WAVELET REGRESSION FOR RANDOM UNIFORM DESIGN

by

Lawrence D. Brown and T. Tony Cai
University of Pennsylvania | Purdue University

Technical Report #97-15

Department of Statistics
Purdue University
West Lafayette, IN USA

August 1997

◇

Wavelet Regression For Random Uniform Design

Lawrence D. Brown
Department of Statistics, University of Pennsylvania
T. Tony Cai
Department of Statistics, Purdue University

August 24, 1997

Abstract

The current research on wavelet regression has been mostly focused on equispaced samples. In general nonequispaced samples require different treatment. And the currently available wavelet methods for nonequispaced samples are relatively difficult to implement.

In the present paper, we consider samples with random uniform design. We show that if the samples have random uniform design, the universal thresholding method can be applied directly to the samples as if they were equispaced. The resulting estimator achieves within a logarithmic factor from the minimax rate of convergence over a family of Hölder classes. Simulation results also show that the mean squared error for samples with random uniform design is comparable to that for samples with equispaced design.

Keywords: wavelets, nonparametric regression, minimax, adaptivity, Hölder class.
AMS 1991 Subject Classification: Primary 62G07, Secondary 62G20.

1 Introduction

Wavelet shrinkage methods have been very successful in nonparametric regression. But so far most of the wavelet regression methods have been focused on equispaced samples. There, data are transformed into empirical wavelet coefficients and threshold rules are applied to the coefficients. The estimators are obtained via the inverse transform of the denoised wavelet coefficients. The most widely used wavelet shrinkage method for equispaced samples is the Donoho-Johnstone's VisuShrink procedure (Donoho & Johnstone (1992), Donoho, Johnstone, Kerkyacharian & Picard (1995)) . The VisuShrink procedure has three steps:

1. Transform the noisy data via the discrete wavelet transform;

2. Denoise the empirical wavelet coefficients by "hard" or "soft" thresholding rules with threshold $\lambda = \epsilon\sqrt{2\log n}$.

3. Estimate function f at the sample points by inverse discrete wavelet transform of the denoised wavelet coefficients.

This procedure is adaptive and easy to implement. The computational cost is of $O(n)$. And with high probability, VisuShrink estimators are at least as smooth as the target function. The estimators produced by the procedure achieve minimax convergence rates up to a logarithmic penalty over a wide range of function classes.

In many statistical applications, however, the samples are nonequispaced. It is shown that the procedure might produce suboptimal estimators if it is applied directly to nonequispaced samples (Cai, 1996). Wavelet methods for samples with nonequispaced designs have been studied by Brown and Cai (1997) and Hall and Turlach (1996). Brown and Cai (1997) introduced a wavelet shrinkage method for samples with fixed nonequispaced designs based on approximation approach. It is shown that the estimator attains near-minimaxity across a range of piecewise Hölder classes. Hall and Turlach (1996) proposed interpolation methods for samples with random designs. They used samples with random uniform design as examples for their methods. Despite the asymptotic near-optimality for these nonequispaced methods, the estimators are computationally much harder to implement than the VisuShrink for equispaced samples.

In the present paper, we consider the special case of samples with random uniform design. We show that in this special case the samples can in fact be treated as if they were equispaced. That is, the VisuShrink procedure of Donoho and Johnstone can be applied directly to the data and the resulting estimator adaptively achieves within a logarithmic factor of the optimal convergence rate across a range of Hölder classes. Therefore, we have a fast estimation procedure for samples with random uniform design. Simulation is conducted to evaluate the numerical performance of the method. It is shown that the mean squared error is comparable to that of the samples with truly equispaced designs.

In Section 2 we describe the method and state the asymptotic optimality property of the estimator. Section 3 summarizes the simulation results. Some relevant results on wavelet approximation is presented in Section 4. Section 5 contains a concise proof of the main results.

2 Methodology

2.1 Wavelets

Let ϕ and ψ denote the orthogonal father and mother wavelet functions. The functions ϕ and ψ are assumed to be compactly supported with associated discrete wavelet transform W. Assume ψ has r vanishing moments and ϕ satisfies $\int \phi = 1$. Let

$$\phi_{jk}(t) = 2^{j/2}\phi(2^j t - k), \quad \psi_{jk}(t) = 2^{j/2}\psi(2^j t - k)$$

And denote the periodized wavelets

$$\phi^p_{jk}(t) = \sum_{l \in Z} \phi_{jk}(t - l), \quad \psi^p_{jk}(t) = \sum_{l \in Z} \phi_{jk}(t - l) \quad \text{for } t \in [0, 1]$$

For the purposes of this paper, we use the periodized wavelet bases on [0, 1]. The collection $\{\phi^p_{j_0 k}, k = 1, ..., 2^{j_0}; \psi^p_{jk}, j \geq j_0, k = 1, ..., 2^j\}$ constitutes such an orthonormal basis of $L_2[0, 1]$. Note that the basis functions are periodized at the boundary. The superscript "p" will be suppressed from the notations for convenience. This basis has an associated exact orthogonal Discrete Wavelet Trasnform (DWT) that transforms data into wavelet coefficient domains.

For a given square-integrable function f on [0, 1], denote

$$\xi_{jk} = \langle f, \phi_{jk} \rangle, \qquad \theta_{jk} = \langle f, \psi_{jk} \rangle$$

So the function f can be expanded into a wavelet series:

$$f(x) = \sum_{k=1}^{2^{j_0}} \xi_{j_0 k} \phi_{j_0 k}(x) + \sum_{j=j_0}^{\infty} \sum_{k=1}^{2^j} \theta_{jk} \psi_{jk}(x) \tag{1}$$

Wavelet transform decomposes a function into different resolution components. In (1), $\xi_{j_0 k}$ are the coefficients at the coarsest level. They represent the gross structure of the function f. And θ_{jk} are the wavelet coefficients. They represent finer and finer structures of the function f as the resolution level j increases.

We note that the DWT is an orthogonal transform, so it transforms i.i.d. Gaussian noise to i.i.d. Gaussian noise and it is norm-preserving. This important property of DWT allows us to transform the problem in the function domain into a problem in the sequence domain of the wavelet coefficients with isometry of risks.

A good introduction to wavelets is given by Strang (1989). For a detailed treatment on wavelets, the readers are refered to Daubechies (1992) and Meyer (1990).

2.2 The Estimator

Consider the nonparametric regression model:

$$y_i = f(x_i) + \epsilon z_i \tag{2}$$

$i = 1, 2, ..., n(= 2^J)$, x_i's are independently uniformly distributed on [0, 1], z_i's are independent $N(0,1)$ variables and independent of x_i's.

The function $f(\cdot)$ is an unknown function of interest. We wish to estimate $f(\cdot)$ globally with small integrated mean squared error:

$$R(\hat{f}, f) = E \int_0^1 (\hat{f}(t) - f(t))^2 \, dt$$

Let $0 \le x_{(1)} < x_{(2)} < ... < x_{(n)} \le 1$ be the order statistics of the x_i's. Now relabel y_i's and z_i's according the order of the x_i's. For convenience, we use the same label. So,

$$y_i = f(x_{(i)}) + \epsilon z_i \tag{3}$$

Now we observed $(x_{(1)}, y_1), (x_{(2)}, y_2), \cdots, (x_{(n)}, y_n)$ with x_i independently uniformly distributed on [0, 1]. So $x'_{(i)}s$ are not equispaced in general. But we pretend that $x_{(i)}$ is $Ex_{(i)} = i/(n+1)$. That is, we pretend to have an equispaced sample:

$$(\frac{1}{n+1}, y_1), (\frac{2}{n+1}, y_2), \cdots, (\frac{n}{n+1}, y_n).$$

We apply Donoho and Johnstone's VisuShrink procedure directly to $y = \{y_1, y_2, ..., y_n\}$.

Let $\tilde{\theta} = W \cdot n^{-1/2} y$ be the discrete wavelet transform of $n^{-1/2}y$. Write

$$\tilde{\theta} = (\tilde{\xi}_{j_0 1}, \cdots, \tilde{\xi}_{j_0 2^{j_0}}, \tilde{\theta}_{j_0 1}, \cdots, \tilde{\theta}_{j_0 2^{j_0}}, \cdots, \tilde{\theta}_{J-1,1}, \cdots, \tilde{\theta}_{J-1,2^{J-1}})^T$$

Here $\tilde{\xi}_{j_0 k}$ are the empirical coefficients of the father wavelets at the lowest resolution level. They represent the gross structure of the function and are usually not thresholded. The coefficients $\tilde{\theta}_{jk} (j = 1, \cdots, J-1, k = 1, \cdots, 2^j)$ are fine structure wavelet terms.

We estimate the function f by

$$\hat{f}_*(x) = \sum_{k=1}^{2^{j_0}} \tilde{\xi}_{j_0 k} \phi_{j_0 k}(x) + \sum_{j=j_0}^{J-1} \sum_{k=1}^{2^j} \tilde{\theta}_{jk} I(|\tilde{\theta}_{jk}| > \epsilon \sqrt{2n^{-1} \log n}) \psi_{jk}(x)$$

If one is interested in estimating the function at the sample points, then the three step procedure can be applied:

1. Transform $y = \{y_1, y_2, ..., y_n\}$ into wavelet domain via discrete wavelet transform W:

$$V = W \cdot y$$

2. Denoise the empirical wavelet coefficients by soft thresholding:

$$\hat{\theta}_{jk} = \eta_\lambda^s(v_{jk}) = sgn(v_{jk})(|v_{jk}| - \lambda)_+, \text{ where } \lambda = \epsilon\sqrt{2\log n}$$

3. Obtain the estimator via the inverse transform of the denoised wavelet coefficients.

$$(f_*\widehat{(x_{(k)})})_{k=1}^n = W^{-1} \cdot \hat{\theta}$$

The estimator is adaptive and easy to implement.

Theorem 1 *Suppose that the sample* $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ *is observed as in (2) and the mother wavelet* ψ *has r vanishing moments. Then the estimator constructed above achieves within a logarithmic factor of the optimal convergence rate over the a range of Hölder classes* $\Lambda^\alpha(M)$ *(defined in Section 4) with* $1/2 \le \alpha \le r$*. That is,*

$$\sup_{f\in\Lambda^\alpha(M)} E\|\hat{f}_* - f\|_2^2 \le C \cdot (\frac{\log n}{n})^{\frac{2\alpha}{1+2\alpha}}(1 + o(1)) \tag{4}$$

$$\sup_{f\in\Lambda^\alpha(M)} \frac{1}{n}\sum E\|\widehat{f_*(x_i)} - f(x_i)\|_2^2 \le C \cdot (\frac{\log n}{n})^{\frac{2\alpha}{1+2\alpha}}(1 + o(1)) \tag{5}$$

for all $M \in (0, \infty)$ *and* $\alpha \in [1/2, r]$.

Remark: The same results hold for hard threshold estimators. The results show that in the case of random design with uniformly distributed x_i's, we can treat it as if they are fixed equispaced design.

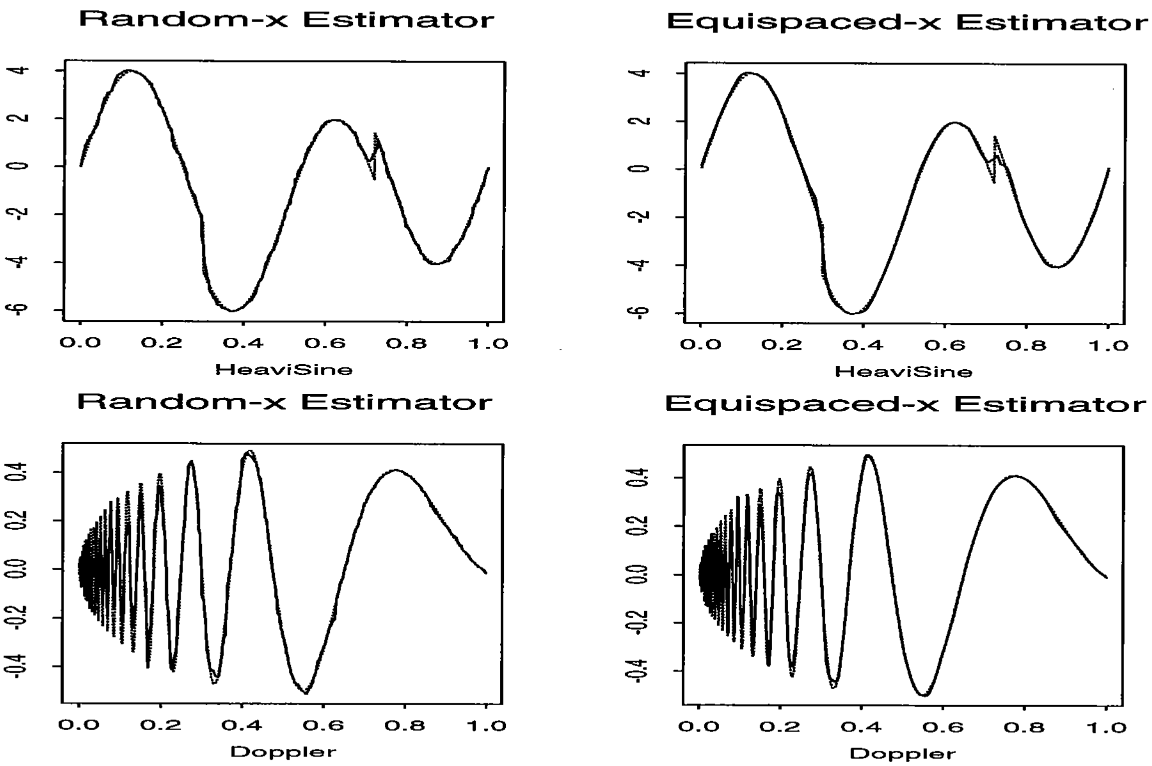
3 Simulations

A simulation study is conducted to compare the estimator based on random-x samples with those estimator based on truly equispaced samples. The results show that the quality of the estimator based on random-x samples are very close to the estimators based on equispaced samples. So simulation confirms our theoretical results.

We studied eight functions representing different level of spatial variability. The test functions are Doppler, HeaviSine, Bumps, Blocks, Angles, Blip, Corner, and Wave. The definition of the test functions is given in the appendix. For each of the eight objects under study, we compare the estimators at two noise levels, one with signal-to-noise ratio SNR = 5 and another with SNR = 7. Sample sizes from $n = 512$ to $n = 8192$ are studied.

We report in Table 1 and Table 2 the mean squared errors over 100 replications of the eight test functions. The wavelet used is the Symmlet "s8". From the tables, we can see the MSEs are comparable for estimators based on random-x samples and for estimators based on equispaced-x samples.

The following plots compare the visual quality of the estimators. The solid line is the estimator and dotted line is the true function. The signal-to-noise ratio is 7, the sample size is 1024, and the wavelet used is the Daubechies Symmlet "s8". For each function, one is based on a sample with uniformly distributed design and another is based on a sample with equispaced design. Both samples have the same noise level. One can see from the plots, the quality of the estimators are comparable.



In the appendix, we also include more plots to compare the visual quality of the two estimators. On each plot of the estimator, the solid line is the estimator and dash line is the true function. Two estimators are plotted on each page, one is based on a sample with uniformly distributed design and another is based on a sample of the same noise level with equispaced design.

4 Wavelet Approximation

Wavelets provide smoothness characterization of function spaces. Many traditional smoothness spaces, for example Hölder spaces, Sobolev spaces and Besov spaces, can be completely characterized by wavelet coefficients. See Meyer [11]. In the present paper, we consider the estimation problem over a range of Hölder classes.

Definition 1 *We define the following Hölder classes $\Lambda^\alpha(M)$:*

$$\begin{aligned} (i). \quad & \text{if } \alpha \le 1, \ \Lambda^\alpha(M) = \{f : |f(x) - f(y)| \le M\,|x-y|^\alpha\} \\ (ii). \quad & \text{if } \alpha > 1, \ \Lambda^\alpha(M) = \{f : |f^{(\lfloor\alpha\rfloor)}(x) - f^{(\lfloor\alpha\rfloor)}(y)| \le M\,|x-y|^{\alpha'} \text{ and } |f'(x)| \le M\} \end{aligned}$$

where $\lfloor\alpha\rfloor$ is the largest integer less than α and $\alpha' = \alpha - \lfloor\alpha\rfloor$.

The wavelet coefficients of functions in a Hölder classes $\Lambda^\alpha(M)$ decay fast.

Lemma 1 *Let $f \in \Lambda^\alpha(M)$ and let the wavelet function ψ has r vanishing moments with $r \ge \alpha$. Let $\theta_{jk} = \langle f, \psi_{jk}\rangle$ be wavelet coefficients of f. Then*

$$|\theta_{jk}| \le C \cdot 2^{-j(1/2+\alpha)} \tag{6}$$

where C is a constant depending on M and the wavelet basis only.

If one has a sampled function $\{f(k/(n+1))\}_{k=1}^n$ with $n = 2^J$, one can utilize a wavelet basis to get a good approximation of the entire function f. Denote $s(\alpha) = min(\alpha, 1)$. We have the following.

Proposition 1 *Suppose that $f \in \Lambda^\alpha(M)$, and let $\xi_{Jk} = \langle f, \phi_{Jk}\rangle$, then*

$$|n^{-\frac{1}{2}} f(\frac{k}{n+1}) - \xi_{Jk}| \le C \cdot n^{-(1/2+s(\alpha))} \tag{7}$$

According to this result, we may use $n^{-\frac{1}{2}} f(\frac{k}{n+1})$ as an approximation of ξ_{Jk}. This means that if an equispaced sampled function is given, we can use a wavelet basis to get an approximation of the entire function f. To be more specific, we can use $f_n(x) = \sum_{k=1}^n n^{-\frac{1}{2}} f(\frac{k}{n+1})\phi_{Jk}(x)$ as an approximation of f. Furthermore, the approximation error can be bounded based on the sample size and the smoothness of the function. We quote the following result from Cai (1996).

Proposition 2 *Suppose that $f \in \Lambda^\alpha(M)$. Let $f_n(x) = \sum_{k=1}^n n^{-\frac{1}{2}} f(\frac{k}{n+1})\phi_{Jk}(x)$ Then the approximation error satisfies*

$$\|f_n - f\|_2^2 \le C n^{-2s(\alpha)} \tag{8}$$

5 Proof

We need some preparations before we prove the theorem. First some well known results on the order statistics of uniform variables.

Lemma 2 *Let x_i be iid uniform random variables on [0, 1]. And let $0 \le x_{(1)} < x_{(2)} < \dots < x_{(n)} \le 1$ be the order statistics. Then $x_{(k)}$ is distributed as Beta(k, n - k + 1). In particular,*

$$Ex_{(k)} = \frac{k}{n+1}, \quad Ex^2_{(k)} = \frac{k+k^2}{(n+1)(n+2)}, \quad Varx_{(k)} = \frac{(n+1)k-k^2}{(n+1)^2(n+2)}$$

Now let us consider the noiseless case. We want to simply use $f(x_{(i)})$ as an approximation of $f(\frac{i}{n+1})$. We are interested in knowing the approximation error. Denote E_1 the conditional expectation given $x_1, x_2, \cdots, x_n$ and denote E_x the expectation with respect to $x_1, x_2, \cdots, x_n$.

Lemma 3 *The upper bound of the approximation error is*

$$\sup_{f\in\Lambda^\alpha(M)} \frac{1}{n}\sum E_x(f(x_{(k)}) - f(\frac{k}{n+1}))^2 \le Cn^{-s(\alpha)} \tag{9}$$

Proof: For $f \in \Lambda^\alpha(M)$, we have $|f(x) - f(y)| \le C\,|x-y|^{s(\alpha)}$. Hence,

$$\begin{aligned}
&\frac{1}{n}\sum_k E_x(f(x_{(k)}) - f(\frac{k}{n+1}))^2 \le \frac{C}{n}\sum_k E_x(x_{(k)} - \frac{k}{n+1})^{2s(\alpha)} \\
&\le \frac{C}{n}\sum_k [E_x(x_{(k)} - \frac{k}{n+1})^2]^{s(\alpha)} = \frac{C}{n}\sum_k [\frac{(n+1)k-k^2}{(n+1)^2(n+2)}]^{s(\alpha)} \\
&\le \frac{C\sum_k k^{s(\alpha)}}{n(n+1)^{s(\alpha)}(n+2)^{s(\alpha)}} \le \frac{C(n+1)^{1+s(\alpha)}}{n(n+1)^{s(\alpha)}(n+2)^{s(\alpha)}} \\
&\le Cn^{-s(\alpha)} \quad \blacksquare
\end{aligned}$$

To prove the main result, we also need the following upper bound of the risk of threshold estimator of a univariate normal mean. Similar bound holds for hard threshold. The proof can be found in Cai (1996).

Lemma 4 *Suppose that $y \sim N(\theta, n^{-1}\epsilon^2)$. Then $\hat{\theta} = \eta^s_\lambda(y)$ with $\lambda = \epsilon\sqrt{2n^{-1}\log n}$ satisfies*

$$E(\hat{\theta}-\theta)^2 \le (2\theta^2 + n^{-2}\epsilon^2) \wedge (2\log n + 1)n^{-1}\epsilon^2 \tag{10}$$

Proof of Theorem 1: We give the proof of (4) only. The proof of (5) is similar. First, some notations. We use ξ_{jk} as coefficients of ϕ_{jk} (the "father wavelets"), and use θ_{jk} as coefficients of ψ_{jk} (the "mother wavelets"). The $\tilde{\xi}_{j_0k}$ are the empirical coefficients at the coarsest level. They represent the gross structure of the function and they are usually not thresholded. The discrete wavelet transform W is an orthogonal transform, so it is norm-preserving. This fact is useful in the proof.

Let $\tilde{f}(x) = \sum_i n^{-\frac{1}{2}} y_i \phi_{Jk}(x)$. Then $\tilde{f}(x)$ can be written as

$$\begin{aligned}
\tilde{f}(x) &= \sum_i [n^{-\frac{1}{2}} f(x_{(i)}) + n^{-\frac{1}{2}} \epsilon z_i] \phi_{Ji}(x) \\
&= \sum_i [\xi_{Ji} + \overbrace{(n^{-\frac{1}{2}} f(\frac{i}{n+1}) - \xi_{Ji})}^{A} + \overbrace{(n^{-\frac{1}{2}} f(x_{(i)}) - n^{-\frac{1}{2}} f(\frac{i}{n+1}))}^{B} + \overbrace{n^{-\frac{1}{2}} \epsilon z_i}^{R}] \phi_{Ji}(x) \\
&= \sum_k [\xi_{j_0 k} + \tilde{a}_{j_0 k} + \tilde{b}_{j_0 k} + \tilde{r}_{j_0 k}] \phi_{j_0 k}(x) + \sum_j \sum_k [\theta_{jk} + a_{jk} + b_{jk} + r_{jk}] \psi_{jk}(x)
\end{aligned}$$

Here the $\xi_{j_0 k}$ and θ_{jk} are the discrete wavelet transform of ξ_{Ji} , and likewise $\tilde{a}_{j_0 k}$ and a_{jk} the transform of the term A, $\tilde{b}_{j_0 k}$ and b_{jk} the transform of B and $\tilde{r}_{j_0 k}$ and r_{jk} the transform of R.

Let $\tilde{\xi}_{j_0 k} = \xi_{j_0 k} + \tilde{a}_{j_0 k} + \tilde{b}_{j_0 k} + \tilde{r}_{j_0 k}$ be the coefficients of gross structure terms. These coefficients are not thresholded. Set

$$\hat{\xi}_{j_0 k} = \tilde{\xi}_{j_0 k}$$

Let $\theta'_{jk} = \theta_{jk} + a_{jk} + b_{jk}$ and let $v_{jk} = \theta'_{jk} + r_{jk}$ be the noisy empirical wavelet coefficients. Then $v_{jk} \sim N(\theta'_{jk}, n^{-1}\epsilon^2)$. Now let $\lambda = \epsilon\sqrt{2n^{-1}\log n}$. And let $\hat{\theta}_{jk} = sign(v_{jk})(|v_{jk}| - \lambda)_+$.

Set the estimator of the regression function f to be

$$\hat{f}_*(x) = \sum_k \hat{\xi}_{j_0 k} \phi_{j_0 k}(x) + \sum_{j=j_0}^{J-1} \sum_k \hat{\theta}_{jk} \psi_{jk}(x)$$

$$E\|\hat{f}_* - f\|_2^2 = E_x(E_1\|\hat{f}_* - f\|_2^2)$$

$$E_1\|\hat{f}_* - f\|_2^2 = \sum_k E_1(\hat{\xi}_{j_0 k} - \xi_{j_0 k})^2 + \sum_{j=j_0}^{J-1} \sum_k E_1(\hat{\theta}_{jk} - \theta_{jk})^2 + \sum_{j=J}^{\infty} \sum_k \theta_{jk}^2$$

From Lemma 1, we have

$$\sum_{j=J}^{\infty} \sum_k \theta_{jk}^2 = O(n^{-2\alpha}) \qquad (11)$$

Also we have

$$\sum_k E_1(\hat{\xi}_{j_0 k} - \xi_{j_0 k})^2 = 2^{j_0} n^{-1}\epsilon^2 + \sum_k (\tilde{a}_{j_0 k} + \tilde{b}_{j_0 k})^2 \le 2^{j_0} n^{-1}\epsilon^2 + 2\sum_k \tilde{a}_{j_0 k}^2 + 2\sum_k \tilde{b}_{j_0 k}^2 \qquad (12)$$

Now consider $E_1(\hat{\theta}_{jk} - \theta_{jk})^2$. Apply Lemma 4,

$$\begin{aligned}
E_1(\hat{\theta}_{jk} - \theta_{jk})^2 &\le 2E_1(\hat{\theta}_{jk} - \theta'_{jk})^2 + 2a_{jk}^2 + 2b_{jk}^2 \\
&\le (2(\theta'_{jk})^2 + n^{-2}\epsilon^2) \wedge (2\log n + 1) n^{-1}\epsilon^2 + 2a_{jk}^2 + 2b_{jk}^2 \\
&\le 8\theta_{jk}^2 \wedge 3n^{-1}\epsilon^2 \log n + 10a_{jk}^2 + 10b_{jk}^2 + n^{-2}\epsilon^2
\end{aligned}$$

Let J_1 satisfy $2^{J_1} = (n/\log n)^{1/(1+2\alpha)}$. Then,

$$\begin{aligned}\sum_{j,k} E_1(\hat{\theta}_{jk} - \theta_{jk})^2 &\leq \sum_{j=j_0}^{J_1-1}\sum_k 3n^{-1}\epsilon^2 \log n + \sum_{j=J_1}^{J-1}\sum_k 8\theta_{jk}^2 + 10\sum_{j=j_0}^{J-1}\sum_k (a_{jk}^2 + b_{jk}^2) + n^{-1}\epsilon^2 \\ &\leq C(n^{-1}\log n)^{\frac{2\alpha}{1+2\alpha}}(1+o(1)) + 10\sum_{j=j_0}^{J-1}\sum_k a_{jk}^2 + 10\sum_{j=j_0}^{J-1}\sum_k b_{jk}^2 \qquad (13)\end{aligned}$$

It follows from Lemma 1 and Lemma 3 that

$$\sum_k \tilde{a}_{j_0k}^2 + \sum_{j=j_0}^{J-1}\sum_k a_{jk}^2 = \sum_i (n^{-\frac{1}{2}} f(\frac{i}{n+1}) - \xi_{Ji})^2 \leq C \cdot n^{-2s(\alpha)} \qquad (14)$$

$$E_x(\sum_k \tilde{b}_{j_0k}^2 + \sum_{j=j_0}^{J-1}\sum_k b_{jk}^2) = \frac{1}{n}\sum_{i=1}^{n} E_x(f(x_{(i)}) - f(\frac{i}{n+1}))^2 \leq Cn^{-s(\alpha)} \qquad (15)$$

For $\alpha \geq \frac{1}{2}$, $s(\alpha) \geq \frac{2\alpha}{1+2\alpha}$. Now it follows from (11) – (15) that

$$E\|\hat{f}_* - f\|_2^2 \leq C \cdot (\frac{\log n}{n})^{\frac{2\alpha}{1+2\alpha}}(1+o(1)) \quad \blacksquare$$

References

[1] Brown, L.D. & Cai, T. (1997). Wavelet Shrinkage For Nonequispaced Samples, Technical Report #97-06, Department of Statistics, Purdue University.

[2] Brown, L.D. & Low, M.G. (1996). Asymptotic Equivalence of Nonparametric Regression and White Noise. *Annals of Statistics.* **24** 2384 - 2398.

[3] Cai, T. (1996). Nonparametric Function Estimation via Wavelets. Ph.D Thesis, Cornell University.

[4] Casella, G. & Strawderman, W.E. (1981). Estimating a Bounded Normal Mean. *Annals of Statistics.* **9** 870-878.

[5] Daubechies, I. (1992). *Ten Lectures on Wavelets* SIAM: Philadelphia.

[6] Donoho, D.L. & Johnstone, I.M. (1992). Ideal Spatial Adaptation via Wavelet Shrinkage. *Biometrika,* ***81***, 425–455.

[7] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. & Picard, D. (1995). Wavelet Shrinkage: Asymptopia?, *J. Roy. Stat. Soc. Ser. B,* ***57***, 301–369.

[8] Hall, P., Penev, S., Kerkyacharian, G. & Picard, D. (1996). Numerical Performance of Block Thresholded Wavelet Estimators, manuscript.

[9] Hall, P. & Turlach, B.A. (1996). Interpolation Methods for Nonlinear Wavelet Regression with Irregularly Spaced Design. *Annals of Statistics.* To appear.

[10] Marron, J.S., Adak, S., Johnstone, I.M., Neumann, M.H. & Patil, P. (1995). Exact Risk Analysis of Wavelet Regression. Manuscript.

[11] Meyer, Y. (1990). *Ondelettes et Opérateurs: I. Ondelettes.* Hermann et Cie, Paris.

[12] Strang, G. (1989). Wavelet And Dilation Equations: A Brief Introduction. *SIAM Review,* 31(4), 614 - 627.

6 Appendix

The test functions come from various sources. The functions Doppler, HeaviSine, Bumps and Blocks are from Donoho and Johnstone [6]; the functions Blip and Wave come from Marron, Adak, Johnstone, Neumann and Patil [10]; and the function Corner is from Hall, Penev, Kerkyacharian and Picard [8]. Formulae of the test functions are as follows.

1. *Doppler.*

$$f(x) = \sqrt{x(1-x)}\sin(2.1\pi/(x+.05))$$

2. *HeaviSine.*

$$f(x) = 4\sin 4\pi x - sgn(x-.3) - sgn(.72-x)$$

3. *Bumps.*

$$f(x) = \sum h_j K((x-x_j)/w_j) \qquad K(x) = (1+|x|)^{-4}.$$

$$\begin{array}{llllllllllll} (x_j) & = (.1, & .13, & .15, & .23, & .25, & .40, & .44, & .65, & .76, & .78, & .81) \\ (h_j) & = (4, & 5, & 3, & 4, & 5, & 4.2, & 2.1, & 4.3, & 3.1, & 5.1, & 4.2) \\ (w_j) & = (.005, & .005, & .006, & .01, & .01, & .03, & .01, & .01, & .005, & .008, & .005) \end{array}$$

4. *Blocks.*

$$f(x) = \sum h_j K(x-x_j) \qquad K(x) = (1+sgn(x))/2.$$

$$\begin{array}{llllllllllll} (x_j) & = (.1, & .13, & .15, & .23, & .25, & .40, & .44, & .65, & .76, & .78, & .81) \\ (h_j) & = (4, & -5, & 3, & -4, & 5, & -4.2, & 2.1, & 4.3, & -3.1, & 5.1, & -4.2) \end{array}$$

5. *Angles.*

$$\begin{aligned} f(x) = \quad & 14xI_{[0,1/7)}(x) + (10-56x)I_{[1/7,3/14)}(x) \\ & (28x-8)I_{[3/14,2/7)}(x) + (2-7x)I_{[2/7,1/2)}(x) \\ & (49x-26)I_{[1/2,4/7)}(x) + (22-35x)I_{[4/7,5/7)}(x) \\ & (28x-23)I_{[5/7,6/7)}(x) + (7-7x)I_{[6/7,1)}(x) \end{aligned}$$

6. *Blip.*

$$f(x) = (0.32 + 0.6x + 0.3e^{-100(x-0.3)^2}) I_{(0,.8]}(x) + (-0.28 + 0.6x + 0.3e^{-100(x-0.3)^2}) I_{[.8,1]}(x)$$

7. *Corner.*

$$f(x) = 10x^3(1-4x^2) I_{(0,.5]}(x) + 3(0.125 - x^3)x^4 I_{(.5,.8]}(x) + (70(x-.8)^3 - 0.4756) I_{(.8,1]}(x)$$

8. *Wave.*

$$f(x) = .5 + .2\cos(4\pi x) + .1\cos(24\pi x)$$

Table 1: Mean Squared Error From 100 Replications

	SNR = 5		SNR = 7	
n	Random-x	Equispaced-x	Random-x	Equispaced-x
Doppler				
512	3.953	2.921	2.846	1.814
1024	2.542	1.890	1.755	1.185
2048	1.619	1.233	1.089	0.767
4096	0.848	0.687	0.577	0.427
8192	0.522	0.445	0.339	0.266
HeaviSine				
512	0.626	0.544	0.455	0.396
1024	0.425	0.399	0.299	0.282
2048	0.283	0.295	0.194	0.196
4096	0.170	0.192	0.115	0.119
8192	0.115	0.123	0.075	0.076
Bumps				
512	8.302	9.271	5.330	5.812
1024	6.121	5.955	3.780	3.576
2048	3.991	3.793	2.490	2.247
4096	2.212	1.978	1.388	1.160
8192	1.341	1.219	0.833	0.706
Blocks				
512	5.387	5.346	3.425	3.395
1024	3.801	3.688	2.353	2.275
2048	2.644	2.550	1.639	1.597
4096	1.547	1.435	0.937	0.887
8192	1.043	0.992	0.647	0.621

Table 2: Mean Squared Error From 100 Replications

	SNR = 5		SNR = 7	
n	Random-x	Equispaced-x	Random-x	Equispaced-x
Angles				
512	0.858	0.537	0.671	0.383
1024	0.533	0.393	0.417	0.290
2048	0.363	0.297	0.266	0.216
4096	0.130	0.112	0.093	0.072
8192	0.080	0.072	0.059	0.047
Blip				
512	0.905	0.847	0.553	0.488
1024	0.610	0.596	0.368	0.341
2048	0.406	0.367	0.244	0.228
4096	0.253	0.248	0.143	0.144
8192	0.155	0.151	0.093	0.088
Corner				
512	0.410	0.364	0.270	0.241
1024	0.244	0.229	0.179	0.167
2048	0.172	0.166	0.129	0.121
4096	0.107	0.103	0.071	0.066
8192	0.068	0.066	0.046	0.043
Wave				
512	2.533	2.236	1.779	1.384
1024	1.666	1.521	1.072	0.795
2048	1.026	0.876	0.632	0.459
4096	0.155	0.090	0.125	0.061
8192	0.092	0.060	0.076	0.045

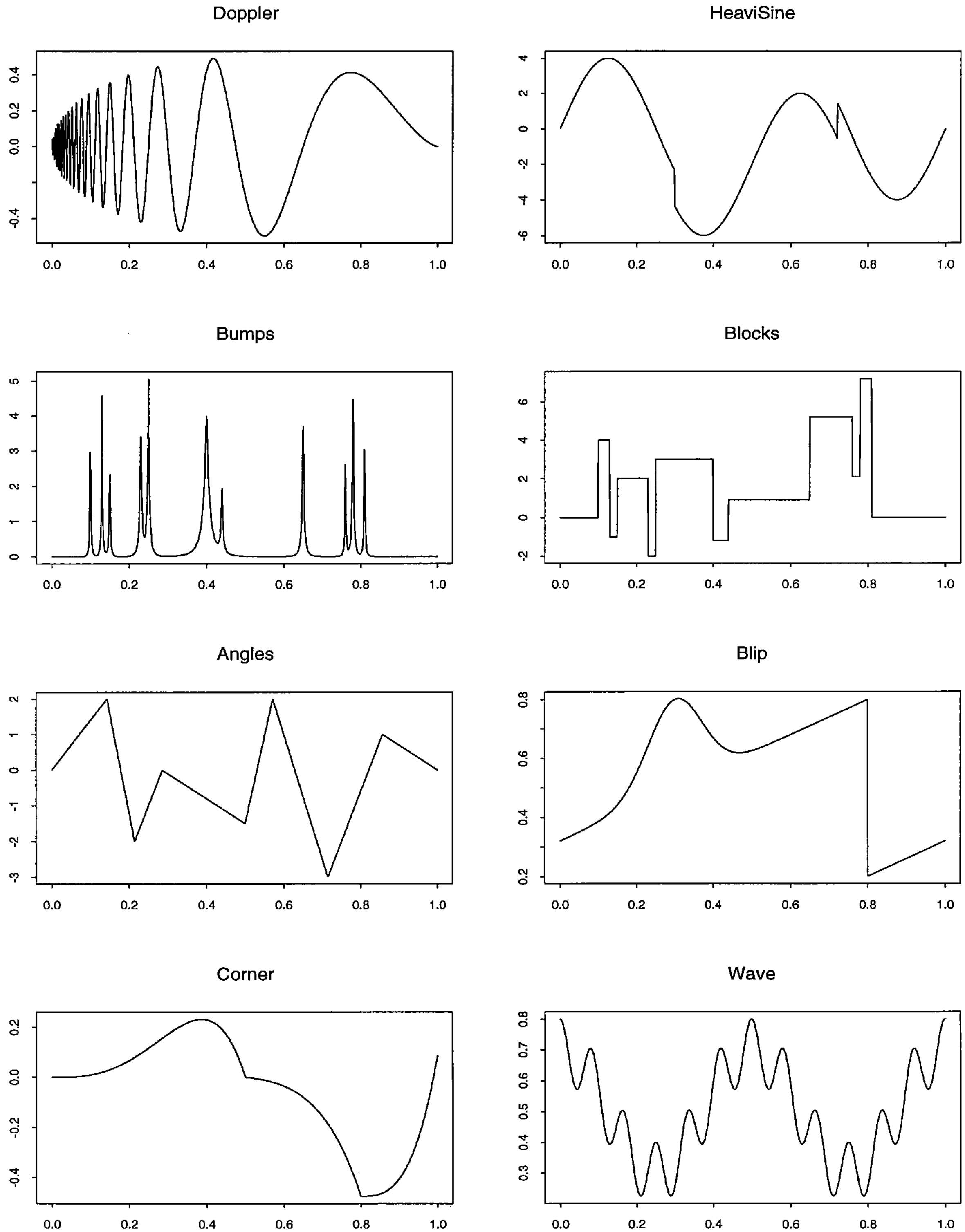


Figure 1: Test Functions

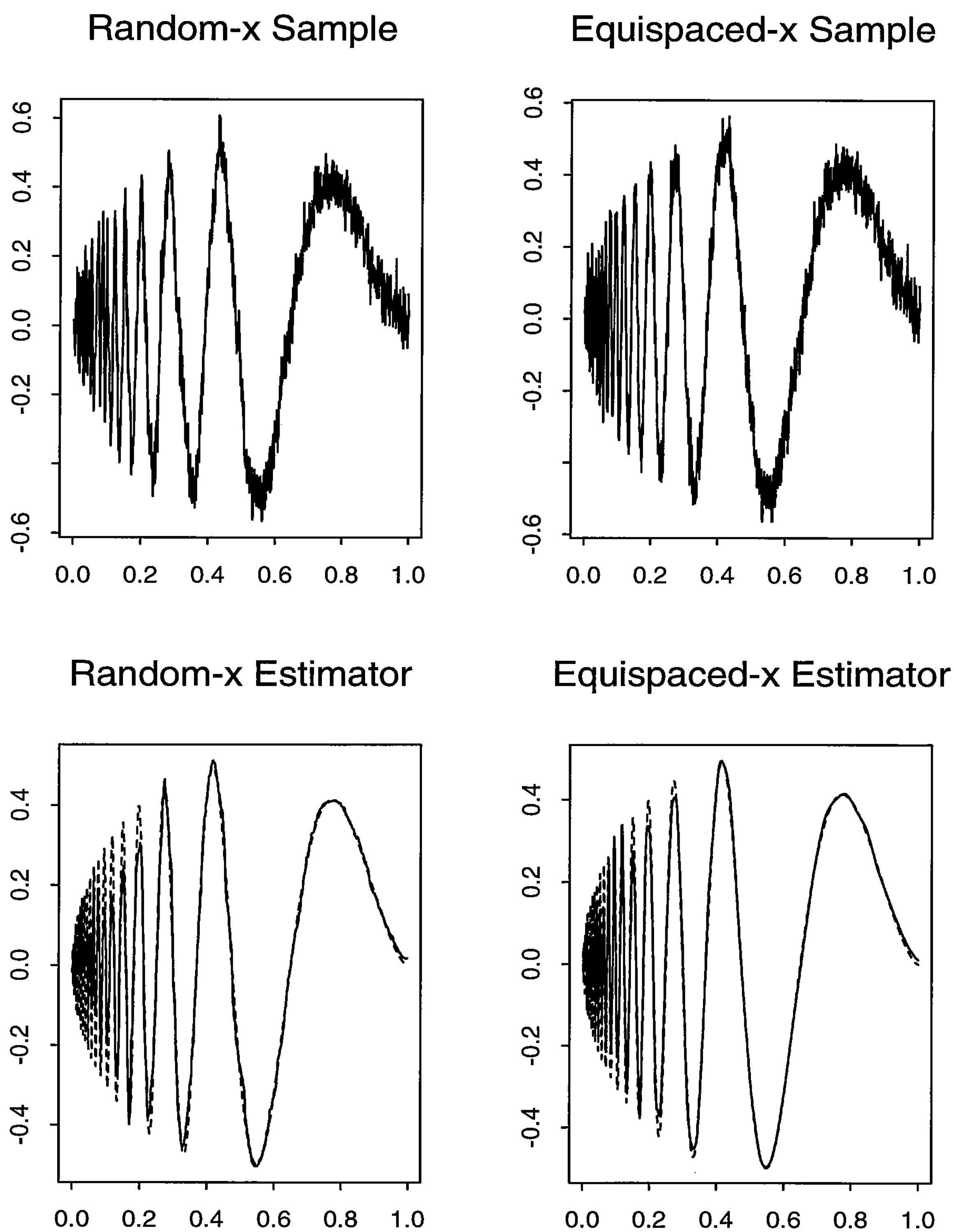


Figure 2: Doppler

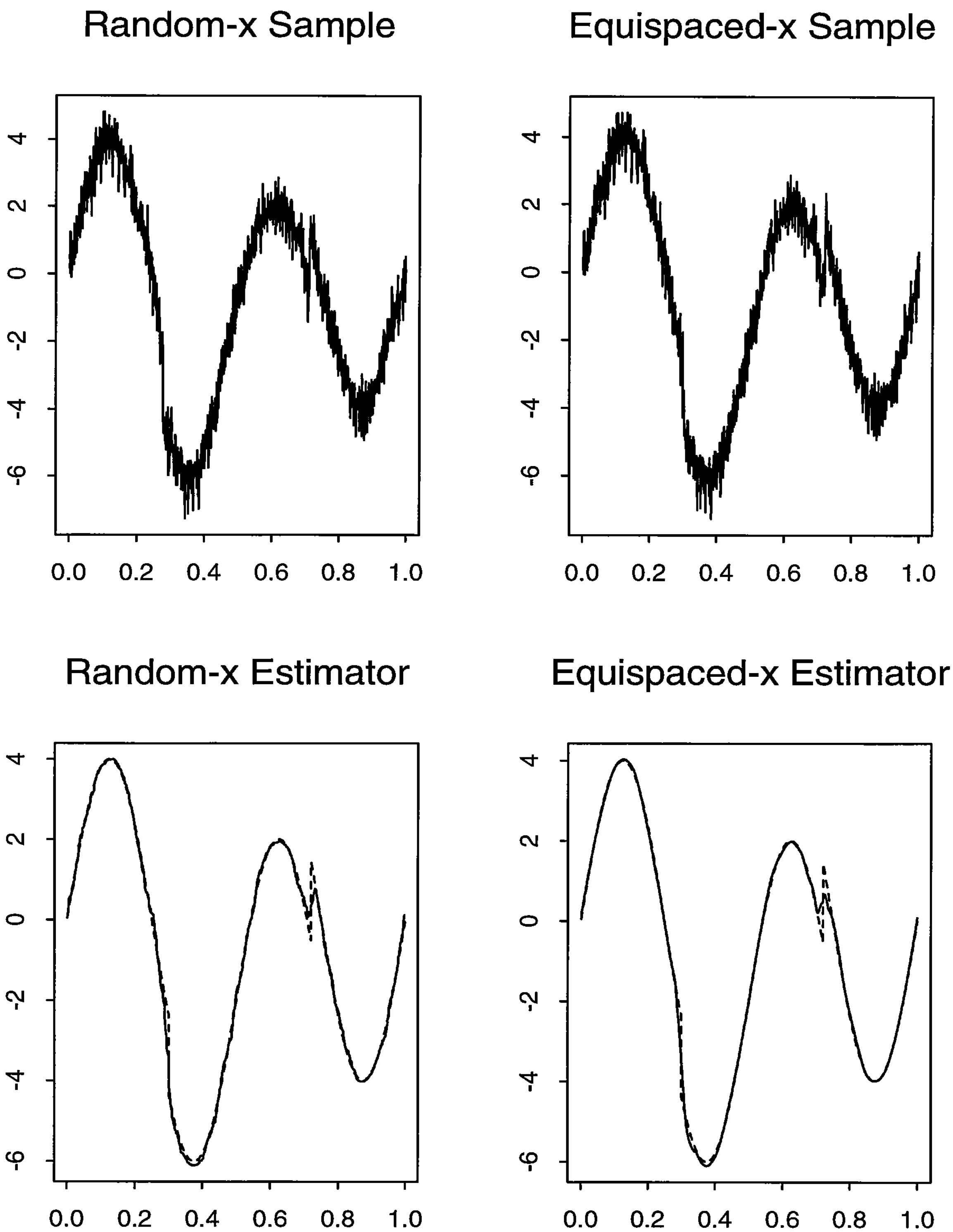


Figure 3: HeaviSine

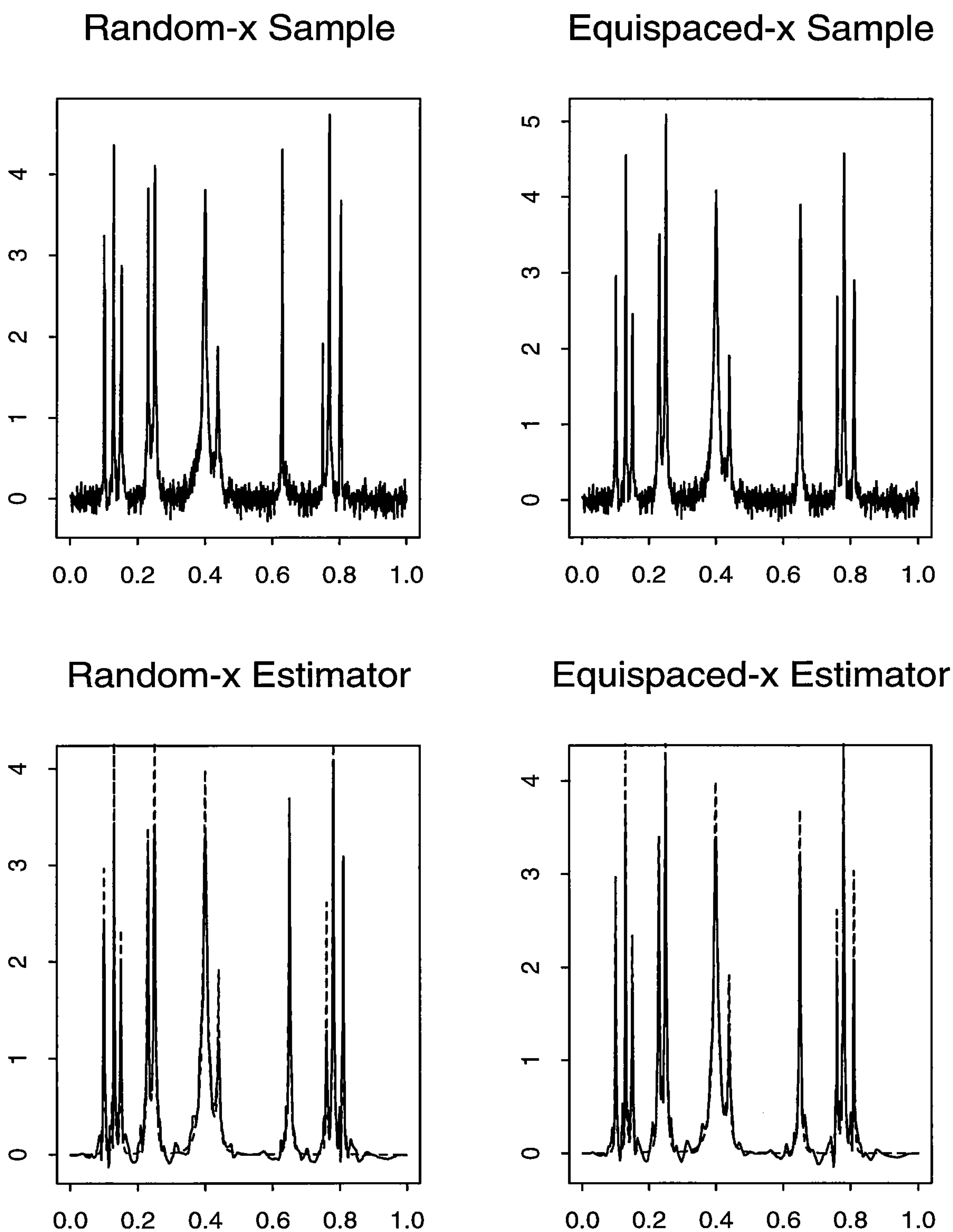


Figure 4: Bumps

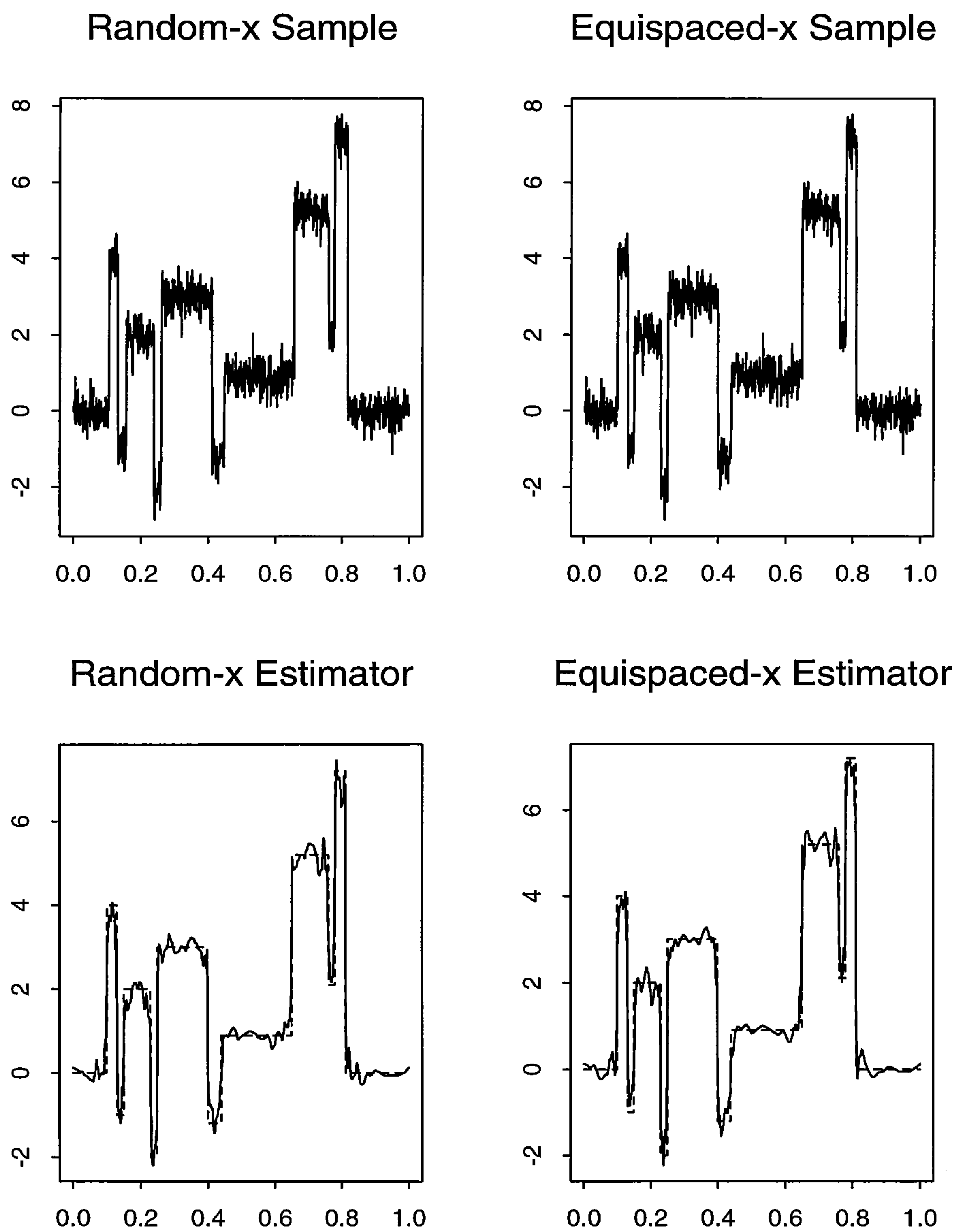


Figure 5: Blocks

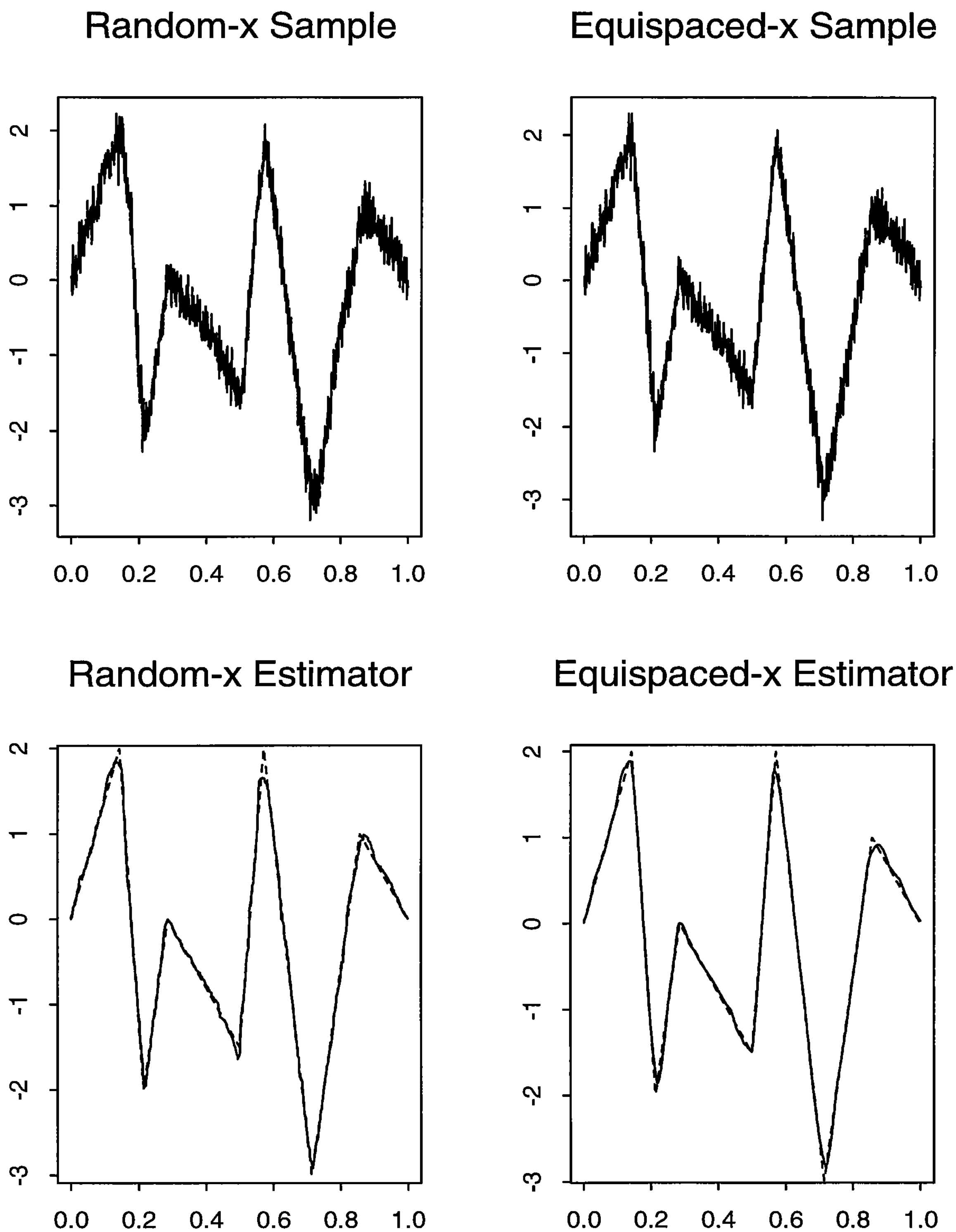


Figure 6: Angles

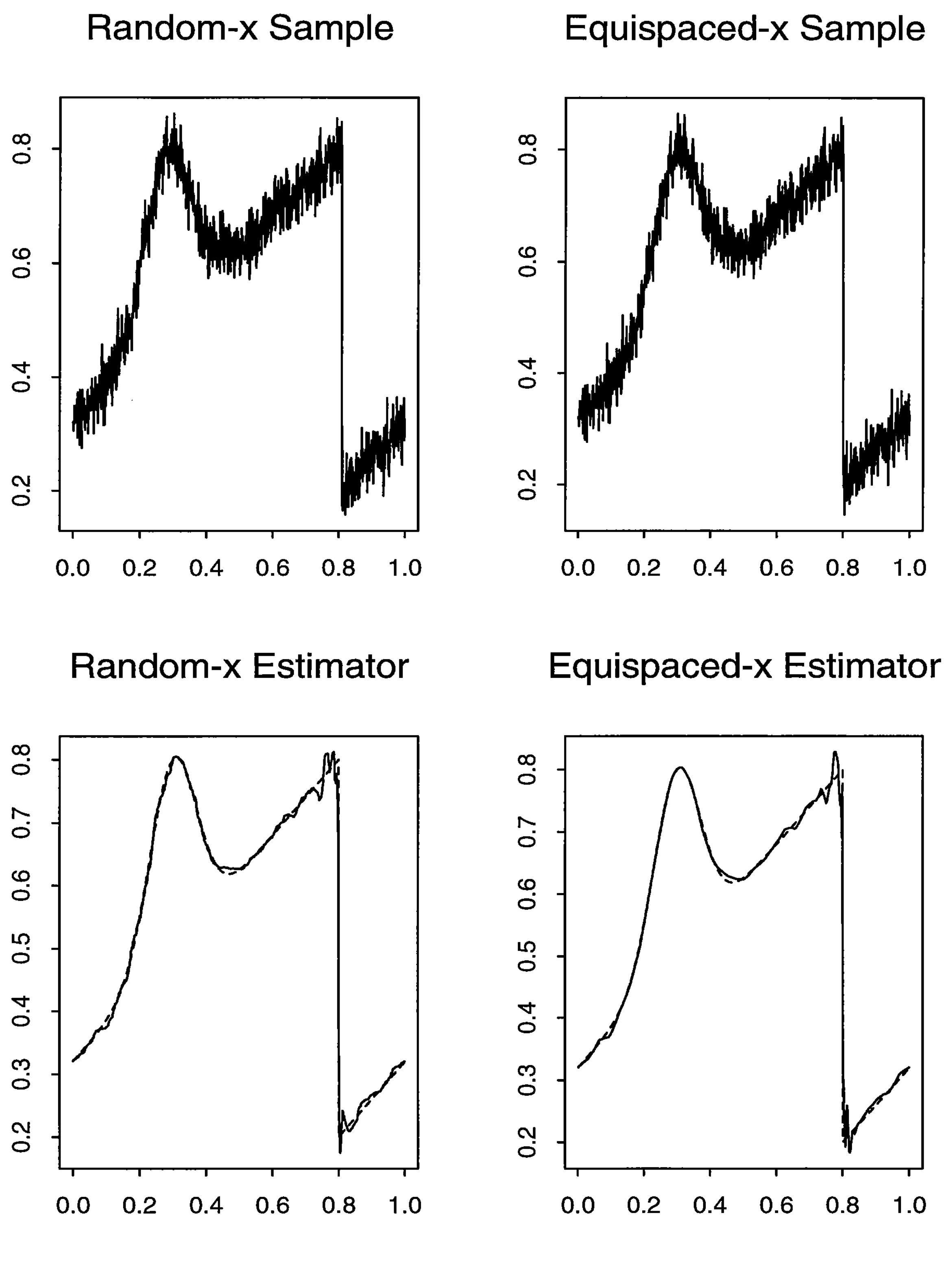


Figure 7: Blip

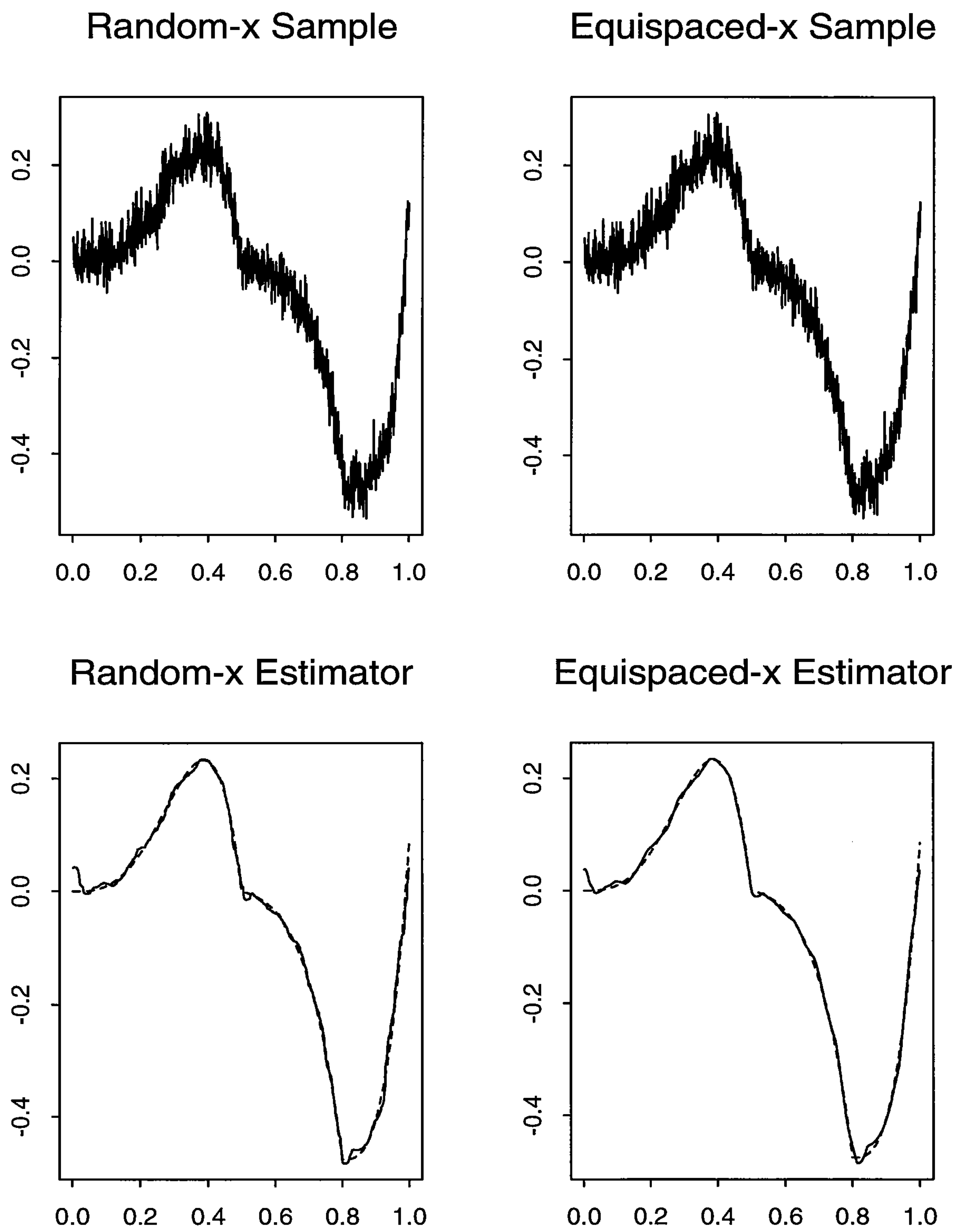


Figure 8: Corner

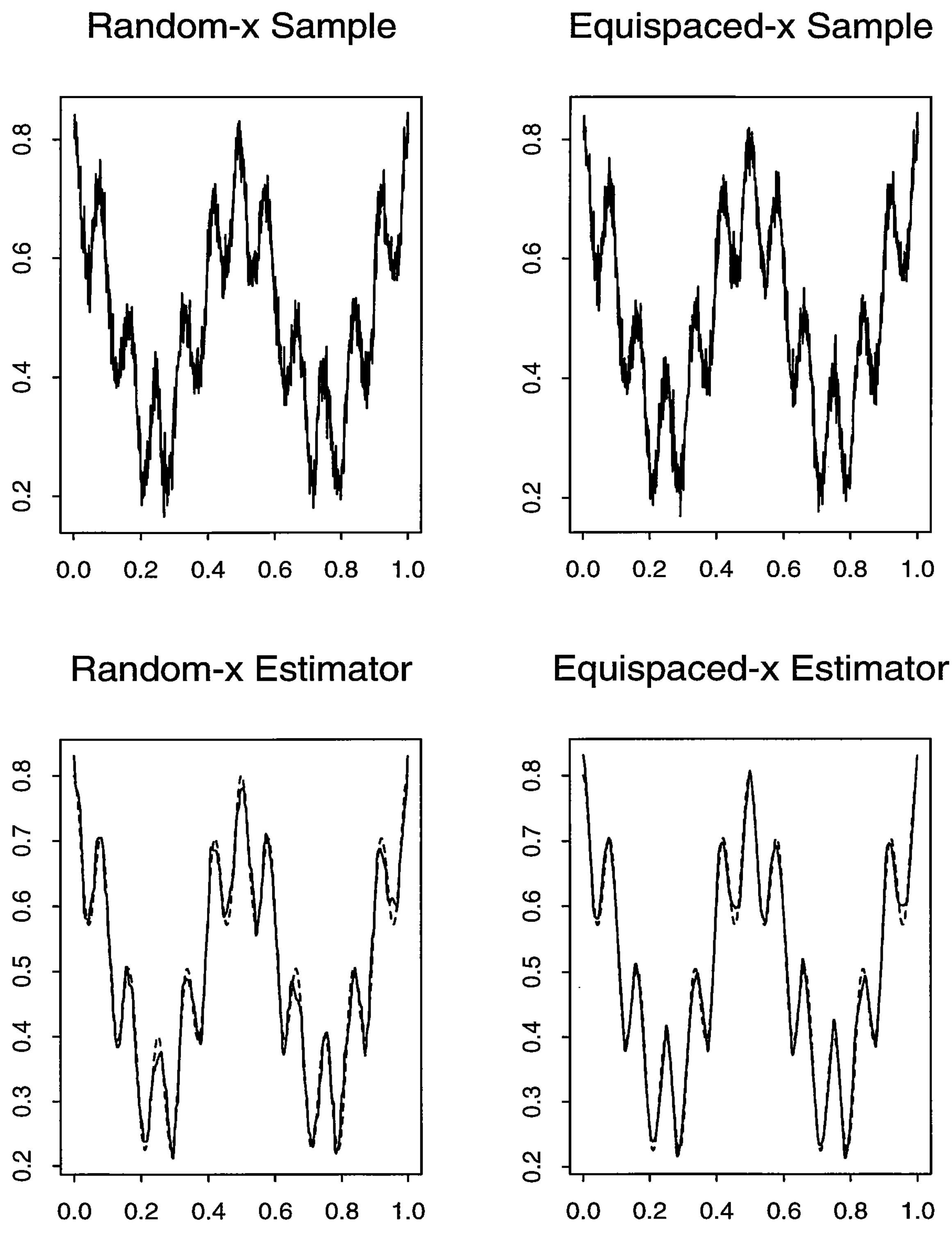


Figure 9: Wave