The Annals of Statistics
2002, Vol. 30, No. 3, 688–707

ASYMPTOTIC EQUIVALENCE THEORY FOR NONPARAMETRIC REGRESSION WITH RANDOM DESIGN

BY LAWRENCE D. BROWN, T. TONY CAI, MARK G. LOW AND CUN-HUI ZHANG

University of Pennsylvania, University of Pennsylvania, University of Pennsylvania and Rutgers University

This paper establishes the global asymptotic equivalence between the nonparametric regression with random design and the white noise under sharp smoothness conditions on an unknown regression or drift function. The asymptotic equivalence is established by constructing explicit equivalence mappings between the nonparametric regression and the white-noise experiments, which provide synthetic observations and synthetic asymptotic solutions from any one of the two experiments with asymptotic properties identical to the true observations and given asymptotic solutions from the other. The impact of such asymptotic equivalence results is that an investigation in one nonparametric problem automatically yields asymptotically analogous results in all other asymptotically equivalent nonparametric problems.

1. Introduction. The purpose of this paper is to establish the global asymptotic equivalence between the nonparametric regression with random design and the white noise under sharp smoothness conditions on an unknown regression or drift function. We establish this asymptotic equivalence by constructing explicit equivalence mappings between the nonparametric regression and the white-noise problems, as in Brown and Low (1996) for their asymptotic equivalence results. The equivalence mapping from the nonparametric regression to the white noise provides synthetic observations of the white noise from the nonparametric regression such that the distributions of the synthetic observations are asymptotically equivalent to those of the true observations of the white noise. For any asymptotic solution to a white-noise problem, the application of the solution to the synthetic observations provides an asymptotic solution to the corresponding nonparametric regression problem with identical asymptotic properties. Likewise, the equivalence mapping from the white noise produces synthetic observations of the nonparametric regression problem and synthetic asymptotic solutions to white-noise problems based on those of the corresponding nonparametric regression problems. The impact of such asymptotic equivalence results is that an investigation in one nonparametric problem automatically yields asymptotically analogous results in

Received September 1999; revised October 2001.
AMS 2000 subject classifications. Primary 62G20; secondary 62G08.
Key words and phrases. Asymptotic equivalence, Le Cam's distance, nonparametric regression, white-noise model.

all other equivalent nonparametric problems. For example, the Pinsker (1980) estimator can be used to produce asymptotically minimax estimators of the whole regression function for Sobolev classes in nonparametric regression with i.i.d. design points under L_2-loss. The results of Fan (1991) on the convergence rates for estimation of quadratic functionals can be directly translated into the nonparametric regression setting. Additional important examples, references and discussion can be found in Brown and Low (1996) and Nussbaum (1996). More recent results for potential applications of the asymptotic equivalence include Donoho and Johnstone (1998), Donoho, Johnstone, Kerkyacharian and Picard (1996), Efromovich (1998) and Tsybakov (1998), among many others.

Recently there have been several papers on the global asymptotic equivalence of certain nonparametric experiments. Brown and Low (1996) established global asymptotic equivalence of the white-noise problem with unknown drift f to the nonparametric regression problem with deterministic design and unknown regression f. Nussbaum (1996) established global asymptotic equivalence of the white-noise problem to the nonparametric density problem with unknown density $g = f^2/4$. In both these instances the global asymptotic equivalence was established under a smoothness assumption: f belongs to the Lipschitz classes with smoothness index $\alpha > 1/2$. It has also been demonstrated that such nonparametric problems are typically asymptotically nonequivalent when the unknown f belongs to larger classes, for example, with smoothness index $\alpha \leq 1/2$. Brown and Low (1996) showed the asymptotic nonequivalence between the white-noise problem and nonparametric regression with deterministic design for $\alpha \leq 1/2$; Efromovich and Samarov (1996) showed that the asymptotic equivalence may fail when $\alpha < 1/4$. Brown and Zhang (1998) showed the asymptotic nonequivalence for $\alpha \leq 1/2$ between any pair of the following four experiments: white noise; density problem; nonparametric regression with random design; and nonparametric regression with deterministic design.

The asymptotic equivalence established in this paper between the nonparametric regression with random design and the white noise applies to compact classes of functions in a Besov space with smoothness index $\alpha = 1/2$, described in Section 3. This result is sharp in the sense that the asymptotic equivalence fails for balls of positive radii in the same Besov space with $\alpha = 1/2$, as shown in Brown and Zhang (1998). It follows that the asymptotic equivalence holds in Lipschitz and Sobolev classes with smoothness index $\alpha > 1/2$, since balls of positive radii in these spaces are contained in compact sets in the Besov space with smoothness index $\alpha = 1/2$. Furthermore, upper bounds of order $n^{-\alpha\wedge(1/2)}$ are provided for the difference between the unknown functions of the synthetic and true versions of the experiments for all $\alpha > 0$, so that certain asymptotic results can be easily translated between the white-noise problem and nonparametric regression even for smoothness index $\alpha < 1/2$. The equivalence mappings constructed here improve upon those based on infinite Haar series expansions in Brown and Zhang (1996), where a stronger Besov metric with smoothness index $\alpha = 1/2$ was used.

In the rest of the section, we formally describe the white-noise and nonparametric regression experiments and global asymptotic equivalence, and then state in Theorem 1 the asymptotic equivalence between the white-noise problem and nonparametric regression for the Lipschitz and Sobolev classes for $\alpha > 1/2$, with some discussion. The equivalence mappings between the two experiments are constructed in Section 2. In Section 3, we derive certain upper bounds for the Hellinger distances between the probability measures of the two experiments under these equivalence mappings; based on these upper bounds, we prove Theorem 1 and establish sharper asymptotic equivalence results. We have the following:

1. *White noise* $\xi_{1,n}$. A Gaussian process $\{Z_n^\star(t),\ 0 \le t \le 1\}$ is observed such that

$$Z_n^\star(t) \equiv \int_0^t f(x)\,dx + \frac{B^\star(t)}{\sqrt{n}}, \qquad 0 \le t \le 1, \tag{1.1}$$

   with a standard Brownian motion $B^\star(t)$ and an unknown $f \in \mathcal{F}_n$. Here we allow the parameter spaces $\mathcal{F}_n$ to depend on n.
2. *Nonparametric regression with random design* $\xi_{2,n}$. Random vectors (Y_i, X_i), $1 \le i \le n$, are observed such that

$$Y_i \equiv f(X_i) + \varepsilon_i, \qquad 1 \le i \le n, \tag{1.2}$$

   where $\{\varepsilon_i,\ 1 \le i \le n\}$ are i.i.d. $N(0,1)$ variables independent of $\{X_i, 1 \le i \le n\}$ and $\{X_i,\ 1 \le i \le n\}$ are i.i.d. uniform random variables on $[0,1]$. Again, $f \in \mathcal{F}_n$. The asymptotic theory in this paper also applies to i.i.d. design points X_i' with any continuous distribution G with the transformation $X_i = G(X_i')$ and the corresponding translation of the conditions on f. We shall focus on the uniform case without loss of generality.
3. *Asymptotic equivalence*. Two sequences of experiments $\{\xi_{1,n}, n \ge 1\}$ and $\{\xi_{2,n}, n \ge 1\}$, with a common parameter space $\mathcal{F}_n$ for each n, are asymptotically equivalent if

$$\Delta(\xi_{1,n}, \xi_{2,n}; \mathcal{F}_n) \to 0 \qquad \text{as } n \to \infty. \tag{1.3}$$

For any two experiments ξ_1 and ξ_2 with a common parameter space Θ, $\Delta(\xi_1, \xi_2; \Theta)$ is Le Cam's distance [cf., e.g., Le Cam (1986) or Le Cam and Yang (1990)] defined as

$$\Delta(\xi_1, \xi_2; \Theta) \equiv \sup_L \max_{j=1,2} \inf_{\delta^{(k)}} \sup_{\delta^{(j)}} \sup_{\theta \in \Theta} \bigl| E_\theta^{(j)} L(\theta, \delta^{(j)}) - E_\theta^{(k)} L(\theta, \delta^{(k)}) \bigr|,$$

where the first supremum is taken over all decision problems with loss function $\|L\|_\infty \le 1$, given the decision problem and $j = 1, 2$, $k \equiv 3 - j$ ($k = 2$ for $j = 1$ and $k = 1$ for $j = 2$) and the minimax value of the maximum difference in risks over Θ is computed over all (randomized) statistical procedures $\delta^{(\ell)}$ for ξ_ℓ, and the expectations $E_\theta^{(\ell)}$ are evaluated in experiments ξ_ℓ with parameter θ, $\ell = j, k$. The statistical interpretation of the Le Cam distance is as follows: if $\Delta(\xi_1, \xi_2; \Theta) < \varepsilon$, then for any decision problem with $\|L\|_\infty \le 1$ and any statistical procedure $\delta^{(j)}$ with the experiment $\xi^{(j)}$, $j = 1, 2$, there exists a (randomized) procedure $\delta^{(k)}$ with $\xi^{(k)}$, $k = 3 - j$, such that the risk of $\delta^{(k)}$ evaluated in $\xi^{(k)}$ nearly matches (within ε) that of $\delta^{(j)}$ evaluated in $\xi^{(j)}$.

For $0 < \alpha \le 1$, the Lipschitz classes are formally defined as

$$(1.4)\quad \mathcal{F}_{\alpha,M}^{(L)} \equiv \{f : \|f\|_\alpha^{(L)} \le M\}, \qquad \|f\|_\alpha^{(L)} \equiv \sup_{0\le x<y\le 1} \frac{|f(x) - f(y)|}{|x-y|^\alpha},$$

and the Sobolev classes are formally defined as

$$(1.5)\quad \mathcal{F}_{\alpha,M}^{(S)} \equiv \{f : \|f\|_\alpha^{(S)} \le M\}, \qquad \|f\|_\alpha^{(S)} \equiv \left[\sum_{j=-\infty}^{\infty} |j|^{2\alpha} |a_j(f)|^2\right]^{1/2},$$

where $a_j(f) \equiv \int_0^1 e^{i2\pi jt} f(t)\, dt$ are the Fourier coefficients of f. For both Lipschitz and Sobolev classes, α is the smoothness index. It follows from Theorem 2 in Section 3 that the asymptotic equivalence also applies to other classes of f, for example, balls of positive radii in Sobolev-type spaces with a seminorm of the form $[\sum_{j=-\infty}^{\infty} b_j^{2\alpha} |\langle f, \psi_j\rangle|^2]^{1/2}$ for certain $b_j/j \to 1$ and basis $\{\psi_j\}$, as long as these balls are contained in compact sets in the Besov space with norm $\|f\|^{(B)}$ in (3.4).

THEOREM 1. *Let $\xi_{1,n}$ and $\xi_{2,n}$ be the white-noise and nonparametric regression experiments with random design as in* (1.1) *and* (1.2). *Then*

$$(1.6)\quad \lim_{n\to\infty} \Delta(\xi_{1,n}, \xi_{2,n}; \mathcal{F}_{\alpha,M}^{(L)}) = 0 \quad \textit{and} \quad \lim_{n\to\infty} \Delta(\xi_{1,n}, \xi_{2,n}; \mathcal{F}_{\alpha,M}^{(S)}) = 0$$

for all $\alpha > 1/2$ and $M < \infty$.

Suppose $\xi_{j,n}$ are characterized by families of probability measures $\{P_f^{(j,n)}, f \in \mathcal{F}_n\}$ in certain sample spaces $\mathcal{X}_{j,n}$. The basic idea for the proof of (1.3) is to explicitly construct (a) randomized versions $\widetilde{\xi}_{j,n}$ of $\xi_{j,n}$, characterized by families of probability measures $\{\widetilde{P}_f^{(j,n)}, f \in \mathcal{F}_n\}$ in certain spaces $\widetilde{\mathcal{X}}_{j,n}$, and (b) equivalence mappings $T_{1,n} : \widetilde{\mathcal{X}}_{1,n} \to \mathcal{X}_{2,n}$ and $T_{2,n} : \widetilde{\mathcal{X}}_{2,n} \to \mathcal{X}_{1,n}$, independent of f, such that

$$(1.7)\quad \lim_{n\to\infty} \sup_{f\in\mathcal{F}_n} H\Big(\widetilde{P}_f^{(k,n)} \circ T_{k,n}^{-1}, P_f^{(j,n)}\Big) = 0, \qquad k = 3 - j,\ j = 1, 2,$$

where $H(\widetilde{P}_f^{(k,n)} \circ T_{k,n}^{-1}, P_f^{(j,n)})$ are the Hellinger distances between measures $\widetilde{P}_f^{(k,n)} \circ T_{k,n}^{-1}$ and $P_f^{(j,n)}$ in the spaces $\mathcal{X}_{j,n}$. The asymptotic equivalence (1.3) is a consequence of (1.7) due to the following reasons: given a loss function L and a statistical procedure $\delta^{(j,n)}$ with the experiment $\xi^{(j,n)}$, the equivalence mappings provide a (randomized) synthetic procedure $\tilde{\delta}^{(k,n)} = \delta^{(j,n)} \circ T_{k,n}$ from the other experiment $\xi^{(k,n)}$, $k = 2$ for $j = 1$ and $k = 1$ for $j = 2$, such that

$$
\begin{aligned}
&\sup_{f\in\mathcal{F}_n} \left| E_f^{(j,n)} L(f, \delta^{(j,n)}) - E_f^{(k,n)} L(f, \tilde{\delta}^{(k,n)}) \right| \\
&\quad = \sup_{f\in\mathcal{F}_n} \left| \int_{\mathcal{X}_{j,n}} L(f, \delta^{(j,n)}) \{ dP_f^{(j,n)} - d\widetilde{P}_f^{(k,n)} \circ T_{k,n}^{-1} \} \right| \qquad (1.8)\\
&\quad \le 2\|L\|_\infty \sup_{f\in\mathcal{F}_n} H^2\Big(\widetilde{P}_f^{(k,n)} \circ T_{k,n}^{-1}, P_f^{(j,n)} \Big)
\end{aligned}
$$

[cf., e.g., Brown and Low (1996) for a proof of the above inequality]. Thus, $\Delta(\xi_{1,n}, \xi_{2,n}; \mathcal{F}_n) \le 2 \max_{j=1,2} \sup_{f\in\mathcal{F}_n} H^2(\widetilde{P}_f^{(3-j,n)} \circ T_{3-j,n}^{-1}, P_f^{(j,n)})$.

In the above discussion, the synthetic statistical procedures mentioned in the beginning of the section are formally defined as $\tilde{\delta}^{(k,n)} = \delta^{(j,n)} \circ T_{k,n}$, $k = 3 - j$. It is worthwhile to emphasize that these synthetic procedures depend on the loss function only through the original procedures $\delta^{(j,n)}$, so that $\tilde{\delta}^{(k,n)}$ and $\delta^{(j,n)}$ share the same asymptotic properties even when evaluated with (infinitely) many loss functions [e.g., $L(f, \delta) = I\{(f(x) - \delta(x))/r_n \le t\}$ for all real (x, t) in the evaluation of asymptotic distributions at rate r_n]. For the proof of Theorem 1, in Section 3, we simply set $\mathcal{F}_n = \mathcal{F}_{\alpha,M}^{(L)}$ or $\mathcal{F}_n = \mathcal{F}_{\alpha,M}^{(S)}$ for all n in (1.7). The equivalence mappings constructed in Section 2 are independent of (α, M). Thus, the asymptotic equivalence results also cover the adaptive situations where a family of problems with different α is considered.

Throughout the sequel, the unknown function f is assumed to belong to $L_2[0,1] \equiv \{f : \|f\| < \infty\}$, where $\|f\| \equiv \{\int_0^1 f^2(t)\,dt\}^{1/2}$.

2. The equivalence mappings. This section describes in detail the mappings $\{T_{j,n}\}$ satisfying (1.7) which provide the asymptotic equivalence claimed in this paper. The fact that these mappings yield asymptotic equivalence will be proved in Sections 3 and 4. For convenience in both construction and proof these mappings are broken into several stages.

For clarity and notational convenience, we shall use boldface capital letters to denote vectors and collections of stochastic processes of continuous time t, denote functions of $Z^\star \equiv \{Z^\star(t) \equiv Z_n^\star(t),\ 0 \le t \le 1\}$ in (1.1) with a * and those of $(\mathbf{X}, \mathbf{Y}) = \{(X_i, Y_i),\ i \le n\}$ in (1.2) without the *. For quantities with two subscripts, we shall use $\mathbf{U}_k$ to denote $\{U_{k,\ell},\ \ell = 0, \dots, 2^k - 1\}$ for any symbol U (e.g., $U = W, W^\star, N, f, \overline{Z}$, etc.), with or without explicit declaration. The parameter n is fixed unless otherwise stated, and the subscript n and the continuous variable t of stochastic processes are often dropped as in $Z^\star$ above.

Equivalence mappings from nonparametric regression. We begin with data $(\mathbf{X}, \mathbf{Y})$ in (1.2) in the nonparametric regression problem. The construction then proceeds in several stages to produce synthetic observations of the white-noise problem, of the form of a random function $Z \equiv \{Z(t) \equiv Z_n(t),\ 0 \le t \le 1\}$. We will show in Sections 3 and 4 that under suitable smoothness conditions on $\mathcal{F}_n$, the Hellinger distance between the distributions of $Z^\star$ in (1.1) and Z converges to zero uniformly in $f \in \mathcal{F}_n$, which implies (1.7) for $j = 1$ and $k = 2$.

The construction is built on a binary representation scheme for $[0, 1]$. To motivate our method, we shall describe key random variables from $Z^\star$ to be approximated in distribution by their counterpart from $(\mathbf{X}, \mathbf{Y})$ throughout the construction. Define

$$(2.1)\quad \overline{\mathbf{Z}}_k^\star \equiv \{\overline{Z}_{k,\ell}^\star, 0 \le \ell < 2^k\}, \qquad \overline{Z}_{k,\ell}^\star \equiv 2^k\left\{Z^\star\left(\frac{\ell+1}{2^k}\right) - Z^\star\left(\frac{\ell}{2^k}\right)\right\}.$$

A naive approach is to approximate directly a discretization of the process $Z^\star$, $\{Z^\star(\ell/2^k) = \sum_{j=0}^{\ell-1} \overline{Z}_{k,j}^\star/2^k,\ 0 \le \ell \le 2^k\}$, at a certain resolution level k by the corresponding averages of $(\mathbf{X}, \mathbf{Y})$, that is, to approximate $\overline{\mathbf{Z}}_k^\star$ in distribution by

$$(2.2)\qquad \overline{\mathbf{Y}}_k \equiv \{\overline{Y}_{k,\ell}, 0 \le \ell < 2^k\}, \qquad \overline{Y}_{k,\ell} \equiv \frac{1}{N_{k,\ell}} \sum_{X_i \in I_{k,\ell}} Y_i,$$

for certain fixed $k = k_1$, where $I_{k,\ell} \equiv [\,\ell/2^k, (\ell+1)/2^k)$ and

$$(2.3)\qquad N_{k,\ell} \equiv \#\{X_i : X_i \in I_{k,\ell}\}, \qquad \ell = 0, \dots, 2^k - 1.$$

Note that $\overline{\mathbf{Z}}_k^\star$ is a vector of independent normal variables and conditionally on $\mathbf{X}$, $\overline{\mathbf{Y}}_k$ is a vector of independent normal variables with

$$(2.4)\quad E_f[\overline{Y}_{k,\ell}|\mathbf{X}] = \eta_{k,\ell} \equiv \frac{1}{N_{k,\ell}} \sum_{X_i \in I_{k,\ell}} f(X_i) \approx 2^k \int_{I_{k,\ell}} f(t)\,dt = E_f \overline{Z}_{k,\ell}^\star$$

and

$$(2.5)\qquad \mathrm{Var}_f(\overline{Y}_{k,\ell}|\mathbf{X}) = 1/N_{k,\ell} \approx 2^k/n = \mathrm{Var}_f(\overline{Z}_{k,\ell}^\star).$$

This direct approximation works for deterministic $\{X_i = i/(n+1),\ i \le n\}$ as in Brown and Low (1996). However, for random $\mathbf{X}$, the approximation $1/N_{k,\ell} \approx 2^k/n$ in (2.5) for the variance is accurate only for large $N_{k,\ell}$, that is, for small k, but a good approximation to $Z^\star$ by its discretization requires fine resolution, that is, large $k = k_1$. [The approximation (2.4) is valid for large k as long as $N_{k,\ell} \ge 1$, based on the smoothness of f.] This dilemma is overcome with the following multiresolution construction: approximate $\overline{Z}_{k_0,\ell}^\star$ by

$$(2.6)\quad \overline{Z}_{k_0,\ell} \equiv \overline{Y}_{k_0,\ell} \mathbb{1}_{\{N_{k_0,\ell}>0\}} + (2^{k_0}/n)^{1/2} \widetilde{U}_{k_0,\ell} \mathbb{1}_{\{N_{k_0,\ell}=0\}}, \qquad 0 \le \ell < 2^{k_0},$$

at a relatively low resolution with small k_0, and approximate the differences

$$(2.7)\qquad W_{k,2\ell}^\star \equiv -W_{k,2\ell+1}^\star \equiv (\overline{Z}_{k,2\ell}^\star - \overline{Z}_{k,2\ell+1}^\star)/2$$

by normalized $\overline{Y}_{k,2\ell} - \overline{Y}_{k,2\ell+1}$ to match the conditional variance given $\mathbf{X}$; that is, approximate (2.7) by

$$W_{k,2\ell} \equiv -W_{k,2\ell+1} \equiv \begin{cases} C_{k,2\ell}(\overline{Y}_{k,2\ell} - \overline{Y}_{k,2\ell+1})/2, & \text{if } N_{k,2\ell}N_{k,2\ell+1} > 0, \\ (2^{k-1}/n)^{1/2}\widetilde{U}_{k,2\ell}, & \text{otherwise}, \end{cases} \tag{2.8}$$

for $0 \le \ell < 2^{k-1}$ and $k = k_0 + 1, \ldots, k_1$, up to a fine resolution with large $k_1 > k_0$, where $C_{k,2\ell} = \sqrt{2^{k+1}/n}\sqrt{N_{k,2\ell}N_{k,2\ell+1}/N_{k-1,\ell}}$ and $\widetilde{\mathbf{U}} = \{\widetilde{U}_{k,\ell}, k \ge k_0, \ell \ge 0\}$ is a sequence of i.i.d. $N(0,1)$ variables independent of $(\mathbf{X},\mathbf{Y})$. Note that $I_{k,\ell} = I_{k+1,2\ell} \cup I_{k+1,2\ell+1}$ and randomization with $\widetilde{\mathbf{U}}$ is used when $\overline{Z}_{k_0,\ell}$ and $W_{k,2\ell}$ cannot be directly derived from $(\mathbf{X},\mathbf{Y})$.

It follows from (2.1) and (2.7) that

$$W^\star_{k,2\ell} = \overline{Z}^\star_{k,2\ell} - \overline{Z}^\star_{k-1,\ell} \quad \text{and} \quad W^\star_{k,2\ell+1} = \overline{Z}^\star_{k,2\ell+1} - \overline{Z}^\star_{k-1,\ell}, \tag{2.9}$$

so that the mapping from $\overline{\mathbf{Z}}^\star_{k_1}$ to $\{\overline{\mathbf{Z}}^\star_{k_0}, \mathbf{W}^\star_k, k_0 < k \le k_1\}$ is one-to-one, where $\mathbf{W}^\star_k \equiv \{W^\star_{k,\ell},\ 0 \le \ell < 2^k\}$. In fact, by (2.9) the inverse mapping is

$$\overline{Z}^\star_{k_1,\ell} = \overline{Z}^\star_{k_0,[2^{k_0}\ell/2^{k_1}]} + \sum_{k=k_0+1}^{k_1} W^\star_{k,[2^k\ell/2^{k_1}]} \tag{2.10}$$

for $0 \le \ell < 2^{k_1}$, where $[a]$ is the integer part of a. Thus, we define

$$\overline{Z}_{k_1,\ell} \equiv \overline{Z}_{k_0,[2^{k_0}\ell/2^{k_1}]} + \sum_{k=k_0+1}^{k_1} W_{k,[2^k\ell/2^{k_1}]}, \qquad 0 \le \ell < 2^{k_1}, \tag{2.11}$$

and construct a synthetic version of the discretization of the white noise by setting

$$Z(\ell/2^{k_1}) \equiv 2^{-k_1}\sum_{j=0}^{\ell-1}\overline{Z}_{k_1,j}, \qquad 0 \le \ell \le 2^{k_1}. \tag{2.12}$$

In the final stage, the construction of Z from (2.12) is done by randomization:

$$Z(t) \equiv \frac{\widetilde{B}_\ell(2^{k_1}t - \ell)}{\sqrt{n2^{k_1}}} + Z\left(\frac{\ell}{2^{k_1}}\right) + \left(t - \frac{\ell}{2^{k_1}}\right)\overline{Z}_{k_1,\ell}, \tag{2.13}$$

for $\ell/2^{k_1} \le t < (\ell+1)/2^{k_1}$, $0 \le \ell < 2^{k_1}$, where $\widetilde{\mathbf{B}} = \{\widetilde{B}_\ell(\cdot), 0 \le \ell < 2^{k_1}\}$ is a sequence of independent Brownian bridge processes independent of $(\mathbf{X},\mathbf{Y},\widetilde{\mathbf{U}})$. This ensures that the conditional distribution of Z given $\overline{\mathbf{Z}}_{k_1} = \mathbf{z}_{k_1}$ is identical to that of $Z^\star$ given $\overline{\mathbf{Z}}^\star_{k_1} = \mathbf{z}_{k_1}$, assuming f is constant on each interval $I_{k_1,\ell}$.

We have completed the construction of the equivalence mapping $T_{2,n} : (\mathbf{X},\mathbf{Y},\widetilde{\mathbf{U}},\widetilde{\mathbf{B}}) \to Z$, up to the specification of $k_0 = k_0(n)$ and $k_1 = k_1(n) > k_0$, that is, the mappings given by (2.2), (2.6), (2.8), (2.11), (2.12) and (2.13) with $(\mathbf{X},\mathbf{Y})$ randomized by $(\widetilde{\mathbf{U}},\widetilde{\mathbf{B}})$. The key to the construction is the approximation

$$\left(\overline{\mathbf{Z}}^\star_{k_0}, \mathbf{W}^\star_k, k_0 < k \le k_1\right) \approx \left(\overline{\mathbf{Z}}_{k_0}, \mathbf{W}_k, k_0 < k \le k_1\right) \qquad \text{in distribution.} \tag{2.14}$$

For the asymptotic equivalence claims, we choose integers k_0 and k_1 satisfying

$$k_0 \to \infty, \qquad 2^{k_0}/\sqrt{n} \to 0, \qquad n/4 \le 2^{k_1} < n/2. \tag{2.15}$$

The choice of k_1 above produces the simplest explicit upper bounds in our statements. Our results still hold, with possibly different constant factors in upper bounds in certain cases, if 2^{k_1} are of the order n. Furthermore, the asymptotic equivalence for the Lipschitz and Sobolev classes holds for more economical choices of k_1 satisfying $n2^{-2\alpha k_1} \to 0$, $1/2 < \alpha \le 1$. See Lemma 1 and the remark thereafter.

Equivalence mappings from the white-noise problem. The preceding steps which map the nonparametric regression into a stochastic process can be reversed to produce an asymptotic equivalence map in the reverse direction. To do so, we begin with $Z^\star \equiv \{Z_n^\star(t),\ 0 \le t \le 1\}$ in (1.1) and some i.i.d. uniform [0, 1] variables $\mathbf{X} \equiv \{X_i : i \le n\}$ independent of $Z^\star$, and then recover $\{\overline{\mathbf{Y}}_k^\star\}$, the counterpart of $\{\overline{\mathbf{Y}}_k\}$ with the white noise problem, from $\{\overline{\mathbf{Z}}_k^\star, \mathbf{W}_k^\star\}$ through (2.14). This immediately yields

$$\overline{\mathbf{Y}}_{k_0}^\star \equiv \{\overline{Y}_{k_0,\ell}^\star, 0 \le \ell < 2^{k_0}\} \equiv \overline{\mathbf{Z}}_{k_0}^\star \tag{2.16}$$

at resolution level $k = k_0$. Furthermore, since

$$\overline{Y}_{k,2\ell+j} = \overline{Y}_{k-1,\ell} + (-1)^j \frac{N_{k,2\ell+1-j}}{N_{k-1,\ell}} (\overline{Y}_{k,2\ell} - \overline{Y}_{k,2\ell+1}), \qquad j = 0, 1, \tag{2.17}$$

$0 \le \ell < 2^{k-1}$, $k = k_0 + 1, \dots$, appropriate $\overline{\mathbf{Y}}_k^\star \equiv \{\overline{Y}_{k,\ell}^\star,\ 0 \le \ell < 2^k\}$ are produced by setting inductively, for $k = k_0 + 1, \dots,$

$$\overline{Y}_{k,2\ell+j}^\star \equiv \overline{Y}_{k-1,\ell}^\star + V_{k,2\ell+j}^\star, \qquad j = 0, 1,\ \ell = 0, \dots, 2^{k-1} - 1, \tag{2.18}$$

where, for the $\mathbf{W}_k^\star$ in (2.7) and $\mathbf{N}_k = \{N_{k,\ell},\ 0 \le \ell < 2^k\}$ in (2.3),

$$V_{k,2\ell+j}^\star \equiv \frac{2W_{k,2\ell+j}^\star N_{k,2\ell+1-j}}{C_{k,2\ell} N_{k-1,\ell}}, \tag{2.19}$$

with the convention $0/0 = 0$ in the case of $N_{k,2\ell+1-j} = 0$, in view of (2.8). For suitable $k_1 > k_0$ [e.g., specified in (2.15)], this gives

$$\overline{Y}_{k_1,\ell}^\star = \overline{Z}_{k_0,[2^{k_0}\ell/2^{k_1}]}^\star + \sum_{k=k_0+1}^{k_1} V_{k,[2^k\ell/2^{k_1}]}^\star, \qquad 0 \le \ell < 2^{k_1}. \tag{2.20}$$

Note that $\overline{Y}_{k_1,\ell}^\star$ is well defined whenever $N_{k_1,\ell} > 0$.

Finally, we construct $\mathbf{Y}^\star = \{Y_i^\star, i \le n\}$, with $Y_i^\star$ being the synthetic Y_i, from $\overline{\mathbf{Y}}_{k_1}^\star \equiv \{\overline{Y}_{k_1,\ell}^\star,\ 0 \le \ell < 2^{k_1}\}$ by randomization. Let $e_i^\star$, $1 \le i \le n$, be i.i.d. $N(0, 1)$ variables independent of $(Z^\star, \mathbf{X})$. Define

$$Y_i^\star \equiv \overline{Y}_{k_1,\ell}^\star + \tilde{e}_i^\star, \qquad \tilde{e}_i^\star \equiv e_i^\star - \sum_{X_i \in I_{k_1,\ell}} e_i^\star / N_{k_1,\ell}, \tag{2.21}$$

for the $N_{k_1,\ell}$ variables with $X_i \in N_{k_1,\ell}$. This ensures that the conditional distribution of $\mathbf{Y}^\star$ given $(\overline{\mathbf{Y}}^\star_{k_1}, \mathbf{X}) = (\mathbf{y}_{k_1}, \mathbf{x})$ matches that of $\mathbf{Y}$ given $(\overline{\mathbf{Y}}_{k_1}, \mathbf{X}) = (\mathbf{y}_{k_1}, \mathbf{x})$, with $\overline{\mathbf{Y}}^\star_{k_1}$ and $\overline{\mathbf{Y}}_{k_1}$ in (2.20) and (2.2), respectively, and assuming f is constant on each interval $I_{k_1,j}$. (Thus if, e.g., $X_1, \dots, X_{N_{k_1,\ell}}$ are the set of values of $\mathbf{X}$ which fall in $I_{k_1,\ell}$, then $Y^\star_1, \dots, Y^\star_{N_{k_1,\ell}}$ are conditionally on $\overline{\mathbf{Y}}^\star_{k_1}$ jointly normal with common mean $\overline{Y}^\star_{k_1,\ell}$ and singular covariance matrix $I - N^{-1}_{k_1,\ell}11'$.) This completes the construction of $T_{1,n} : (Z^\star, \mathbf{X}, \tilde{\mathbf{e}}^\star) \to (\mathbf{X}, \mathbf{Y}^\star)$, given by (2.1), (2.7) and (2.19)–(2.21), with $Z^\star$ randomized by $(\mathbf{X}, \tilde{\mathbf{e}}^\star)$, where $\tilde{\mathbf{e}}^\star = \{\tilde{e}^\star_i,\ 1 \le i \le n\}$ is as in (2.21).

REMARK. The choice of $k_1 = \infty$ will result in slightly different equivalence mappings, based on infinite Haar series expansions, with the same asymptotic properties. Equivalence mappings based on infinite Haar series expansions can be found in Brown and Zhang (1996), using a different approximation of $W^\star_{k,\ell}$ than (2.8). The advantage of the current version is the conditional independence of $W_{k,\ell}$ given $\mathbf{X}$ between different resolution levels, which leads to a weaker Besov norm in the upper bound and a clearer presentation.

3. Asymptotic equivalence. In this section we shall prove a stronger version of Theorem 1 based on the mappings defined in Section 2. As mentioned in the Introduction, we shall establish in Theorem 2 below the global asymptotic equivalence between the white noise and nonparametric regression for compact sets in a Besov space with smoothness index $\alpha = 1/2$. We shall also provide direct comparisons between the synthetic and true observations in Theorems 3 and 4, with upper bounds for their differences for general $\alpha > 0$. An upper bound for the Hellinger distances between probability distributions in the two experiments, stated in Lemma 1 below, is crucial in our proofs.

Let $(Z^\star, \mathbf{X}, \tilde{\mathbf{e}}^\star)$ and $(\mathbf{X}, \mathbf{Y}, \widetilde{\mathbf{U}}, \widetilde{\mathbf{B}})$ be respectively the randomizations of (1.1) and (1.2) described in Section 2 [cf. (2.21), (2.6), (2.8) and (2.13)]. Throughout the section, for any $\mathcal{U}$-valued mappings $\mathbf{U}^\star$ from $(Z^\star, \mathbf{X}, \tilde{\mathbf{e}}^\star)$ and $\mathbf{U}$ from $(\mathbf{X}, \mathbf{Y}, \widetilde{\mathbf{U}}, \widetilde{\mathbf{B}})$ (e.g., $\mathbf{U} = \overline{\mathbf{Z}}_k, \mathbf{W}_k, \overline{\mathbf{Y}}_k$, etc.), we shall denote by $H_f(\mathbf{U}^\star, \mathbf{U})$ the Hellinger distance between the distributions of $\mathbf{U}^\star$ and $\mathbf{U}$ in $\mathcal{U}$ when f is the true unknown function in (1.1) and (1.2), and we shall denote by $H_{f,\mathbf{X}}(\mathbf{U}^\star, \mathbf{U})$ the conditional version of $H_f(\mathbf{U}^\star, \mathbf{U})$ given $\mathbf{X}$. Since the distribution of $\mathbf{X}$ is identical for both randomized experiments,

$$H^2_f(\mathbf{U}^\star, \mathbf{U}) \le H^2_f\big((\mathbf{U}^\star, \mathbf{X}), (\mathbf{U}, \mathbf{X})\big) = E H^2_{f,\mathbf{X}}(\mathbf{U}^\star, \mathbf{U}). \tag{3.1}$$

According to (1.7), we shall find upper bounds on

$$D_n(f) \equiv \max\big\{H^2_f(Z^\star, Z), H^2_f\big((\mathbf{X}, \mathbf{Y}^\star), (\mathbf{X}, \mathbf{Y})\big)\big\}, \tag{3.2}$$

where $Z^\star,\ Z,\ \mathbf{X},\ \mathbf{Y}$ and $\mathbf{Y}^\star$ are given by (1.1), (2.13), (1.2) and (2.21).

Let $\bar{f}_k$ be the piecewise average of f at resolution level k, that is, the piecewise constant function defined by

$$\bar{f}_k \equiv \bar{f}_k(t) \equiv \sum_{\ell=0}^{2^k-1} f_{k,\ell} \mathbb{1}_{\{t \in I_{k,\ell}\}}, \qquad f_{k,\ell} = 2^k \int_{I_{k,\ell}} f(t)\,dt, \tag{3.3}$$

with the intervals $I_{k,\ell}$ in (2.2). For $f \in L_2[0,1]$ define the Besov norm

$$\|f\|^{(B)} \equiv \|f\|^{(B)}_{1/2,2,2}, \qquad \|f\|^{(B)}_{\alpha,p,q} \equiv \left\{ \sum_{k=0}^{\infty} (2^{k\alpha} \|\bar{f}_k - \bar{f}_{k+1}\|_p)^q \right\}^{1/q}, \tag{3.4}$$

based on the Haar system, where $\|f\|_p$ is the $L_p[0,1]$-norm and α is the smoothness index.

THEOREM 2. *Let $D_n(f)$ be given by* (3.2) *and let $k_1 = k_1(n)$ satisfy $2^{k_1} \geq n/4$. Then, for $f \in L_2[0,1]$,*

$$D_n(f) \leq \frac{3}{2}\{\|f - \bar{f}_{k_0}\|^{(B)}\}^2 + \frac{2^{2k_0+1}}{n}. \tag{3.5}$$

Consequently, for all compact sets $\mathcal{F}$ in the Besov space $\{f : \|f\|^{(B)} < \infty\}$,

$$\Delta(\xi_{1,n}, \xi_{2,n}; \mathcal{F}) \leq 2 \sup_{f \in \mathcal{F}} D_n(f) \to 0, \tag{3.6}$$

provided that k_0 and k_1 are chosen as in (2.15).

In many applications, it is more convenient to express the synthetic observations in the form of the true ones.

THEOREM 3. *Let $Z \equiv \{Z(t) \equiv Z_n(t),\ 0 \leq t \leq 1\}$ be the synthetic white noise from* (1.2), *given by* (2.13) *with $k_0 = 0$ and $n/4 \leq 2^{k_1} < n/2$. Then Z can be decomposed into*

$$Z_n(t) = G_n(t) + \frac{B(t)}{\sqrt{n}}, \qquad 0 \leq t \leq 1, \tag{3.7}$$

where $G_n(t) \equiv E_f[Z_n(t)|\mathbf{X}]$ and $\{B(t),\ 0 \leq t \leq 1\}$ is a Brownian motion process independent of $\mathbf{X}$ (and thus of G_n). Furthermore, $G_n(\cdot)$ is a piecewise linear function with derivative $g_n(\cdot)$ such that

$$n^{1/2}\sqrt{E_f\|g_n - f\|^2} \leq \sqrt{5}\|f\|^{(B)}_{1/2,2,2} \tag{3.8}$$

and with the norms in (1.4), (1.5), (3.4), $0 < \alpha < 1/2$,

$$\begin{aligned} n^{\alpha}\sqrt{E_f\|g_n - f\|^2} &\leq \left[\frac{5/2}{2^{(k_1+1)(1-2\alpha)}} \left\{ \|f - \bar{f}_0\|^2 + \sum_{k=0}^{k_1} 2^k \|f - \bar{f}_k\|^2 \right\} \right]^{1/2} \\ &\leq \min\{C_\alpha^{(L)}\|f\|_\alpha^{(L)}, C_\alpha^{(S)}\|f\|_\alpha^{(S)}, C_\alpha^{(B)}\|f\|^{(B)}_{\alpha,2,2}\} \end{aligned} \tag{3.9}$$

for certain finite constants $C_\alpha^{(L)}$, $C_\alpha^{(S)}$ and $C_\alpha^{(B)}$.

THEOREM 4. *Let* $(X_i, Y_i^\star)$, $1 \le i \le n$, *be the synthetic observations of the nonparametric regression from the white noise* (1.1), *given by* (2.21) *with* $k_0 = 0$ *and* $n/4 \le 2^{k_1} < n/2$. *Then*

$$Y_i^\star = g_n^\star(X_i) + \varepsilon_i^\star, \qquad 0 \le i \le n, \tag{3.10}$$

where $g_n^\star(\cdot)$ *is piecewise constant at resolution level* k_1, $g_n^\star(X_i) = E_f[Y_i^\star|\mathbf{X}]$, *and* $\{\varepsilon_i^\star, 1 \le i \le n\}$ *are i.i.d.* $N(0,1)$ *variables independent of* $g_n^\star$. *Furthermore,* $E_f \sum_{i=1}^n (g_n^\star(X_i) - f(X_i))^2 \le n E_f \|g_n^\star - f\|^2$ *and* (3.8) *and* (3.9) *hold with* g_n *replaced by* $g_n^\star$.

Theorems 1 and 2 are immediate consequences of the following lemma.

LEMMA 1. *Let* $D_n(f)$ *be given by* (3.2). *Then, for all* $f \in L_2[0,1]$ *and* $k_0 = k_0(n) > 0$,

$$\begin{aligned} D_n(f) \le{} & \frac{7}{8} 2^{k_0} \|f - \bar{f}_{k_0}\|^2 + \frac{5}{8} \sum_{k=k_0}^{k_1} 2^k \|f - \bar{f}_k\|^2 + \frac{2^{2k_0+1}}{n} \\ & + \left(\frac{n}{4} - 2^{k_1}\right) \|f - \bar{f}_{k_1}\|^2, \end{aligned} \tag{3.11}$$

and, for $k_0 = 0$,

$$D_n(f) \le \frac{5}{8} \left\{ \|f - \bar{f}_0\|^2 + \sum_{k=0}^{k_1} 2^k \|f - \bar{f}_k\|^2 \right\} + \left(\frac{n}{4} - 2^{k_1}\right) \|f - \bar{f}_{k_1}\|^2.$$

REMARK. In the rest of the proof, we take $k_1 = k_1(n)$ satisfying $2^{k_1} \ge n/4$, so that the last terms in the upper bounds in Lemma 1 can be omitted for simplicity. The asymptotic equivalence results are still valid for other choices of k_1 as long as $n\|f - \bar{f}_{k_1}\|^2 \to 0$, which holds for the Lipschitz and Sobolev classes when $n 2^{-2\alpha k_1} \to 0$, $1/2 < \alpha \le 1$.

We shall prove Lemma 1 and Theorems 3 and 4 in Section 4. The proof of Theorem 2 and the statement and proof of a stronger version of Theorem 1 are given here based on Lemma 1.

PROOF OF THEOREM 2. Due to the decomposition $\|f - \bar{f}_k\|^2 = \|f - \bar{f}_{k+1}\|^2 + \|\bar{f}_k - \bar{f}_{k+1}\|^2$, the sum in (3.11) can be written as

$$\begin{aligned} \sum_{k=k_0}^{k_1} 2^k \|f - \bar{f}_k\|^2 &= \sum_{k=k_0}^{k_1} 2^k \sum_{j=k}^{\infty} \|\bar{f}_j - \bar{f}_{j+1}\|^2 \\ &= 2 \sum_{j=k_0}^{\infty} 2^{j \wedge k_1} \|\bar{f}_j - \bar{f}_{j+1}\|^2 - 2^{k_0} \|f - \bar{f}_{k_0}\|^2. \end{aligned}$$

Similarly, the Besov norm of $f - \bar{f}_k$ can be written as $\{\|f - \bar{f}_k\|^{(B)}\}^2 = \sum_{j=k}^{\infty} 2^j \times \|\bar{f}_j - \bar{f}_{j+1}\|^2$. Thus, we obtain (3.5) by inserting the above identities into (3.11). A set $\mathcal{F}$ in the Besov space is compact if and only if $\sup_{f\in\mathcal{F}} \|f - \bar{f}_k\|^{(B)} \to 0$ as $k \to \infty$. Thus, (3.6) follows from (1.8) and (3.5), as $k_0 \to \infty$ and $2^{2k_0}/n \to 0$ in (2.15). This completes the proof. □

THEOREM 5. *Let $\xi_{1,n}$ and $\xi_{2,n}$ be as in* (1.1) *and* (1.2) *and* $1/2 < \alpha_n \le 1$. *Then*

$$\lim_{n\to\infty} \Delta\Big(\xi_{1,n}, \xi_{2,n}; \mathcal{F}^{(L)}_{\alpha_n,M_n}\Big) = 0 \quad \textit{and} \quad \lim_{n\to\infty} \Delta\Big(\xi_{1,n}, \xi_{2,n}; \mathcal{F}^{(S)}_{\alpha_n,M_n}\Big) = 0,$$

provided that $c_n \equiv (1 - 2^{1-2\alpha_n})^{-1} M_n^2/n^{\alpha_n - 1/2} \to 0$.

PROOFS OF THEOREMS 1 AND 5. Let $\mathcal{F}_n$ be either $\mathcal{F}^{(L)}_{\alpha_n,M_n}$ or $\mathcal{F}^{(S)}_{\alpha_n,M_n}$. By (1.4) and (1.5), there exists a universal constant C_* such that $\|f - \bar{f}_k\| \le C_* \max\{\|f\|_\alpha^{(L)}, \|f\|_\alpha^{(S)}\}/2^{k\alpha}$ for all $0 < \alpha \le 1$. Thus,

$$\begin{aligned}\Delta(\xi_{1,n}, \xi_{2,n}; \mathcal{F}_n) &\le 2 \sup_{f\in\mathcal{F}_n} D_n(f)\\ &\le \frac{2^{2k_0+2}}{n} + 2C_*^2 2^{k_0(1-2\alpha_n)} M_n^2 \bigg(\frac{7}{8} + \frac{5/8}{1 - 2^{1-2\alpha_n}}\bigg),\end{aligned}$$

which is bounded by $4(2^{k_0}/\sqrt{n})^2 + 3C_*^2(\sqrt{n}/2^{k_0})c_n \to 0$ for k_0 satisfying $\sqrt{c_n} \le 2^{k_0}/\sqrt{n} < 2\sqrt{c_n}$. This proves Theorem 5. Also, it follows from the above proof that Theorem 1 holds with k_0 in (2.15) independent of α. □

4. Proofs of Lemma 1 and Theorems 3 and 4. The proofs of Lemma 1 and Theorems 3 and 4, given at the end of the section, are based on several lemmas which are of some independent interest. In addition to (3.3), the quantities $v_{k,\ell}$ and the identities below are used in the proofs:

$$v_{k,\ell} \equiv \int_{I_{k,\ell}} \{f(t) - f_{k,\ell}\}^2\,dt = \int_{I_{k,\ell}} f^2(t)\,dt - 2^{-k} f_{k,\ell}^2, \tag{4.1}$$

$$\|f - \bar{f}_k\|^2 = \int_0^1 \{f(t) - \bar{f}_k(t)\}^2\,dt = \sum_{\ell=0}^{2^k-1} v_{k,\ell} \tag{4.2}$$

and due to $f_{k-1,\ell} = (f_{k,2\ell} + f_{k,2\ell+1})/2$, for $\ell = 0, \dots, 2^{k-1} - 1$,

$$(f_{k,2\ell} - f_{k,2\ell+1})^2 = 2^{k+1}(v_{k-1,\ell} - v_{k,2\ell} - v_{k,2\ell+1}). \tag{4.3}$$

Let us reverse (2.13) by setting $\widetilde{\mathbf{B}}^\star \equiv \{\widetilde{B}^\star_\ell, 1 \le \ell < 2^{k_1}\}$ with

$$\widetilde{B}^\star_\ell(t) \equiv \sqrt{n 2^{k_1}} \bigg\{ Z^\star\bigg(\frac{t+\ell}{2^{k_1}}\bigg) - Z^\star\bigg(\frac{\ell}{2^{k_1}}\bigg) - \frac{t}{2^{k_1}} \overline{Z}^\star_{k_1,\ell} \bigg\}. \tag{4.4}$$

Likewise, we produce the counterpart $\tilde{\mathbf{e}}$ of $\tilde{\mathbf{e}}^\star = \{\tilde{e}_i^\star,\ i \le n\}$ in (2.21) with

$$\tilde{\mathbf{e}} \equiv \{\tilde{e}_i, 1 \le i \le n\}, \qquad \tilde{e}_i = Y_i - \overline{Y}_{k_1,\ell} \qquad \forall X_i \in I_{k_1,\ell}. \tag{4.5}$$

LEMMA 2. *Let* $Z^\star$, Z, $\mathbf{X}$, $\mathbf{Y}$ *and* $\mathbf{Y}^\star$ *be as in* (3.2). *Then*

$$H^2_{f,\mathbf{X}}(Z^\star, Z) \le H^2_{f,\mathbf{X}}(\widetilde{\mathbf{B}}^\star, \widetilde{\mathbf{B}}) + H^2_{f,\mathbf{X}}(\overline{\mathbf{Z}}^\star_{k_1}, \overline{\mathbf{Z}}_{k_1}), \tag{4.6}$$

$$H^2_{f,\mathbf{X}}\big((\mathbf{X}, \mathbf{Y}^\star), (\mathbf{X}, \mathbf{Y})\big) \le H^2_{f,\mathbf{X}}(\tilde{\mathbf{e}}^\star, \tilde{\mathbf{e}}) + H^2_{f,\mathbf{X}}(\overline{\mathbf{Y}}^\star_{k_1}, \overline{\mathbf{Y}}_{k_1}), \tag{4.7}$$

with the variables in (2.1), (2.11), (2.20) *and* (2.2). *Moreover*,

$$\begin{aligned} &H^2_{f,\mathbf{X}}(\overline{\mathbf{Y}}^\star_{k_1}, \overline{\mathbf{Y}}_{k_1}) \le H^2_{f,\mathbf{X}}(\overline{\mathbf{Z}}^\star_{k_1}, \overline{\mathbf{Z}}_{k_1}) \\ &\quad \le \sum_{\ell=0}^{2^{k_0}-1} H^2_{f,\mathbf{X}}(\overline{Z}^\star_{k_0,\ell}, \overline{Z}_{k_0,\ell}) + \sum_{k=k_0+1}^{k_1} \sum_{\ell=0}^{2^{k-1}-1} H^2_{f,\mathbf{X}}(W^\star_{k,2\ell}, W_{k,2\ell}), \end{aligned} \tag{4.8}$$

with the variables in (2.1), (2.6), (2.7) *and* (2.8) [*cf.* (2.14)].

PROOF. The proof is based on two well-known properties of the Hellinger distance: (1) the squared Hellinger distance of product measures is less than the sum of the squared Hellinger distances on the marginal measures; (2) $H_{f,\mathbf{X}}(\mathbf{U}^\star, \mathbf{U}) \ge H_{f,\mathbf{X}}(\widetilde{\mathbf{U}}^\star, \widetilde{\mathbf{U}})$ if $\widetilde{\mathbf{U}}^\star = T(\mathbf{U}^\star)$ and $\widetilde{\mathbf{U}} = T(\mathbf{U})$ for a single mapping T, with $H_{f,\mathbf{X}}(\mathbf{U}^\star, \mathbf{U}) = H_{f,\mathbf{X}}(\widetilde{\mathbf{U}}^\star, \widetilde{\mathbf{U}})$ for invertible T.

It follows from (4.4) and (2.1) that $\widetilde{\mathbf{B}}^\star$ is independent of $\overline{\mathbf{Z}}^\star_{k_1}$ given $\mathbf{X}$, so that (4.6) holds as the mapping $Z^\star \to (\widetilde{\mathbf{B}}^\star, \overline{\mathbf{Z}}^\star_{k_1})$ is the inverse of (2.12)–(2.13). Similarly, (4.7) follows from the independence between $\tilde{\mathbf{e}}$ and $\overline{\mathbf{Y}}_{k_1}$ given $\mathbf{X}$ [cf. (4.5) and (2.2)] and the inverse relationship between the mapping in (2.21) and the mapping in (4.5) and (2.2). Finally, (4.8) is the consequence of the following facts corresponding to properties (1) and (2) of the Hellinger distance: (1a) by (2.1) and (2.7) the combined vector

$$\mathbf{U}^\star \equiv \big(\overline{Z}^\star_{k_0,\ell}, 0 \le \ell < 2^{k_0}, W^\star_{k,2\ell}, 0 \le \ell < 2^{k-1}, k_0 < k \le k_1\big)$$

is composed of independent normal random variables independent of (and thus given) $\mathbf{X}$; (1b) by (2.2), (2.6) and (2.8) the combined vector

$$\mathbf{U} \equiv \big(\overline{Z}_{k_0,\ell}, 0 \le \ell < 2^{k_0}, W_{k,2\ell}, 0 \le \ell < 2^{k-1}, k_0 < k \le k_1\big)$$

is composed of independent normal random variables given $\mathbf{X}$; (2a) the mappings from $\mathbf{U}^\star$ to $(\overline{\mathbf{Z}}^\star_{k_0}, \mathbf{W}^\star_k,\ k_0 < k \le k_1)$, that is, $W^\star_{k,2\ell} = -W^\star_{k,2\ell+1}$ in (2.7), and then to $\overline{\mathbf{Z}}^\star_{k_1}$ in (2.10) are invertible and identical to those from $\mathbf{U}$ to $(\overline{\mathbf{Y}}_{k_0}, \mathbf{W}_k,\ k_0 < k \le k_1)$ in (2.8) and then to $\overline{\mathbf{Z}}_{k_1}$ in (2.11); (2b) the mapping $(\overline{\mathbf{Z}}^\star_{k_0}, \mathbf{W}^\star_k,\ k_0 < k \le k_1)$ to $\overline{\mathbf{Y}}^\star_{k_1}$ in (2.19)–(2.20) is identical to the mapping from $(\overline{\mathbf{Y}}_{k_0}, \mathbf{W}_k,\ k_0 < k \le k_1)$ to $\overline{\mathbf{Y}}_{k_1}$ in (2.17). This completes the proof. $\square$

The computations about the expectations of the right-hand sides of (4.6), (4.7) and (4.8), provided in Lemmas 5, 6 and 7 below, are based on the facts about the Hellinger distance between normal variables summarized in Lemmas 3 and 4 below.

LEMMA 3. *Let U_j be $N(\mu_j, \sigma_j^2)$ variables, $j = 1, 2$. Then*

$$H^2(U_1, U_2) \le 2\left(\frac{\sigma_1^2}{\sigma_2^2} - 1\right)^2 + \frac{(\mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}. \tag{4.9}$$

PROOF. With p_j being the density of U_j, $H(U_1, U_2)$ is

$$\int \left(\sqrt{p_1(t)} - \sqrt{p_2(t)}\right)^2 dt = 2\left(1 - \sqrt{\frac{2\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2}} \exp\left[-\frac{(\mu_1 - \mu_2)^2}{4(\sigma_1^2 + \sigma_2^2)}\right]\right),$$

which implies (4.9) as $1 - \sqrt{2\sigma/(1 + \sigma^2)} \le (1 - \sigma)^2 \le (\sigma^2 - 1)^2$ for $\sigma = \sigma_1/\sigma_2$. $\square$

LEMMA 4. *Let $\widetilde{B}_0$ be a Brownian bridge process. Then*

$$H^2(\widetilde{B}_0 + h, \widetilde{B}_0) \le \tfrac{1}{4} \int_0^1 \{h'(t)\}^2 \, dt \tag{4.10}$$

for all differentiable h with $h(0) = h(1)$ and $h' \in L_2[0, 1]$. Let ε_i' be i.i.d. $N(0, 1)$ variables and let c_i be constants. Then

$$H^2\big((\tilde{\varepsilon}_1', \dots, \tilde{\varepsilon}_m'), (\tilde{\varepsilon}_1', \dots, \tilde{\varepsilon}_m') + (\widetilde{c}_1, \dots, \widetilde{c}_m)\big) \le \tfrac{1}{4} \sum_{i=1}^m \widetilde{c}_i^2, \tag{4.11}$$

where $\tilde{\varepsilon}_i' = \varepsilon_i' - \sum_{i=1}^m \varepsilon_i'/m$ and $\widetilde{c}_i = c_i - \sum_{i=1}^m c_i/m$.

PROOF. Set $\tilde{\varepsilon}' = (\tilde{\varepsilon}_1', \dots, \tilde{\varepsilon}_m')$ and $\widetilde{c} = (\widetilde{c}_1, \dots, \widetilde{c}_m)$. Let ϕ_j, $j = 1, \dots, m - 1$, be orthonormal vectors in the $(m - 1)$-dimensional space $\{(a_1, \dots, a_m): \sum_j a_j = 0\}$. Then $\tilde{\varepsilon}' = \sum_{j=1}^{m-1} U_j \phi_j$ and $\widetilde{c} = \sum_{j=1}^{m-1} a_j \phi_j$, where $U_j = \langle \tilde{\varepsilon}', \phi_j \rangle$ are i.i.d. $N(0, 1)$ and $a_j = \langle \widetilde{c}, \phi_j \rangle$. Thus, by property (1) of the Hellinger distance in the proofs of Lemmas 2 and 3,

$$H^2(\tilde{\varepsilon}', \tilde{\varepsilon}' + \widetilde{c}) \le \sum_{j=1}^{m-1} H^2(U_j, U_j + a_j) \le \sum_{j=1}^{m-1} \frac{a_j^2}{4} = \sum_{i=1}^m \frac{\widetilde{c}_i^2}{4}.$$

The proof of (4.10) is omitted as it is an infinite dimensional version of (4.11) based on expansions with orthonormal basis in the space $\{g \in L^2[0, 1] : \int_0^1 g(t) \, dt = 0\}$. $\square$

LEMMA 5. *Let* $\widetilde{\mathbf{B}}^\star$, $\widetilde{\mathbf{B}}$, $\tilde{\mathbf{e}}^\star$ *and* $\tilde{\mathbf{e}}$ *be given by* (4.4), (2.13), (2.21) *and* (4.5), *respectively. Then*

$$\max\{EH^2_{f,\mathbf{X}}(\widetilde{\mathbf{B}}^\star,\widetilde{\mathbf{B}}), EH^2_{f,\mathbf{X}}(\tilde{\mathbf{e}}^\star,\tilde{\mathbf{e}})\} \le \frac{n}{4}\|f-\bar f_{k_1}\|^2. \tag{4.12}$$

REMARK. Since $\widetilde{\mathbf{B}}^\star$ and $\widetilde{\mathbf{B}}$ are independent of $\mathbf{X}$, $H_{f,\mathbf{X}}(\widetilde{\mathbf{B}}^\star,\widetilde{\mathbf{B}}) = H_f(\widetilde{\mathbf{B}}^\star,\widetilde{\mathbf{B}})$.

PROOF OF LEMMA 5. Conditionally on $\mathbf{X}$ the random vectors $\tilde e^\star_{(\ell)} \equiv (\tilde e^\star_i, X_i \in I_{k_1,\ell})$, $0 \le \ell < 2^{k_1}$, are independent of each other by (2.21), and the random vectors $\tilde e_{(\ell)} \equiv (\tilde e_i,\ X_i \in I_{k_1,\ell})$, $0 \le \ell < 2^{k_1}$, are independent of each other by (4.5). Furthermore, $E_f[\tilde e^\star_i|\mathbf{X}] = 0$ and $E_f[\tilde e_i|\mathbf{X}] = f(X_i) - \eta_{k_1,\ell}$, where $\eta_{k,\ell}$, given in (2.4), is the average of $f(X_i)$ for $X_i \in I_{k,\ell}$. Thus, by property (1) of the Hellinger distance in the proof of Lemma 2 and (4.11) of Lemma 4,

$$H^2_{f,\mathbf{X}}(\tilde{\mathbf{e}}^\star,\tilde{\mathbf{e}}) \le \sum_{\ell=0}^{2^{k_1}-1} H^2_{f,\mathbf{X}}(\tilde{\mathbf{e}}^\star_{(\ell)},\tilde{\mathbf{e}}_{(\ell)}) \le \sum_{\ell=0}^{2^{k_1}-1}\sum_{X_i\in I_{k_1,\ell}} \bigl(f(X_i)-\eta_{k_1,\ell}\bigr)^2/4.$$

Since $\eta_{k_1,\ell}$ are the minimizers of the inner sum on the right-hand side above,

$$H^2_{f,\mathbf{X}}(\tilde{\mathbf{e}}^\star,\tilde{\mathbf{e}}) \le \sum_{\ell=0}^{2^{k_1}-1}\sum_{X_i\in I_{k_1,\ell}} \bigl(f(X_i)-f_{k_1,\ell}\bigr)^2/4 = \sum_{i=1}^{n}\bigl(f(X_i)-\bar f_{k_1}(X_i)\bigr)^2/4.$$

This implies $EH^2_{f,\mathbf{X}}(\tilde{\mathbf{e}}^\star,\tilde{\mathbf{e}}) \le n\|f-\bar f_{k_1}\|^2/4$ after taking the expectation on both sides, as $E(f(X_i)-\bar f_{k_1}(X_i))^2 = \|f-\bar f_{k_1}\|^2$. The proof of the upper bound for $EH_{f,\mathbf{X}}(\widetilde{\mathbf{B}}^\star,\widetilde{\mathbf{B}})$ is omitted as it is an $L_2[0,1]$-version of the above using (4.10) of Lemma 4. $\square$

LEMMA 6. *Let* $\overline{Z}^\star_{k_0,\ell}$ *and* $\overline{Z}_{k_0,\ell}$ *be as in* (2.1) *and* (2.6). *Then, for* $k_0 > 0$,

$$\sum_{\ell=0}^{2^{k_0}-1} EH^2_{f,\mathbf{X}}(\overline{Z}^\star_{k_0,\ell},\overline{Z}_{k_0,\ell}) \le \frac{2^{2k_0+1}}{n} + 2^{k_0-1}\|f-\bar f_{k_0}\|^2. \tag{4.13}$$

For $k_0 = 0$, $EH^2_{f,\mathbf{X}}(\overline{Z}^\star_{0,0},\overline{Z}_{0,0}) \le E\{E_f[\overline{Z}_{0,0}|\mathbf{X}] - E_f\overline{Z}^\star_{0,0}\}^2/\{4\,\mathrm{Var}_f(\overline{Z}_{0,0}|\mathbf{X})\} = \|f-\bar f_0\|^2/4$.

PROOF. We shall only prove (4.13) for $k_0 > 0$. The proof for $k_0 = 0$, with $N_{k_0,0} = n$, is simpler and omitted. Since $\overline{Z}^\star_{k_0,\ell}$ and $\overline{Z}_{k_0,\ell}$ given $\mathbf{X}$ are both normal variables, by (2.4), (2.5) and Lemma 3,

$$H^2_{f,\mathbf{X}}(\overline{Z}^\star_{k_0,\ell},\overline{Z}_{k_0,\ell}) \le 2\biggl(\frac{2^{k_0}/n}{1/N_{k_0,\ell}} - 1\biggr)^2 + \frac{(f_{k_0,\ell}-\eta_{k_0,\ell})^2}{2(2^{k_0}/n + 1/N_{k_0,\ell})}, \tag{4.14}$$

where $f_{k_0,\ell}$ is as in (3.3). Since $\eta_{k_0,\ell} = \sum_{X_i \in I_{k_0,\ell}} f(X_i)/N_{k_0,\ell}$ is the average of $N_{k_0,\ell}$ i.i.d. variables $f(X_i)$ with conditional mean $f_{k_0,\ell}$ given $N_{k_0,\ell}$,

$$E_f\big[(f_{k_0,\ell} - \eta_{k_0,\ell})^2 \mid N_{k_0,\ell}\big] = \frac{\mathrm{Var}(f(X_i) \mid X_i \in I_{k_0,\ell})}{N_{k_0,\ell}} = \frac{2^{k_0} v_{k_0,\ell}}{N_{k_0,\ell}}, \tag{4.15}$$

where $v_{k_0,\ell}$ is as in (4.1). Thus, by (4.14) and algebra,

$$E\big[H^2_{f,\mathbf{X}}(\overline{Z}^\star_{k_0,\ell}, \overline{Z}_{k_0,\ell}) \mid N_{k_0,\ell}\big] \le 2\left(\frac{2^{k_0}}{n}\right)^2 \left(N_{k_0,\ell} - \frac{n}{2^{k_0}}\right)^2 + 2^{k_0-1} v_{k_0,\ell}.$$

This inequality is also valid for $N_{k_0,\ell} = 0$ since the Hellinger distance is always bounded by 2. Taking the expectation on both sides above and then summing over ℓ, we obtain

$$\sum_{\ell=0}^{2^{k_0}-1} E H^2_{f,\mathbf{X}}(\overline{Z}^\star_{k_0,\ell}, \overline{Z}_{k_0,\ell}) \le \frac{2^{2k_0+1}}{n} + 2^{k_0-1} \sum_{\ell=0}^{2^{k_0}-1} v_{k_0,\ell}.$$

This implies (4.13) by (4.2). $\square$

LEMMA 7. *Let $W^\star_{k,2\ell}$ and $W_{k,2\ell}$ be as in* (2.7) *and* (2.8). *Then, for $k_0 < k \le k_1$,*

$$\begin{aligned} \sum_{\ell=0}^{2^{k-1}-1} E H^2_{f,\mathbf{X}}(W^\star_{k,2\ell}, W_{k,2\ell}) &\le \sum_{\ell=0}^{2^k-1} E\left(\frac{\{E_f[W_{k,\ell}|\mathbf{X}] - E_f W^\star_{k,\ell}\}^2}{4\,\mathrm{Var}_f(W_{k,\ell}|\mathbf{X})}\right) \\ &\le 2^{k-1}\|f - \bar{f}_{k-1}\|^2 - \frac{3}{8} 2^k \|f - \bar{f}_k\|^2. \end{aligned} \tag{4.16}$$

PROOF. By (2.1), (2.2), (2.4), (2.5), (2.7) and (2.8),

$$E_f[W^\star_{k,2\ell}|\mathbf{X}] = \frac{1}{2}(f_{k,2\ell} - f_{k,2\ell+1}), \qquad \mathrm{Var}_f[W^\star_{k,2\ell}|\mathbf{X}] = \frac{2^{k-1}}{n}, \tag{4.17}$$

and, with $C_{k,2\ell} = \sqrt{2^{k+1}/n}\sqrt{N_{k,2\ell}N_{k,2\ell+1}/N_{k-1,\ell}}$ as in (2.8),

$$E_f[W_{k,2\ell}|\mathbf{X}] = \frac{C_{k,2\ell}}{2}(\eta_{k,2\ell} - \eta_{k,2\ell+1}), \qquad \mathrm{Var}_f[W_{k,2\ell}|\mathbf{X}] = \frac{2^{k-1}}{n}, \tag{4.18}$$

with the convention $C_{k,2\ell} = 0$ for $N_{k,2\ell}N_{k,2\ell+1} = 0$. Since both vectors $(\overline{Z}^\star_{k,2\ell}, \overline{Z}^\star_{k,2\ell+1})$ and $(\overline{Y}_{k,2\ell}, \overline{Y}_{k,2\ell+1})$ given $\mathbf{X}$ are composed of independent normal random variables, $W^\star_{k,2\ell}$ and $W_{k,2\ell}$ are both conditionally normally distributed, so that, by Lemma 3 and (4.17) and (4.18),

$$H^2_{f,\mathbf{X}}(W^\star_{k,2\ell}, W_{k,2\ell}) \le \frac{n}{2^{k+3}}\big\{C_{k,2\ell}(\eta_{k,2\ell} - \eta_{k,2\ell+1}) - (f_{k,2\ell} - f_{k,2\ell+1})\big\}^2. \tag{4.19}$$

Given $(N_{k,2\ell}, N_{k,2\ell+1})$, $\eta_{k,2\ell+j}$ are averages of collections of $N_{k,2\ell+j}$ i.i.d. random variables with means $f_{k,2\ell+j}$, $j = 0, 1$, as in the proof of Lemma 6.

Moreover, these two collections of variables are conditionally independent of each other. Thus, the conditional expectation of the right-hand side of (4.19) can be broken into a sum of variance terms as in (4.15) and a squared bias term, that is,

$$\begin{aligned}\frac{2^{k+3}}{n}&E\big[H^2_{f,\mathbf{X}}(W^\star_{k,2\ell},W_{k,2\ell})|N_{k,2\ell},N_{k,2\ell+1}\big]\\ &\le C^2_{k,2\ell}\left\{\frac{2^k v_{k,2\ell}}{N_{k,2\ell}}+\frac{2^k v_{k,2\ell+1}}{N_{k,2\ell+1}}\right\}+(1-C_{k,2\ell})^2(f_{k,2\ell}-f_{k,2\ell+1})^2.\end{aligned}\tag{4.20}$$

A suitable upper bound for the random factor $(1-C_{k,2\ell})^2$ of the squared bias term above can be found based on $N_{k-1,\ell}/4-N_{k,2\ell}N_{k,2\ell+1}/N_{k-1,\ell}=(N_{k,2\ell}-N_{k-1,\ell}/2)^2/N_{k-1,\ell}$, due to $N_{k,2\ell}+N_{k,2\ell+1}=N_{k-1,\ell}$, and the inequalities $(\sqrt{b}-\sqrt{a})^2\le(b-a)^2/b$ and $(\sqrt{b}-\sqrt{a})^2\le|b-a|$ for all nonnegative a and b; to wit

$$\begin{aligned}(1-C_{k,2\ell})^2&=(2^{k+1}/n)\Big(\sqrt{n/2^{k+1}}-\sqrt{N_{k,2\ell}N_{k,2\ell+1}/N_{k-1,\ell}}\Big)^2\\ &\le\frac{2^{k+2}}{n}\left\{\left(\sqrt{\frac{n}{2^{k+1}}}-\sqrt{\frac{N_{k-1,\ell}}{4}}\right)^2+\left(\sqrt{\frac{N_{k-1,\ell}}{4}}-\sqrt{\frac{N_{k,2\ell}N_{k,2\ell+1}}{N_{k-1,\ell}}}\right)^2\right\}\\ &\le\frac{2^k}{n}\frac{2^{k-1}}{n}\left(\frac{n}{2^{k-1}}-N_{k-1,\ell}\right)^2+\frac{2^{k+2}}{n}\frac{(N_{k,2\ell}-N_{k-1,\ell}/2)^2}{N_{k-1,\ell}}\mathbb{1}_{\{N_{k-1,\ell}>0\}}.\end{aligned}$$

Taking expectations on both sides, we find

$$E(1-C_{k,2\ell})^2\le\frac{2^k}{n}\left(1-\frac{1}{2^{k-1}}\right)+\frac{2^k}{n}\le\frac{2^{k+1}}{n},\tag{4.21}$$

since $N_{k-1,\ell}\sim\mathrm{Bin}(n,2^{1-k})$ and $(N_{k,2\ell}|N_{k-1,\ell}=m)\sim\mathrm{Bin}(m,1/2)$. The variance term in (4.20) can be written as

$$C^2_{k,2\ell}\left\{\frac{2^k v_{k,2\ell}}{N_{k,2\ell}}+\frac{2^k v_{k,2\ell+1}}{N_{k,2\ell+1}}\right\}=\frac{2^{2k+1}}{n}\left(\frac{N_{k,2\ell+1}}{N_{k-1,\ell}}v_{k,2\ell}+\frac{N_{k,2\ell}}{N_{k-1,\ell}}v_{k,2\ell+1}\right),$$

which implies

$$EC^2_{k,2\ell}\left\{\frac{2^k v_{k,2\ell}}{N_{k,2\ell}}+\frac{2^k v_{k,2\ell+1}}{N_{k,2\ell+1}}\right\}\le\frac{2^{2k}}{n}(v_{k,2\ell}+v_{k,2\ell+1}).\tag{4.22}$$

Inserting (4.21) and (4.22) into the expectation of (4.20) and then evoking (4.3), we find

$$\begin{aligned}EH^2_{f,\mathbf{X}}(W^\star_{k,2\ell},W_{k,2\ell})&\le\frac{2^k}{8}(v_{k,2\ell}+v_{k,2\ell+1})+\frac{1}{4}(f_{k,2\ell}-f_{k,2\ell+1})^2\\ &=2^{k-1}v_{k-1,\ell}-\frac{3}{8}2^k(v_{k,2\ell}+v_{k,2\ell+1}).\end{aligned}$$

Summing over $0\le\ell<2^{k-1}$, we obtain (4.16) via (4.2). $\square$

PROOF OF LEMMA 1. We shall only prove the upper bound for $H_f(Z^\star, Z)$ in the case of $k_0 > 0$, since the proof of the upper bound for $H_f((\mathbf{X}, \mathbf{Y}^\star), (\mathbf{X}, \mathbf{Y}))$ is identical and the proof for $k_0 = 0$ is simpler. Inserting the inequalities in Lemmas 5, 6 and 7 into the expectations of (4.6) and (4.8) of Lemma 2, we find

$$\begin{aligned}
EH_{f,\mathbf{X}}(Z^\star, Z) &\le EH^2_{f,\mathbf{X}}(\widetilde{\mathbf{B}}^\star, \widetilde{\mathbf{B}}) + \sum_{\ell=0}^{2^{k_0}-1} EH^2_{f,\mathbf{X}}(\overline{Z}^\star_{k_0,\ell}, \overline{Z}_{k_0,\ell}) \\
&\quad + \sum_{k=k_0+1}^{k_1} \sum_{\ell=0}^{2^{k-1}-1} EH^2_{f,\mathbf{X}}(W^\star_{k,2\ell}, W_{k,2\ell}) \\
&\le \frac{n}{4}\|f - \bar{f}_{k_1}\|^2 + \frac{2^{2k_0+1}}{n} + \frac{2^{k_0}}{2}\|f - \bar{f}_{k_0}\|^2 \\
&\quad + \sum_{k=k_0+1}^{k_1} \left(2^{k-1}\|f - \bar{f}_{k-1}\|^2 - \frac{3}{8}2^k\|f - \bar{f}_k\|^2\right).
\end{aligned}$$

This gives the $EH_{f,\mathbf{X}}(Z^\star, Z)$ portion of (3.11) by algebra. $\square$

PROOF OF THEOREM 3. For $k_0 = 0$ and given $\mathbf{X}$, the covariance structure of $(\overline{\mathbf{Z}}_0, \mathbf{W}_k,\ 1 \le k \le k_1, \widetilde{\mathbf{B}})$ is identical to that of $(\overline{\mathbf{Z}}^\star_0, \mathbf{W}^\star_k,\ 1 \le k \le k_1, \widetilde{\mathbf{B}}^\star)$, as shown in the proofs of Lemmas 5, 6 and 7. Since the equivalence mappings are linear given $\mathbf{X}$, Z must have the same conditional covariance structure as $Z^\star$, so that (3.7) holds. Since (2.13) is done through randomization, $G_n(t)$ is a continuous piecewise linear function with derivative $g_{n,k_1,\ell} = E_f[Z_{k_1,\ell}|\mathbf{X}]$ in the interval $I_{k_1,\ell}$, and $n\|g_n - f\|^2 = n\|g_n - \bar{f}_{k_1}\|^2 + n\|f - \bar{f}_{k_1}\|^2$. It follows from (2.1) and (2.12) that

$$n\|g_n - \bar{f}_{k_1}\|^2 = \sum_{\ell=0}^{2^{k_1}-1} \frac{\{g_{n,k_1,\ell} - f_{k_1,\ell}\}^2}{2^{k_1}/n} = \sum_{\ell=0}^{2^{k_1}-1} \frac{\{E_f[\overline{Z}_{k_1,\ell}|\mathbf{X}] - E_f\overline{Z}^\star_{k_1,\ell}\}^2}{\mathrm{Var}_f(\overline{Z}_{k_1,\ell}|\mathbf{X})}.$$

Since $\mathbf{U} = (\overline{Z}_{0,0}, W_{k,2\ell},\ 0 \le \ell < 2^{k-1},\ 1 \le k \le k_1)$ is a vector of independent variables given $\mathbf{X}$, the mapping from $\mathbf{Z}_{k_1}$ to $\mathbf{U}$ is orthogonal and the right-hand side above must be

$$\begin{aligned}
&\frac{\{E_f[\overline{Z}_{0,0}|\mathbf{X}] - E_f\overline{Z}^\star_{0,0}\}^2}{\mathrm{Var}_f(\overline{Z}_{0,0}|\mathbf{X})} + \sum_{k=1}^{k_1}\sum_{\ell=0}^{2^{k-1}-1} \frac{\{E_f[W_{k,2\ell}|\mathbf{X}] - E_f W^\star_{k,2\ell}\}^2}{\mathrm{Var}_f(W_{k,2\ell}|\mathbf{X})} \\
&\quad = n\|g_n - f\|^2 - n\|f - \bar{f}_{k_1}\|^2.
\end{aligned}$$

Taking the expectation on both sides above and then using Lemmas 6 and 7, we obtain

$$\begin{aligned}
&nE_f\|g_n - f\|^2 \\
&\quad \le n\|f - \bar f_{k_1}\|^2 + \|f - \bar f_0\|^2 + 4\sum_{k=1}^{k_1}\left(2^{k-1}\|f - \bar f_{k-1}\|^2 - \frac{3}{8}2^k\|f - \bar f_k\|^2\right) \\
&\quad \le \frac{5}{2}\left\{\|f - \bar f_0\|^2 + \sum_{k=0}^{k_1} 2^k\|f - \bar f_k\|^2\right\} + (n - 2^{k_1+1})\|f - \bar f_{k_1}\|^2.
\end{aligned}$$

This implies (3.8) and the first inequality of (3.9). The second inequality of (3.9) is a well-known fact (cf. the proof of Theorems 1 and 5). □

PROOF OF THEOREM 4. For each ℓ and those $X_i \in I_{k_1,\ell}$, $g_n^\star(X_i) = E_f[Y_i^\star|\mathbf{X}] = E_f[\overline{Y}^\star_{k_1,\ell}|\mathbf{X}]$ does not depend on i, due to the independence of $e_i^\star$ and $\mathbf{X}$ in (2.21). Since the mapping from $(\overline{Z}^\star_{0,0}, W^\star_{k,2\ell},\ 0 \le \ell < 2^{k-1},\ 1 \le k \le k_1)$ to $\overline{Y}^\star_{k_1}$ is orthogonal given $\mathbf{X}$ and $\mathrm{Var}_f(\overline{Y}^\star_{k_1,\ell}|\mathbf{X}) = \mathrm{Var}_f(\overline{Y}_{k_1,\ell}|\mathbf{X}) = 1/N_{k_1,\ell}$,

$$\begin{aligned}
&\sum_{i=1}^n \big(g_n^\star(X_i) - f(X_i)\big)^2 \\
&\quad = \sum_{N_{k_1,\ell}>0}\left[\frac{\{E_f[\overline{Y}^\star_{k_1,\ell}|\mathbf{X}] - E_f[\overline{Y}_{k_1,\ell}|\mathbf{X}]\}^2}{\mathrm{Var}_f(\overline{Y}^\star_{k_1,\ell}|\mathbf{X})} + \sum_{X_i\in I_{k_1,\ell}}\{f(X_i) - \eta_{k_1,\ell}\}^2\right] \\
&\quad \le \frac{\{E_f[\overline{Z}_{0,0}|\mathbf{X}] - E_f\overline{Z}^\star_{0,0}\}^2}{\mathrm{Var}_f(\overline{Z}^\star_{0,0}|\mathbf{X})} + \sum_{k=1}^{k_1}\sum_{\ell=0}^{2^{k-1}-1}\frac{\{E_f[W_{k,2\ell}|\mathbf{X}] - E_f W^\star_{k,2\ell}\}^2}{\mathrm{Var}_f(W^\star_{k,2\ell}|\mathbf{X})} \\
&\qquad + \sum_{i=1}^n\{f(X_i) - \bar f_{k_1}(X_i)\}^2.
\end{aligned}$$

Note that $E_f[\overline{Y}_{k_1,\ell}|\mathbf{X}] = \eta_{k_1,\ell}$ as in (2.4). The proof is completed as the rest follows directly from the corresponding parts in those of Lemma 5 and Theorem 3. □

REFERENCES

BROWN, L. D. and LOW, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24** 2384–2398.

BROWN, L. D. and ZHANG, C.-H. (1996). Coupling inequalities for some random design matrices and asymptotic equivalence of nonparametric regression and white noise. Technical Report 96-020, Dept. Statistics, Rutgers Univ.

BROWN, L. D. and ZHANG, C.-H. (1998). Asymptotic nonequivalence of nonparametric experiments when the smoothness index is $1/2$. *Ann. Statist.* **26** 279–287.
DONOHO, D. L. and JOHNSTONE, I. M. (1998). Minimax estimation via wavelet shrinkage. *Ann. Statist.* **26** 879–921.
DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24** 508–539.
EFROMOVICH, S. (1998). Simultaneous sharp estimation of functions and their derivatives. *Ann. Statist.* **26** 273–278.
EFROMOVICH, S. and SAMAROV, A. (1996). Asymptotic equivalence of nonparametric regression and white noise model has its limits. *Statist. Probab. Lett.* **28** 143–145.
FAN, J. (1991). On the estimation of quadratic functionals. *Ann. Statist.* **19** 1273–1294.
LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.
LE CAM, L. and YANG, G. L. (1990). *Asymptotics in Statistics*: *Some Basic Concepts*. Springer, New York.
NUSSBAUM, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430.
PINSKER, M. S. (1980). Optimal filtering of square integrable signals in Gaussian white noise. *Problems Inform. Transmission* **16** 120–133.
TSYBAKOV, A. B. (1998). Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Ann. Statist.* **26** 2420–2469.

L. D. BROWN
DEPARTMENT OF STATISTICS
THE WHARTON SCHOOL
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104
E-MAIL: lbrown@wharton.upenn.edu

T. T. CAI
DEPARTMENT OF STATISTICS
THE WHARTON SCHOOL
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104
E-MAIL: tcai@wharton.upenn.edu

M. G. LOW
DEPARTMENT OF STATISTICS
THE WHARTON SCHOOL
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104

C.-H. ZHANG
DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
PISCATAWAY, NEW JERSEY 08854