The Annals of Statistics
1998, Vol. 26, No. 5, 1760–1782

LOCAL ASYMPTOTICS FOR REGRESSION SPLINES AND CONFIDENCE REGIONS[1]

BY S. ZHOU, X. SHEN AND D. A. WOLFE

Ohio State University

In this paper, we study the local behavior of regression splines. In particular, explicit expressions for the asymptotic pointwise bias and variance of regression splines are obtained. In addition, asymptotic normality for regression splines is established, leading to the construction of approximate confidence intervals and confidence bands for the regression function.

1. Introduction. Consider the regression problem of estimating $f(x)$ based on data sampled from the model

$$y_i = f(x_i) + \varepsilon_i, \qquad i = 1, \dots, n, \tag{1}$$

where the ε_i's are uncorrelated with $E\varepsilon_i = 0$, and $E\varepsilon_i^2 = \sigma^2 > 0$, and the design points $\{x_i\}_{i=1}^n$ are either deterministic or random. Without loss of generality, we assume that $x_i \in [0,1]$, $i = 1, \dots, n$. Our goal is to estimate $f(x)$ and to construct a confidence region for it.

To estimate the regression function, we consider spline approximations. A spline is defined as a piecewise polynomial that is smoothly connected at its joints (knots). More specifically, for any fixed integer $m > 1$, denote $S(m, \underline{t})$ to be the set of spline functions with knots $\underline{t} = \{0 = t_0 < t_1 < \cdots < t_{k_0+1} = 1\}$. Then for $m = 1$, $S(m, \underline{t})$ is the set of step functions with jumps at the knots and, for $m \geq 2$,

$$S(m, \underline{t}) = \big\{ s \in C^{m-2}[0,1] \colon s(x) \text{ is a polynomial of degree } (m-1) \text{ on each subinterval } [t_i, t_{i+1}] \big\}.$$

To estimate $f(x)$, we use the least squares criterion. The regression spline estimator of order m for $f(x)$ is defined to be the least squares minimizer $\hat{f}(x) \in S(m, \underline{t})$ corresponding to

$$\sum_{i=1}^{n} \big(y_i - \hat{f}(x_i) \big)^2 = \min_{s(x) \in S(m, \underline{t})} \sum_{i=1}^{n} \big(y_i - s(x_i) \big)^2. \tag{2}$$

There is a large amount of literature on regression splines. In the univariate case, Agarwal and Studden (1980) and Huang and Studden (1993) considered the rates of convergence and the connection between splines and

Received December 1996; revised April 1998.

[1] Supported in part by NSA Grant MDA904-97-1-0030 and NSF Grant DMS-9802358.

AMS 1991 *subject classifications*. Primary 62G07; secondary 62G15, 62G20.

Key words and phrases. Regression splines, least squares, Bernoulli polynomial, asymptotic bias, asymptotic variance, asymptotic normality, confidence interval and confidence band.

kernels. In the multivariate case, Stone (1985, 1986, 1994), Stone and Koo (1986) and Friedman (1991) obtained the rates of convergence for the multivariate regression splines and developed computation methods. For partial linear models, Shi and Li (1994) studied the rates of convergence for M-type regression splines. For survival models, Kooperberg, Stone and Truong (1995) studied the rates of convergence and computation methods. For splines defined by a robust criterion such as the absolute deviation criterion, Shi and Li (1995), Koenker, Ng and Portnoy (1994) and Portnoy (1997) studied the rates of convergence and computation methods. Adaptive splines with variable orders and knots were studied by Shen and Hu (1995).

Previous research on regression splines focused mainly on estimation methods and convergence properties of the regression splines in various function classes to which the regression function belongs. The focus of this paper is on inference for regression splines. In particular, we study the asymptotic distribution of the regression spline $\hat{f}(x)$. We establish asymptotic normality for a properly standardized $\hat{f}(x)$ at a rate slower than $n^{-1/2}$ depending on the amount of smoothness of the regression function; that is, $[\hat{f}(x) - f(x) - b(x)]\sqrt{\mathrm{Var}(\hat{f}(x))}$ converges to $N(0,1)$, where $b(x)$ is the asymptotic bias and $\mathrm{Var}(\hat{f}(x))$ is the variance. In addition, we show that $b(x)$ can be expressed as a scaled Bernoulli polynomial and $\mathrm{Var}(\hat{f}(x))$ is approximated by a quadratic form. Based on these results, an asymptotic confidence interval and an asymptotic confidence band for the regression function are constructed. An extension to the case of heteroscedastic errors is also considered. In addition, a simulation study is carried out to exemplify the small sample behavior of these confidence regions.

This paper is organized as follows. Section 2 presents the results for asymptotic bias and variance in the fixed and random design cases. Section 3 provides the result for asymptotic normality. Section 4 discusses construction of confidence regions. The simulation study for the proposed confidence regions is reported in Section 5. Technical proofs are given in Section 6.

2. Asymptotic bias and variance. It is convenient to express functions in $S(m, \underset{\sim}{t})$ in terms of B-splines. For any fixed m and $\underset{\sim}{t}$, let

$$N_{i,m}(x) = (t_i - t_{i-m})[t_{i-m}, \dots, t_i](t-x)_+^{m-1}, \qquad i = 1, \dots, k = k_0 + m,$$

where $[t_{i-m}, \dots, t_i]g$ denotes the mth-order divided difference of the function g and $t_i = t_{\min(\max(i,0),k_0+1)}$ for any $i = 1-m, \dots, k$. Then $\{N_{i,m}(\cdot)\}_{i=1}^k$ form a basis for $S(m, \underset{\sim}{t})$ [see Schumaker (1981), page 124]; that is, for any $s(x) \in S(m, \underset{\sim}{t})$, there exists an $\underset{\sim}{a}$ such that $s(x) = \underset{\sim}{a}'\mathbf{N}_m(x)$, where $\mathbf{N}_m(x) = (N_{1,m}(x), \dots, N_{k,m}(x))'$. For convenience, in the sequel, $\mathbf{N}_m(\cdot)$ and $\{N_{i,m}(\cdot)\}_{i=1}^k$ will be abbreviated as $\mathbf{N}(\cdot)$ and $\{N_i(\cdot)\}_{i=1}^k$, respectively. To study the asymptotic bias and variance of $\hat{f}(x)$, we need to specify some conditions. Here we assume that

$$\max_{1 \le i \le k_0} |h_{i+1} - h_i| = o(k_0^{-1}) \quad \text{and} \quad h \Big/ \min_{1 \le i \le k_0} h_i \le M, \tag{3}$$

where $h_i = t_i - t_{i-1}$, $h = \max_{1 \le i \le k_0} h_i$ and $M > 0$ is a predetermined constant. Such an assumption assures that $M^{-1} < k_0 h < M$, which is necessary for numerical computations.

REMARK 1. We see next that the local asymptotic bias and variance of $\hat{f}(x)$ are both independent of the magnitude of M. The assumption in (3) is a weak restriction on the knot distribution. A commonly used condition is that knots are generated from a positive continuous density [see, e.g., Agrawal and Studden (1980)]. However, the assumption in (3) is adequate for our purpose.

In the case where the design points $\{x_i^n\}_{i=1}$ are deterministic, assume that

$$\sup_{x \in [0,1]} \left| Q_n(x) - Q(x) \right| = o\left(k_0^{-1}\right), \tag{4}$$

where $Q_n(x)$ is the empirical distribution function of $\{x_i^n\}_{i=1}$, and $Q(x)$ is a distribution with a positive continuous density $q(x)$.

Theorem 2.1 provides expressions for the asymptotic bias and variance of the regression estimator $\hat{f}(x)$ for the fixed design. It implies that the asymptotic bias can be expressed locally in terms of a scaled Bernoulli polynomial and the asymptotic variance is related to the location of the knots and the design density $q(x)$ when the number of knots is not too large. Of course, the value of k_0 controls the trade-off between the bias and the variance of $\hat{f}(x)$.

THEOREM 2.1 (Asymptotic bias and variance: fixed design case). *Suppose* (3) *and* (4) *hold. If* $f \in C^m[0,1]$ *and* $k_0 = o(n)$, *then, for any* $x \in (t_i, t_{i+1}]$, $i = 0, \dots, k_0$,

$$E\left(\hat{f}(x)\right) - f(x) = b(x) + o(h^m), \tag{5}$$

$$\operatorname{Var}\left(\hat{f}(x)\right) = \frac{\sigma^2}{n} \mathbf{N}'(x) G^{-1}(q) \mathbf{N}(x) + o\left((nh)^{-1}\right), \tag{6}$$

where $\mathbf{N}'(x)$ *is the transpose of* $\mathbf{N}(x)$,

$$b(x) = -\frac{f^{(m)}(x) h_i^m}{m!} B_m\left(\frac{x - t_i}{h_i}\right) \quad \text{and} \quad G(q) = \int_0^1 \mathbf{N}(x)\mathbf{N}'(x) q(x)\, dx.$$

Here $B_m(\cdot)$ *is the mth Bernoulli polynomial which is inductively defined as follows*:

$$B_0(x) = 1, \qquad B_i(x) = \int_0^x i B_{i-1}(z)\, dz + b_i,$$

where $b_i = -i \int_0^1 \int_0^x B_{i-1}(z)\, dz\, dx$ *is the ith Bernoulli number* [*see Barrow and Smith* (1978)].

In practice, there are many situations, including many survey or other observational studies, where the design points are not provided or controlled by the experimenters. As a result, one may consider that the design points are random. For random designs, we assume that the design points are sampled from a continuous distribution. The next theorem provides the conditional asymptotic bias and variance for $\hat{f}(x)$ for this random design setting.

THEOREM 2.2 (Asymptotic bias and variance: random design case). *Suppose* (3) *holds and the design points* $\{x_i\}_{i=1}^n$ *are randomly sampled from* $Q(x)$. *If* $f \in C^m[0,1]$ *and* $k_0 = o(n^r)$ *for some* $r \in (0, 1/2]$, *then, for any* $x \in (t_i, t_{i+1}]$, $i = 0, \dots, k_0$,

$$E\big(\hat{f}(x)|\underline{x}\big) - f(x) = b(x) + o_p(h^m),$$

$$\mathrm{Var}\big(\hat{f}(x)|\underline{x}\big) = \frac{\sigma^2}{n}\mathbf{N}'(x)G^{-1}(q)\mathbf{N}(x) + o_p\big((nh)^{-1}\big).$$

REMARK 2. The asymptotic bias in (5) does not depend on the design distribution $Q(x)$. This is a reflection that the regression spline is a local smoother and the fact that the design distribution is close to uniform locally for large n and k_0. The asymptotic variance in (6) depends on both the knot and the design distributions as we should expect. In addition, if the knot sequence converges to a limiting distribution with positive continuous density $p(x)$, analysis of the leading term on the right-hand side of (6) leads to the conclusion that the pointwise variance of $\hat{f}(x)$ is asymptotically proportional to $q(x)$ and $p^{-2}(x)$ for any $x \in (0,1)$.

REMARK 3. In many situations, the error random variables ε_i are not homogenous; that is, the $\{\varepsilon_i\}_{i=1}^n$ are uncorrelated with mean 0 and $\mathrm{Var}(\varepsilon_i) = w(x_i)\sigma^2$, where $w(\cdot)$ is a positive continuous weight function on $[0,1]$. In such settings, it is more appropriate to consider a weighted sum of squares criterion, such as minimizing

$$\sum_{i=1}^{n} w^{-1}(x_i)\big(y_i - s(x_i)\big)^2. \tag{7}$$

Similarly, we can obtain

$$E\big(\hat{f}(x)\big) - f(x) = b(x) + o(h^m),$$

$$\mathrm{Var}\big(\hat{f}(x)\big) = \frac{\sigma^2}{n}\mathbf{N}'(x)G_w^{-1}(q)\mathbf{N}(x) + o\big((nh)^{-1}\big),$$

where $G_w(q) = \int_0^1 w(x)\mathbf{N}(x)\mathbf{N}'(x)q(x)\,dx$ and $\hat{f}(x)$ is now the minimizer of (7).

REMARK 4. When $k_0 = Cn^{1/(2m+1)}$ for some $C > 0$, it follows by (5) and Lemma 6.6 of Section 5 that

$$\|b(\cdot)\|_{L_\infty} \equiv \max_{x\in[0,1]} |b(t)| = O(n^{-m/(2m+1)})$$

and

$$\max_{x\in[0,1]} \operatorname{Var}\big(\hat{f}(x)\big) = O\big(n^{-2m/(2m+1)}\big).$$

Therefore $\hat{f}(x) - f(x) = O_P(n^{-m/(2m+1)})$ uniformly for any $x \in [0,1]$, which says that regression splines do not suffer from boundary effects [Gasser and Miller (1984)]. However, we note that the variance of $\hat{f}(x)$ near the boundary of $[0,1]$ is much larger than in the interior. This phenomenon can be easily explained by the fact that fewer observations are collected near the boundary. In addition, for any probability measure μ, we have that

$$\text{(8)} \qquad \text{IMSE} = \sup_{f\in C(m,\delta)} \int_0^1 E\big(\hat{f}(x) - f(x)\big)^2 \, d\mu(x) = O\big(n^{-2m/(2m+1)}\big),$$

where $C(m,\delta) = \{f \in C^m[0,1]\colon |f^{(m)}| \le \delta\}$ for some $\delta > 0$. This agrees with the result of Stone (1982).

REMARK 5. From Theorems 2.1 and 2.2, the asymptotic optimal knot placement can be derived. This has been done in Agarwal and Studden (1980) for the fixed design under the additional assumption that the knots are generated from a density. Their result can be generalized to our setting using similar arguments.

3. Asymptotic normality. In this section, we study the asymptotic distribution of a properly standardized $\hat{f}(x)$.

THEOREM 3.1 (Asymptotic normality). *In addition to the assumptions in Theorem* 2.1, *suppose that the* $\{\varepsilon_i\}_{i=1}^n$ *are independently and identically distributed with mean* 0 *and variance* σ^2, *and* $k_0 \ge Cn^{1/(2m+1)}$ *for some constant* $C > 0$. *Then, for any fixed* $x \in [0,1]$,

$$\text{(9)} \qquad \frac{\hat{f}(x) - (f(x) + b(x))}{\sqrt{\operatorname{Var}\big(\hat{f}(x)\big)}} \to_d N(0,1).$$

Under the assumptions in Theorem 2.2, (9) *continues to hold with* $\operatorname{Var}(\hat{f}(x))$ *replaced by* $\operatorname{Var}(\hat{f}(x)|\underline{x})$.

REMARK 6. In the case of heteroscedastic errors, under the additional assumptions that $w(\cdot)$ is continuous on $[0,1]$ and the $\{\varepsilon_i/\sqrt{w(x_i)}\}_{i=1}^n$ are independently and identically distributed with mean 0 and variance σ^2, we have

$$\frac{\hat{f}(x) - (f(x) + b(x))}{\sqrt{\operatorname{Var}\big(\hat{f}(x)\big)}} \to_d N(0,1),$$

where $k_0 \ge Cn^{1/(2m+1)}$.

4. Confidence regions. We now apply the results in Sections 2 and 3 to construct confidence regions for $f(x)$ when the common variance σ^2 is known.

THEOREM 4.1 (Confidence interval). *In addition to the assumptions in Theorem* 3.1, *assume that* $k_0 n^{-1/(2m+1)} \to \infty$. *Then for any fixed* $x \in [0,1]$, *a* $100(1-\alpha)\%$ *asymptotic confidence interval for* $f(x)$ *is*

$$\hat{f}(x) \pm z_{\alpha/2}\sqrt{\operatorname{Var}(\hat{f}(x))},$$

where $z_{\alpha/2}$ *is the* $(1-\alpha/2)$*th normal percentile and*

$$\operatorname{Var}(\hat{f}(x)) = \sigma^2 \mathbf{N}'(x) G_{k,n}^{-1} \mathbf{N}(x)/n$$

with $G_{k,n} = n\mathbf{X}\mathbf{X}'$ *and* $\mathbf{X} = n^{-1}(\mathbf{N}(x_1), \dots, \mathbf{N}(x_n))$.

The choice of k_0 in Theorem 4.1 gives an undersmoothed $\hat{f}(x)$, as we see from Remark 4. As a result, we cannot directly use the knots selected according to GCV [Craven and Wahba (1979)] or other similar methods in the construction of confidence intervals for $f(x)$. To get around this problem, we consider an alternative two-step procedure. In the first step, we use an m^*th, $m^* < m$, order spline to select the knots to assure that the number of knots satisfies the required condition in Theorem 4.1. In the second step, we fit an mth-order spline using the knots selected in step 1 to construct confidence intervals for $f(x)$. See Section 5 for more details.

In many studies, it is desirable to obtain a simultaneous confidence band for $f(x)$ over all $x \in [0,1]$. Since regression splines are linear estimators, we can use standard simultaneous inference theory in linear regression to construct such confidence bands for $E\hat{f}(x)$ [see, e.g., Johansen and Johnstone (1990)]. Although these bands will not necessarily maintain the desired coverage probability for small sample sizes, the effect of the bias of $\hat{f}(x)$ decreases as n becomes larger. As a result, these bands will provide the desired coverage probability asymptotically, as stated precisely in the next theorem.

THEOREM 4.2 (Confidence band). *In addition to the assumptions in Theorem* 2.1 *or* 2.2, *assume that the* $\{\varepsilon_i\}_{i=1}^n$ *are i.i.d.* $N(0,\sigma^2)$, *with* σ^2 *known, and that* $k_0 \ge Cn^{1/(2m+1)}$. *For large* n,

$$P\left(\sup_{0\le x\le 1} \frac{|\hat{f}(x) - f(x)|}{\sqrt{\operatorname{Var}(\hat{f}(x))}} \le c(\alpha)\right) \ge 1-\alpha,$$

where $\alpha \in (0,1)$ *and* $c(\alpha)$ *satisfies*

$$\frac{\alpha}{2} = \frac{|\gamma|}{2\pi}\exp(-c^2(\alpha)/2) + 1 - \Phi(c(\alpha)). \tag{10}$$

Here

$$|\gamma| = \int_0^1 \left\{ \frac{d}{dx} \left(\frac{G_{k,n}^{-1/2}\mathbf{N}'(x)}{\left(\mathbf{N}'(x)G_{k,n}^{-1}\mathbf{N}(x)\right)^{1/2}} \right) \times \frac{d}{dx} \left(\frac{G_{k,n}^{-1/2}\mathbf{N}(x)}{\left(\mathbf{N}'(x)G_{k,n}^{-1}\mathbf{N}(x)\right)^{1/2}} \right) \right\}^{1/2} dx. \tag{11}$$

In other words, a $100(1-\alpha)\%$ *asymptotic confidence band for* $f(x)$ *is*

$$\hat{f}(x) \pm c(\alpha)\sqrt{\operatorname{Var}\big(\hat{f}(x)\big)}.$$

REMARK 7. When σ^2 is unknown, Theorem 4.2 continues to hold if $\operatorname{Var}(\hat{f}(x))$ is replaced by $\hat{\sigma}^2\mathbf{N}'(x)G_{k,n}^{-1}\mathbf{N}(x)$, where $\hat{\sigma}^2$ is any $n^{1/2}$-consistent estimator of σ^2. Such an estimator can be found, for example, in Gasser, Sroka and Jennen-Steinmetz (1986) and Hall, Kay and Titterington (1990).

REMARK 8. In the case where the $\{\varepsilon_i\}_{i=1}^n$ are dependent, Theorem 4.2 continues to hold if the minimum and maximum eigenvalues of $\operatorname{Cov}(\underline{\varepsilon})$ are bounded away from 0 and ∞.

5. Simulation study. To illustrate the proposed method for constructing of the confidence interval at a point and the confidence band, we perform a simulation study using model (1) with a test function

$$f(x) = \exp\big(-32(x-0.5)^2\big) + 2x - 1.$$

In the simulation, the x_i are equally spaced in $[0,1]$ and the ε_i are taken to be i.i.d. $N(0,(0.1)^2)$. Sample sizes 50, 100, 200, 400, 600 and 800 are considered and the coverage probability based on 1000 runs is reported for each sample size. The estimated curve is obtained by fitting a cubic spline using the knots selected by a second-order spline function according to GCV with the forward addition procedure. In the construction of confidence regions, we used the estimator for σ^2 proposed by Hall, Kay and Titterington (1990):

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n-2}\left(0.809y_i - 0.5y_{i+1} - 0.309y_{i+2}\right)^2. \tag{12}$$

Figure 1 displays the pointwise coverage probabilities of confidence intervals for different sample sizes. It suggests that these pointwise confidence intervals provide reasonable coverage probabilities for these sample sizes. However, there is considerable variation in the pointwise coverage probabilities for smaller sample sizes. We believe that this is due to the fact that the pointwise bias of the regression spline depends on the knot locations (see Theorems 2.1 and 2.2). When sample sizes are large, the asymptotic effect takes control and forces the pointwise coverage probability close to 95% at each point. It is interesting to note that the coverage probability at a large

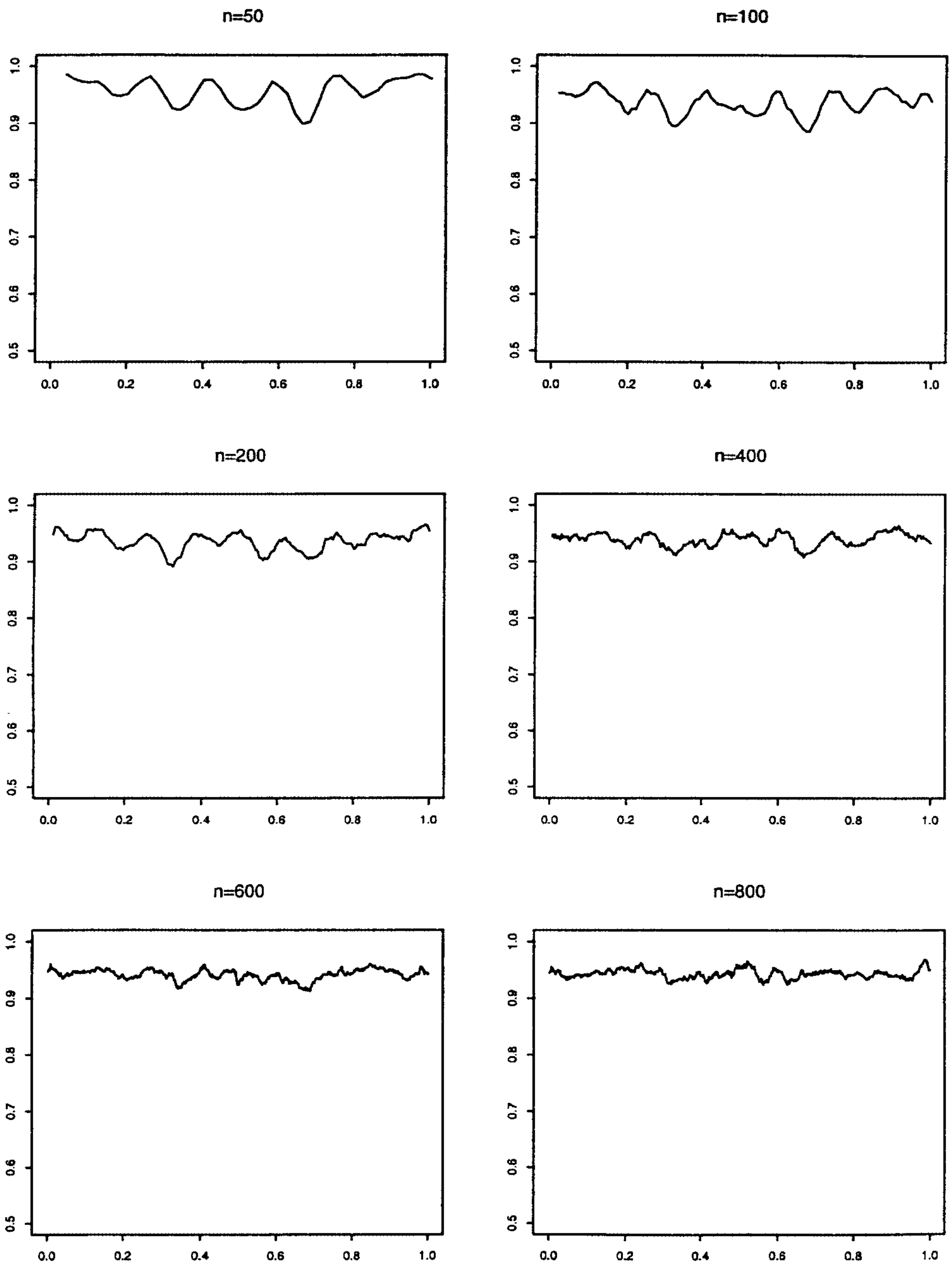

FIG. 1. *Empirical coverage probabilities of pointwise confidence intervals for sample sizes* n = 50, 100, 200, 400, 600 *and* 800.

number of points exceeds 95% for $n = 50$. This is due to the positive bias of $\hat{\sigma}^2$ in (12), which helps increase the coverage probability. For the same reason, the coverage probabilities of the confidence bands for small sample sizes ($n = 50, 100$) are larger than that for $n = 200$ (see Figure 3). We need to point out that, unlike Bayesian confidence intervals for smoothing splines where the coverage is provided in average sense, here the coverage is in the classical pointwise sense.

Figure 2 displays simulated data, the true regression function, the estimated function and the corresponding 95% confidence band based on Theorem 4.2. Figure 3 plots the coverage probability versus different sample sizes. It shows that the coverage probability increases as n increases for $n \geq 200$,

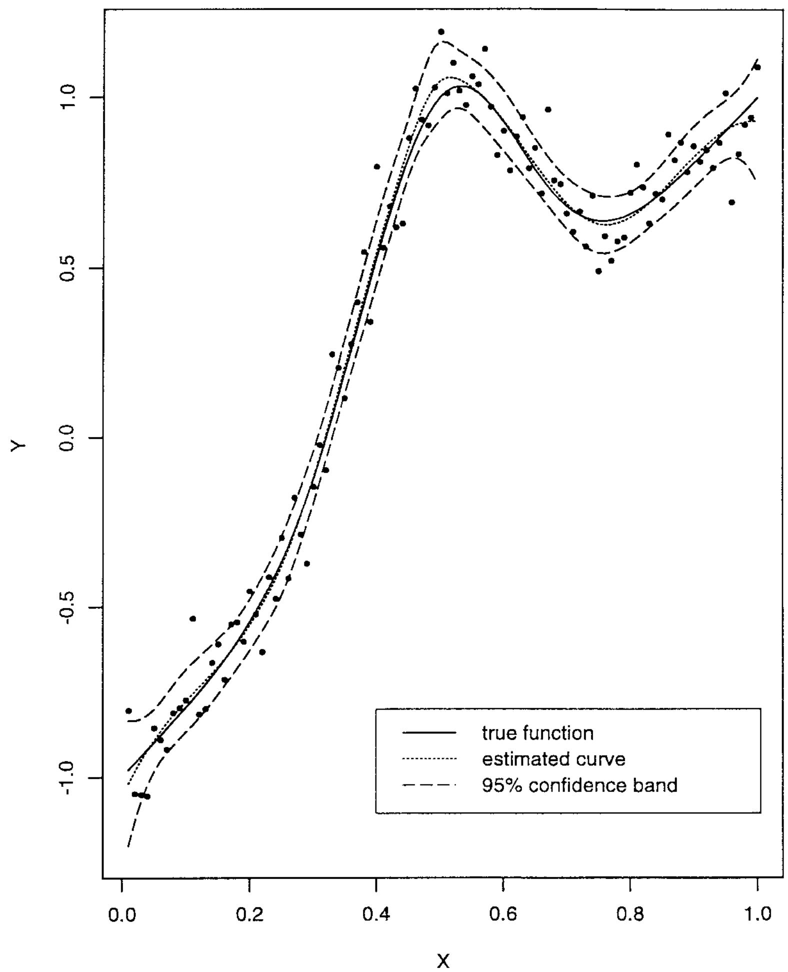

FIG. 2. *Simulated data for $n = 100$ with the true regression function, the estimated curve and 95% confidence band.*

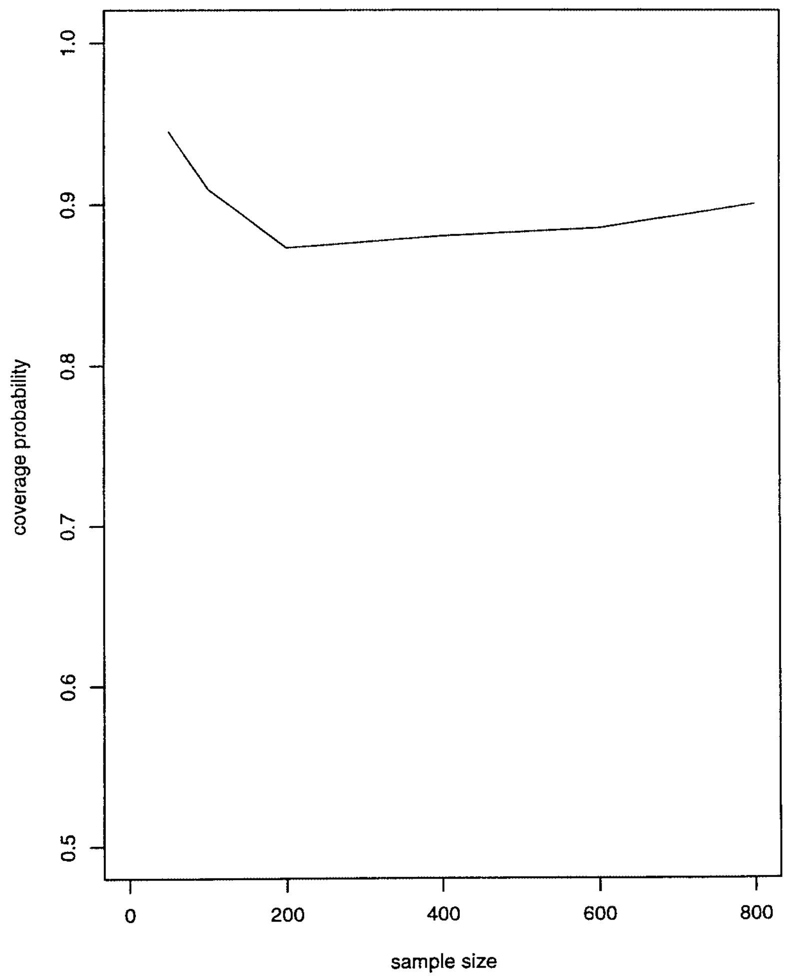

FIG. 3. *Empirical coverage probabilities for confidence bands.*

which agrees with the result of Theorem 4.2. However, the coverage is lower than the ideal 95% even for $n = 800$. This is due to the fact that $c(\alpha)$ is increasing at a very slow rate $O(\sqrt{\log n})$ and therefore the bias effect can be large even for $n = 800$. Consequently, a suitable bias correction may improve the performance for small sample sizes and is worthy of further investigation.

6. Proofs of theorems. Before proving Theorem 2.1, we state several technical lemmas about the matrices $G_{k,n}$ and $G_{k,n}^{-1}$. In the proof, we use the following result of de Boor [(1978), page 155]: There exists a constant $c_0 > 0$, depending only on m, such that, for any $s(x) = \sum_{i=1}^k a_i N_i(x) \in S(m, \underline{t})$,

$$c_0 \left(\sum_{i=1}^k a_i^2 (t_i - t_{i-m}) \right)^{1/2} \leq \|s\|_{L_2} \equiv \left(\int_0^1 s^2(x)\, dx \right)^{1/2} \leq \left(\sum_{i=1}^k a_i^2 (t_i - t_{i-m}) \right)^{1/2}. \tag{13}$$

First we derive an analogue of (13).

LEMMA 6.1. *There exist constants* $0 < c_1 \leq c_2 < \infty$ (*independent of* n *or* k_0) *such that, for any* $s(x) = \sum_{i=1}^k a_i N_i(x) \in S(m, \underline{t})$,

$$(c_1 + o(1)) h \sum_{i=1}^k a_i^2 \leq \int_0^1 s^2(x)\, dQ_n(x) \leq (c_2 + o(1)) h \sum_{i=1}^k a_i^2. \tag{14}$$

PROOF. To prove the right-hand inequality, note that

$$\int_0^1 s^2(x)\, dQ(x) = \int_0^1 s^2(x) q(x)\, dx \leq q_{\max} \int_0^1 s^2(x)\, dx = q_{\max} \| s(x) \|_{L_2}^2,$$

where $q_{\max} = \max_{x \in [0,1]} q(x) < \infty$. It follows from (13) that

$$\int_0^1 s^2(x)\, dQ(x) \leq q_{\max} \int_0^1 s^2(x)\, dx \leq q_{\max} \sum_{i=1}^k a_i^2 (t_i - t_{i-m}) \leq q_{\max} m h \sum_{i=1}^k a_i^2. \tag{15}$$

Using integration by parts, we have from (4) that

$$\begin{aligned} &\left| \int_0^1 s^2(x)\, d(Q_n - Q)(x) \right| \\ &\quad = \left| s^2(x)(Q_n - Q)(x)|_0^1 - \int_0^1 2(Q_n - Q)(x) s^{(1)}(x) s(x)\, dx \right| \\ &\quad \leq o(h) \int_0^1 \left| s^{(1)}(x) \right| s(x)\, dx, \end{aligned}$$

where $s^{(1)}(x)$ is the derivative of $s(x)$. In the last inequality, the fact that $(Q_n - Q)(1) = (Q_n - q)(0) = 0$ has been used. By the Hölder inequality, it follows that

$$\left| \int_0^1 s^2(x)\, d(Q_n - Q)(x) \right| \leq o(h) \| s(x) \|_{L_2} \| s^{(1)}(x) \|_{L_2}.$$

From de Boor (1972), page 54, $s^{(1)}(x) = \sum_{i=1}^{k-1} a_i^{(1)} N_{i,m-1}(x)$, where $a_i^{(1)} = (m-1)(a_{i+1} - a_i)/(t_i - t_{i-m+1})$. By (13) and (3), we have

$$\left| \int_0^1 s^2(x)\, d(Q_n - Q)(x) \right|$$

$$\leq o(h)\left(\sum_{i=1}^{k} a_i^2 (t_i - t_{i-m}) \right)^{1/2}$$

$$\times \left(\sum_{i=1}^{k-1} ((m-1)(a_{i+1} - a_i))^2 / (t_i - t_{i-m+1}) \right)^{1/2}$$

$$\leq o(h)(mh)^{1/2} \left(\sum_{i=1}^{k} a_i^2 \right)^{1/2} (m-1) \left(\min_{1 \leq i \leq k_0} h_i \right)^{-1/2} \tag{16}$$

$$\times \left(\sum_{i=1}^{k-1} (a_{i+1} - a_i)^2 \right)^{1/2}$$

$$= o(h) \left(h \Big/ \min_{1 \leq i \leq k_0} h_i \right)^{1/2} \left(\sum_{i=1}^{k} a_i^2 \right)^{1/2} \left(\sum_{i=1}^{k-1} (a_{i+1} - a_i)^2 \right)^{1/2}$$

$$\leq o(h) \sum_{i=1}^{k} a_i^2.$$

In the last inequality, the fact that $\sum_{i=1}^{k-1} (a_{i+1} - a_i)^2 \leq 4 \sum_{i=1}^{k} a_i^2$ has been used. This, together with (15), implies the right-hand inequality of (14).

To prove the left-hand inequality, note that

$$\int_0^1 s^2(x)\, dQ(x) \geq q_{\min} \int_0^1 s^2(x)\, dx = q_{\min} \| s(x) \|_{L_2}^2,$$

where $q_{\min} = \min_{x \in [0,1]} q(x) > 0$. It follows from (3) and (13) that

$$\int_0^1 s^2(x)\, dQ(x) \geq q_{\min} c_0^2 \sum_{i=1}^{k} a_i^2 (t_i - t_{i-m})$$

$$\geq q_{\min} c_0^2 \min_{1 \leq i \leq k_0} h_i \sum_{i=1}^{k} a_i^2 \geq q_{\min} c_0^2 M^{-1} h \sum_{i=1}^{k} a_i^2.$$

By (16), we have

$$\int_0^1 s^2(x)\, dQ_n(x) = o(h) \sum_{i=1}^{k} a_i^2 + \int_0^1 s^2(x)\, dQ(x)$$

$$\geq \left(q_{\min} c_0^2 M^{-1} h + o(h) \right) \sum_{i=1}^{k} a_i^2.$$

This completes the proof of Lemma 6.1. □

Lemma 6.2 provides lower and upper bounds for the eigenvalues of $G_{k,n}$.

LEMMA 6.2.

$$(c_1 + o(1))h \le \lambda_{\min} \le \lambda_{\max} \le (c_2 + o(1))h, \tag{17}$$

where $\lambda_{\min}$ *and* $\lambda_{\max}$ *are the minimum and maximum eigenvalues of* $G_{k,n}$, *respectively*.

PROOF. Note that

$$\lambda_{\max} = \max_{\sum_{i=1}^k a_i^2 = 1} \underline{a}' G_{k,n} \underline{a} \quad \text{and} \quad \lambda_{\min} = \min_{\sum_{i=1}^k a_i^2 = 1} \underline{a}' G_{k,n} \underline{a}.$$

By the definition of $G_{k,n}$, we have

$$\begin{aligned} \underline{a}' G_{k,n} \underline{a} &= n \underline{a}' \mathbf{X}\mathbf{X}' \underline{a} \\ &= \frac{1}{n} (\underline{a}' \mathbf{N}(x_1), \dots, \underline{a}' \mathbf{N}(x_n))(\underline{a}' \mathbf{N}(x_1), \dots, \underline{a}' \mathbf{N}(x_n))' \\ &= \frac{1}{n} \sum_{j=1}^{n} \left(\sum_{i=1}^{k} a_i N_i(x_i) \right)^2 = \int_0^1 s^2(x)\, dQ_n(x), \end{aligned} \tag{18}$$

where $s(x) = \sum_{i=1}^k a_i N_i(x)$. It follows from (18) and Lemma 6.1 that

$$\lambda_{\max} = \max_{\sum a_i^2 = 1} \int_0^1 s^2(x)\, dQ_n(x) \le (c_2 + o(1))h$$

and

$$\lambda_{\min} = \min_{\sum_{i=1}^k a_i^2 = 1} \int_0^1 s^2(x)\, dQ_n(x) \ge (c_1 + o(1))h.$$

This completes the proof. □

Lemma 6.3 states that the elements of $G_{k,n}^{-1}$ decay exponentially.

LEMMA 6.3. *There exist constants* $c_3 \in (0, \infty)$ *and* $\gamma \in (0,1)$ *such that*, *for large* n,

$$|\alpha_{ij}| \le c_3 h^{-1} \gamma^{|i-j|},$$

where α_{ij} *is the* (i,j)*th element of* $G_{k,n}^{-1}$. *In addition*,

$$\|G_{k,n}^{-1}\|_\infty \equiv \max_{1 \le i \le k} \sum_{j=1}^{k} |\alpha_{ij}| = O(h^{-1}). \tag{19}$$

PROOF. Noting that $G_{k,n}$ is a band matrix, we apply Theorem 2.2 of Demko (1977) to $\lambda_{\max}^{-1} G_{k,n}$. First we verify the required conditions. Note that $\lambda_{\max}$ is the maximum eigenvalue of $G_{k,n}$. It follows that

$$\|\lambda_{\max}^{-1} G_{k,n}\|_2 = \lambda_{\max}^{-1} \|G_{k,n}\|_2 \equiv \lambda_{\max}^{-1} \max_{\sum_{i=1}^k z_i^2 = 1} \left(\underline{z}' G_{k,n}^2 \underline{z} \right)^{1/2} \le 1.$$

On the other hand,

$$\|\lambda_{\max} G_{k,n}^{-1}\|_2 = (\lambda_{\max}/\lambda_{\min})\|\lambda_{\min} G_{k,n}^{-1}\|_2 \le (\lambda_{\max}/\lambda_{\min}).$$

By Lemma 6.2, for large n, $c_1 h/2 \le \lambda_{\min} \le \lambda_{\max} \le 2c_2 h$ and $(\lambda_{\max}/\lambda_{\min}) \le 4c_2/c_1$. By Theorem 2.2 of Demko (1977), there exist $c_* > 0$ and $\gamma \in (0,1)$ depending only on $4c_2/c_1$ and m such that $|\lambda_{\max}\alpha_{ij}| \le c^* \gamma^{|i-j|}$. Therefore

$$|\alpha_{ij}| \le c^* \lambda_{\max}^{-1} \gamma^{|i-j|} \le c_3 h^{-1} \gamma^{|i-j|},$$

where $c_3 = 2c^*/c_1$. This completes the proof of Lemma 6.3. □

We now introduce a result of Barrow and Smith (1978) [see (2.7) of that paper] on the approximation error of spline functions in $S(m, \underline{t})$:

$$\inf_{s(x) \in S(m,\underline{t})} \|f(x) + b^*(x) - s(x)\|_{L_\infty} = o(h^m), \tag{20}$$

where

$$b^*(x) = -\frac{f^{(m)}(t_i) h_i^m}{m!} B_m\left(\frac{x - t_i}{h_i}\right).$$

For other reasons, Barrow and Smith assumed that the knots are generated according to a positive density in their paper. However, the proof of (20) uses only the assumption of (3). From (20), it follows that there exists an $s_f(x) \in S(m, \underline{t})$ such that

$$f + b^*(x) - s_f(x) = o(h^m). \tag{21}$$

Proof of (5) of Theorem 2.1. By (2), it follows from the standard least squares calculation that

$$\hat{f}(x) = \mathbf{N}'(x) G_{k,n}^{-1} \mathbf{X}\mathbf{Y}, \tag{22}$$

where $\mathbf{Y} = (y_1, \dots, y_n)'$. From (21), we have

$$\begin{aligned} E(\hat{f}(x)) - f(x) &= [s_f(x) - f(x)] + [E(\hat{f}(x)) - s_f(x)] \\ &= b^*(x) + [E(\hat{f}(x)) - s_f(x)] + o(h^m). \end{aligned}$$

Because $f \in C^m[0,1]$, we have $f^{(m)}(x) = f^{(m)}(t_i) + o(1)$ and therefore $b(x) = b^*(x) + o(h^m)$. It follows that

$$E(\hat{f}(x)) - s_f(x) = b(x) + [E(\hat{f}(x)) - s_f(x)] + o(h^m). \tag{23}$$

Hence it suffices to show that $E(\hat{f}(x)) - s_f(x) = o(h^m)$. Note that since $s_f(x) \in S(m, \underline{t})$, it follows from (22) that

$$E(\hat{f}(x)) - s_f(x) = \mathbf{N}'(x) G_{k,n}^{-1} \mathbf{X}(f(\underline{x}) - s_f(\underline{x})) = \mathbf{N}'(x) G_{k,n}^{-1} \underline{r}, \tag{24}$$

where $\underline{r} = (r_1, \dots, r_k)$ with $r_i = \int_0^1 N_i(x)(f - s_f)(x)\, dQ_n(x)$. Under assumptions (3) and (4), $\max_{1 \le i \le k} r_i = o(h^{m+1})$, as shown in the proof of Lemma 6.10 of Agarwal and Studden (1980). Note that $N_i(x) \le 1$ for any $x \in [0,1]$ and $1 \le i \le k$. By (24) and Lemma 6.3, we have

$$\left\| E(\hat{f}(x)) - s_f(x) \right\|_{L_\infty} \le \|G_{k,n}^{-1}\|_\infty o(h^{m+1}) = o(h^m). \tag{25}$$

Hence (5) follows from (23) and (25). □

PROOF OF (6) OF THEOREM 2.1. When $m = 1$, (6) can be easily verified. We now discuss the case of $m \geq 2$. From (22), we have

$$\operatorname{Var}\big(\hat{f}(x)\big) = \frac{\sigma^2}{n}\mathbf{N}'(x)G_{k,n}^{-1}\mathbf{N}(x). \tag{26}$$

Noting that $N_j(x) = 0$ when $x \notin [t_{j-m}, t_j]$ [de Boor (1972), page 52], we have

$$N_j(x) = 0 \quad \text{if } j < i_x \text{ or } j > i_x + m - 1, \tag{27}$$

where i_x is the integer such that $x \in [t_{i_x-1}, t_{i_x}]$. Then

$$\operatorname{Var}\big(\hat{f}(x)\big) = \frac{\sigma^2}{n}\sum_{i=i_x}^{i_x+m}\sum_{j=i_x}^{i_x+m}\alpha_{ij}N_i(x)N_j(x), \tag{28}$$

where α_{ij} is the (i, j)th element of $G_{k,n}^{-1}$. By Lemma 6.4,

$$\alpha_{ij}N_i(x)N_j(x) - \alpha_{ij}(q)N_i(x)N_j(x) = o(h^{-1}),$$

where $\alpha_{ij}(q)$ is the (i, j)th element of $G^{-1}(q)$. Hence

$$\operatorname{Var}\big(\hat{f}(x)\big) = \frac{\sigma^2}{n}\sum_{i=i_x}^{i_x+m}\sum_{j=i_x}^{i_x+m}\alpha_{ij}(q)N_i(x)N_j(x) + o(nh^{-1}).$$

From (27), we know that

$$\frac{\sigma^2}{n}\sum_{i=i_x}^{i_x+m}\sum_{j=i_x}^{i_x+m}\alpha_{ij}(q)N_i(x)N_j(x) = \frac{\sigma^2}{n}\mathbf{N}'(x)G^{-1}(q)\mathbf{N}(x),$$

which completes the proof of (6). □

LEMMA 6.4.

$$\max_{1\leq i,j\leq k}\big|\alpha_{ij} - \alpha_{ij}(q)\big| = o(h^{-1}). \tag{29}$$

PROOF OF LEMMA 6.4. We first show that the elements of $G_{k,n} - G(q)$ are of order $o(h)$. Let g_{ij} and $g_{ij}(q)$ be the (i, j)th element of $G_{k,n}$ and $G(q)$, respectively. Then

$$g_{ij}(q) - g_{ij} = \int_0^1 N_i(z)N_j(z)\,d(Q - Q_n)(z). \tag{30}$$

Using integration by parts and (27), we have

$$\begin{aligned} g_{ij}(q) - g_{ij} &= N_i(z)N_j(z)(Q_n - Q)(z)\big|_0^1 \\ &\quad - \int_{t_{\max(i,j)-m}}^{t_{\min(i,j)}}(Q_n - Q)(z) \\ &\quad \times\big(N_i^{(1)}(z)N_j(z) + N_i(z)N_j^{(1)}(z)\big)\,dz \end{aligned} \tag{31}$$

for any $|i - j| < m$, where $N_i^{(1)}(z)$ is the derivative of $N_i(z)$. Note that $(Q_n - Q)(1) = (Q_n - Q)(0) = 0$, $k_0 h < M$ and $\|N_i^{(1)}\|_{L_\infty} \leq ck_0$, $i = 1, \ldots, k$, for some constant $c > 0$ independent of k_0 [see de Boor (1978), page 138]. By (4), we have

$$\begin{aligned} &\left|g_{ij}(q) - g_{ij}\right| \\ &\quad = o(h)\int_{t_{\max(i,j)-m}}^{t_{\min(i,j)}} \left(N_i^{(1)}(z)N_j(z) + N_i(z)N_j^{(1)}(z)\right)dz \qquad (32)\\ &\quad \le o(h)2ck_0mh = o(h). \end{aligned}$$

Let $B = (b_{ij}) = G(q)G_{k,n}^{-1}$. It follows from (32) that

$$\begin{aligned} b_{ij} &= \sum_{l=1}^{k} g_{il}(q)\alpha_{lj} = \sum_{l=\max(i-m,1)}^{\min(k,i+m-1)} g_{il}(q)\alpha_{lj} = \sum_{l=\max(i-m,1)}^{\min(k,i+m-1)} (g_{il} + o(h))\alpha_{ij} \\ &= \delta_{ij} + o(h)\sum_{l=\max(i-m,1)}^{\min(k,i+m-1)} \alpha_{lj}, \end{aligned}$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. Using the fact that $\gamma^{|l-j|} \le \gamma^{|i-j|-m}$ for any $l = \max(i-m,1),\dots,\min(k, i+m-1)$ we have from Lemma 6.3 that

$$|b_{ij} - \delta_{ij}| \le o(1)m\gamma^{|i-j|-m} = o(1)\gamma^{|i-j|}. \qquad (33)$$

From (33) and the fact that $G_{k,n}^{-1} - G^{-1}(q) = G^{-1}(q)(B - I)$, it follows that

$$\left|\alpha_{ij} - \alpha_{ij}(q)\right| = \left|\sum_{l=1}^{k} \alpha_{il}(q)(b_{lj} - \delta_{lj})\right| = \sum_{l=1}^{k}\alpha_{il}(q)o(1)\gamma^{|l-j|}. \qquad (34)$$

Using arguments similar to those in the proof of Lemma 6.3, we have that there exist constants $C_q > 0$ and $\gamma_q \in (0,1)$ (independent of k_0 and n) such that

$$\left|\frac{t_i - t_{i-m}}{m}\alpha_{ij}(q)\right| \le C_q\gamma_q^{|i-j|}.$$

Hence

$$\left|\alpha_{ij}(q)\right| < C_q M^{-1}h^{-1}\gamma_q^{|i-j|}. \qquad (35)$$

It follows from (34) and (35) that

$$\begin{aligned} \left|\alpha_{ij} - \alpha_{ij}(q)\right| &\le \sum_{l=1}^{k} C_q M^{-1}h^{-1}\gamma_q^{|i-j|}o(1)\gamma^{|l-j|} \\ &= o(h^{-1})\sum_{l=1}^{k}\gamma_q^{|i-l|}\gamma^{|l-j|} \le o(h^{-1})\sum_{l=1}^{k}\gamma_d^{|i-l|+|l-j|} \\ &= o(h^{-1})\gamma_d^{|i-j|}\left(|i-j| + \left(\sum_{l=1}^{\min(i,j)} + \sum_{l=1}^{k-\max(i,j)}\right)\gamma_d^l\right) \\ &= o(h^{-1})(|i-j| + 2/\gamma_d)\gamma_d^{|i-j|} = o(h^{-1}), \end{aligned}$$

where $\gamma_d = \max(\gamma, \gamma_q) < 1$. In the last equality, we have used the fact that

$$\max_{1\le i,j\le k}(|i-j| + 2/\gamma_d)\gamma_d^{|i-j|} < c_5,$$

where $c_5 < \infty$ is a constant (independent of k and n). This completes the proof of Lemma 6.4. □

PROOF OF THEOREM 2.2. In the case where the design points $\{x_i\}_{i=1}^n$ are sampled from a distribution $Q(x)$, it follows from the Glivenko–Cantelli theorem [see, e.g., Gaenssler and Wellner (1981)] that

$$\max_{0 \le x \le 1} |Q_n(x) - Q(x)| = O_p(n^{-1/2}).$$

Using arguments similar to those in the proof of Theorem 2.1, we complete the proof of Theorem 2.2. □

LEMMA 6.5. *If A and B are $l \times l$ nonnegative matrices, then*

$$\lambda_{\min}^A \mathrm{Tr}(B) \le \mathrm{Tr}(AB) \le \lambda_{\max}^A \mathrm{Tr}(B),$$

where $\lambda_{\min}^A$ and $\lambda_{\max}^A$ are the minimum and maximum eigenvalues of A, respectively.

PROOF. There exists an orthonormal matrix C such that

$$A = CD^A C',$$

where $D^A = \mathrm{diag}(\lambda_1^A, \dots, \lambda_l^A)$ and $\{\lambda_i^A\}_{i=1}^l$ are the eigenvalues of A. Then we have

$$\mathrm{Tr}(AB) = \mathrm{Tr}(CD^A C'B) = \mathrm{Tr}(D^A C'BC).$$

Because $C'BC$ is also nonnegative definite, we have

$$\lambda_{\min}^A \mathrm{Tr}(C'BC) \le \mathrm{Tr}(AB) \le \lambda_{\max}^A \mathrm{Tr}(C'BC).$$

Lemma 6.5 then follows from the fact that $\mathrm{Tr}(C'BC) = \mathrm{tr}(BCC') = \mathrm{Tr}(B)$. □

LEMMA 6.6.

$$\big((c_2 m)^{-1} + o(1)\big)\sigma^2 (nh)^{-1} \le \mathrm{Var}\big(\hat{f}(x)\big) \le \big(c_1^{-1} + o(1)\big)\sigma^2 (nh)^{-1}. \tag{36}$$

PROOF. By (22) and Lemma 6.5, we have

$$\begin{aligned}
\mathrm{Var}\big(\hat{f}(x)\big) &= n^{-1}\sigma^2 \mathbf{N}'(x) G_{k,n}^{-1} \mathbf{N}(x) = n^{-1}\sigma^2 \mathrm{Tr}\big(G_{k,n}^{-1}\mathbf{N}(x)\mathbf{N}'(x)\big) \\
&\le n^{-1}\sigma^2 \lambda_{\min}^{-1} \mathrm{Tr}(\mathbf{N}(x)\mathbf{N}'(x)) = n^{-1}\sigma^2 \lambda_{\min}^{-1} \sum_{i=1}^k N_i^2(x) \\
&\le n^{-1}\sigma^2 \lambda_{\min}^{-1} \left(\sum_{i=1}^k N_i(x)\right)^2 = n^{-1}\sigma^2 \lambda_{\min}^{-1}.
\end{aligned}$$

It then follows from Lemma 6.2 that

$$\mathrm{Var}\big(\hat{f}(x)\big) \le \frac{\big(c_1^{-1} + o(1)\big)\sigma^2}{nh}.$$

Using similar arguments, we obtain

$$\operatorname{Var}\big(\hat{f}(x)\big) \ge n^{-1}\sigma^2\lambda_{\max}^{-1}\sum_{i=1}^{k} N_i^2(x).$$

By (27), we have

$$\sum_{i=1}^{k} N_i^2(x) = \sum_{i=i_x}^{i_x+m-1} \big(N_i(x)\big)^2 \ge \frac{1}{m}\left(\sum_{i=i_x}^{i_x+m-1} N_i(x)\right)^2 = \frac{1}{m}.$$

Thus

$$\operatorname{Var}\big(\hat{f}(x)\big) \ge \frac{\sigma^2}{mn\lambda_{\max}}.$$

The left-hand inequality of (36) then follows from Lemma 6.2, completing the proof of Lemma 6.6. □

PROOF OF THEOREM 3.1. From (36) and Theorem 2.1, we have

$$\frac{E\hat{f}(x)}{\sqrt{\operatorname{Var}\big(\hat{f}(x)\big)}} - \frac{f(x)+b(x)}{\sqrt{\operatorname{Var}\big(\hat{f}(x)\big)}} = \frac{o(k_0^{-m})}{\sqrt{k_0/n}} = o\big(n^{1/2}k_0^{-(m+1/2)}\big) = o(1).$$

Thus (9) follows if

$$\frac{\hat{f}(x) - E\hat{f}(x)}{\sqrt{\operatorname{Var}\big(\hat{f}(x)\big)}} \to_d N(0,1).$$

From (22), we have

$$\hat{f}(x) - E\hat{f}(x) = \mathbf{N}'(x)G_{k,n}^{-1}\mathbf{X}\underline{\varepsilon} = \sum_{i=1}^{n} a_i\varepsilon_i, \tag{37}$$

where $a_i = \mathbf{N}'(x)G_{k,n}^{-1}\mathbf{N}(x_i)/n$. To check the required Lindeberg–Feller conditions, it suffices to verify that

$$\max_{1\le i\le n}\big(a_i^2\big) = o\left(\sum_{i=1}^{n} a_i^2\right) = o\Big(\operatorname{Var}\big(\hat{f}(x)\big)\Big). \tag{38}$$

By Lemma 6.5, we have

$$\begin{aligned}
a_i^2 n^2 &= \mathbf{N}'(x)G_{k,n}^{-1}\mathbf{N}(x_i)\mathbf{N}'(x_i)G_{k,n}^{-1}\mathbf{N}(x)\\
&= \operatorname{Tr}\big(\mathbf{N}(x_i)\mathbf{N}'(x_i)G_{k,n}^{-1}\mathbf{N}(x)\mathbf{N}'(x)G_{k,n}^{-1}\big)\\
&\le e_i \operatorname{Tr}\big(G_{k,n}^{-1}\mathbf{N}(x)\mathbf{N}'(x)G_{k,n}^{-1}\big)\\
&\le e_i\lambda_{\min}^{-2}\operatorname{Tr}(\mathbf{N}(x)\mathbf{N}'(x)),
\end{aligned}$$

where e_i is the maximum eigenvalue of $\mathbf{N}(x_i)\mathbf{N}'(x_i)$. By definition, for any $x \in [0,1]$, $0 \le N_i(x) \le 1$ and $\sum_{i=1}^{k} N_i(x) = 1$, $1 \le i \le k$. This implies that $\operatorname{Tr}(\mathbf{N}(x)\mathbf{N}'(x)) \le 1$ and $e_i \le 1$. Therefore

$$a_i^2 \le \lambda_{\min}^{-2}/n^2.$$

It follows from (17) and (36) that

$$\frac{a_i^2}{\mathrm{Var}(\hat{f}(x))} \le \frac{n^{-2}}{[(c_1 + o(1))h]^2} \frac{nh}{(c_2^{-1} + o(1))\sigma^2} = \frac{(c_1 + o(1))^2}{(c_2^{-1} + o(1))hn\sigma^2}.$$

Then (38) follows from the assumption that $k_0/n \to 0$, $hn \to \infty$. Thus the proof of (9) of Theorem 3.1 is complete. For random designs, the result can be established similarly. □

PROOF OF THEOREM 4.1. Using arguments similar to those in the proof of Theorem 3.1, we have

$$\frac{\hat{f}(x) - f(x)}{\sqrt{\mathrm{Var}(\hat{f}(x))}} \to_d N(0,1).$$

Therefore the desired asymptotic confidence interval can be constructed. □

PROOF OF THEOREM 4.2. From (22), we have

$$\hat{f}(x) - E(\hat{f}(x)) = \gamma(x)\underline{\varepsilon},$$

where $\gamma(x) = \mathbf{N}'(x)G_{k,n}^{-1}\mathbf{X}$. From Johansen and Johnstone (1990) [see (2.16) on page 661], we have, for any $c \in (0,1)$,

$$P\left(\sup_{0 \le x \le 1} \frac{|\hat{f}(x) - E(\hat{f}(x))|}{\sqrt{\mathrm{Var}(\hat{f}(x))}} \ge c\right) \ge \frac{|\gamma|}{\pi} \exp\left(\frac{-c^2}{2}\right) + 2(1 - \Phi(c)).$$

It follows from (10) that

$$P\left(\sup_{0 \le x \le 1} \frac{|\hat{f}(x) - E(\hat{f}(x))|}{\sqrt{\mathrm{Var}(\hat{f}(x))}} \le c(\alpha)\right) \ge 1 - \alpha.$$

Hence

$$P\left(\sup_{0 \le x \le 1} \frac{|\hat{f}(x) - f(x)|}{\sqrt{\mathrm{Var}(\hat{f}(x))}} \le c(\alpha) + \sup_{0 \le x \le 1} \frac{|E(\hat{f}(x)) - f(x)|}{\sqrt{\mathrm{Var}(\hat{f}(x))}}\right) \ge 1 - \alpha.$$

By Lemmas 6.6 and 6.7, Theorem 2.1 and the assumption that $k_0 \ge Cn^{1/(2m+1)}$ [i.e., $h = O(n^{-1/(2m+1)})$], we have

$$\begin{aligned}(c(\alpha))^{-1}&\left(c(\alpha) + \sup_{0 \le x \le 1} \frac{|E(\hat{f}(x)) - f(x)|}{\sqrt{\mathrm{Var}(\hat{f}(x))}}\right)\\ &= 1 + O\left((\log(h^{-1}))^{-1/2} n^{1/2} h^{m+1/2}\right) \to 1.\end{aligned}$$

It follows that, for large n,

$$P\left(\sup_{0\le x\le 1}\frac{|\hat{f}(x)-f(x)|}{\sqrt{\operatorname{Var}(\hat{f}(x))}}\le c(\alpha)\right)\ge 1-\alpha,$$

completing the proof. □

LEMMA 6.7. *For any fixed* $\alpha\in(0,1)$, *there exists a constant* $c_6>0$ (*independent of* k_0 *and* n) *such that*

$$c(\alpha)\ge (c_6+o(1))\sqrt{\log(h^{-1})}.$$

PROOF. From (10) and Lemma 6.6, it suffices to show that

$$|\gamma|\ge (c_7+o(1))h^{-1}, \tag{39}$$

where $c_7>0$ is a constant (independent of k_0 and n). From de Boor [(1972), page 54], we know that

$$d\mathbf{N}(x)/dx = D\mathbf{N}_{m-1}(x),$$

where

$$D=(m-1) \times\begin{pmatrix} -(t_1-t_{1-m+1})^{-1} & 0 & \cdots & 0 & 0\\ (t_1-t_{1-m+1})^{-1} & -(t_2-t_{2-m+1})^{-1} & \cdots & 0 & 0\\ 0 & (t_2-t_{2-m+1})^{-1} & \cdots & 0 & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & \cdots & 0 & (t_{k-1}-t_{k-m})^{-1}\end{pmatrix}. \tag{40}$$

It follows from (11) that

$$|\gamma|=\int_0^1\frac{\left(\mathbf{N}'(x)G_{k,n}^{-1}A'G_{k,n}^{-1}AG_{k,n}^{-1}\mathbf{N}(x)\right)^{1/2}}{\left(\mathbf{N}'(x)G_{k,n}^{-1}\mathbf{N}(x)\right)^{3/2}}\,dx, \tag{41}$$

where

$$A=D\mathbf{N}_{m-1}(x)\mathbf{N}'(x)-\mathbf{N}(x)\mathbf{N}'_{m-1}(x)D'.$$

Note that $\mathbf{N}'(x)G_{k,n}^{-1}\mathbf{N}(x)=n\operatorname{Var}(\hat{f}(x))\sigma^{-2}$. By Lemma 6.6,

$$\left((c_2m)^{-1}+o(1)\right)h^{-1}\le \mathbf{N}'(x)G_{k,n}^{-1}\mathbf{N}(x)\le\left(c_1^{-1}+o(1)\right)h^{-1}. \tag{42}$$

By Lemma 6.5,

$$\begin{aligned}\mathbf{N}'(x)G_{k,n}^{-1}A'G_{k,n}^{-1}AG_{k,n}^{-1}\mathbf{N}(x) &\ge \lambda_{\max}^{-1}\operatorname{Tr}\left(AG_{k,n}^{-1}\mathbf{N}(x)\mathbf{N}'(x)G_{k,n}^{-1}A'\right)\\ &=\lambda_{\max}^{-1}\mathbf{N}'(x)G_{k,n}^{-1}A'AG_{k,n}^{-1}\mathbf{N}(x).\end{aligned}$$

It follows from (41), (42) and the above inequality that (39) holds if

$$\int_0^1\left(\mathbf{N}'(x)G_{k,n}^{-1}A'AG_{k,n}^{-1}\mathbf{N}(x)\right)^{1/2}dx\ge (c_8+o(1))h^{-2}, \tag{43}$$

where $c_8 > 0$ is a constant (independent of k_0 and n). To verify (43), let $\rho(x) = \mathbf{N}'(x)G_{k,n}^{-1}A'AG_{k,n}^{-1}\mathbf{N}(x)$. After some calculations, we have

$$\rho(x) = \big(g_1(x)\mathbf{N}(x) - g_2(x)D\mathbf{N}_{m-1}(x)\big)' \times \big(g_1(x)\mathbf{N}(x) - g_2(x)D\mathbf{N}_{m-1}(x)\big), \tag{44}$$

where

$$g_1(x) = \mathbf{N}'_{m-1}(x)D'G_{k,n}^{-1}\mathbf{N}(x) \quad \text{and} \quad g_2(x) = \mathbf{N}'(x)G_{k,n}^{-1}\mathbf{N}(x).$$

We now construct a vector which is orthogonal to $\mathbf{N}(x)$. Let $L'(x) = (l_1(x), l_2(x), \ldots, l_k(x))$, where

$$l_i(x) = \begin{cases} 1 - N_{i_x}(x), & i = i_x, \\ -N_{i_x}(x), & i = i_x + 1, \ldots, i_x + m - 1, \\ 0, & i < i_x \text{ or } i \geq i_x + m, \end{cases}$$

and i_x is the integer such that $x \in [t_{i_x - 1}, t_{i_x}]$. Note that

$$L'(x)L(x) = \big(1 - N_{i_x}(x)\big)^2 + (m-1)N_{i_x}^2(x) \leq m. \tag{45}$$

From (27) and (45), we have

$$\begin{aligned} L'(x)\mathbf{N}(x) &= \sum_{j=i_x}^{i_x+m-1} l_j(x)N_j(x) \\ &= \big(1 - N_{i_x}(x)\big)N_{i_x}(x) - N_{i_x}(x)\sum_{j=i_x+1}^{i_x+m-1} N_j(x) \\ &= \big(1 - N_{i_x}(x)\big)N_{i_x}(x) - N_{i_x}(x)\big(1 - N_{i_x}(x)\big) = 0. \end{aligned} \tag{46}$$

By (40),

$$\begin{aligned} L'(x)D\mathbf{N}_{m-1}(x) &= (m-1)\sum_{j=1}^{k-1}\left(\frac{l_j(x) - l_{j+1}(x)}{t_j - t_{j-m+1}}\right)N_{j,m-1}(x) \\ &= (m-1)\sum_{j=i_x}^{i_x+m-2}\left(\frac{l_j(x) - l_{j+1}(x)}{t_j - t_{j-m+1}}\right)N_{j,m-1}(x) \\ &= (m-1)\frac{1 - N_{i_x}(x) + N_{i_x}(x)}{t_{i_x} - t_{i_x-m+1}}N_{i_x,m-1}(x) \\ &= \frac{(m-1)N_{i_x,m-1}(x)}{t_{i_x} - t_{i_x-m+1}}. \end{aligned} \tag{47}$$

It follows from (42), (44), (45), (46), (47) and Hölder's inequality that

$$\begin{aligned}\rho(x)m &= \rho(x)L'(x)L(x) \geq \big(L'(x)\big[g_1(x)\mathbf{N}(x) - g_2(x)D\mathbf{N}_{m-1}(x)\big]\big)^2\\ &= g_2^2(x)\big(L'(x)D\mathbf{N}_{m-1}(x)\big)^2\\ &= g_2^2(x)\big(t_{i_x} - t_{i_x-m+1}\big)^{-2}(m-1)^2 N_{i_x,m-1}^2(x)\\ &\geq \Big[\big((c_2m)^{-1} + o(1)\big)h^{-2}\Big]\Big[(m-1)^2M^2h^{-2}\Big]N_{i_x,m-1}^2(x).\end{aligned}$$

Hence

$$\int_0^1(\rho(x))^{1/2}\,dx \geq \big[c_2^{-1}(m-1)Mm^{-3/2} + o(1)\big]h^{-2}\int_0^1 N_{i_x,m-1}(x)\,dx. \tag{48}$$

Using a recurrence relation of B-splines [see, e.g., de Boor (1972), page 52], we have, for any $x \in [t_{i-1}, t_i]$, $i = 1, \dots, k_0 + 1$,

$$N_{i,m-1}(x) = (t_i - x)^{m-2}\prod_{j=1}^{m-2}\big(t_i - t_{i+j-m+1}\big)^{-1} \geq h^{-m+2}(t_i - x)^{m-2}/(m-2)!.$$

It follows that

$$\begin{aligned}\int_0^1 N_{i_x,m-1}(x)\,dx &= \sum_{i=1}^{k_0+1}\int_{t_{i-1}}^{t_i} N_{i,m-1}(x)\,dx \geq k_0(h/M)^{m-1}h^{-m+2}/(m-1)!\\ &\geq \big[M^{m-1}(m-1)!\big]^{-1}.\end{aligned}$$

Equation (43) then follows from (48) and the above inequality, completing the proof. □

Acknowledgments. We would like to thank the editors and the referees for their valuable comments and suggestions.

REFERENCES

AGARWAL, G. G. and STUDDEN W. J. (1980). Asymptotic integrated mean square error using least squares and bias minimizing splines. *Ann. Statist.* **8** 1307–1325.

BARROW, D. L. and SMITH, P. W. (1978). Asymptotic properties of best $L_2[0,1]$ approximation by spline with variable knots. *Quart. Appl. Math.* **36** 293–304.

CRAVEN, P. and WAHBA, G. (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the generalized cross-validation. *Numer. Math.* **31** 337–403.

DE BOOR, C. (1972). On calculating with B-splines. *J. Approx. Theory* **6** 50–62.

DE BOOR, C. (1978). *A Practical Guide to Splines*. Springer, New York.

DEMKO, S. (1977). Inverses of band matrices and local convergence of spline projection. *SIAM J. Numer. Anal.* **14** 616–619.

FRIEDMAN, J. H. (1991). Multivariate adaptive regression splines (with discussion). *Ann. Statist.* **19** 1–141.

GAENSSLER, P. and WELLNER, J. A. (1981). Glivenko–Cantelli theorems. *Encyclopedia of Statistical Science* **3** 442–445. Wiley, New York.

GASSER, T. and MÜLLER, H. G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.* **11** 171–185.

GASSER, T., SROKA, L. and JENNEN-STEINMETZ, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* **73** 625–633.

HALL, P., KAY, J. W. and TITTERINGTON, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77** 521–528.

HUANG, S. Y. and STUDDEN, W. Y. (1993). An equivalent kernel method for least squares spline regression. *Statist. Decisions* **3** 179–201 (Supplement).

JOHANSEN, S. and JOHNSTONE, M. (1990). Hotelling's theorem on the volume of tubes: some illustrations in simultaneous inference and data analysis. *Ann. Statist.* **18** 652–684.

KOENKER, R., NG, P. and PORTNOY, S. (1994). Non-parametric estimation of conditional quantile functions. *Biometrika* **81** 673–680.

KOOPERBERG, C., STONE, C. and TRUONG, Y. K. (1995). The L_2 rate of convergence for hazard regression. *Scand. J. Statist.* **22** 143–157.

NADARAYA, E. A. (1964). On estimating regression. *Theory Probab. Appl.* **9** 141–142.

PORTNOY, S. (1997). Local asymptotics for quantile smoothing splines. *Ann. Statist.* **25** 414–434.

SCHUMAKER, L. L. (1981). *Spline Functions: Basic Theory*. Wiley, New York.

SHEN, X. and HU, C. (1995). Universal sieves and spline adaptation. Technical Report 543, Dept. Statist., Ohio State Univ.

SHI, P. and LI, G. (1994). On the rate of convergence of "minimum L_1-norm" estimates in a partly linear model. *Comm. Statist. Theory Methods* **23** 175–196.

SHI, P. and LI, G. (1995). Global convergence rates of *B*-spline *M*-estimators in nonparametric regression. *Statist. Sinica* **5** 303–318.

STONE, C. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.

STONE, C. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13** 689–705.

STONE, C. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.* **14** 590–606.

STONE, C. (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *Ann. Statist.* **22** 118–184.

STONE, C. and KOO, C. Y. (1986). Additive splines in statistics. In *Proceedings of the Statistical Computing Section* 45–48. Amer. Statist. Assoc., Washington, DC.

S. ZHOU
DEPARTMENT OF STATISTICS
OHIO STATE UNIVERSITY
COLUMBUS, OHIO 43210-1247
E-MAIL: zhou@stat.ohio-state.edu

X. SHEN
D. A. WOLFE
DEPARTMENT OF STATISTICS
OHIO STATE UNIVERSITY
COLUMBUS, OHIO 43210-1247