may conclude that the suggested data-driven error density estimator is simultaneously sharp-minimax over analytic and Sobolev density classes. Third, the inequality (11) verifies that the data-driven error density estimator matches Pinsker oracle under the MISE criteria for large sample sizes. Fourth, we conclude that residuals (10) do proxy underlying errors in the above-defined optimal sense. Fifth, it is important to stress that the made assumptions involve no interplay between smoothness of the conditional error density and smoothness of the triplet of nuisance functions (design density, regression and scale). In particular, even if the error density is analytic (infinitely differentiable), then it suffices for the design density, regression and scale functions to be only differentiable for preserving the very fast $\ln(n)n^{-1}$ rate of the error density estimator's MISE convergence. This allows us to conclude that, under a mild assumption, residuals can be robust and optimal proxies for unobserved regression errors, and this conclusion supports the customary methodology of the applied residual analysis. Sixth, following Neter et al. (1996), Hart (1999), Cheng (2002, 2004) and Müller et al. (2004), we can conclude that the proposed error density estimator allows one to solve many classical applied regression problems like prediction, change-point inference, goodness-of-fit testing, estimation of functionals of the error density, etc. Finally, let us recall that Akritas and Van Keilegom (2001) estimated the cdf of an (independent of predictor) regression error, and then the suggested estimate was used for prediction and goodness-of-fit tests. Following Efromovich (2004), it is reasonable to conjecture that, by taking a corresponding integral of the error density estimator, it is possible to obtain a second-order efficient estimator of the cdf, and then establish corresponding optimal properties for goodness-of-fit tests. This along with other natural applications of the error density estimator will be presented elsewhere.

2.6 Examples

We begin with the analysis of a real dataset. The top diagram in Fig. 1 exhibits results of a controlled experiment conducted by BIFAR, a company with businesses in equipment and chemicals for wastewater treatment plants. The regression curve

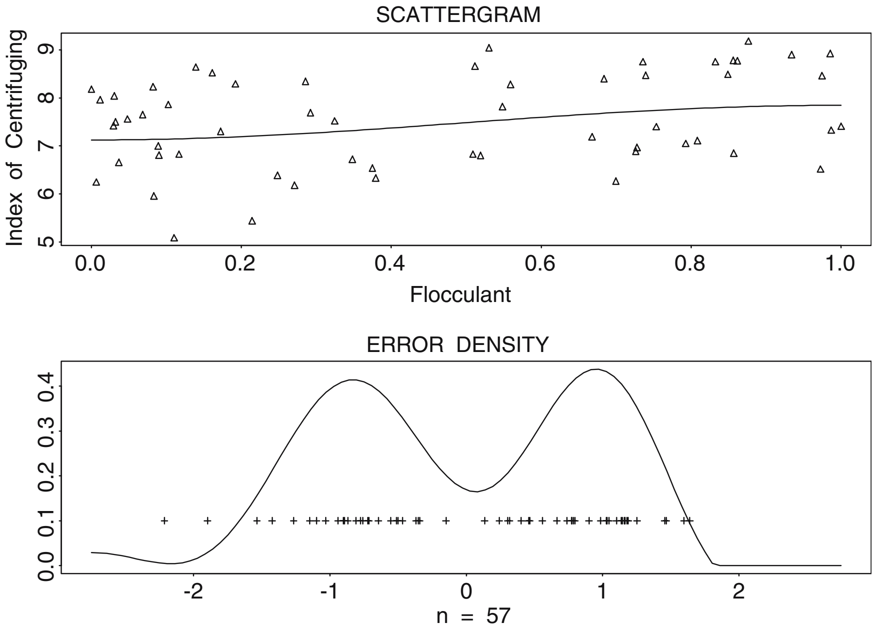

Fig. 1 BIFAR's experiment of centrifuging a mixture of several wastes with the controlled variable being the quantity of used flocculant. The data is rescaled by BIFAR. Observations and estimated regression errors are shown by *triangles* and *crosses* in the top and bottom diagrams, respectively. The nonparametric regression and the error density estimates are shown by the *solid lines*

indicates that the process of centrifuging is improved as more flocculant is added; here software of Efromovich (1999) is used. It is difficult to see something unusual in this scattergram, but the estimated error density, exhibited in the bottom diagram, clearly tells us that there are two pronounced clusters of sludges which are affected differently by the centrifuging. This shows us that the error density estimation is an important univariate characteristic on its own, which can shed a new light on the association between the predictor and response.

Figure 2 helps us to understand how the error density estimator works because a simulated example is considered. The left column of the diagram exhibits a classical regression analysis: we can visualize the regression and scale estimates as well as the rescaled residuals. Note that the estimates are far from being perfect, and this is typical for the considered sample size $n = 50$. At the same time, the estimates do a good job for the data at hand because the rescaled residuals look homogeneous. Then it is of special interest to look at how the suggested error density estimate will perform for such a complicated case with relatively large measurement errors caused by the poor estimation of the regression and scale functions. The right column of the diagram shows us the error density analysis. First of all, the top diagram exhibits the "true" errors, which are available to the oracle but not to the statistician. The middle diagram exhibits the estimated errors, which are the same as is in the left-bottom diagram. The bottom diagram shows that the shape of the error density estimate, which is not as bad as one could expect (after visualization of the regression and scale estimates), but it is definitely worse than the oracle's estimate based on the "true" errors. Here the estimate and the oracle are based on the same density estimator of Efromovich (1999, s.3.1) with the only difference being the used errors. The interested reader can find more simulations as well as results of an intensive Monte Carlo study in Efromovich (2006).

3 Several technical results

This section presents several results needed for the proof of Theorem 1; they will also allow us to shed light on the assumptions.

In what follows C's and C_k's denote generic positive and finite constants, $g^{(l)}(x)$ denotes the lth derivative, if the interval of integration is not indicated then it is assumed to be [0, 1], and $J := \sum_{k=1}^{K} L_k$. Similarly to (4), denote by $\hat{\theta}_j := n_2^{-1} \sum_{l=n-n_2+1}^{n} \varphi_j(\hat{\epsilon}_l)$ the jth plugged-in estimate of Fourier coefficient θ_j and then by $\hat{\mu}_k$ a shrinkage coefficient defined in (5) with $\bar{\theta}_j$ replaced by $\hat{\theta}_j$ and n by n_2.

Lemma 1 (a) *Suppose that $X_1, X_2, \ldots, X_n$ are iid according to the probability density $p(x)$ satisfying Assumption A. Then for any positive integer k the design density estimator $\hat{p}(x)$, defined in (6), satisfies*

$$\max_{x\in[0,1]} E(p(x) - \hat{p}(x))^{2k} \le C_k \ln^{2k}(n) n^{-2k/3}. \tag{12}$$

(b) *Suppose that Assumption A holds, then for any positive integer k the regression estimator $\hat{m}(x)$, defined in (7), satisfies*

$$\max_{x\in[0,1]} E(m(x) - \hat{m}(x))^{2k} \le C_k \ln^{4k+1}(n) n^{-2k/3}, \tag{13}$$

and the scale estimator $\hat{\sigma}(x)$, defined in (9), satisfies

$$\max_{x\in[0,1]} E(\sigma^2(x) - \hat{\sigma}^2(x))^{2k} \leq C_k \ln^{4k+1}(n) n^{-2k/3}. \tag{14}$$

Further, let us consider, as defined in (8), estimate $\hat{\kappa}_s$ of Fourier coefficients $\kappa_s = \int m(x)\varphi_s(x)\mathrm{d}x$ as well an estimate $\hat{\nu}_s$ of Fourier coefficients ν_s of the squared scale function where

$$\nu_s := \int \sigma^2(x)\varphi_s(x)\mathrm{d}x, \quad \hat{\nu}_s := \int \hat{\sigma}^2(x)\varphi_s(x)\mathrm{d}x. \tag{15}$$

Lemma 2 *Suppose that Assumption A holds. Then for any $s \in \{0, 1, \dots, S\}$*

$$\max\left(E(\kappa_s - \hat{\kappa}_s)^4, E(\nu_s - \hat{\nu}_s)^4\right) \leq C\left[n_1^{-2} + S^{-4}(1 + S - s)^{-4}\right]. \tag{16}$$

Denote

$$\begin{aligned} H_l &:= H(X_l, \xi_l) \\ &:= \frac{m(X_l) - \hat{m}(X_l)}{b\hat{\sigma}(X_l)} + \xi_l \frac{\sigma(X_l) - \hat{\sigma}(X_l)}{b\hat{\sigma}(X_l)} =: V(X_l) + \xi_l W(X_l). \end{aligned} \tag{17}$$

The following proposition describes properties of statistics V and W.

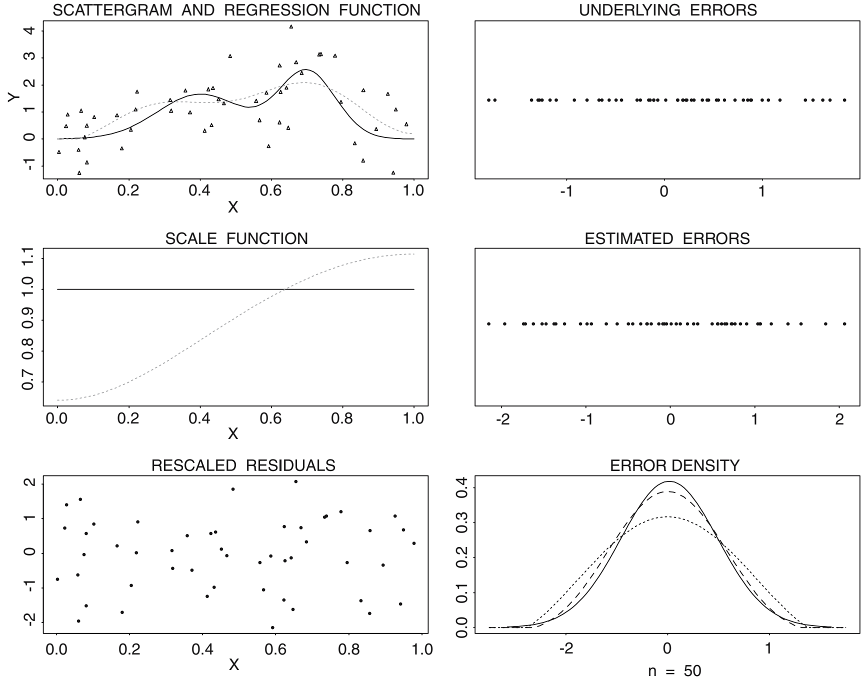

Fig. 2 Simulated example. The *solid lines* show the underlying functions, the *dotted lines* show the estimates, and the *dashed line* shows the oracle's error density estimate

Lemma 3 *Suppose that Assumptions A and B hold. Then for any positive integer k*

$$E(V(X_n))^{2k} + E(W(X_n))^{2k} \leq C_k \ln^{4k+1}(n) n^{-2k/3}, \tag{18}$$

and for a differentiable on $[0, 1]$ *function* $g(x)$*, whose derivative is square integrable,*

$$E\left[\int_0^1 g(x)V(x)\mathrm{d}x\right]^4 + E\left[\int_0^1 g(x)W(x)\mathrm{d}x\right]^4 \leq Cn_1^{-2}. \tag{19}$$

Our final result describes properties of Fourier coefficients of differentiable functions as well as an important properties of H_l defined in (17).

Lemma 4 (a) *Let a function* $g(u)$ *be differentiable on* $[0, 1]$ *and* $\int_0^1 [g^{(1)}(u)]^2 \mathrm{d}u < \infty$*. Then*

$$\sum_{j=1}^{\infty} (\pi j)^2 \left[\int_0^1 g(u)\varphi_j(u)\mathrm{d}u\right]^2 = \int_0^1 \left[g^{(1)}(u)\right]^2 \mathrm{d}u. \tag{20}$$

If additionally $g(0) = g(1) = 0$*, then*

$$\sum_{j=0}^{\infty} (\pi j)^2 \left[\int_0^1 g(u)2^{1/2} \sin(\pi j u)\mathrm{d}u\right]^2 = \int_0^1 \left[g^{(1)}(u)\right]^2 \mathrm{d}u, \tag{21}$$

and if also the function g is twice differentiable and $\int_0^1 \left[g^{(2)}(u)\right]^2 \mathrm{d}x < \infty$*, then*

$$\sum_{j=0}^{\infty} (\pi j)^4 \left[\int_0^1 g(u)2^{1/2} \sin(\pi j u)\mathrm{d}u\right]^2 = \int_0^1 \left[g^{(2)}(u)\right]^2 \mathrm{d}u, \tag{22}$$

and

$$\max_j (1 + j)^4 \left[\int g(u)\varphi_j(u)\mathrm{d}u\right]^2 \leq C. \tag{23}$$

(b) *Let* $g(x), x \in [0, 1]$ *be a function with a square integrable derivative. Then under Assumption B for any natural k*

$$\sum_{j=0}^{\infty} \sum_{s=0}^{\infty} (1+j)^4 (1+s)^2 \left[\int_0^1 \int_0^1 \varphi_s(x)g(x)u^k \psi(u|x) \sin(\pi j u)\mathrm{d}u\mathrm{d}x\right]^2 \leq C_k, \tag{24}$$

and

$$\max_{x\in[0,1]} \sum_{j=0}^{\infty} (1 + j)^4 \left[\int_0^1 u^k \psi(u|x) \sin(\pi j u)\mathrm{d}u\right]^2 \leq C_k. \tag{25}$$

(c) *Under assumption B for any natural k*

$$\max_{x\in[0,1]} \max\left(\left[\int_0^1 u^k\psi(u|x)\sin(\pi ju)\mathrm{d}u\right]^2, \left[\int_0^1 u^k\psi(u|x)\cos(\pi ju)\mathrm{d}u\right]^2\right) \le (1+j)^{-4}C_k. \tag{26}$$

(d) *Let Assumptions A and B hold. Then*

$$E\left(\sum_{j\in B_k} j^2\left[\int_0^1\int_0^1 p(x)H(x,a+bu)\psi(u|x)\sin(\pi ju)\mathrm{d}u\mathrm{d}x\right]^2\right)^2 \le Cn_1^{-2}k^{-6}. \tag{27}$$

Let us comment on these results and assumptions.

Remark 1 All traditional trigonometric approximation theorems, that imply either the above-mentioned or any other approximation results, are established for periodic/circular functions on $[-1, 1]$ and a classical sine–cosine basis (Butzer and Nessel, 1971; Efromovich, 1999). In this article only cosine elements of the classical basis are used, the considered interval is $[0, 1]$, and no periodicity/circularity is assumed. Let us explain the approach taken via the regression function example. Consider in place of $m(x)$ a corresponding periodic and even function $m^*(x) := m(-x), x \in [-1, 0), m^*(x) = m(x), x \in [0, 1]$ and $m^*(x) = m^*(x+2), x \in (-\infty, \infty)$. Then all sine Fourier coefficients vanish and we are left with only cosine elements. Under Assumption A the function $m^*(x)$ is Lipschitz of order 1 on a real line [or equivalently on the circumference of a circle of radius $(2\pi)^{-1}$] and thus all corresponding results for such functions and their cosine Fourier coefficients hold. There is also another way to look at the issue via a direct analysis of $m(x), x \in [0, 1]$. Neither the considered regression function $m(x), x \in [0, 1]$ nor its derivative is assumed to be periodic/circular on $[0, 1]$, that is, it is not assumed that $m(0) = m(1)$ and/or $m^{(1)}(0) = m^{(1)}(1)$. This prevents us from employing the classical approximation theory. On the other hand, one can define a new function $g(x) := m(x) + a_1x + a_2x^2$ where, a_1 and a_2 are such that $g(x)$ and $g^{(1)}(x)$ are periodic, namely $g(0) = g(1)$ and $g^{(1)}(0) = g^{(1)}(1)$. Then Assumption A immediately implies that a_1 and a_2 are finite. Denote by $\lambda_j := \int g(x)\varphi_j(x)\mathrm{d}x$ Fourier coefficients of the new function $g(x)$ and by $\nu_j := \int(a_1x + a_2x^2)\varphi_j(x)\mathrm{d}x$ Fourier coefficients of $a_1x + a_2x^2$. Note that Fourier coefficients of $m(x)$ satisfy $\kappa_j = \lambda_j - \nu_j$ and then, for instance, $|\sum_{s>S}\kappa_s\varphi_s(x)| \le |\sum_{s>S}\lambda_s\varphi_s(x)| + |\sum_{s>S}\nu_s\varphi_s(x)|$. The first sum is at most $C\ln(S)S^{-1}$ uniformly over $x \in [0, 1]$ because g and $g^{(1)}$ are periodic and bounded on $[0, 1]$ (Butzer and Nessel, 1971, s.2.4). For the second sum we note that according to Efromovich (1999, p.32), $\nu_s = 2^{1/2}(\pi s)^{-2}[(a_1 + 2a_2)\cos(\pi s) - a_1]$. This implies that $\max_{x\in[0,1]}|\sum_{s>S}\nu_s\varphi_s(x)| \le CS^{-1}$. Combining the results we get a sufficient for this article upper bound $\max_{x\in[0,1]}|\sum_{s>S}\kappa_s\varphi_s(x)| \le C\ln(S)S^{-1}$. In a similar manner other known approximation results can be employed whenever a function and its derivative are involved. A word of caution: a boundary condition may be needed if second and/or higher derivatives are involved; see examples in Efromovich (1999, s.2.2).

Remark 2 Let us explain how the suggested estimation of the scale function can be viewed as a special regression problem. Remember that estimation of $\sigma^2(x)$ is based on n_1 pairs $\{(X_l, Y_l^*), l = 2n_1 + 1, \dots, 3n_1\}$ where,

$$\begin{aligned} Y_l^* &:= [Y_l - \hat{m}(X_l)]^2 = [m(X_l) + \sigma(X_l)\xi_l - \hat{m}(X_l)]^2 \\ &= \sigma^2(X_l) + \{(\xi_l^2 - 1)\sigma^2(X_l) + 2\xi_l(m(X_l) - \hat{m}(X_l))\sigma(X_l) \\ &\quad + [m(X_l) - \hat{m}(X_l)]^2\} \\ &=: \sigma^2(X_l) + \zeta_l, \quad l = 2n_1 + 1, \dots, 3n_1. \end{aligned}$$

If we formally compare these observations with the regression ones $\{(X_l, Y_l = m(X_l) + \sigma(X_l)\xi_l), l = n_1 + 1, \dots, 2n_1\}$, then Y_l^* plays the role of Y_l, $\sigma^2(X_l)$ plays the role of $m(X_l)$, and ζ_l plays the role of $\sigma(X_l)\xi_l$. The main difference between the two regressions is that ζ_l is no longer zero mean but, as we shall see shortly, its mean is sufficiently (for our purposes) close to zero. To explain this, let us introduce a new notation $\hat{\mu}(X_l) + \hat{v}(X_l)\eta_l := \zeta_l$, where $\hat{\mu}(X_l) := [m(X_l) - \hat{m}(X_l)]^2$, $\hat{v}(X_l) := [\mathrm{Var}(\zeta_l | X_l, (X,Y)_1^{2n_1})]^{1/2}$, and $(X,Y)_1^{2n_1} := \{(X_1, Y_1), \dots, (X_{2n_1}, Y_{2n_1})\}$. Then $E(\eta_l | X_l, (X,Y)_1^{2n_1}) = 0$ and we can write

$$Y_l^* = \sigma^2(X_l) + [\hat{\mu}(X_l) + \hat{v}(X_l)\eta_l], \quad l = 2n_1 + 1, \dots, 3n_1. \tag{28}$$

It is plainly verified via using (13) that for any positive integer k

$$\max_x E\{\hat{\mu}^{2k}(x)\} \le C_k \ln^{8k+1}(n) n^{-4k/3}, \tag{29}$$

and for any $l = 2n_1 + 1, \dots, 3n_1$ we get

$$\begin{aligned} \hat{v}^2(x) &= E\left\{[(\xi_l^2 - 1)\sigma^2(x) + 2\xi_l(m(x) - \hat{m}(x))\sigma(x)]^2 | X_l = x, (X,Y)_1^{2n_1}\right\} \\ &= \sigma^4(x) E\left\{(\xi_l^4 - 1) | X_l = x\right\} \\ &\quad + 4\sigma^2(x) E\left\{(m(x) - \hat{m}(x))^2 | X_l = x, (X,Y)_1^{2n_1}\right\} \\ &\quad + 4\sigma^3(x) E\left\{\xi_l^3 | X_l = x\right\} E\left\{(m(x) - \hat{m}(x)) | (X,Y)_1^{2n_1}\right\}. \end{aligned}$$

The latter, together with Assumption A, implies that for any positive integer k

$$\max_x E\{\hat{v}^{2k}(x)\} < C_k. \tag{30}$$

As a result, for all our purposes the artificial scale regression model (28) can be treated similarly to the "true" regression model (2).

Remark 3 The proof of Lemma 1 in Sect. 5 indicates that the factor $\ln^{4k+1}(n)$ in (13)–(14) can be replaced by $b_n^{c_k} \ln^{4k}(n)$ with some $c_k < \infty$.

Remark 4 It follows from the proof of Lemma 2 in Sect. 5 that if the design density estimator is based on $(1+q)S$, $q > 0$ estimated Fourier coefficients instead of just S, then the right hand side of (16) is simplified into Cn_1^{-2} and we get the classical parametric rate of convergence.

4 Proof of Theorem 1

Let us assume that Lemmas 1–4 are valid; they will be verified in Sect. 5. Parseval identity implies that

$$E\int(\hat{f}_P(u,\hat{\mathbf{Z}})-f(u))^2du=E\sum_{k=1}^{K}\sum_{j\in B_k}(\hat{\mu}_k\hat{\theta}_j-\theta_j)^2+\sum_{k>K}\sum_{j\in B_k}\theta_j^2. \quad (31)$$

Because Pinsker oracle uses n observations and only n_2 residuals are plugged in, it is convenient to consider a modified Pinsker oracle based only on last n_2 errors $\mathbf{Z}^* := (\epsilon_{n-n_2+1},\ldots,\epsilon_n)$. Using oracle inequalities of Efromovich (1985) and the plain $n_2 \geq [1-3(b_n^{-1}+n^{-1})]n$ we get

$$E\int(\hat{f}_P(u,\mathbf{Z}^*)-f(u))^2\mathrm{d}u \leq (1+Cb_n^{-1})E\int(\hat{f}_P(u,\mathbf{Z})-f(u))^2\mathrm{d}u. \quad (32)$$

If we use (32) in (11) then it becomes clear that it suffices to compare the data-driven error density estimate with Pinsker oracle based on $\mathbf{Z}^*$. As a result, from now on we denote by $\bar{\theta}_j$ and $\bar{\mu}_k$ the corresponding oracle's component defined in (4)–(5) and based on $\mathbf{Z}^*$ in place of $\mathbf{Z}$; note that now n_2 is used in those formulae in place of n.

Keeping the new notation in mind, we continue the analysis of MISE of the estimator. Using Cauchy inequality and a plain algebra we get

$$\begin{aligned}E\int(\hat{f}_P(u,\hat{\mathbf{Z}})-f(u))^2\mathrm{d}u &\leq (1+\ln^{-1}(b_n))E\int(\hat{f}_P(u,\mathbf{Z}^*)-f(u))^2\mathrm{d}u\\ &\quad+2(1+\ln(b_n))\left[\sum_{k=1}^{K}\sum_{j\in B_k}E\left\{\bar{\mu}_k^2\left(\hat{\theta}_j-\bar{\theta}_j\right)^2\right\}\right.\\ &\quad\left.+\sum_{k=1}^{K}E\left\{\left(\hat{\mu}_k-\bar{\mu}_k\right)^2\sum_{j\in B_k}\hat{\theta}_j^2\right\}\right]. \end{aligned} \quad (33)$$

Now we are evaluating the terms in (33) in turn. Write

$$\begin{aligned}\hat{\theta}_j-\bar{\theta}_j &= n_2^{-1}\sum_{l=3n_1+1}^{n}\left[\varphi_j\Big([Y_l-\hat{m}(X_l)]/b\hat{\sigma}(X_l)-a/b\Big)-\varphi_j(\epsilon_l)\right]\\ &= n_2^{-1}\sum_{l=3n_1+1}^{n}\left[\varphi_j\left(\epsilon_l+\frac{m(X_l)-\hat{m}(X_l)}{b\hat{\sigma}(X_l)}+\xi_l\frac{\sigma(X_l)-\hat{\sigma}(X_l)}{b\hat{\sigma}(X_l)}\right)-\varphi_j(\epsilon_l)\right]. \end{aligned} \quad (34)$$

Then Taylor's expansion (17) and a simple algebra yield

$$
\begin{aligned}
(\hat{\theta}_j - \bar{\theta}_j)^2 \le C \Bigg[& j^2 n_2^{-2} \left\{ \sum_{l=3n_1+1}^{n} H_l \sin(\pi j \epsilon_l) \right\}^2 \\
& + j^4 n_2^{-2} \left\{ \sum_{l=3n_1+1}^{n} H_l^2 \cos(\pi j \epsilon_l) \right\}^2 \\
& + j^6 n_2^{-2} \left\{ \sum_{l=3n_1+1}^{n} H_l^3 \sin(\pi j \epsilon_l) \right\}^2 \\
& + j^8 n_2^{-2} \left\{ \sum_{l=3n_1+1}^{n} H_l^4 \cos(\pi j \epsilon_l) \right\}^2 \\
& + j^{10} n_2^{-2} \left\{ \sum_{l=3n_1+1}^{n} |H_l|^5 \right\}^2 \Bigg] \\
=: & \; C \left[j^2 \tilde{A}_1(j) + j^4 \tilde{A}_2(j) + j^6 \tilde{A}_3(j) + j^8 \tilde{A}_4(j) + j^{10} \tilde{A}_5(j) \right].
\end{aligned}
\tag{35}
$$

Let us consider in turn the five terms on the right side of (35). Remember that we are assuming that n is large enough so $\min(n_1, n_2) > 4$. We begin with the analysis of $\tilde{A}_1(j)$. Write $E\tilde{A}_1(j) = n_2^{-2} \sum_{l,t=3n_1+1}^{n} E\{H_l H_t \sin(\pi j \epsilon_l) \sin(\pi j \epsilon_t)\} \le n_2^{-1} E\{H_n^2\} + |E\{H_n H_{n-1} \sin(\pi j \epsilon_n) \sin(\pi j \epsilon_{n-1})\}|$. Because pairs $(X_{n-1}, \epsilon_{n-1})$ and (X_n, ϵ_n) are independent, the second expectation can be written in the following form [remember notation (17)]:

$$
\begin{aligned}
& E\left\{H_n H_{n-1} \sin(\pi j \epsilon_n) \sin(\pi j \epsilon_{n-1})\right\} \\
& \quad = E\left[\int \int p(x) H(x, bu + a) \psi(u|x) \sin(\pi j u) \mathrm{d}u \mathrm{d}x \right]^2 \\
& \quad = E\left[\int p(x) V(x) \int \psi(u|x) \sin(\pi j u) \mathrm{d}u \mathrm{d}x \right. \\
& \qquad \left. + \int p(x) W(x) \int (bu + a) \psi(u|x) \sin(\pi j u) \mathrm{d}u \mathrm{d}x \right]^2 \\
& \quad \le 2E\left[\int p(x) V(x) \int \psi(u|x) \sin(\pi j u) \mathrm{d}u \mathrm{d}x \right]^2 \\
& \qquad + 2E\left[\int p(x) W(x) \int (bu + a) \psi(u|x) \sin(\pi j u) \mathrm{d}u \mathrm{d}x \right]^2 \\
& \quad =: 2A_{11}(j) + 2A_{12}(j).
\end{aligned}
$$

We begin with the evaluation of $A_{11}(j)$. Denote $\int \psi(u|x)\sin(\pi j u)\mathrm{d}u =: v_j(x)$ and write

$$\begin{aligned}
A_{11}(j) &= E\left[\int p(x)V(x)v_j(x)\mathrm{d}x\right]^2 = E\left[\int \frac{(m(x)-\hat{m}(x))p(x)v_j(x)}{b\hat{\sigma}(x)}\mathrm{d}x\right]^2 \\
&= E\Bigg[\int \Bigg(\frac{(m(x)-\hat{m}(x))p(x)v_j(x)}{b\sigma(x)} \\
&\quad + \frac{(m(x)-\hat{m}(x))(\sigma(x)-\hat{\sigma}(x))p(x)v_j(x)}{b\sigma(x)\hat{\sigma}(x)}\Bigg)\mathrm{d}x\Bigg]^2 \\
&\le 2E\left[\int \frac{(m(x)-\hat{m}(x))p(x)v_j(x)}{b\sigma(x)}\mathrm{d}x\right]^2 \\
&\quad + 2E\left[\int \frac{(m(x)-\hat{m}(x))(\sigma(x)-\hat{\sigma}(x))p(x)v_j(x)}{b\sigma(x)\hat{\sigma}(x)}\mathrm{d}x\right]^2 \\
&=: 2A_{111}(j) + 2A_{112}(j).
\end{aligned}$$

To evaluate $A_{111}(j)$ we note that $m(x)-\hat{m}(x) = \sum_{s=0}^{S}(\kappa_s-\hat{\kappa}_s)\varphi_s(x) + \sum_{s>S}\kappa_s\varphi_s(x)$, where $\kappa_s := \int m(x)\varphi_s(x)\mathrm{d}x$. Denote $v_{js} := \int \varphi_s(x)b^{-1}\sigma^{-1}(x)p(x)v_j(x)\mathrm{d}x$ and write using Hölder inequality:

$$\begin{aligned}
A_{111}(j) &= E\left[\sum_{s=0}^{S}(\kappa_s-\hat{\kappa}_s)v_{js} + \sum_{s>S}\kappa_s v_{js}\right]^2 \\
&\le 2E\left\{\sum_{s=0}^{S}(\kappa_s-\hat{\kappa}_s)^2(1+s)^{-2}\right\}\sum_{s=0}^{S}(1+s)^2 v_{js}^2 \\
&\quad + 2\left[\sum_{s>S}s^{-2}\kappa_s^2\right]\left[\sum_{s>S}s^2 v_{js}^2\right] \\
&\le C\left[n_1^{-1}\sum_{s=0}^{S}(1+s)^2 v_{js}^2 + S^{-4}\sum_{s>S}s^2 v_{js}^2\right] \le Cn_1^{-1}\sum_{s=0}^{\infty}(1+s)^2 v_{js}^2.
\end{aligned}$$

In the next to last inequality we used Lemma 2 and based on Lemma 3 relation $\sum_{s>S}s^{-2}\kappa_s^2 \le CS^{-4}\sum_{s>S}s^2\kappa_s^2 \le S^{-4}\int(m^{(1)}(x))^2\mathrm{d}x < CS^{-4}$. Then (24) allows us to conclude that $\sum_{j=1}^{J} j^2 A_{111}(j) \le Cn^{-1}\sum_{j=1}^{J}\sum_{s=0}^{\infty} j^2(1+s)^2 v_{js}^2 \le Cn_1^{-1} \le Cb_n n^{-1}$. Consider $A_{112}(j)$,

$$
\begin{aligned}
A_{112}(j) &= E\Bigg[\int \frac{(m(x)-\hat{m}(x))(\sigma(x)-\hat{\sigma}(x))}{b\sigma^2(x)} p(x)v_j(x)\mathrm{d}x \\
&\quad + \int \frac{(m(x)-\hat{m}(x))(\sigma(x)-\hat{\sigma}(x))^2}{b\sigma^2(x)\hat{\sigma}(x)} p(x)v_j(x)\mathrm{d}x\Bigg]^2 \\
&\le 2E\left[\int \frac{(m(x)-\hat{m}(x))(\sigma(x)-\hat{\sigma}(x))}{b\sigma^2(x)} p(x)v_j(x)\mathrm{d}x\right]^2 \\
&\quad + 2E\left[\int \frac{(m(x)-\hat{m}(x))(\sigma(x)-\hat{\sigma}(x))^2}{b\sigma^2(x)\hat{\sigma}(x)} p(x)v_j(x)\mathrm{d}x\right]^2 \\
&=: 2A_{1121}(j) + 2A_{1122}(j).
\end{aligned}
$$

Using Cauchy–Schwarz inequality and Lemma 1 we get

$$
\begin{aligned}
A_{1121}(j) &= E\left[\int \frac{(m(x)-\hat{m}(x))(\sigma^2(x)-\hat{\sigma}^2(x))}{b\sigma^2(x)(\sigma(x)+\hat{\sigma}(x))} p(x)v_j(x)\mathrm{d}x\right]^2 \\
&\le CE^{1/2}\left\{\int (m(x)-\hat{m}(x))^4\mathrm{d}x\right\}E^{1/2}\left\{\int \left(\sigma^2(x)-\hat{\sigma}^2(x)\right)^4\mathrm{d}x\right\}\int v_j^2(x)\mathrm{d}x \\
&\le C\ln^{10}(n)n^{-4/3}\int v_j^2(x)\mathrm{d}x.
\end{aligned}
$$

According to (26) we have $\sum_{j=1}^J j^2 \int v_j^2(x)\mathrm{d}x < C$ and then $\sum_{j=1}^J j^2 A_{1121}(j) \le Cn^{-1}$. Using $\hat{\sigma}(x) > b_n^{-1}$ and $|v_j(x)| \le \int \psi(u|x)|\sin(\pi ju)|\mathrm{d}u \le 1$, we get $A_{1122}(j) \le Cb_n^2E^{1/2}\{\int(m(x)-\hat{m}(x))^4\mathrm{d}x\}E^{1/2}\{\int(\sigma^2(x)-\hat{\sigma}^2(x))^8\mathrm{d}x\} \le C\ln^{14}(n)n^{-2}$. This implies $\sum_{j=1}^J j^2A_{1122}(j) \le Cn^{-1}$. Combining the obtained results we get $\sum_{j=1}^J j^2A_{11}(j) \le Cb_nn^{-1}$. Then, to evaluate $A_{12}(j)$, we make a preliminary calculation:

$$
\begin{aligned}
\frac{\sigma(x)-\hat{\sigma}(x)}{\hat{\sigma}(x)} &= \frac{\sigma^2(x)-\hat{\sigma}^2(x)}{(\sigma(x)+\hat{\sigma}(x))\hat{\sigma}(x)} \\
&= \frac{\sigma^2(x)-\hat{\sigma}^2(x)}{(\sigma(x)+\hat{\sigma}(x))\sigma(x)} + \frac{(\sigma^2(x)-\hat{\sigma}^2(x))(\sigma(x)-\hat{\sigma}(x))}{(\sigma(x)+\hat{\sigma}(x))\hat{\sigma}(x)\sigma(x)} \\
&= \frac{\sigma^2(x)-\hat{\sigma}^2(x)}{2\sigma^2(x)} + \frac{(\sigma^2(x)-\hat{\sigma}^2(x))(\sigma(x)-\hat{\sigma}(x))}{2\sigma^2(x)(\sigma(x)+\hat{\sigma}(x))} \\
&\quad + \frac{(\sigma^2(x)-\hat{\sigma}^2(x))(\sigma(x)-\hat{\sigma}(x))}{(\sigma(x)+\hat{\sigma}(x))\hat{\sigma}(x)\sigma(x)} \\
&= \frac{\sigma^2(x)-\hat{\sigma}^2(x)}{2\sigma^2(x)} + \frac{(\sigma^2(x)-\hat{\sigma}^2(x))^2}{2\sigma^2(x)(\sigma(x)+\hat{\sigma}(x))^2} \\
&\quad + \frac{(\sigma^2(x)-\hat{\sigma}^2(x))^2}{(\sigma(x)+\hat{\sigma}(x))^2\hat{\sigma}(x)\sigma(x)}.
\end{aligned}
$$

Using this result, the term $A_{12}(j)$ is evaluated similarly to the above-presented evaluation of $A_{11}(j)$, and we get $\sum_{j=1}^J j^2[A_{11}(j)+A_{12}(j)] \le Cb_nn^{-1}$. This, together with Lemma 3, yields

$$\sum_{j=1}^{J} j^2 E\left\{\tilde{A}_1(j)\right\} \le n_2^{-1} E\left\{H_n^2\right\} \sum_{j=1}^{J} j^2 + \sum_{j=1}^{J} j^2 (A_{11}(j) + A_{12}(j))$$
$$\le C n^{-1} \ln^5(n) n^{-2/3} \left(b_n n^{1/5}\right)^3 + C b_n n^{-1} \le C b_n n^{-1}.$$

Let us evaluate $E\{\tilde{A}_2(j)\}$. We have $E\{\tilde{A}_2(j)\} = n_2^{-2} \sum_{l,t=3n_1+1}^{n} E\{H_l^2 H_t^2 \cos(\pi j \epsilon_l) \cos(\pi j \epsilon_t)\} \le n_2^{-1} E\{H_n^4\} + |E\{H_n^2 H_{n-1}^2 \cos(\pi j \epsilon_n) \cos(\pi j \epsilon_{n-1})\}|$. Equation (26) allows us to write

$$E\left\{H_n^2 H_{n-1}^2 \cos(\pi j \epsilon_n) \cos(\pi j \epsilon_{n-1})\right\}$$
$$= E\left[\int\int H^2(x, a + bu) p(x) \psi(u|x) \cos(\pi j u) \mathrm{d}u \mathrm{d}x\right]^2$$
$$\le C j^{-4} b_n^4 E\left\{\int (m(x) - \hat{m}(x))^4 \mathrm{d}x + \int (\sigma(x) - \hat{\sigma}(x))^4 \mathrm{d}x\right\}.$$

We conclude, with the help of Lemma 1, that $\sum_{j=1}^{J} j^4 E\{\tilde{A}_2(j)\} \le C n^{-1}$.

Now we are considering $E\{\tilde{A}_3(j)\}$. Write

$$E\{\tilde{A}_3(j)\} = n_2^{-2} E\left[\sum_{l=3n_1+1}^{n} H_l^3 \sin(\pi j \epsilon_l)\right]^2$$
$$\le n_2^{-1} E\{H_n^6\} + |E\{H_n^3 H_{n-1}^3 \sin(\pi j \epsilon_n) \sin(\pi j \epsilon_{n-1})\}|$$
$$= n_2^{-1} E\{H_n^6\} + E\left[\int\int H^3(x, a + bu) p(x) \psi(u|x) \sin(\pi j u) \mathrm{d}u \mathrm{d}x\right]^2.$$

Then a simple calculation, based on using Lemma 3, implies $\sum_{j=1}^{J} j^6 E\{\tilde{A}_3(j)\} \le C \sum_{j=1}^{J} j^6 \ln^{14}(n) n^{-6/3} [n_2^{-1} + j^{-4}] \le C n^{-1}$. Similarly we get $\sum_{j=1}^{J} j^8 E\{\tilde{A}_4(j)\} \le C \sum_{j=1}^{J} j^8 \ln^{14}(n) n^{-8/3} [n_2^{-1} + j^{-4}] \le C n^{-1}$. To estimate $\tilde{A}_5(j)$ we note, using Lemma 3, that $E\{\tilde{A}_5(j)\} \le C \ln^{21}(n) n^{-10/3}$, and then $\sum_{j=1}^{J} j^{10} E\{\tilde{A}_5(j)\} \le C \ln^{21}(n) n^{-10/3} J^{11} \le C \ln^{21}(n) n_1^{-10/3} n^{11/5} b_n^{11} \le C n^{-1}$. Combining the obtained results and utilizing the plain inequality $\bar{\mu}_j^2 \le 1$ we get the following upper bound for the first sum on the right side of (33)

$$\sum_{k=1}^{K} \sum_{j \in B_k} E\left\{\bar{\mu}_k^2 \left(\hat{\theta}_j - \bar{\theta}_j\right)^2\right\} \le \sum_{j=0}^{J} E\left\{\left(\hat{\theta}_j - \bar{\theta}_j\right)^2\right\} \le C b_n n^{-1}. \qquad (36)$$

Now let us consider the second sum on the right side of (33). Choose a particular block with index k, denote $\bar{\Theta}_k := L_k^{-1} \sum_{j \in B_k} (\bar{\theta}_j^2 - n_2^{-1})$, $\hat{\Theta}_k := L_k^{-1} \sum_{j \in B_k} (\hat{\theta}_j^2 - n_2^{-1})$ and write

$$(\hat{\mu}_k - \bar{\mu}_k)^2 \sum_{j \in B_k} \hat{\theta}_j^2 = L_k \left[\frac{\hat{\Theta}_k}{\hat{\Theta}_k + n_2^{-1}} - \frac{\bar{\Theta}_k}{\bar{\Theta}_k + n_2^{-1}} \right]^2 \left(\hat{\Theta}_k + n_2^{-1} \right)$$
$$\times I \left(\hat{\Theta}_k > t_k n_2^{-1} \right) I \left(\bar{\Theta}_k > t_k n_2^{-1} \right)$$
$$+ \frac{\bar{\Theta}_k^2}{\left(\bar{\Theta}_k + n_2^{-1} \right)^2} \sum_{j \in B_k} \hat{\theta}_j^2 I \left(\hat{\Theta}_k \le t_k n_2^{-1} \right) I \left(\bar{\Theta}_k > t_k n_2^{-1} \right)$$
$$+ \frac{L_k \hat{\Theta}_k^2}{\hat{\Theta}_k + n_2^{-1}} I \left(\hat{\Theta}_k > t_k n_2^{-1} \right) I \left(\bar{\Theta}_k \le t_k n_2^{-1} \right)$$
$$=: F_1(k) + F_2(k) + F_3(k). \tag{37}$$

To evaluate $F_1(k)$ we begin with the consideration of

$$F_1^*(k) := L_k \left[\frac{\hat{\Theta}_k}{\hat{\Theta}_k + n_2^{-1}} - \frac{\bar{\Theta}_k}{\bar{\Theta}_k + n_2^{-1}} \right]^2 \left(\hat{\Theta}_k + n_2^{-1} \right)$$
$$= \frac{L_k n_2^{-2} (\hat{\Theta}_k - \bar{\Theta}_k)^2}{(\hat{\Theta}_k + n_2^{-1})(\bar{\Theta}_k + n_2^{-1})^2}.$$

Using Cauchy inequality we can write for any $c_k \ge 1$ that

$$(\hat{\Theta}_k - \bar{\Theta}_k)^2 = L_k^{-2} \left[\sum_{j \in B_k} (\hat{\theta}_j^2 - \bar{\theta}_j^2) \right]^2$$
$$\le L_k^{-2} \left[2c_k \sum_{j \in B_k} (\hat{\theta}_j - \bar{\theta}_j)^2 + c_k^{-1} \sum_{j \in B_k} \bar{\theta}_j^2 \right]^2$$
$$\le 4 L_k^{-2} c_k^2 \left[\sum_{j \in B_k} (\hat{\theta}_j - \bar{\theta}_j)^2 \right]^2 + 2 L_k^{-2} c_k^{-2} \left[\sum_{j \in B_k} \bar{\theta}_j^2 \right]^2. \tag{38}$$

Plainly $\sum_{j \in B_k} \bar{\theta}_j^2 = L_k(\bar{\Theta}_k + n_2^{-1})$, and then

$$F_1^*(k) \le 4 n_2^{-2} \frac{(c_k^2/L_k) \left[\sum_{j \in B_k} \left(\hat{\theta}_j - \bar{\theta}_j \right)^2 \right]^2}{\left(\hat{\Theta}_k + n_2^{-1} \right) \left(\bar{\Theta}_k + n_2^{-1} \right)^2} + 2 n_2^{-2} \frac{c_k^{-2} L_k}{\hat{\Theta}_k + n_2^{-1}} =: F_{11}^*(k) + F_{12}^*(k).$$

Set $c_k^2 := L_k k^{1+d}$, $0 < d < 1$ and $F_{12}(k) := F_{12}^*(k) I(\hat{\Theta}_k > t_k n_2^{-1}) I(\bar{\Theta}_k > t_k n_2^{-1})$. Then

$$\sum_{k=1}^{K} F_{12}(k) \le 2 n_2^{-1} \sum_{k=1}^{\infty} \frac{k^{-1-d} n_2^{-1} I(\hat{\Theta}_k > t_k n_2^{-1})}{\hat{\Theta}_k + n_2^{-1}} \le C n^{-1}. \tag{39}$$

Analysis of $F^*_{11}(k)$ is more involved. Denote $F_{11}(k) := F^*_{11}(k)I(\hat{\Theta}_k > t_k n_2^{-1})$ $I(\bar{\Theta}_k > t_k n_2^{-1})$ and note that

$$E\{F_{11}(k)\} \le C n_2 k^{1+d} E\left[\sum_{j\in B_k}(\hat{\theta}_j - \bar{\theta}_j)^2\right]^2. \tag{40}$$

Using (35) we can write

$$\begin{aligned}\left[\sum_{j\in B_k}(\hat{\theta}_j - \bar{\theta}_j)^2\right]^2 &\le C\left[\sum_{j\in B_k}(j^2\tilde{A}_1(j) + j^4\tilde{A}_2(j) + j^6\tilde{A}_3(j) + j^8\tilde{A}_4(j)\right.\\ &\qquad\left. + j^{10}\tilde{A}_5(j))\right]^2\\ &\le C\left[\left(\sum_{j\in B_k} j^2\tilde{A}_1(j)\right)^2 + \left(\sum_{j\in B_k} j^4\tilde{A}_j(j)\right)^2\right.\\ &\qquad + \left(\sum_{j\in B_k} j^6\tilde{A}_3(j)\right)^2 + \left(\sum_{j\in B_k} j^8\tilde{A}_4(j)\right)^2\\ &\qquad\left. + \left(\sum_{j\in B_k} j^{10}\tilde{A}_5(j)\right)^2\right]\\ &=: C[\check{A}_1(k) + \check{A}_2(k) + \check{A}_3(k) + \check{A}_4(k) + \check{A}_5(k)]. \end{aligned} \tag{41}$$

We are considering these five terms in turn. Recall that H_l was defined in (17) and write $\check{A}_1(k) = \left[\sum_{j\in B_k} j^2 n_2^{-2}(\sum_{l=3n_1+1}^n H_l \sin(\pi j \epsilon_l))^2\right]^2 = n_2^{-4}\sum_{j,i\in B_k} j^2 i^2$ $\sum_{l_1,l_2,l_3,l_4=3n_1+1}^n H_{l_1}H_{l_2}H_{l_3}H_{l_4} \times \sin(\pi j\epsilon_{l_1})\sin(\pi j\epsilon_{l_2})\sin(\pi i\epsilon_{l_3})\sin(\pi i\epsilon_{l_4})$. Then a direct calculation yields

$$\begin{aligned}E\{\check{A}_1(k)\} &\le C\left[n_2^{-3}\left(\sum_{j\in B_k} j^2\right)^2 E\{H_n^4\}\right.\\ &\quad + n_2^{-2}\sum_{j,i\in B_k} j^2 i^2\left|E\left\{H_n^3 V(X_{n-1})\int\psi(u|X_{n-1})\sin(\pi j u)\mathrm{d}u\right\}\right|\\ &\quad + n_2^{-2}\sum_{j,i\in B_k} j^2 i^2\left|E\left\{H_n^3 W(X_{n-1})\int(a+bu)\psi(u|X_{n-1})\sin(\pi j u)\mathrm{d}u\right\}\right|\\ &\quad + n_2^{-2}\sum_{j,i\in B_k} j^2 i^2 E\left\{H_n^2 H_{n-1}^2\right\}\end{aligned}$$

$$
\begin{aligned}
&+n_2^{-1}\sum_{j,i\in B_k} j^2 i^2\left|E\left\{H_n^2\int\int\psi(u|X_{n-1})\psi(v|X_{n-2})\right.\right.\\
&\qquad\qquad\times H(X_{n-1},a+bu)H(X_{n-2},a+bv)\\
&\qquad\qquad\left.\left.\times\sin(\pi ju)\sin(\pi iv)\mathrm{d}u\mathrm{d}v\right\}\right|\\
&+n_2^{-1}\sum_{j,i\in B_k} j^2 i^2\left|E\left\{H_n^2\int\int\psi(u|X_{n-1})\psi(v|X_{n-2})\right.\right.\\
&\qquad\qquad\times H(X_{n-1},a+bu)H(X_{n-2},a+bv)\\
&\qquad\qquad\left.\left.\times\sin(\pi ju)\sin(\pi jv)\mathrm{d}u\mathrm{d}v\right\}\right|\\
&+\sum_{j,i\in B_k} j^2 i^2\left|E\left\{\int\int\int\int\psi(u_n|X_n)\psi(u_{n-1}|X_{n-1})\right.\right.\\
&\qquad\qquad\times\psi(u_{n-2}|X_{n-2})\psi(u_{n-3}|X_{n-3})H(X_n,a+bu_n)\\
&\qquad\qquad\times H(X_{n-1},a+bu_{n-1})H(X_{n-2},a+bu_{n-2})\\
&\qquad\qquad\times H(X_{n-3},a+bu_{n-3})[\sin(\pi ju_n)\sin(\pi ju_{n-1})\sin(\pi iu_{n-2})\\
&\qquad\qquad\left.\left.\left.\times\sin(\pi iu_{n-3})]\mathrm{d}u_n\mathrm{d}u_{n-1}\mathrm{d}u_{n-2}\mathrm{d}u_{n-3}\right\}\right|\right]\\
&=: C[D_1(k)+D_2(k)+D_3(k)+D_4(k)+D_5(k)+D_6(k)+D_7(k)].
\end{aligned}
$$

We need to evaluate seven terms in the last expression. The first one can be evaluated with the help of Lemma 1 and the following technical statement.

Remark 5 Recall that we consider $k = 1, 2, \ldots, K$ and that the length of kth block is $L_k = k^2$. This implies the following relations: $K \le Cb_n n^{1/15}$; $j \le Ck^3$ whenever $j \in B_k$; $\sum_{j\in B_k} j^r \le Ck^{2+3r}$. Also note that if $g(u)$ is a function whose second derivative is square integrable on [0,1] and the function is vanishing on the boundary points, then (22) implies that $\sum_{j\in B_k} j^2 \max_{s\in\{0,1\}}\{[\int u^s g(u)\sin(\pi ju)\mathrm{d}u]^2\} \le Ck^{-6}$.

Using Lemma 3 and Remark 5 we get $D_1(k) \le Cn_2^{-3}(\sum_{j\in B_k} j^2)^2\ln^9(n)\, n^{-4/3} \le Cn^{-3}$. Let us consider $D_2(k)$. Recall our notation $v_j(x) = \int\psi(u|x)\sin(\pi ju)\mathrm{d}u$ and write using Remark 5 and Lemmas 3 and 4,

$$
\begin{aligned}
D_2(k) &\le n_2^{-2}\sum_{i\in B_k} i^2 E\left\{\left|H_n^3 V(X_{n-1})\right|\sum_{j\in B_k} j^2\left|v_j(X_{n-1})\right|\right\}\\
&\le n_2^{-2}\sum_{i\in B_k} i^2 E\left\{\left|H_n^3 V(X_{n-1})\right|\left[\sum_{j\in B_k} j^4 v_j^2(X_{n-1})\right]^{1/2} L_k^{1/2}\right\}\\
&\le Cn_2^{-2}\ln^9(n)n^{-4/3}L_k^{1/2}\sum_{i\in B_k} i^2 \le Cn^{-3}n^{-1/3}\ln^{10}(n)k^9 \le Cn^{-2}n^{-1/3}.
\end{aligned}
$$

The term $D_3(k)$ is estimated absolutely similarly and we get $D_3(k) \le Cn^{-2}n^{-1/3}$. To evaluate $D_4(k)$ we use Lemma 3, Remark 5 and then get $D_4(k) \le Cn_2^{-2}\ln^9(n)$

$n^{-4/3}(\sum_{i\in B_k} i^2)^2 \le Cn^{-2}n^{-4/3}\ln^9(n)k^{16} \le Cn^{-2}n^{-3/15}$. Let us consider the fifth term. Denote $v_j^*(x) := \int(a+bu)\psi(u|x)\sin(\pi ju)\mathrm{d}u$. Then $D_5(k) = n_2^{-1} \sum_{i,j\in B_k} j^2i^2|E\{H_n^2[V(X_{n-1})V(X_{n-2}) \times v_j(X_{n-1})v_i(X_{n-2}) + 2V(X_{n-1}) W(X_{n-2})v_j(X_{n-1})v_i^*(X_{n-2})+W(X_{n-1})W(X_{n-2})\times v_j^*(X_{n-1})v_i^*(X_{n-2})]\}|$. Then a simple calculation based on using Lemmas 3 and 4 yields $D_5(k) \le Cn_2^{-1}\ln^9(n) n^{-4/3}\sum_{i,j\in B_k} j^2i^2\rho_j\rho_i$, where $\{\rho_j\}$'s denote generic sequences satisfying $\sum_{j=1}^{\infty} j^4\rho_j^2 \le C$. Because $\sum_{i,j\in B_k} j^2i^2\rho_j\rho_i = \left(\sum_{j\in B_k} j^2\rho_j\right)^2 \le L_k\sum_{j\in B_k} j^4\rho_j^2 \le Ck^2$, we conclude that $D_5(k) \le Cn^{-2}n^{-3/15}\ln^{10}(n)$. Absolutely similarly, only now using the relation $\sum_{i,j\in B_k} j^2i^2\rho_j^2 = (\sum_{i\in B_k} i^2)(\sum_{j\in B_k} j^2\rho_j^2) \le Ck^8k^{-6}\sum_{j\in B_k} j^4\rho_j^2 \le Ck^2$, we establish that $D_6(k) \le Cn^{-2}n^{-3/15}\ln^{10}(n)$. Using Lemma 4, $D_7(k)$ can be evaluated as follows:

$$\begin{aligned}
D_7(k) &= \sum_{i,j\in B_k} i^2j^2E\left\{\left[\int\int p(x)H(x,a+bu)\psi(u|x)\sin(\pi ju)\mathrm{d}u\mathrm{d}x\right]^2\right.\\
&\quad\left.\times\left[\int\int p(x)H(x,a+bu)\psi(u|x)\sin(\pi iu)\mathrm{d}u\mathrm{d}x\right]^2\right\}\\
&= E\left(\sum_{j\in B_k} j^2\left[\int\int p(x)H(x,a+bu)\psi(u|x)\sin(\pi ju)\mathrm{d}u\mathrm{d}x\right]^2\right)^2\\
&\le Cn_1^{-2}k^{-6}.
\end{aligned}$$

We have evaluated all seven terms, and combining the results we get

$$E\left\{\check{A}_1(k)\right\} \le C[n^{-2}n^{-3/15}\ln^{10}(n) + n_1^{-2}k^{-6}]. \tag{42}$$

In the same way it is established that $\check{A}_2(k)$, $\check{A}_3(k)$ and $\check{A}_4(k)$ are also bounded by the right side of (42). Let us consider $\check{A}_5(k)$. Using Lemma 3 and Remark 5 we can write

$$\begin{aligned}
E\{\check{A}_5(k)\} &\le E\left[\sum_{j\in B_k} j^{10}n_2^{-2}\left(\sum_{l=3n_1+1}^{n}|H_l|^5\right)^2\right]^2 \le \sum_{j,i\in B_k} j^{10}i^{10}E\{H_n^{20}\}\\
&\le C\ln^{41}(n)n^{-20/3}\left(\sum_{j\in B_k} j^{10}\right)^2\\
&\le Cn^{-2}\ln^{41}(n)n^{-14/3}k^{64} \le Cn^{-2}n^{-1/3}.
\end{aligned}$$

Combining the obtained results in (41) we get

$$E\left[\sum_{j\in B_k}(\hat{\theta}_j - \bar{\theta}_j)^2\right]^2 \le C\left[n^{-2}n^{-3/15}\ln^{10}(n) + n_1^{-2}k^{-6}\right]. \tag{43}$$

Using this result to evaluate $E\{F_{11}(k)\}$ in (40) implies $\sum_{k=1}^{K} E\{F_{11}(k)\} \le Cn^{-1}b_n^2$, and together with (39) we get $E\left\{\sum_{k=1}^{K} F_1(k)\right\} \le Cn^{-1}b_n^2$.

Now we are considering the second term $F_2(k)$ in (37). Write

$$\begin{aligned}
E\{F_2(k)\} &= E\left\{\frac{\bar{\Theta}_k^2}{(\bar{\Theta}_k + n_2^{-1})^2} L_k(\hat{\Theta}_k + n_2^{-1}) I(\hat{\Theta}_k \le t_k n_2^{-1}) I(\bar{\Theta}_k > t_k n_2^{-1})\right\} \\
&\le Cn_2^{-1} L_k E\left\{\frac{\bar{\Theta}_k^2}{(\bar{\Theta}_k + n_2^{-1})^2}\Big[I\left(t_k n_2^{-1} < \bar{\Theta}_k \le 2t_k n_2^{-1}\right)\right. \\
&\quad \left. + I(\bar{\Theta}_k > 2t_k n_2^{-1}) I(\bar{\Theta}_k - \hat{\Theta}_k > \bar{\Theta}_k/2)\Big]\right\} \\
&\le Cn_2^{-1} L_k t_k^2 E\{I\left(t_k n_2^{-1} < \bar{\Theta}_k \le 2t_k n_2^{-1}\right)\} \\
&\quad + Cn_2^{-1} L_k E\{I(\bar{\Theta}_k - \hat{\Theta}_k > \bar{\Theta}_k/2) I(\bar{\Theta}_k > 2t_k n_2^{-1})\} \\
&=: G_1 + G_2.
\end{aligned}$$

Note that using (38) and recalling notation $c_k^2 = L_k k^{1+d}$, $0 < d < 1$ we get

$$\begin{aligned}
\left(\hat{\Theta}_k - \bar{\Theta}_k\right)^2 &\le CL_k^{-2}\left[c_k \sum_{j \in B_k} (\hat{\theta}_j - \bar{\theta}_j)^2 + c_k^{-1} \sum_{j \in B_k} \bar{\theta}_j^2\right]^2 \\
&\le CL_k^{-2} c_k^2 \left[\sum_{j \in B_k} (\hat{\theta}_j - \bar{\theta}_j)^2\right]^2 + Cc_k^{-2}\left(\bar{\Theta}_k + n_2^{-1}\right)^2. \quad (44)
\end{aligned}$$

This allows us to evaluate G_1 and G_2. Using Chebyshev inequality and (43) we get

$$\begin{aligned}
G_2 &\le Cn_2^{-1} L_k \left[\frac{c_k^2 E\left\{[\sum_{j \in B_k} (\hat{\theta}_j - \bar{\theta}_j)^2]^2\right\}}{L_k^2 t_k^2 n_2^{-2}}\right. \\
&\quad \left. + c_k^{-2} E\left\{\frac{(\bar{\Theta}_k + n_2^{-1})^2}{\bar{\Theta}_k^2} I(\bar{\Theta}_k > 2t_k n_2^{-1})\right\}\right] \\
&\le Cn_2^{-1} L_k \left[\frac{c_k^2 (n^{-2} n^{-3/15} \ln^8(n) + n_1^{-2} k^{-6})}{L_k^2 t_k^2 n_2^{-2}} + c_k^{-2} t_k^{-2}\right] \\
&\le Ck^{1+d} t_k^{-2}\left(n^{-1} n^{-3/15} \ln^{10}(n) + n_1^{-1} b_n k^{-6}\right) + Cn_2^{-1} t_k^{-2} k^{-1-d}.
\end{aligned}$$

Let us evaluate the term G_1. Denote $\Theta_k := L_k^{-1} \sum_{j \in B_k} \theta_j^2$, $\theta_j = \int f(u)\varphi_j(u)\mathrm{d}u$ and write

$$
\begin{aligned}
& E\left\{I(t_k n_2^{-1} < \bar{\Theta}_k \leq 2t_k n_2^{-1})\right\} \\
\leq & \; I((1/2)t_k n_2^{-1} < \Theta_k \leq 4t_k n_2^{-1}) \\
& + E\left\{I(\bar{\Theta}_k - \Theta_k > (1/2)t_k n_2^{-1})\right\} I\left(\Theta_k \leq (1/2)t_k n_2^{-1}\right) \\
& + E\left\{I(\Theta_k - \bar{\Theta}_k > (1/2)\Theta_k)\right\} I\left(\Theta_k > 4t_k n_2^{-1}\right) \\
\leq & \; \frac{t_k + 2}{t_k} \frac{\Theta_k}{\Theta_k + n_2^{-1}} I\left((1/2)t_k n_2^{-1} < \Theta_k \leq 4t_k n_2^{-1}\right) \\
& + \frac{CL_k^{-2} n_2^{-2} (\Theta_k + n_2^{-1})^2}{t_k^4 n_2^{-4}} I\left(\Theta_k \leq (1/2)t_k n_2^{-1}\right) \\
& + \frac{CL_k^{-2} n_2^{-2} (\Theta_k + n_2^{-1})^2}{\Theta_k^4} I\left(\Theta_k > 4t_k n_2^{-1}\right) \\
\leq & \; Ct_k^{-1} \frac{\Theta_k}{\Theta + n_2^{-1}} I\left((1/2)t_k n_2^{-1} < \Theta_k \leq 4t_k n_2^{-1}\right) + Ct_k^{-4} L_k^{-2}. \qquad (45)
\end{aligned}
$$

In the next to last inequality we used Lemma 1 from Efromovich (1985) which asserts that $E(\bar{\Theta}_k - \Theta_k)^4 \leq CL_k^{-2} n_2^{-2} (\Theta_k + n_2^{-1})^2$. Now let us remember a blockwise Wiener oracle, discussed in Sect. 2, that knows regression errors and an estimated density of errors and employs optimal shrinkage coefficients $\mu_k^* = \Theta_k/(\Theta_k + n^{-1})$, $1 \leq k < \infty$ based on Fourier coefficients of the known (to the oracle) estimated density of errors. This oracle is the benchmark for Pinsker oracle, and its MISE is proportional to $n_2^{-1} \sum_{k=1}^{\infty} L_k \Theta_k/(\Theta_k + n_2^{-1})$ (Efromovich, 1985, 1999). Hence, combining the results we get

$$
\begin{aligned}
\sum_{k=1}^{K} E\{F_2(k)\} \leq & \; Cn_2^{-1} \sum_{k=1}^{K} t_k L_k \frac{\Theta_k}{\Theta_k + n_2^{-1}} I\left((1/2)t_k n_2^{-1} < \Theta_k \leq 4t_k n_2^{-1}\right) \\
& + Cb_n^2 n^{-1} \\
\leq & \; Cn_2^{-1} b_n^2 + C \sum_{k > b_n^{2/3}} t_k L_k \frac{\Theta_k}{\Theta_k + n_2^{-1}} I\left((1/2)t_k n_2^{-1} \Theta_k \leq 4t_k n_2^{-1}\right) \\
& + Cb_n^2 n^{-1} \\
\leq & \; C \ln^{-1}(b_n) E \int (\hat{f}_P(u, \mathbf{Z}^*) - f(u))^2 du + Cb_n^2 n^{-1}.
\end{aligned}
$$

Finally, let us consider $F_3(k)$ in (37). Write, $F_3(k) \leq L_k \hat{\Theta}_k I(\hat{\Theta}_k > 2t_k n_2^{-1}) I(\bar{\Theta}_k \leq t_k n_2^{-1}) + 2t_k L_k \hat{\Theta}_k I(t_k n_2^{-1} < \hat{\Theta}_k \leq 2t_k n_2^{-1}) I(\bar{\Theta}_k \leq t_k n_2^{-1}) =: F_{31}(k) + F_{32}(k)$.

Then the relation $F_{31}(k) \le 2L_k|\hat{\Theta}_k - \bar{\Theta}_k| I(\hat{\Theta}_k - \bar{\Theta}_k > t_k n_2^{-1}) I(\bar{\Theta}_k \le t_k n_2^{-1})$, Chebyshev inequality, (43) and (44) imply

$$E\{F_{31}(k)\} \le C L_k n_1^{-2} \Big[L_k^{-2} c_k^2 (n^{-3/15} \ln^{10}(n) + k^{-6}) + c_k^{-2} \Big] / (t_k n_2^{-1})$$
$$\le C n_1^{-2} n_2 t_k^{-1} \left[k^{1+d} (n^{-3/5} \ln^{10}(n) + k^{-6}) + k^{-1-d} \right],$$

with the plain corollary $\sum_{k=1}^{K} E\{F_{31}(k)\} \le C n_1^{-2} n_2 < C b_n^2 n^{-1}$. To evaluate $F_{32}(k)$ we write

$$F_{32}(k) = 2t_k L_k \hat{\Theta}_k I \left(t_k n_2^{-1} < \hat{\Theta}_k \le 2t_k n_2^{-1} \right) I \left(\bar{\Theta}_k \le (1/2) t_k n_2^{-1} \right)$$
$$+ 2t_k L_k \hat{\Theta}_k I \left(t_k n_2^{-1} < \hat{\Theta}_k \le 2t_k n_2^{-1} \right) I \left((1/2) t_k n_2^{-1} < \bar{\Theta}_k < t_k n_2^{-1} \right)$$
$$=: F_{321}(k) + F_{322}(k).$$

The term $F_{321}(k)$ is evaluated similarly to $F_{31}(k)$, and then $\sum_{k=1}^{K} E\{F_{321}(k)\} < C b_n^2 n^{-1}$. The term $F_{322}(k)$ can be estimated as follows. First, we note that $\hat{\Theta}_k I(t_k n_2^{-1} < \hat{\Theta}_k \le 2t_k n_2^{-1}) \le 2t_k n_2^{-1}$. Second, we use (45) and then a simple algebra implies $\sum_{k=1}^{K} E\{F_{322}(k)\} \le C \ln^{-1}(b_n) E \int (\hat{f}_P(u, \mathbf{Z}^*) - f(u))^2 \mathrm{d}u + C b_n^2 n^{-1}$. Combining the results we get $\sum_{k=1}^{K} E\{F_3(k)\} \le C \ln^{-1}(b_n) E \int (\hat{f}_P(u, \mathbf{Z}^*) - f(u))^2 \mathrm{d}u + C b_n^2 n^{-1}$. Using the obtained results in (37) we get

$$E \left\{ \sum_{k=1}^{K} (\hat{\mu}_k - \bar{\mu}_k)^2 \sum_{j \in B_k} \hat{\theta}_j^2 \right\} \le C \ln^{-1}(b_n) E \left\{ \int (\hat{f}_P(u, \mathbf{Z}^*) - f(u)^2 \mathrm{d}u \right\} + C b_n^2 n^{-1}.$$

Using this and (36) in (33) proves Theorem 1. □

5 Proof of Lemmas 1–4

Proof of Lemma 1 We begin with verification of (12). For $n > n_0$, where n_0 depends only on $\min_{x \in [0,1]} p(x)$, we can write $|p(x) - \hat{p}(x)| \le |p(x) - \tilde{p}(x)|$. Here $\tilde{p}(x) := 1 + \sum_{s=1}^{S} \tilde{\pi}_s \varphi_s(x)$ and $\tilde{\pi}_s = n_1^{-1} \sum_{l=1}^{n_1} \varphi_s(X_l)$. Thus it suffices to evaluate the risk of the nontruncated $\tilde{p}$ (the original series design density estimate). Write for those n:

$$E \left\{ (p(x) - \hat{p}(x))^{2k} \right\} \le E \left\{ (p(x) - \tilde{p}(x))^{2k} \right\}$$
$$= E \left[\sum_{s=1}^{S} (\pi_s - \tilde{\pi}_s) \varphi_s(x) + \sum_{s > S} \pi_s \varphi_s(x) \right]^{2k}$$
$$\le C_k E \left[\sum_{s=1}^{S} (\pi_s - \tilde{\pi}_s) \varphi_s(x) \right]^{2k} + \left[\sum_{s > S} \pi_s \varphi_s(x) \right]^{2k}.$$

The second term is at most $C_k \ln^{2k}(n) S^{-2k}$ due to the assumption about the bounded derivative of the design density; remember Remark 1. To evaluate the first term we write $\pi_s - \tilde{\pi}_s = \pi_s - n_1^{-1} \sum_{l=1}^{n_1} \varphi_s(X_l) = n_1^{-1} \sum_{l=1}^{n_1} (\pi_s - \varphi_s(X_l))$. Using this expression we get

$$E\left[\sum_{s=1}^{S}(\pi_s - \tilde{\pi}_s)\varphi_s(x)\right]^{2k} = n_1^{-2k} E\left[\sum_{l=1}^{n_1}\sum_{s=1}^{S}(\pi_s - \varphi_s(X_l))\varphi_s(x)\right]^{2k}$$
$$= n_1^{-2k} \sum_{l_1,\ldots,l_{2k}=1}^{n_1} E\left\{\prod_{i=1}^{2k}\sum_{s=1}^{S}(\pi_s - \varphi_s(X_{l_i}))\varphi_s(x)\right\}.$$

Let us make several comments. First, recall that predictors are independent and $E\varphi_s(X_l) = \pi_s$. Thus the last expectation is zero whenever at least one index l_i has no match. Second, following Efromovich (1999, s.2.4), introduce

$$\sum_{s=0}^{S}\varphi_s(u) = c_1 + c_2 D_S(u), \quad D_S(u) := \frac{\sin(\pi(2S+1)u/2)}{\sin(\pi u/2)}, \tag{46}$$

where c_1 and c_2 are some absolute constants and $D_S(u)$ is the Dirichlet kernel which has the following familiar property (note that the integral is improper)

$$|D_S(u)| \le 2S + 1, \quad \int_{-\infty}^{\infty} |D_S(u)|\mathrm{d}u \le C\ln(S). \tag{47}$$

Then using (46) and (47) together with $\max_{x,S} |\sum_{s=1}^{S} \pi_s\varphi_s(x)| < C$ (the latter is based on Assumption A and Remark 1), we get $E[\sum_{s=1}^{S}(\pi_s - \varphi_s(X_l))\varphi_s(x)]^2 \le C + ED_S^2(X_1 - x) + ED_S^2(X_1 + x) \le C\ln(n)S \le C\ln(n)n^{1/3}$, and $|\sum_{s=1}^{S}(\pi_s - \varphi_s(X_l))\varphi_s(x)| \le CS$. Using these comments together with a simple calculation implies

$$\max_x E\left[\sum_{s=1}^{S}(\pi_s - \tilde{\pi}_s)\varphi_s(x)\right]^{2k} \le C_k \ln^{2k}(n) n^{-2k/3}. \tag{48}$$

Combining the results we establish (12).

Now we are verifying (13). According to Butzer and Nessel (1971, s.2.4) and Remark 1, under Assumption A we can write $m(x) = \sum_{s=0}^{\infty} \kappa_s\varphi_s(x)$, $x \in [0, 1]$ with $\kappa_s := \int_0^1 m(x)\varphi_s(x)\mathrm{d}x$ and then

$$E(m(x) - \hat{m}(x))^{2k} = E\left[\sum_{s=0}^{S}(\kappa_s - \hat{\kappa}_s)\varphi_s(x) + \sum_{s>S}\kappa_s\varphi_s(x)\right]^{2k}$$
$$\le C_k E\left[\sum_{s=0}^{S}(\kappa_s - \hat{\kappa}_s)\varphi_s(x)\right]^{2k}$$
$$+ C_k\left[\sum_{s>S}\kappa_s\varphi_s(x)\right]^{2k}, \tag{49}$$

where the second term decreases not slower than $C_k \ln^{2k}(S)S^{-2k}$ uniformly over $x \in [0, 1]$. Now let us consider the first term on the right side of (49). Write

$$\hat{\kappa}_s = n_1^{-1} \sum_{l=n_1+1}^{2n_1} \frac{Y_l \varphi_s(X_l)}{\hat{p}(X_l)} = n_1^{-1} \sum_{l=n_1+1}^{2n_1} \frac{m(X_l)\varphi_s(X_l)}{\hat{p}(X_l)}$$
$$+ n_1^{-1} \sum_{l=n_1+1}^{2n_1} \xi_l \frac{\sigma(X_l)\varphi_s(X_l)}{\hat{p}(X_l)}$$
$$= \kappa_s + n_1^{-1} \sum_{l=n_1+1}^{2n_1} \left(\frac{m(X_l)\varphi_s(X_l)}{\hat{p}(X_l)} - \kappa_s \right) + n_1^{-1} \sum_{l=n_1+1}^{2n_1} \xi_l \frac{\sigma(X_l)\varphi_s(X_l)}{\hat{p}(X_l)}.$$

Using this we get

$$\begin{aligned}
\sum_{s=0}^{S} (\hat{\kappa}_s - \kappa_s)\varphi_s(x) &= n_1^{-1} \sum_{l=n_1+1}^{2n_1} \sum_{s=0}^{S} \left(\frac{m(X_l)\varphi_s(X_l)}{\hat{p}(X_l)} - \kappa_s \right) \varphi_s(x) \\
&\quad + n_1^{-1} \sum_{l=n_1+1}^{2n_1} \xi_l \frac{\sigma(X_l)}{\hat{p}(X_l)} \sum_{s=0}^{S} \varphi_s(X_l)\varphi_s(x) \\
&= n_1^{-1} \sum_{l=n_1+1}^{2n_1} \sum_{s=0}^{S} \left(\frac{m(X_l)\varphi_s(X_l)}{p(X_l)} - \kappa_s \right) \varphi_s(x) \\
&\quad + n_1^{-1} \sum_{l=n_1+1}^{2n_1} \sum_{s=0}^{S} \frac{m(X_l)\varphi_s(X_l)\varphi_s(x)(p(X_l) - \hat{p}(X_l))}{p(X_l)\hat{p}(X_l)} \\
&\quad + n_1^{-1} \sum_{l=n_1+1}^{2n_1} \xi_l \frac{\sigma(X_l)}{\hat{p}(X_l)} \sum_{s=0}^{S} \varphi_s(X_l)\varphi_s(x) \\
&=: A_1(x) + A_2(x) + A_3(x). \qquad (50)
\end{aligned}$$

Consider these three terms in turn. Denote $\tilde{a}_l(x) := \sum_{s=0}^{S} (m(X_l)\varphi_s(X_l)p^{-1}(X_l) - \kappa_s)\varphi_s(x)$, and then $E\{A_1^{2k}(x)\} = n_1^{-2k} \sum_{l_1,\dots,l_{2k}=n_1+1}^{2n_1} E\{\prod_{i=1}^{2k} \tilde{a}_{l_i}(x)\}$. Recall that predictors are independent and $E\{\tilde{a}_l(x)\} = 0$. Then a term $E\{\prod_{i=1}^{2k} \tilde{a}_{l_i}(x)\}$ is zero whenever values of the indexes $(l_1, l_2, \dots, l_{2k})$ do not appropriately match. Further, using (46) and (47) we can write $|\tilde{a}_l(x)| = |m(X_l)p^{-1}(X_l) \sum_{s=0}^{S} \varphi_s(X_l) \varphi_s(x) - \sum_{s=0}^{S} \kappa_s \varphi_s(x)| \leq C|c_1 + (c_2/2)D_S(X_l + x) + (c_2/2)D_S(X_l - x)| + |\sum_{s=0}^{S} \kappa_s \varphi_s(x)|$. Note that $|\sum_{s=0}^{S} \kappa_s \varphi_s(x)| < C$ uniformly over x and S because these are partial Fourier sums of the function $m(x)$ which has a bounded derivative; recall Remark 1.

Using these results together with a straightforward calculation we get $E\{A_1^{2k}(x)\} \leq Cn_1^{-k} E\left\{\prod_{l=1}^{k} [\tilde{a}_l(x)]^2\right\} + Cn_1^{-k-1} \ln^{k-1}(S)S^{k+1} \leq C \ln^k(n) n_1^{-k} S^k \leq C \ln^{k+1}(n)n^{-2k/3}$. Now let us consider $A_2(x)$ in (50). Write

$$
\begin{aligned}
A_2(x) &= n_1^{-1} \sum_{l=n_1+1}^{2n_1} \frac{m(X_l)(p(X_l)-\hat{p}(X_l))}{p(X_l)\hat{p}(X_l)} \sum_{s=0}^{S} \varphi_s(X_l)\varphi_s(x) \\
&\le Cb_n n_1^{-1} \sum_{l=n_1+1}^{2n_1} \Big| p(X_l) - \hat{p}(X_l) \Big| \\
&\quad \times \Big| c_1 + (c_2/2)D_S(X_l - x) + (c_2/2)D_S(X_l + x) \Big|.
\end{aligned}
$$

Then

$$
\begin{aligned}
E\{A_2^{2k}(x)\} &\le Cb_n^{2k} n_1^{-2k} E \left\{ \sum_{l_1,\ldots,l_{2k}=n_1+1}^{2n_1} \prod_{i=1}^{2k} \Big| p(X_{l_i}) - \hat{p}(X_{l_i}) \Big| \right. \\
&\qquad \left. \times \Big| C + D_S(X_{l_i} - x) + D_S(X_{l_i} + x) \Big| \right\} \\
&= Cb_n^{2k} n_1^{-2k} E \left\{ \sum_{l_1,\ldots,l_{2k}=n_1+1}^{2n_1} E \left\{ \prod_{i=1}^{2k} |p(X_{l_i}) - \hat{p}(X_{l_i})| \Big| X_{l_1}, \ldots, X_{l_{2k}} \right\} \right. \\
&\qquad \left. \times \prod_{j=1}^{2k} [1 + |D_S(X_{l_j} - x)| + |D_S(X_{l_j} + x)|] \right\} \\
&\le Cb_n^{2k} n_1^{-2k} E \\
&\quad \times \left\{ \sum_{l_1,\ldots,l_{2k}=n_1+1}^{2n_1} \prod_{i=1}^{2k} E^{1/(2k)} \left\{ |p(X_{l_i}) - \hat{p}(X_{l_i})|^{2k} \Big| X_{l_1}, \ldots, X_{l_{2k}} \right\} \right. \\
&\quad \left. \times \prod_{j=1}^{2k} [1 + |D_S(X_{l_j} - x)| + |D_S(X_{l_j} + x)|] \right\}.
\end{aligned}
$$

Using (12), (46) and (47) we get

$$
\begin{aligned}
E\{A_2^{2k}(x)\} &\le Cb_n^{2k} n_1^{-2k} \ln^{2k}(n) n_1^{-2k/3} E \\
&\quad \times \left\{ \sum_{l_1,\ldots,l_{2k}=n_1+1}^{2n_1} \prod_{j=1}^{2k} \left[1 + |D_S(X_{l_j} - x)| + |D_S(X_{l_j} + x)| \right] \right\} \\
&\le Cb_n^{2k} n_1^{-2k} \ln^{2k}(n) n_1^{-2k/3} n_1^{2k} \ln^{2k}(S) \le Cb_n^{2k} \ln^{4k}(n) n_1^{-2k/3}.
\end{aligned}
$$

Denote $\hat{a}_l(x) := [\sigma(X_l)/\hat{p}(X_l)] \sum_{s=0}^{S} \varphi_s(X_l)\varphi_s(x)$ and get

$$
E\{A_3^{2k}(x)\} = n_1^{-2k} \sum_{l_1,\ldots,l_{2k}=n_1+1}^{2n_1} E \left\{ E \left\{ \prod_{i=1}^{2k} \xi_{l_i} | X_{n_1+1}, \ldots, X_{2n_1} \right\} \prod_{j=1}^{2k} \hat{a}_{l_j}(x) \right\}. \tag{51}
$$

Recall that pairs $(X_1, \xi_1), \ldots, (X_n, \xi_n)$ are independent and $E\{\xi_l|X_l\} = 0$. Also note that $|\hat{a}_l(x)| \leq Cb_n|c_1 + (c_2/2)D_S(X_l - x) + (c_2/2)D_S(X_l + x)|$. These remarks, together with (46) and (47), yield

$$E\{A_3^{2k}(x)\} \leq Cn_1^{-k} b_n^{2k} \ln^k(n) S^k \leq C \ln^{k+1}(n) n^{-2k/3}. \tag{52}$$

Combining the obtained results in (50) with (49) proves (13).

Let us check (14). For all $n > n_0'$, where n_0' depends only on $\min_{x\in[0,1]} \sigma^2(x)$, we have $|\sigma^2(x) - \hat{\sigma}^2(x)| \leq |\sigma^2(x) - \tilde{\sigma}^2(x)|$. Thus it suffices to verify (14) for the nontruncated scale estimate $\tilde{\sigma}^2(x)$. To make the proof shorter, we shall use Remark 2 and steps of the above-presented proof of (13) for the regression function $m(x)$; note that according to (28) the only difference between the two models is that the role of $\sigma(X_l)\xi_l$ is now played by $\hat{\mu}(X_l) + \hat{v}(X_l)\eta_l$. Here the relation (49) holds for the considered "scale-regression" model, and the second term, in its right side, is again bounded by $C \ln^{2k}(S) S^{-2k}$ because $\sigma^2(x)$ has a bounded derivative. The relation (50) also holds with $\sigma(X_l)\xi_l$ replaced by $\hat{\mu}(X_l) + \hat{v}(X_l)\eta_l$. Then the only difference in the proof below (50), which should be explained, is the analysis of the third term $A_3(x)$. Let us express this term using our new notation, namely for the scale function case the third term becomes $A_3(x) := n_1^{-1} \sum_{l=2n_1+1}^{3n_1} [\hat{\mu}(X_l) + \hat{v}(X_l)\eta_l]\hat{p}^{-1}(X_l) \sum_{s=0}^{S} \varphi_s(X_l)\varphi_s(x)$. Write

$$A_3^{2k}(x) = n_1^{-2k} \sum_{l_1,\ldots,l_{2k}=2n_1+1}^{3n_1} \prod_{i=1}^{2k} \left(\frac{\hat{\mu}(X_{l_i}) + \hat{v}(X_{l_i})\eta_{l_i}}{\hat{p}(X_{l_i})} \sum_{s=0}^{S} \varphi_s(X_{l_i})\varphi_s(x) \right).$$

For any set G of $2k$ elements chosen with replacement from $\{2n_1 + 1, \ldots, 3n_1\}$, if at least one of its elements is not matched, $E\left\{ \prod_{i\in G} \eta_{l_i} | X_{2n_1+1}, \ldots, X_{3n_1}, (X, Y)_1^{2n_1} \right\} = 0$. Using this together with (29) and (30), and then following along (51) and (52) with the use of (13) we get $\max_{x\in[0,1]} EA_3^{2k}(x) \leq C_k \ln^{k+1}(n) n^{-2k/3}$. Combining obtained results establishes (14). □

Proof of Lemma 2 We begin with the evaluation of $E(\kappa_s - \hat{\kappa}_s)^4$; note that this would be a simple task if the design density p was known or if the density was assumed to be smoother. Write

$$\begin{aligned}
\kappa_s - \hat{\kappa}_s &= n_1^{-1} \sum_{l=n_1+1}^{2n_1} (\kappa_s - Y_l\varphi_s(X_l)\hat{p}^{-1}(X_l)) \\
&= n_1^{-1} \sum_{l=n_1+1}^{2n_1} (\kappa_s - Y_l\varphi_s(X_l)p^{-1}(X_l)) \\
&\quad + n_1^{-1} \sum_{l=n_1+1}^{2n_1} Y_l\varphi_s(X_l) \frac{\hat{p}(X_l) - p(X_l)}{\hat{p}(X_l)p(X_l)}.
\end{aligned}$$

This yields

$$(\kappa_s - \hat{\kappa}_s)^4 \le C \left[n_1^{-1} \sum_{l=n_1+1}^{2n_1} (\kappa_s - Y_l \varphi_s(X_l) p^{-1}(X_l)) \right]^4$$
$$+ C \left[n_1^{-1} \sum_{l=n_1+1}^{2n_1} \sigma(X_l) \xi_l \varphi_s(X_l) \frac{\hat{p}(X_l) - p(X_l)}{\hat{p}(X_l) p(X_l)} \right]^4$$
$$+ C \left[n_1^{-1} \sum_{l=n_1+1}^{2n_1} m(X_l) \varphi_s(X_l) \frac{\hat{p}(X_l) - p(X_l)}{\hat{p}(X_l) p(X_l)} \right]^4 =: A_1 + A_2 + A_3.$$

To evaluate A_1 we recall that pairs $(X_l, Y_l), l = n_1 + 1, \ldots, X_{2n_1}$ are independent and $\kappa_s = E\{Y_l \varphi_j(X_l) p^{-1}(X_l)\}$. The second term is considered similarly because ξ_l's are independent and zero mean gives the predictor. This implies $E\{A_1 + A_2\} \le C n_1^{-2}$. Now let us consider A_3. Denote $g_s(x) := m(x)\varphi_s(x)$ and write

$$\begin{aligned} &n_1^{-1} \sum_{l=n_1+1}^{2n_1} g_s(X_l) \frac{\hat{p}(X_l) - p(X_l)}{\hat{p}(X_l) p(X_l)} \\ &= n_1^{-1} \sum_{l=n_1+1}^{2n_1} g_s(X_l) \frac{\hat{p}(X_l) - p(X_l)}{p^2(X_l)} - n_1^{-1} \sum_{l=n_1+1}^{2n_1} g_s(X_l) \frac{(\hat{p}(X_l) - p(X_l))^2}{\hat{p}(X_l) p^2(X_l)} \\ &=: G_1 - G_2. \end{aligned} \tag{53}$$

Note that $g_s(x)$ is bounded on [0, 1]. Then using (12), together with a simple calculation, implies $EG_2^4 \le C n_1^{-4} b_n^4 n_1^4 \max_{x \in [0,1]} E(\hat{p}(x) - p(x))^8 \le C b_n^4 \ln^8(n)$ $n_1^{-8/3} \le C n_1^{-2}$. Evaluation of G_1 is more involved. First of all, we can write

$$\begin{aligned} G_1 &= n_1^{-1} \sum_{l=n_1+1}^{2n_1} g_s(X_l) \frac{\hat{p}(X_l) - \tilde{p}(X_l)}{p^2(X_l)} + n_1^{-1} \sum_{l=n_1+1}^{2n_1} g_s(X_l) \frac{\tilde{p}(X_l) - p(X_l)}{p^2(X_l)} \\ &=: G_{11} + G_{12}, \end{aligned}$$

where $\tilde{p}(x) = 1 + \sum_{j=1}^{S} \tilde{\pi} \varphi_j(x), \tilde{\pi}_j = n_1^{-1} \sum_{l=1}^{n_1} \varphi_j(X_l)$. G_{11} is straightforwardly evaluated using Chebyshev's approach and (12); recall that (12) was proved for $\tilde{p}(x)$ as well. Then denote $\delta := \min_{x \in [0,1]} p(x)/2$ and write for all sufficiently large n such that $b_n^{-1} \le \delta$,

$$\begin{aligned} E\{G_{11}^4\} &\le CE \left\{ \int (\hat{p}(x) - \tilde{p}(x))^4 \mathrm{d}x \right\} \\ &\le CE \int_{\{x\colon |\tilde{p}(x) - p(x)| > \delta\}} (\tilde{p}(x) - p(x))^4 \mathrm{d}x \\ &\le C\delta^{-4} E \left\{ \int (\tilde{p}(x) - p(x))^8 \mathrm{d}x \right\} \le C n_1^{-7/3}. \end{aligned}$$

To evaluate G_{12} we write $G_{12} = n_1^{-1} \sum_{l=n_1+1}^{2n_1} g_s(X_l) p^{-2}(X_l) [\sum_{j=1}^{S} (\tilde{\pi}_j - \pi_j)\varphi_j (X_l) + \sum_{j>S} \pi_j \varphi_j(X_l)]$. This yields

$$E\{G_{12}^4\} \leq 4n_1^{-4} E \left[\sum_{l=n_1+1}^{2n_1} g_s(X_l) p^{-2}(X_l) \sum_{j=0}^{S} (\tilde{\pi}_j - \pi_j)\varphi_j(X_l) \right]^4 + 4n_1^{-4} E \left[\sum_{l=n_1+1}^{2n_1} g_s(X_l) p^{-2}(X_l) \sum_{j>S} \pi_j \varphi_j(X_l) \right]^4 =: G_{121} + G_{122}.$$

To evaluate G_{121} we write

$$G_{121} = 4n_1^{-4} E \left\{ \sum_{l_1,l_2,l_3,l_4=n_1+1}^{2n_1} \prod_{i=1}^{4} \left(g_s(X_{l_i}) p^{-2}(X_{l_i}) \sum_{j=1}^{S} (\tilde{\pi}_j - \pi_j)\varphi_j(X_{l_i}) \right) \right\}. \tag{54}$$

Let us begin the analysis by considering a particular case where all $\{l_1, l_2, l_3, l_4\}$ are different. Denote $r_{sj} := \int g_s(x) p^{-1}(x) \varphi_j(x) \mathrm{d}x$ and write for this case $E\left\{\prod_{i=1}^{4} g_s(X_{l_i}) p^{-2}(X_{l_i}) \sum_{j=1}^{S} (\tilde{\pi}_j - \pi_j)\varphi_j(X_{l_i})\right\} = E\Big[\int g_s(x) p^{-1}(x) \sum_{j=1}^{S} (\tilde{\pi}_j - \pi_j)\varphi_j(x) dx\Big]^4 E\Big[\sum_{j=1}^{S} (\tilde{\pi} - \pi) r_{sj}\Big]^4$. Note that $\tilde{\pi}_j - \pi_j = n^{-1} \sum_{t=1}^{n_1} (\varphi_j(X_t) - \pi_j)$, and then

$$\begin{aligned} E\left[\sum_{j=1}^{S} (\tilde{\pi}_j - \pi_j) r_{sj}\right]^4 &= n_1^{-4} E\left[\sum_{t=1}^{n_1} \sum_{j=1}^{S} (\varphi_j(X_t) - \pi_j) r_{sj}\right]^4 \\ &= n_1^{-4} E\left[\sum_{t=1}^{n_1} q_{sS}(X_t)\right]^4 \\ &= n_1^{-4} \sum_{t_1,t_2,t_3,t_4=1}^{n_1} E\left\{\prod_{i=1}^{4} q_{sS}(X_{t_i})\right\}. \end{aligned}$$

Here $q_{sS}(x) := \sum_{j=1}^{S} (\varphi_j(x) - \pi_j) r_{sj}$. Note that $E\{q_{sS}(X)\} = 0$ because $E\{\varphi_j(X)\} = \pi_j$. Also, we can show that $\max_{x,s,S} |q_{sS}(x)| \leq C < \infty$. To check this inequality we first note that $\sum_{j=1}^{\infty} |\pi_j| < \infty$ due to (20) proved in Efromovich (2001). Second, let $m(x)p^{-1}(x) =: \sum_{j=0}^{\infty} \lambda_j \varphi_j(x)$, then (20) implies that $\sum_{j=0}^{\infty} |\lambda_j| < C < \infty$. Then we get $m(x)p^{-1}(x)\varphi_s(x) = 2^{-1/2} \sum_{j=0}^{\infty} \lambda_j [\varphi_{j+s}(x) + \varphi_{j-s}(x)]$ and thus

$$r_{sj} = 2^{-1/2} [\lambda_{j-s} + \lambda_{j+s}]. \tag{55}$$

The latter implies $\max_s \sum_{j=0}^{\infty} |r_{sj}| < \infty$, and thus we get the verified inequality for $q_{sS}(x)$.

Using the obtained results we conclude that

$$E\left[\sum_{j=1}^{\infty}(\tilde{\pi}_j-\pi_j)r_{sj}\right]^4 \le Cn_1^{-2}. \tag{56}$$

In a similar way we can consider a case where $l_1 = l_2$ and $\{l_2, l_3, l_4\}$ are different in (54). Write $E\left\{\prod_{i=1}^4 g_s(X_{l_i})p^{-2}(X_{l_i})\sum_{j=1}^S(\tilde{\pi}_j-\pi_j)\varphi_j(X_{l_i})\right\} = E\left\{\int g_s^2(X_l) p^{-3}(x)[\sum_{j=1}^S(\tilde{\pi}_j-\pi_j)\varphi_j(x)]^2\mathrm{d}x\,[\sum_{j=1}^S(\tilde{\pi}_j-\pi_j)r_{sj}]^2\right\} \le CE^{1/2}\{\int\left[\sum_{j=1}^S (\tilde{\pi}_j-\pi_j)\varphi_j(x)\right]^4\mathrm{d}x\}E^{1/2}\left[\sum_{j=1}^S(\tilde{\pi}_j-\pi_j)r_{sj}\right]^4 \le C\ln^2(n)n_1^{-5/3}$. In the last inequality we used (48) and (56). For the analysis of all other cases it suffices to utilize a rough inequality based on (48),

$$\max_{l_1,l_2,l_3,l_4}\left|E\left\{\prod_{i=1}^4 g_s(X_{l_i})p^{-2}(X_{l_i})\sum_{j=1}^S(\tilde{\pi}_j-\pi_j)\varphi_j(X_{l_i})\right\}\right| \le C\ln^4(n)n_1^{-4/3}.$$

Combining the results and using a simple calculation we establish that $E\{G_{121}\} \le Cn_1^{-2}$.

Now we are considering G_{122}. Write $G_{122} = 4n_1^{-4}E\{\sum_{l_1,l_2,l_3,l_4=n_1+1}^{2n_1}\prod_{i=1}^4 g_s (X_{l_i})p^{-2}(X_{l_i})\times\sum_{j>S}\pi_j\varphi_j(X_{l_i})\}$. Let us again begin with a particular case where all four indexes $\{l_1, l_2, l_3, l_4\}$ are different. Recalling notation $r_{sj} = \int g_s(x)p^{-1}(x) \varphi_j(x)\mathrm{d}x$ and using Cauchy–Schwarz inequality we get $E\{\prod_{i=1}^4 g_s(X_{l_i})p^{-2}(X_{l_i}) \sum_{j>S}\pi_j\varphi_j(X_{l_i})\} = [\sum_{j>S}\pi_j r_{sj}]^4 \le [\sum_{j>S}\pi_j^2\sum_{j>S}r_{sj}^2]^2$. According to Proposition 1 of Efromovich (2001), we have $\sum_{j>S}\pi_j^2 \le CS^{-2}$ and also, due to (55), $\sum_{j>S}r_{sj}^2 \le \sum_{j>S}(\lambda_{j-s}^2+\lambda_{j+s}^2) \le C[(1+S-s)^{-2}+S^{-2}] \le C(1+S-s)^{-2}$. For a particular case where $l_1 = l_2$ and $\{l_2, l_3, l_4\}$ are different we get

$$\begin{aligned}
&E\left\{\prod_{i=1}^4 g_s(X_{l_i})p^{-2}(X_{l_i})\sum_{j>S}\pi_j\varphi_j(X_{l_i})\right\}\\
&= \left(\int g_s^2(x)p^{-3}(x)\left[\sum_{j>S}\pi_j\varphi_j(x)\right]^2\mathrm{d}x\right)\\
&\quad\times\left[\int g_s(x)p^{-1}(x)\left[\sum_{j>S}\pi_j\varphi_j(x)\right]\mathrm{d}x\right]^2\\
&\le C\int\left[\sum_{j>S}\pi_j\varphi_j(x)\right]^2\mathrm{d}x\left[\int\left|\sum_{j>S}\pi_j\varphi_j(x)\right|\mathrm{d}x\right]^2 \le C\ln^2(n)S^{-4}.
\end{aligned}$$

These relations together with $\max_{x,s}|g_s(x)p^{-2}(x)\sum_{j>S}\pi_j\varphi_j(x)| \le C\ln(n) n_1^{-1/3}$ yield $E\{G_{122}\} \le C[n_1^{-2} + S^{-4}(1+S-s)^{-4}]$. Combining the results we verify (16) for $\kappa_s - \hat{\kappa}_s$. Let us check (16) for its second term. Recall that the nontruncated squared–scale estimate $\tilde{\sigma}^2(x)$ is defined below line (9), and denote its Fourier coefficient as $\tilde{\nu}_s := \int \tilde{\sigma}^2(x)\varphi_s(x)dx$. Then $(\nu_s - \hat{\nu}_s)^4 \le 4(\nu_s - \tilde{\nu}_s)^4 + 4(\tilde{\nu}_s - \hat{\nu}_s)^4 =: 4F_1 + 4F_2$. Let us begin with the analysis of F_2. Set $\delta := \min_{x\in[0,1]} \sigma^2(x)/2$, and from now on we are considering only sufficiently large n such that $b_n^{-1} \le \delta$ and $b_n - \max_{x\in[0,1]} \sigma^2(x) \ge \delta$. Then we can write

$$\begin{aligned}
E\{F_2^4\} &= E\left[\int (\tilde{\sigma}^2(x) - \hat{\sigma}^2(x))\varphi_s(x)\mathrm{d}x\right]^4 \\
&\le 4E\int (\tilde{\sigma}^2(x) - \hat{\sigma}^2(x))^4 \mathrm{d}x \\
&\le 4E \int_{\{x:\, |\tilde{\sigma}^2(x) - \sigma^2(x)| > \delta\}} (\tilde{\sigma}^2(x) - \sigma^2(x))^4 \mathrm{d}x \\
&\le 4\delta^{-4} E\int (\tilde{\sigma}^2(x) - \sigma^2(x))^8 \mathrm{d}x \le Cn^{-7/3}.
\end{aligned}$$

In the last inequality we used (14) together with the remark that (14) was proved for $\tilde{\sigma}^2(x)$ as well as for $\hat{\sigma}^2(x)$. Further,

$$\begin{aligned}
E\{F_1^4\} &= n_1^{-4} E\left[\sum_{l=2n_1+1}^{3n_1} (\nu_s - Y_l^* \varphi_s(X_l)\hat{p}^{-1}(X_l))\right]^4 \\
&\le 4n_1^{-4} E\left[\sum_{l=2n_1+1}^{3n_1} (\nu_s - Y_l^* \varphi_s(X_l) p^{-1}(X_l))\right]^4 \\
&\quad + 4n_1^{-4} E\left[\sum_{l=2n_1+1}^{3n_1} Y_l^* \varphi_s(X_l)(\hat{p}^{-1}(X_l) - p^{-1}(X_l))\right]^4 \\
&\le 4n_1^{-4} E\left[\sum_{l=2n_1+1}^{3n_1} \big(\nu_s - Y_l^* \varphi_s(X_l) p^{-1}(X_l)\big)\right]^4 \\
&\quad + 16n_1^{4} E\left[\sum_{l=2n_1+1}^{3n_1} (\hat{\mu}(X_l) + \hat{v}(X_l)\eta_l)\varphi_s(X_l)(\hat{p}^{-1}(X_l) - p^{-1}(X_l))\right]^4 \\
&\quad + 16n_1^{-4} E\left[\sum_{l=2n_1+1}^{3n_1} \sigma^2(X_l)\varphi_s(X_l)(\hat{p}^{-1}(X_l) - p^{-1}(X_l))\right]^4 \\
&=: F_{11} + F_{12} + F_{13}.
\end{aligned}$$

In the three terms are analogs of the above-evaluated $E\{A_1\}$, $E\{A_1\}$ and $E\{A_3\}$, then

$$F_{11} \le Cn_1^{-4} E\left[\sum_{l=2n_1+1}^{3n_1} (\nu_s - \sigma^2(X_l)\varphi_s(X_l)p^{-1}(X_l))\right]$$
$$+ Cn_1^{-4}\left[\sum_{l=2n_1+1}^{3n_1} (m(X_l) - \hat{m}(X_l))^2 \varphi_s(X_l)p^{-1}(X_l)\right]^4$$
$$+ Cn_1^{-4} E\left[\sum_{l=2n_1+1}^{3n_1} \hat{v}(X_l)\eta_l \varphi_s(X_l)p^{-1}(X_l)\right]^4 =: F_{111} + F_{112} + F_{113}.$$

Note that $\nu_s = E\{\sigma^2(X_l)\varphi_s(X_l)p^{-1}(X_l)\}$ according to (15), and that the functions involved are bounded; this implies that $F_{111} \le Cn_1^{-2}$. Using (13) we get $F_{112} \le Cn_1^{-7/3}$. By recalling that $E\{\eta_l|X_l, (X, Y)_1^{2n_1}\} = 0, l = 2n_1 + 1, \dots, 3n_1$ we get $F_{113} \le Cn_1^{-2}$.

To evaluate F_{12} we are using (12), (13) and $E\{\eta_l|X_l, (X, Y)_1^{2n_1}\} = 0, l = 2n_1 + 1, \dots, 3n_1$; these results together with a direct calculation yield $F_{12} \le Cn_1^{-2}$. Finally, the evaluation of F_{13} is identical to the above-conducted evaluation of $E\{A_3\}$ with the only difference that $\sigma^2(x)$ is used in place of $m(x)$, and note that these two functions satisfy the same smoothness assumption; this yields $F_{13} \le C[n_1^{-2} + S^{-4}(1 + S - s)^{-4}]$. □

Proof of Lemma 3 Inequality (18) follows (13) and (14), Remark 3, Cauchy–Schwarz inequality and $\hat{\sigma}^{-1}(x) \le b_n$. Let us verify (19). Write

$$V(x) = \frac{m(x) - \hat{m}(x)}{b\sigma(x)} + \frac{(\sigma(x) - \hat{\sigma}(x))(m(x) - \hat{m}(x))}{b\hat{\sigma}(x)\sigma(x)} =: V_1(x) + V_2(x),$$

and

$$W(x) = \frac{\sigma^2(x) - \hat{\sigma}^2(x)}{2b\sigma^2(x)} + \left[\frac{(\sigma(x) - \hat{\sigma}(x))^2}{b\hat{\sigma}(x)\sigma(x)} + \frac{(\sigma^2(x) - \hat{\sigma}^2(x))^2}{2b\sigma^2(x)(\sigma(x) + \hat{\sigma}(x))^2}\right]$$
$$=: W_1(x) + W_2(x).$$

For a generic function $g(x)$, which is differentiable and its derivative $g^{(1)}(x)$ is square integrable on [0, 1], we get $\int[m(x) - \hat{m}(x)]g(x)\mathrm{d}x = \sum_{s=0}^{S}(\kappa_s - \hat{\kappa}_s)\int g(x)\varphi_s(x)\mathrm{d}x + \sum_{s>S}\kappa_s \int g(x)\varphi_s(x)\mathrm{d}x$. Denote $w_s := \int g(x)\varphi_s(x)\mathrm{d}x$. Proposition 1 in Efromovich (2001) implies that the inequality $\sum_{s\ge1} s^2 w_s^2 < C$ holds. This together with Hölder inequality yields

$$
\begin{aligned}
E\left[\int [m(x)-\hat{m}(x)]g(x)\mathrm{d}x\right]^4 &= E\left[\sum_{s=0}^{S}(\kappa_s-\hat{\kappa}_s)w_s+\sum_{s>S}\kappa_s w_s\right]^4\\
&\le CE\left[\sum_{s=0}^{S}(1+s)^{-2}(\kappa_s-\hat{\kappa}_s)^2\sum_{s=0}^{S}(1+s)^2 w_s^2\right.\\
&\qquad\left.+\sum_{s>S}\kappa_s^2\sum_{s>S}w_s^2\right]^2\\
&\le C\sum_{s=0}^{S}(1+s)^{-3}\ln^2(2+s)E(\kappa_s-\hat{\kappa}_s)^4+Cn^{-8/3}.
\end{aligned}
\tag{57}
$$

In the last inequality we used Proposition 1 in Efromovich (2001) and $\sum_{s=0}^{S}[(1+s)\ln^2(2+s)]^{-1} < C$. Then using Lemma 2 we get

$$
\begin{aligned}
&E\left[\int (m(x)-\hat{m}(x))g(x)\mathrm{d}x\right]^4\\
&\le C\sum_{s=0}^{S}(1+s)^{-3}\ln^2(2+s)[n_1^{-2}+S^{-4}(1+S-s)^{-4}]+Cn^{-8/3}\\
&\le Cn_1^{-2}+CS^{-4}\sum_{s=0}^{S}(1+s)^{-3}\ln^2(2+s)(1+S-s)^{-4}\\
&\le Cn_1^{-2}+CS^{-8}\sum_{0\le s\le S/2}(1+s)^{-2}+CS^{-7}\ln^2(2+S)\\
&\quad\times\sum_{S/2\le s\le S}(1+S-s)^{-4}\le Cn_1^{-2}.
\end{aligned}
$$

This yields $E[\int g(x)V_1(x)\mathrm{d}x]^4 \le Cn_1^{-2}$. Further, using Cauchy–Schwarz inequality, (13) and (14) and $\hat{\sigma}(x) \ge b_n^{-1}$ we get a rough (but sufficient) upper bound $E[\int g(x)V_2(x)\mathrm{d}x]^4 \le Cn_1^{-7/3}$. Combining the results yield $E[\int g(x)V(x)\mathrm{d}x]^4 \le Cn_1^{-2}$. Now we need to consider $E[\int g(x)W_i(x)\mathrm{d}x]^4$, $i=1,2$. The term involving W_2 is plainly evaluated with the help of Cauchy–Schwarz inequality along with (13) and (14); this yield $E[\int g(x)W_2(x)\mathrm{d}x]^4 \le Cn_1^{-7/3}$. To evaluate the term with W_1, it suffices to consider $E[\int(\hat{\sigma}^2(x)-\sigma^2(x))g(x)\mathrm{d}x]^4$ for a generic $g(x)$ satisfying the condition of Lemma 3. This task is identical to the earlier performed evaluation of the left side of (57). □

Proof of Lemma 4 Relation (20) is proved in Proposition 1 of Efromovich (2001). Using Parseval identity and integration by parts we are verifying (21): $\int_0^1[g^{(1)}(u)]^2 \mathrm{d}u = [\int_0^1 g^{(1)}(u)\mathrm{d}u]^2+\sum_{j=1}^{\infty}[\int_0^1 g^{(1)}(u)2^{1/2}\cos(\pi ju)\mathrm{d}u]^2 = [g(1)-g(0)]^2+\sum_{j=1}^{\infty}\left[g(u)2^{1/2}\cos(\pi ju)\Big|_0^1+\pi j\int_0^1 g(u)2^{1/2}\sin(\pi ju)\mathrm{d}u\right]^2 = \sum_{j=1}^{\infty}(\pi j)^2[\int_0^1$

$g(u)2^{1/2}\sin(\pi ju)\mathrm{d}u]^2$. To establish (22) we use a similar technique: $\int_0^1[g^{(2)}(u)]^2 \mathrm{d}u = \sum_{j=1}^{\infty}\Big[\int_0^1 g^{(2)}(u)2^{1/2}\sin(\pi ju)\mathrm{d}u\Big]^2 = \sum_{j=1}^{\infty}\Big[g^{(1)}(u)2^{1/2}\sin(\pi ju)\Big|_0^1 - \pi j\int_0^1 g^{(1)}(u)2^{1/2}\cos(\pi ju)\mathrm{d}u\Big]^2 = \sum_{j=1}^{\infty}(\pi j)^2\Big[\int g^{(1)}(u)2^{1/2}\cos(\pi ju)\mathrm{d}u\Big]^2 = \sum_{j=1}^{\infty}(\pi j)^4\Big[\int_0^1 g(u)2^{1/2}\sin(\pi ju)\mathrm{d}u\Big]^2$.

Inequality (23) is proved, for instance, in Efromovich (1999, s.2.2). Part (a) is verified.

Now we are considering part (b). Let $\psi(u,x)$ be a function such that $\psi(0,x) = \psi(1,x) = 0$ for all $x \in [0,1]$ and also $\int_0^1\int_0^1[(\partial/\partial x\partial^2/\partial u^2)\psi(u,x)]^2\mathrm{d}u\mathrm{d}x < \infty$. Parseval identity implies:

$$\int_0^1\int_0^1\Big[\frac{\partial}{\partial x}\frac{\partial^2}{\partial u^2}\psi(u,x)\Big]^2\mathrm{d}u\mathrm{d}x$$
$$= \sum_{j=0}^{\infty}\sum_{s=0}^{\infty}\left[\int_0^1\int_0^1\left(\frac{\partial}{\partial x}\frac{\partial^2}{\partial u^2}\psi(u,x)\right)2^{1/2}\sin(\pi ju)2^{1/2}\sin(\pi sx)\mathrm{d}u\mathrm{d}x\right]^2.$$

To evaluate the integral we need to make several preliminary calculations. Write

$$\int_0^1\left(\frac{\partial}{\partial x}\frac{\partial^2}{\partial u^2}\psi(u,x)\right)2^{1/2}\sin(\pi sx)\mathrm{d}x$$
$$= \left(\frac{\partial^2}{\partial u^2}\psi(u,x)\right)2^{1/2}\sin(\pi sx)\Big|_{x=0}^{x=1}$$
$$-(\pi s)\int_0^1\left(\frac{\partial^2}{\partial u^2}\psi(u,x)\right)2^{1/2}\cos(\pi sx)\mathrm{d}x$$
$$= -(\pi s)\int_0^1\left(\frac{\partial^2}{\partial u^2}\psi(u,x)\right)2^{1/2}\cos(\pi sx)\mathrm{d}x.$$

This result, together with Fubini theorem, implies that

$$\int_0^1\int_0^1\Big(\frac{\partial}{\partial x}\frac{\partial^2}{\partial u^2}\psi(u,x)\Big)2^{1/2}\sin(\pi ju)2^{1/2}\sin(\pi sx)\mathrm{d}u\mathrm{d}x$$
$$= -(\pi s)\int_0^1 2^{1/2}\cos(\pi sx)\left[\int_0^1\Big(\frac{\partial^2}{\partial u^2}\psi(u,x)\Big)2^{1/2}\sin(\pi ju)\mathrm{d}u\right]\mathrm{d}x.$$

Consider the inner integral,

$$\int_0^1 \left(\frac{\partial^2}{\partial u^2}\psi(u,x)\right) 2^{1/2}\sin(\pi ju)\mathrm{d}u$$

$$= \left(\frac{\partial}{\partial u}\psi(u,x)\right) 2^{1/2}\sin(\pi ju)\Big|_{u=0}^{u=1} - (\pi j)\int_0^1 \left(\frac{\partial}{\partial u}\psi(u,x)\right) 2^{1/2}\cos(\pi ju)\mathrm{d}u$$

$$= -(\pi j)\psi(u,x)2^{1/2}\cos(\pi ju)\Big|_{u=0}^{u=1} + (\pi j)^2\int_0^1 \psi(u,x)2^{1/2}\sin(\pi ju)\mathrm{d}u.$$

Recall the assumption $\psi(0,x) = \psi(1,x) = 0, x \in [0,1]$, and then combining the results we get $\int_0^1\int_0^1 \Big((\partial^2/\partial u^2)\psi(u,x)\Big)2^{1/2}\sin(\pi ju)2^{1/2}\sin(\pi sx)\mathrm{d}u\mathrm{d}x = -(\pi s)(\pi j)^2\int_0^1\int_0^1 \psi(u,x)2^{1/2}\cos(\pi sx) \times 2^{1/2}\sin(\pi ju)\mathrm{d}u\mathrm{d}x$. Combining the obtained results we get $\int_0^1\int_0^1 \Big[(\partial/\partial x)(\partial^2/\partial u^2)\psi(u,x)\Big]^2\mathrm{d}u\mathrm{d}x = \sum_{j=0}^{\infty}\sum_{s=0}^{\infty} (\pi j)^4(\pi s)^2\Big[\int_0^1\int_0^1 \psi(u,x)2^{1/2}\sin(\pi ju)2^{1/2}\cos(\pi sx)\mathrm{d}u\mathrm{d}x\Big]^2$, which verifies (24). Further, relation (25) follows from (22). Part (b) is verified.

Let us consider (26). The part with $\int_0^1 u^k f(u|x)\sin(\pi ju)\mathrm{d}u$ follows from (25). The part with $\int_0^1 u^k f(u|x)\cos(\pi ju)\mathrm{d}u$ follows from (2.2.7) in Efromovich (1999). This verifies (26). Now we are considering (27). Using (22) and Cauchy–Schwarz inequality we get $\sum_{j\in B_k} j^2\Big[\int_0^1[\int_0^1 p(x)H(x,a+bu)\psi(u|x)\mathrm{d}x] \sin(\pi ju)\mathrm{d}u\Big]^2 \le j_k^{-2}\int_0^1\Big[(\partial^2/\partial u^2)\int_0^1 p(x)H(x,a+bu)\psi(u|x)\mathrm{d}x\Big]^2\mathrm{d}u$, where j_k is the minimal index from the block B_k. Note that $j_k \ge Ck^3$ according to Remark 5. Also, using notation from (17) we can write $\frac{\partial^2}{\partial u^2}\int_0^1 p(x)H(x,a+bu)\psi(u|x)\mathrm{d}x = \frac{\partial^2}{\partial u^2}\int_0^1 p(x)[V(x)+(a+bu)W(x)]\psi(u|x)\mathrm{d}x = \int_0^1 p(x)V(x)[(\partial^2/\partial u^2)\psi(u|x)] \mathrm{d}x + \int_0^1 p(x)W(x)[(\partial^2/\partial u^2)(a+bu)\psi(u|x)]\mathrm{d}x$. Combining the obtained results and using Cauchy–Schwarz inequality we get

$$E\left(\sum_{j\in B_k} j^2\left[\int_0^1\int_0^1 p(x)H(x,a+bu)\psi(u|x)\sin(\pi ju)\mathrm{d}u\mathrm{d}x\right]^2\right)^2$$

$$\le Ck^{-6}E\left\{\int_0^1\left[\int_0^1 p(x)V(x)\left[\frac{\partial^2}{\partial u^2}\psi(u|x)\right]\mathrm{d}x\right.\right.$$

$$\left.\left.+\int_0^1 p(x)W(x)\left[\frac{\partial^2}{\partial u^2}(a+bu)\psi(u|x)\right]\mathrm{d}x\right]^2\mathrm{d}u\right\}^2$$

$$\leq Ck^{-6}E\left\{\int_0^1\left[\int_0^1 p(x)\left[\frac{\partial^2}{\partial u^2}\psi(u|x)\right]V(x)\mathrm{d}x\right]^2\mathrm{d}u\right\}^2$$

$$+Ck^{-6}E\left\{\int_0^1\left[\int_0^1 p(x)\left[\frac{\partial^2}{\partial u^2}(a+bu)\psi(u|x)\right]W(x)\mathrm{d}x\right]^2\mathrm{d}u\right\}^2$$

$$\leq Ck^{-6}\int_0^1\left\{E\left[\int_0^1 p(x)\left[\frac{\partial^2}{\partial u^2}\psi(u|x)\right]V(x)\mathrm{d}x\right]^4\right.$$

$$\left.+E\left[\int_0^1 p(x)\left[\frac{\partial^2}{\partial u^2}(a+bu)\psi(u|x)\right]W(x)dx\right]^4\right\}du\leq Ck^{-6}n_1^{-2}.$$

Here the last inequality holds due to (19) and Assumption B. □

Acknowledgements Supported in part by NSF Grant DMS-0243606 and DMS-0604558. Suggestions of AE and referees are appreciated.

References

Akritas, M. G., & Van Keilegom, I. (2001). Non-parametric estimation of the residual distribution. *Scandinavian Journal of Statistics*, *28*, 549–567.
Bickel, P. J., & Ritov, Y. (2003). Nonparametric estimators that can be "plugged–in". *The Annals of Statistics*, *31*, 1033–1053.
Brown, L. D., Low, M. G., & Zhao, L. H. (1997). Superefficiency in nonparametric function estimation. *The Annals of Statistics*, *25*, 2607–2625.
Butzer, P. L., & Nessel, R. J. (1971). *Fourier analysis and approximation*. New York: Academic press.
Cheng, F. (2002). *Error density and distribution function estimation in nonparametric regression models*. Ph.D. Thesis. Department of Statistics and Probability, Michigan State Univeristy.
Cheng, F. (2004). Weak and strong uniform consistency of a kernel error density estimator in nonparametric regression. *Journal of Statistical Planning and Inference*, *116*, 95–108.
Donoho, D., & Johnstone, D. (1995). Adaptation to unknown smoothness via wavelet shrinkage, *Journal of the American Statistical Association*, *90*, 1200–1224.
Efromovich, S. (1985). Nonparametric estimation of a density with unknown smoothness. *Theory of Probability and its Applications*, *30*, 557–568.
Efromovich, S. (1998). Simultaneous sharp estimation of functions and their derivatives. *The Annals of Statistics*, *26*, 273–278.
Efromovich, S. (1999), *Nonparametric curve estimation: methods, theory and applications*. Berlin Heidelberg New York: Springer.
Efromovich, S. (2001). Density estimation under random censorship and order restrictions: from asymptotic to small sample sizes. *Journal of the American Statistical Association*, *94*, 667–685.
Efromovich, S. (2004). Distribution estimation for biased data. *Journal of Statistical Planning and Inference*, *124*, 1–43.
Efromovich, S. (2005). Estimation of the density of regression errors. *The Annals of Statistics*, *33*, 2194–2227.
Efromovich, S. (2006). Adaptive estimation of the error density in nonparametric regression with small sample size. *Journal of Statistical Planning and Inference* (in press).
Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. New York: Marcel Dekker.

Fan, J., & Gijbels, I. (1996). *Local polynomial modeling and its applications*. New York: Chapman & Hall.
Hart, J. (1997). *Nonparametric smoothing and lack-of-fit tests*. Berlin Heidelberg New York: Springer.
Müller, U. U., & Schick, A., Wefelmeyer, W. (2004), Estimating linear functionals of the error distribution in nonparametric regression. *Journal of Statistical Planning and Inference*, *116* 75–94.
Neter, J., Kutner, M., Nachtsheim, C., & Wasserman, W. (1996). *Applied linear models*, (4th ed.). Boston: McGraw-Hill.
Wasserman, L. (2005). *All of nonparametric statistics*. Berlin Heidelberg New York: Springer.
Zhang, C.-H. (2005). General empirical Bayes wavelet methods and exactly adaptive minimax estimation. *The Annals of Statistics*, *33*, 54–100.