JOURNAL OF MULTIVARIATE ANALYSIS **60**, 1–19 (1997)
ARTICLE NO. MV961642

Local Linear Estimation in Partly Linear Models

Scott A. Hamilton

Genentech, Inc.

and

Young K. Truong

The University of North California at Chapel Hill

Let $(\mathbf{X}, B, Y)$ denote a random vector such that B and Y are real-valued, and $\mathbf{X} \in \mathbb{R}^2$. Local linear estimates are used in the partial regression method for estimating the regression function $E(Y \mid \mathbf{X}, B) = \alpha B + m(\mathbf{X})$, where α is an unknown parameter, and $m(\cdot)$ is a smooth function. Under appropriate conditions, asymptotic distributions of estimates of α and $m(\cdot)$ are established. Moreover, it is shown that these estimates achieve the best possible rates of convergence in the indicated semi-parametric problems.

1. INTRODUCTION

Let $(\mathbf{X}, B, Y)$ denote a random vector such that B and Y are real-valued, and $\mathbf{X} \in \mathbb{R}^2$. In partly linear models, the regression function is given by

$$E(Y \mid \mathbf{X}, B) = \alpha B + m(\mathbf{X}), \tag{1.1}$$

where α is an known parameter and $m(\cdot)$ is a real-valued smooth function on $\mathbb{R}^2$. Model (1.1) is useful in many applications. For instance, it is an extension of the classical analysis of covariance models (Scheffé [17]), where B is a linear effect of special interest, and $\mathbf{X}$ is a vector of covariates that have an unspecified effect on the outcome. Alternatively, model (1.1) can be used to alleviate the *curse of dimensionality* in nonparametric regression. The objective of the present paper is to find a root-n consistent

Received August 31, 1995; revised May 23, 1996.

AMS 1991 subject classifications: primary 62G07; secondary 62G20.

Key words and phrases: partial linear models, semi-parametric models, design-adaptive, nonparametric regression, local polynomial estimator, optimal rate of convergence.

0047-259X/97 $25.00

estimate of α, and to provide a smooth estimate of $m(\cdot)$ that achieves the usual nonparametric rate of convergence.

To motivate our approach, we consider the partly linear model

$$Y = \alpha B + s(X) + \varepsilon,$$

where $s(\cdot)$ is a real-valued smooth function, X is a real-valued non-random effect and ε is a random error with mean zero and finite variance. Wahba [25] described penalized spline estimates of α and $s(\cdot)$. See also, Engle *et al.* [5], Green *et al.* [12], and Shiau *et al.* [18]. Using the same approach, Heckman [13] proved the asymptotic normality of the estimate of α in a balanced design setup. Rice [15] showed that when B and X are correlated, the asymptotic bias of the estimate of α dominated the variance and the root-n rate could only be achieved if the estimate of $s(\cdot)$ was undersmoothed. Consequently, the estimate was not optimal. A similar result is reported in Eubank and Whitney [6]. On the other hand, optimal rates of convergence can be achieved by adopting the partial regression approach proposed independently by Denby [4] and Speckman [19], even when B and X are related. In fact, this has been demonstrated by Speckman [19] using the kernel method; by Chen [1] using a piecewise constant smoother; by Chen and Shiau [2, 3] using smoothing splines; and by Eubank, Hart and Speckman [7] using a trigonometric series approach.

The partial regression method can be improved further by using local linear smoothers. Local linear estimates, or more generally, local polynomial estimates were studied by Stone [20–22]. These estimates have been compared favorably to the kernel-based methods since they have appealing asymptotic bias and variance terms that are not adversely affected by estimation at the boundary, see Fan [8, 9], and Fan and Gijbels [10]. Using a minimax argument, Fan [9] showed that within the class of linear estimators which includes kernel and spline estimates, the local linear estimates achieve the best possible constant and rates of convergence. In light of these desirable properties, we modified the partial regression method by using local linear smoothers to estimate $m(\cdot)$ and α. Under (1.1) and appropriate conditions, we show that these estimates achieve the indicated rates of convergence, and that they also have asymptotic normal distributions.

The rest of the paper is organized as follows. Section 2 describes the partial regression method using local linear smoothers. Asymptotic properties of the estimates are given in Section 3. In particular, root-n consistency of the parametric estimate and the usual nonparametric rate of convergence for smooth estimates are discussed. Proofs are given in Section 4.

2. METHOD

Suppose (1.1) holds. Let $\mathbf{x} = (x_1, x_2)' \in \mathbb{R}^2$ and let $u(\mathbf{x})$ denote the regression function of B on $\mathbf{X}$, so that $u(\mathbf{x}) = E(B \mid \mathbf{X} = \mathbf{x})$. (In this paper, A' denote the transpose of a matrix A.) We start the method by describing the local linear estimate of $u(\mathbf{x})$.

Let $(\mathbf{X}_1, B_1, Y_1), ..., (\mathbf{X}_n, B_n, Y_n)$ denote a random sample from the distribution of $(\mathbf{X}, B, Y)$. Let $K(\cdot)$ denote a kernel function on $\mathbb{R}^2$ and $\mathbf{H}$ denotes a 2×2 symmetric, positive-definite matrix. Set $K_{\mathbf{H}}(\mathbf{x}) = |\mathbf{H}|^{-1} K(\mathbf{H}^{-1}\mathbf{x})$. Here $|\mathbf{H}|$ is the determinant of $\mathbf{H}$, and $\mathbf{H}$ is referred to as the bandwidth matrix (see Wand and Jones [26]). Let $\mathbf{X}(\mathbf{x})$ denote the $n \times 3$ matrix with the ith row given by the vector

$$\mathbf{X}_i'(\mathbf{x}) = [1, \{\mathbf{H}^{-1}(\mathbf{X}_i - \mathbf{x})\}'], \qquad \mathbf{X}_i = (X_{i1}, X_{i2})', \qquad i = 1, ..., n.$$

Let $\mathbf{W}(\mathbf{x})$ denote the $n \times n$ diagonal matrix given by

$$\mathbf{W}(\mathbf{x}) = \operatorname{diag}[K_{\mathbf{H}}(\mathbf{X}_1 - \mathbf{x}), ..., K_{\mathbf{H}}(\mathbf{X}_n - \mathbf{x})].$$

Set $\mathbf{a}' = (1, 0, 0)$ and let $\mathbf{S}'(\mathbf{x})$ denote the $1 \times n$ row vector given by

$$\mathbf{S}'(\mathbf{x}) = \mathbf{a}'[\mathbf{X}'(\mathbf{x})\, \mathbf{W}(\mathbf{x})\, \mathbf{X}(\mathbf{x})]^{-1}\, \mathbf{X}'(\mathbf{x})\, \mathbf{W}(\mathbf{x}).$$

Set $\mathbf{B}' = (B_1, ..., B_n)$. Then the local linear estimator of $u(\mathbf{x})$ is given by

$$\hat{u}(\mathbf{x}) = \mathbf{a}'[\mathbf{X}'(\mathbf{x})\, \mathbf{W}(\mathbf{x})\, \mathbf{X}(\mathbf{x})]^{-1}\, \mathbf{X}'(\mathbf{x})\, \mathbf{W}(\mathbf{x})\, \mathbf{B} = \mathbf{S}'(\mathbf{x})\, \mathbf{B}.$$

See Stone [21, 22], and a sightly different formulation by Ruppert and Wand [16].

We next describe the estimates of α and $m(\mathbf{x})$. Let $\mathbf{S}$ denote the $n \times n$ matrix with the jth row equal to $\mathbf{S}'(\mathbf{X}_j)$, $j = 1, ..., n$. Let $\mathbf{I} = \mathbf{I}_n$ denote the $n \times n$ identity matrix and $\mathbf{Y}' = (Y_1, ..., Y_n)$. Set

$$\tilde{\mathbf{B}} = (\mathbf{I} - \mathbf{S})\, \mathbf{B} \qquad \text{and} \qquad \tilde{\mathbf{Y}} = (\mathbf{I} - \mathbf{S})\, \mathbf{Y}.$$

Let $\hat{\alpha}$ denote the solution to $\min_\alpha (\tilde{\mathbf{Y}} - \alpha \tilde{\mathbf{B}})'(\tilde{\mathbf{Y}} - \alpha \tilde{\mathbf{B}})$, so that $\hat{\alpha} = (\tilde{\mathbf{B}}'\tilde{\mathbf{B}})^{-1} \tilde{\mathbf{B}}'\tilde{\mathbf{Y}}$. Set

$$\hat{m}(\mathbf{x}) = \mathbf{a}'[\mathbf{X}'(\mathbf{x})\, \mathbf{W}(\mathbf{x})\, \mathbf{X}(\mathbf{x})]^{-1}\, \mathbf{X}'(\mathbf{x})\, \mathbf{W}(\mathbf{x})(\mathbf{Y} - \hat{\alpha}\mathbf{B}) = \mathbf{S}'(\mathbf{x})\, (\mathbf{Y} - \hat{\alpha}\mathbf{B}).$$

We use $\hat{\alpha}$ and $\hat{m}(\mathbf{x})$ to estimate α and $m(\mathbf{x})$ of model (1.1). Note that the kernel-based method described by Speckman [19] is a special case of the above procedure.

3. RESULTS

In the remainder of this paper, let $\|\mathbf{v}\| = \{\sum_j v_j^2\}^{1/2}$ denote the usual Euclidean norm of a vector $\mathbf{v}$. More generally, for a matrix $\mathbf{A} = (a_{ij})$, we denote its norm by $\|\mathbf{A}\| = \{\mathrm{tr}(\mathbf{A}'\mathbf{A})\}^{1/2} = \{\sum_{ij} |a_{ij}|^2\}^{1/2}$.

The following conditions are standard in nonparametric regression.

Condition 1. The functions $m(\cdot)$ and $u(\cdot)$ have bounded second partial derivatives.

The above condition is required for bounding the bias terms of the local linear estimates. The next two conditions are required for bounding the variance terms of our estimates.

Condition 2. The random vector $\mathbf{X}' = (X_1, X_2)$ has a continuous density function $f(\cdot)$ supported on a compact set $C_f \subset \mathbb{R}^2$. Moreover, $f(\cdot)$ is bounded away from zero and infinity on C_f.

Set $\sigma^2(b, \mathbf{x}) = \mathrm{var}(Y \,|\, B = b, \mathbf{X} = \mathbf{x})$ and $\sigma_B^2(\mathbf{x}) = \mathrm{var}(B \,|\, \mathbf{X} = \mathbf{x})$.

Condition 3. $\sigma_B^2(\mathbf{x})$ and $\sigma^2(b, \mathbf{x})$ are bounded from above. That is,

$$\sup_{b,\mathbf{x}} \sigma^2(b, \mathbf{x}) < \infty \qquad \text{and} \qquad \sup_{\mathbf{x}} \sigma_B^2(\mathbf{x}) < \infty.$$

The following condition ensures that the bias and variance of our estimates go to zero asymptotically.

Condition 4. The bandwidth matrix $\mathbf{H}$ is symmetric and positive-definite. Moreover,

$$\|\mathbf{H}\| \to 0, \qquad n^{-1}\,|\mathbf{H}|^{-2} \to 0, \qquad \text{and} \qquad |\mathbf{H}|^{-1}\,\|\mathbf{H}\|^4 \to 0 \qquad \text{as} \quad n \to \infty.$$

Suppose $m(\cdot)$ has a bounded second derivative. Then the above condition is easily satisfied by considering $\mathbf{H} = \mathrm{diag}(h, h)$ with $h \sim n^{-1/6}$. [The bandwidth is chosen to have the form $h \sim n^{-1/(2p+d)}$, here p is the order of smoothness and d is the dimensionality. In our context, $p = d = 2$.]

The following condition is required for the kernel function. It is easily satisfied by taking products of symmetric univariate kernels with compact support.

Condition 5. The kernel function $K(\cdot)$ is a continuous bivariate kernel function with compact support such that $\int K(\mathbf{u})\, d\mathbf{u} = 1$. Also, all odd order moments of K vanish; that is, $\int u_1^s u_2^t K(\mathbf{u})\, d\mathbf{u} = 0$ for non-negative integers s, t such that their sum is odd. Moreover, $\int \mathbf{u}\mathbf{u}' K(\mathbf{u})\, d\mathbf{u} = \mu_2(K)\,\mathbf{I}_2$, where $\mu_2(K) > 0$ and $\mathbf{I}_2$ is the 2×2 identity matrix.

The next condition is required for proving asymptotic normality of $\hat{\alpha}$ and $\hat{m}(\mathbf{x})$ (see Theorems 3 and 4).

Condition 6. $\sigma_B^2(\mathbf{x})$ and $\sigma^2(b, \mathbf{x})$ are bounded away from zero. That is,

$$0 < \inf_{\mathbf{x}} \sigma_B^2(\mathbf{x}) \qquad \text{and} \qquad 0 < \inf_{b, \mathbf{x}} \sigma^2(b, \mathbf{x}). \tag{3.1}$$

Moreover, for $0 < \delta < 1/2$,

$$\sup_{\mathbf{x}} E(|B - u(\mathbf{X})|^{2+\delta} \,|\, \mathbf{X} = \mathbf{x}) < \infty, \tag{3.2}$$

and

$$\sup_{b, \mathbf{x}} E(|Y - \alpha B - m(\mathbf{X})|^{2+\delta} \,|\, B = b, \mathbf{X} = \mathbf{x}) < \infty. \tag{3.3}$$

Note that Condition 6 implies Condition 3.
Set

$$\mathbf{M} = [m(\mathbf{X}_1), ..., m(\mathbf{X}_n)]', \qquad \hat{\mathbf{M}} = \mathbf{S}(\mathbf{Y} - \hat{\alpha}\mathbf{B}) = [\hat{m}(\mathbf{X}_1), ..., \hat{m}(\mathbf{X}_n)]',$$

and

$$\mathbf{V} = \text{diag}[\sigma^2(\mathbf{X}_1, B_1), ..., \sigma^2(\mathbf{X}_n, B_n)].$$

In a number of results below, we need to condition on $\mathbf{X}_1, ..., \mathbf{X}_n$, $B_1, ..., B_n$. In order to simplify the notation, we let $E_n(Z)$ and $\text{var}_n(Z)$ denote the conditional expectation and variance of a random variable Z given $\mathbf{X}_1, ..., \mathbf{X}_n$, $B_1, ..., B_n$; that is, $E_n(Z) = E(Z \,|\, \mathbf{X}_1, ..., \mathbf{X}_n, B_1, ..., B_n)$ and $\text{var}_n(Z) = \text{var}(Z \,|\, \mathbf{X}_1, ..., \mathbf{X}_n, B_1, ..., B_n)$. We also write $P_n(A) = P(A \,|\, \mathbf{X}_1, ..., \mathbf{X}_n, B_1, ..., B_n)$.

Our first result deals with the bias and variance of $\hat{\alpha}$.

THEOREM 1. *Suppose Conditions* 1–5 *hold. Then*

$$E_n(\hat{\alpha} - \alpha) = (\tilde{\mathbf{B}}'\tilde{\mathbf{B}})^{-1}\, \tilde{\mathbf{B}}'(\mathbf{I} - \mathbf{S})\, \mathbf{M},$$
$$\text{var}_n(\hat{\alpha}) = (\tilde{\mathbf{B}}'\tilde{\mathbf{B}})^{-1}\, \tilde{\mathbf{B}}'(\mathbf{I} - \mathbf{S})\, \mathbf{V}(\mathbf{I} - \mathbf{S})'\, \tilde{\mathbf{B}}(\tilde{\mathbf{B}}'\tilde{\mathbf{B}})^{-1}.$$

Moreover,

$$|\text{bias}_n(\hat{\alpha})| \equiv |E_n(\hat{\alpha} - \alpha)| = O_p(n^{-1/2}\, \|\mathbf{H}\|^2) \qquad \textit{and} \qquad \text{var}_n(\hat{\alpha}) = O_p(n^{-1}).$$

Proof of Theorem 1 is given in subsection 4.1; The above result shows that the squared bias of $\hat{\alpha}$ is asymptotically negligible compared with its variance without requiring $m(\cdot)$ to be undersmoothed, even when B and $\mathbf{X}$ are correlated. The next result deals with the bias and variance of $\hat{m}(\cdot)$.

THEOREM 2. *Suppose Conditions* 1–5 *hold. Then*

$$|E_n[\hat{m}(\mathbf{X}_i) - m(\mathbf{X}_i)]| = O_p(\|\mathbf{H}\|^2)$$

and

$$\mathrm{var}_n[\hat{m}(\mathbf{X}_i)] = O_p(n^{-1}\,|\mathbf{H}|^{-1}) \qquad i = 1, ..., n.$$

Proof of Theorem 2 is given in subsection 4.2; Observe that the bias is obtained without the differentiability condition of the density function $f(\mathbf{x})$. To achieve such a bound, kernel-based estimates would require the density function $f(\mathbf{x})$ to be differentiable.

Let $\Phi(t)$ denote the standard normal distribution function. Also, set $\mathrm{SD}_n(\hat{\alpha}) = \{\mathrm{var}_n(\hat{\alpha})\}^{1/2}$. The next result deals with the asymptotic distribution of $\hat{\alpha}$.

THEOREM 3. *Suppose Conditions* 1, 2, 4–6 *hold. Then*

$$P_n\left(\frac{\hat{\alpha} - \alpha}{\mathrm{SD}_n(\hat{\alpha})} \leqslant t\right) = \Phi(t) + o_p(1).$$

Consequently,

$$P\left(\frac{\hat{\alpha} - \alpha}{\mathrm{SD}_n(\hat{\alpha})} \leqslant t\right) = \Phi(t) + o(1).$$

Proof of Theorem 3 is given in subsection 4.3. The above result and Theorem 1 show that the estimate $\hat{\alpha}$ achieves the desirable root-n rate of convergence, even when it is jointly estimated with a smooth function $m(\cdot)$.

Set $\mathrm{SD}_n[\hat{m}(\mathbf{x})] = \{\mathrm{var}_n[\hat{m}(\mathbf{x})]\}^{1/2}$. The following result describes the pointwise rate of convergence and the asymptotic distribution of $\hat{m}(\mathbf{x})$.

THEOREM 4. *Suppose Conditions* 1, 2, 4–6 *hold. Then*

$$|\hat{m}(\mathbf{x}) - m(\mathbf{x})| = O_p(n^{-1/2}\,|\mathbf{H}|^{-1/2} + \|\mathbf{H}\|^2). \tag{3.4}$$

Moreover,

$$P_n\left(\frac{\hat{m}(\mathbf{x}) - E_n\hat{m}(\mathbf{x})}{\mathrm{SD}_n[\hat{m}(\mathbf{x})]} \leqslant t\right) = \Phi(t) + o_p(1). \tag{3.5}$$

Consequently,

$$P\left(\frac{\hat{m}(\mathbf{x}) - E_n \hat{m}(\mathbf{x})}{\mathrm{SD}_n[\hat{m}(\mathbf{x})]} \leqslant t\right) = \Phi(t) + o(1).$$

Proof of Theorem 4 is given in subsection 4.4. Note that if the bandwidth matrix is chosen so that $\mathbf{H} = \mathrm{diag}[h, h]$ and $h \sim n^{-1/6}$, then (3.4) implies that $|\hat{m}(\mathbf{x}) - m(\mathbf{x})| = O_p(n^{-1/3})$, which means that $\hat{m}(\mathbf{x})$ achieves the optimal rate of convergence. See Stone [21].

Also, if $\mathrm{var}(Y \mid B, \mathbf{X}) = \mathrm{var}(Y - \alpha B \mid \mathbf{X}) = \sigma^2(\mathbf{X})$, and if the conditional density function of $Y - \alpha B$ given $\mathbf{X} = \mathbf{x}$ is continuous in $\mathbf{x}$. Then it can be shown that

$$\mathrm{var}_n[\hat{m}(\mathbf{x})] = n^{-1} |\mathbf{H}|^{-1} f^{-1}(\mathbf{x})\, \sigma^2(\mathbf{x}) \int K^2(\mathbf{v})\, d\mathbf{v}(1 + o_p(1)).$$

(See Fan [8], and Fan *et al.* [11].) This is the asymptotic variance of local linear estimates. Thus Theorem 4 generalizes the usual asymptotic result of local linear estimates to partly linear models.

Further Remarks. 1. Theorems 1–4 generalize the results of Speckman [19] in several directions. First, $\mathbf{X}$ is a random vector of covariates. Second, by using the local linear estimates, the improved bias term of $\hat{m}(\mathbf{x})$ (see Theorem 2) is achieved without introducing extra smoothness conditions on the density function of $\mathbf{X}$. Third, the asymptotic distribution of $\hat{m}(\mathbf{x})$ is established. Finally, our approach is applicable to regression problems involving heteroscedasticity since the conditional variance of Y given B, $\mathbf{X}$ is not required to be constant.

2. Theorems 3 and 4 show that our estimates achieve the *dimension reduction principle* as discussed by Stone [23]. Namely, model (1.1) contains two additive terms. The first is a linear term whose estimate has a rate of convergence $n^{-1/2}$, as described in Theorem 3. The second term is the functional component m which is estimated by $\hat{m}$ with the optimal rate of convergence $n^{-p/(2p+d)} = n^{-1/3}$, as discussed after Theorem 4. Both rates are faster than $n^{-p/(2p+3)} = n^{-2/7}$, which is the usual rate of convergence for estimating $E(Y \mid B, \mathbf{X})$ without using (1.1).

3. Theorems 1–4 also hold for $\mathbf{X} \in \mathbb{R}^d$ and $B \in \mathbb{R}^m$, provided $u(\cdot)$ and $m(\cdot)$ are estimated by local polynomials of degrees p with $p > d/2$. Here p is the order of smoothness of m and d is the dimension of $\mathbf{X}$ as discussed previously. This constraint comes from $n^{-1} |\mathbf{H}|^{-2} \to 0$ given in Condition 4. To see this, suppose $\mathbf{H}$ is a diagonal smoothing bandwidth $\mathrm{diag}(h, ..., h)$. Then $|\mathbf{H}| = h^d \sim n^{-d/(2p+d)}$. Hence, $n^{-1} |\mathbf{H}|^{-2} \to 0$ can only be satisfied if $p > d/2$. A similar situation involving time series is discussed in Truong and Stone [24].

4. PROOFS

We start this section with a sequence of lemmas. The first lemma deals with the existence of $[\mathbf{X}'(\mathbf{x})\,\mathbf{W}(\mathbf{x})\,\mathbf{X}(\mathbf{x})]^{-1}$.

LEMMA 1. *Suppose Conditions* 2, 4 *and* 5 *hold. Then there is a positive constant* c_1 *such that*

$$\lim_n P(n\{[\mathbf{X}'(\mathbf{x})\,\mathbf{W}(\mathbf{x})\,\mathbf{X}(\mathbf{x})]^{-1}\}_{ij} \leqslant c_1 \text{ for } \mathbf{x} \in C_f \text{ and } i, j = 1, 2, 3) = 1.$$

Proof. Observe that

$$\mathbf{X}'(\mathbf{x})\,\mathbf{W}(\mathbf{x})\,\mathbf{X}(\mathbf{x})$$
$$= \begin{pmatrix} \sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) & \sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})\,\mathbf{H}^{-1}(\mathbf{X}_i - \mathbf{x})' \\ \sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})\,\mathbf{H}^{-1}(\mathbf{X}_i - \mathbf{x}) & \sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})\,\mathbf{H}^{-1}(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})'\,\mathbf{H}^{-1} \end{pmatrix}.$$

It follows from Conditions 2, 4 and 5 that

$$E\left(n^{-1} \sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})\right) = f(\mathbf{x}) \int K(\mathbf{v})\, d\mathbf{v}(1 + o(1)).$$

Similarly,

$$E\left(n^{-1} \sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})\,\mathbf{H}^{-1}(\mathbf{X}_i - \mathbf{x})\right) = f(\mathbf{x}) \int \mathbf{v}K(\mathbf{v})\, d\mathbf{v}(1 + o(1)) = o(1),$$

and

$$E\left(n^{-1} \sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X_i} - \mathbf{x})\,\mathbf{H}^{-1}(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})'\,\mathbf{H}^{-1}\right)$$
$$= f(\mathbf{x})\,\mu_2(K)\,\mathbf{I}_2(1 + o(1)).$$

According to the law of large numbers,

$$n^{-1}[\mathbf{X}'(\mathbf{x})\,\mathbf{W}(\mathbf{x})\,\mathbf{X}(\mathbf{x})] \xrightarrow{p} \begin{pmatrix} f(\mathbf{x}) & \mathbf{0}' \\ \mathbf{0} & f(\mathbf{x})\,\mu_2(K)\,\mathbf{I}_2 \end{pmatrix}.$$

The desired result follows from Hoeffding's inequality [14]. (See the argument in Lemma 1 of Truong and Stone [24].) ∎

Let $w_{j,n}(\mathbf{x})$, $j=1, 2$, denote the diagonal elements of the matrix

$$n^{-1} \sum_{i=1}^{n} \mathbf{H}^{-1}(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})' \mathbf{H}^{-1} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}).$$

It follows from the above argument that there are positive constants c_2 and c_3 such that

$$\lim_n P\left(c_2^{-1} \leqslant n^{-1} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) \leqslant c_2, \mathbf{x} \in C_f\right) = 1, \tag{4.1}$$

and

$$\lim_n P(c_3^{-1} \leqslant w_{j,n}(\mathbf{x}) \leqslant c_3, \mathbf{x} \in C_f, j=1, 2) = 1. \tag{4.2}$$

Denote the ith entry of $\mathbf{S}(\mathbf{x})$ by

$$S(\mathbf{X}_i, \mathbf{x}) = S_i(\mathbf{x}) = \mathbf{a}'[\mathbf{X}'(\mathbf{x})\, \mathbf{W}(\mathbf{x})\, \mathbf{X}(\mathbf{x})]^{-1}\, \mathbf{X}_i(\mathbf{x})\, K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}), \quad i = 1, ..., n.$$

The next result gives an upper bound for $\|\mathbf{S}(\mathbf{x})\|$.

LEMMA 2. *Suppose Conditions* 2, 4 *and* 5 *hold. Then there are positive constants* c_4 *and* c_5 *such that*

$$\lim_n P(\Psi_n) = 1,$$

where

$$\Psi_n = \left\{ \max_{\mathbf{x} \in C_f} \sum_{i=1}^{n} S^2(\mathbf{X}_i, \mathbf{x}) \leqslant c_4 n^{-1} |\mathbf{H}|^{-1}, \right.$$
$$\left. \max_{\mathbf{x} \in C_f} \max_{1 \leqslant i \leqslant n} |S(\mathbf{X}_i, \mathbf{x})| \leqslant c_5 n^{-1} |\mathbf{H}|^{-1} \right\}.$$

Proof. Observe that

$$\sum_{i=1}^{n} S^2(\mathbf{X}_i, \mathbf{x}) = \mathbf{a}'[\mathbf{X}'(\mathbf{x})\, \mathbf{W}(\mathbf{x})\, \mathbf{X}(\mathbf{x})]^{-1}\, \mathbf{X}'(\mathbf{x})\, \mathbf{W}^2(\mathbf{x})\, \mathbf{X}(\mathbf{x})$$
$$\times [\mathbf{X}'(\mathbf{x})\, \mathbf{W}(\mathbf{x})\, \mathbf{X}(\mathbf{x})]^{-1}\, \mathbf{a}.$$

By Condition 5 and (4.1), each entry of $\mathbf{X}'(\mathbf{x})\, \mathbf{W}^2(\mathbf{x})\, \mathbf{X}(\mathbf{x})$ is bounded by a constant times $n\, |\mathbf{H}|^{-1}$. Thus the desired result follows from Lemma 1. ▮

Set $\mathbf{U} = [u(\mathbf{X}_1), ..., u(\mathbf{X}_n)]'$. The next lemma will be used in the proof of Lemma 4.

Lemma 3. *Suppose Conditions* 1, 2, 4 *and* 5 *hold. There is a positive constant* c_6 *such that*

$$E(\|\mathbf{S}(\mathbf{B}-\mathbf{U})\|)^2 \mid \mathbf{X}_1, ..., \mathbf{X}_n) \leqslant c_6 \, |\mathbf{H}|^{-1} \qquad \textit{on } \Psi_n.$$

That is, $P(\{E(\|\mathbf{S}(\mathbf{B}-\mathbf{U})\|^2 \mid \mathbf{X}_1, ..., \mathbf{X}_n) > c_6 |\mathbf{H}|^{-1}\} \cap \Psi_n) = 0.$

Proof. By Lemma 2, there is a positive constant c_6 such that

$$\begin{aligned} E(\|\mathbf{S}(\mathbf{B}-\mathbf{U})\| \mid \mathbf{X}_1, ..., \mathbf{X}_n) &= \sum_{j=1}^{n} \sum_{i=1}^{n} [S(\mathbf{X}_i, \mathbf{X}_j)]^2 \, \sigma_B^2(\mathbf{X}_i) \\ &\leqslant \sup_{\mathbf{x} \in C_f} \sigma_B^2(\mathbf{x}) \sum_{j=1}^{n} \sum_{i=1}^{n} [S(\mathbf{X}_i, \mathbf{X}_j)]^2 \\ &\leqslant c_6 \, |\mathbf{H}|^{-1} \qquad \text{on } \Psi_n. \quad \blacksquare \end{aligned}$$

It follows from $[\mathbf{X}'(\mathbf{x}) \, \mathbf{W}(\mathbf{x}) \, \mathbf{X}(\mathbf{x})]^{-1} \, [\mathbf{X}'(\mathbf{x}) \, \mathbf{W}(\mathbf{x}) \, \mathbf{X}(\mathbf{x})] = I_3$ that

$$\sum_{j=1}^{n} S(\mathbf{X}_j, \mathbf{x}) = \sum_{j=1}^{n} \mathbf{a}'[\mathbf{X}'(\mathbf{x}) \, \mathbf{W}(\mathbf{x}) \, \mathbf{X}(\mathbf{x})]^{-1} \, \mathbf{X}_j(\mathbf{x}) \, K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) = 1,$$

and

$$\sum_{j=1}^{n} \mathbf{a}'[\mathbf{X}'(\mathbf{x}) \, \mathbf{W}(\mathbf{x}) \, \mathbf{X}(\mathbf{x})]^{-1} \, \mathbf{X}_j(\mathbf{x}) \, K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) \, \mathbf{H}^{-1}(\mathbf{X}_j - \mathbf{x}) = (0, 0)'.$$

In particular,

$$\sum_{j=1}^{n} S(\mathbf{X}_j, \mathbf{X}_i) = 1 \quad \text{and} \quad \sum_{j=1}^{n} S(\mathbf{X}_j, \mathbf{X}_i) \, \mathbf{H}^{-1}(\mathbf{X}_j - \mathbf{X}_i) = (0, 0)', \qquad i = 1, ..., n.$$

To prepare the next result, set

$$Du(\mathbf{x}) = \left(\frac{\partial u(\mathbf{x})}{\partial x_1}, \frac{\partial u(\mathbf{x})}{\partial x_2} \right)' \qquad \text{and} \qquad u^*(\mathbf{v}; \mathbf{x}) = u(\mathbf{x}) + (\mathbf{v} - \mathbf{x})' \, Du(\mathbf{x}).$$

That is, $u^*(\mathbf{v}; \mathbf{x})$ is the first order Taylor polynomial of $u(\cdot)$ at $\mathbf{x}$. Then

$$\sum_{j=1}^{n} S(\mathbf{X}_j, \mathbf{x}) \, u^*(\mathbf{X}_j; \mathbf{x}) = u(\mathbf{x}). \tag{4.3}$$

Lemma 4. *Suppose Conditions* 1–5 *hold. Then*

$$n^{-1} \tilde{\mathbf{B}}' \tilde{\mathbf{B}} \xrightarrow{p} \int \sigma_B^2(\mathbf{x}) \, f(\mathbf{x}) \, d\mathbf{x}.$$

Proof. Set $\hat{\mathbf{U}} = \mathbf{SB} = [\hat{u}(\mathbf{X}_1), ..., \hat{u}(\mathbf{X}_n)]'$. Then $\tilde{\mathbf{B}}'\tilde{\mathbf{B}} = \mathbf{B}'(\mathbf{I}-\mathbf{S})'(\mathbf{I}-\mathbf{S})\,\mathbf{B} = (\mathbf{B}-\hat{\mathbf{U}})'\,(\mathbf{B}-\hat{\mathbf{U}})$. Thus

$$\tilde{\mathbf{B}}'\tilde{\mathbf{B}} = \sum_{i=1}^{n} [B_i - u(\mathbf{X}_i)]^2 + \sum_{i=1}^{n} [\hat{u}(\mathbf{X}_i) - u(\mathbf{X}_i)]^2 - 2\sum_{i=1}^{n} [B_i - u(\mathbf{X}_i)][\hat{u}(\mathbf{X}_i) - u(\mathbf{X}_i)]. \tag{4.4}$$

By Condition 3 and the law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n} [B_i - u(\mathbf{X}_i)]^2 \xrightarrow{p} \int \sigma_B^2(\mathbf{x})\, f(\mathbf{x})\, d\mathbf{x}. \tag{4.5}$$

According to (4.3),

$$\begin{aligned}\hat{u}(\mathbf{X}_i) - u(\mathbf{X}_i) &= \sum_{j=1}^{n} S(\mathbf{X}_j, \mathbf{X}_i)[B_j - u(\mathbf{X}_j)] \\ &\quad + \sum_{j=1}^{n} S(\mathbf{X}_j, \mathbf{X}_i)[u(\mathbf{X}_j) - u^*(\mathbf{X}_j\,;\mathbf{X}_i)] \\ &= \mathbf{S}'(\mathbf{X}_i)(\mathbf{B}-\mathbf{U}) + \mathbf{S}'(\mathbf{X}_i)(\mathbf{U}-\mathbf{U}^*(\mathbf{X}_i)),\end{aligned}$$

where $\mathbf{U}^*(\mathbf{x}) = [u^*(\mathbf{X}_1\,;\mathbf{x}), ..., u^*(\mathbf{X}_n\,;\mathbf{x})]'$. Thus

$$\sum_{i=1}^{n} [\hat{u}(\mathbf{X}_i) - u(\mathbf{X}_i)]^2 \leqslant 2\sum_{i=1}^{n} [\mathbf{S}'(\mathbf{X}_i)(\mathbf{B}-\mathbf{U})]^2 + 2\sum_{i=1}^{n} [\mathbf{S}'(\mathbf{X}_i)(\mathbf{U}-\mathbf{U}^*(\mathbf{X}_i))]^2. \tag{4.6}$$

By Lemmas 2 and 3,

$$\frac{1}{n}\sum_{i=1}^{n} [\mathbf{S}'(\mathbf{X}_i)(\mathbf{B}-\mathbf{U})]^2 = \frac{1}{n}\|\mathbf{S}(\mathbf{B}-\mathbf{U})\|^2 = O_p(n^{-1}\,|\mathbf{H}|^{-1}). \tag{4.7}$$

According to Taylor expansion, Conditions 1 and 5, there is a positive constant c_7 such that

$$|u(\mathbf{X}_j) - u^*(\mathbf{X}_j\,;\mathbf{X}_i)| \leqslant c_7\,\|\mathbf{H}\|^2 \qquad \text{for} \quad \mathbf{H}^{-1}(\mathbf{X}_j - \mathbf{X}_i) \in C_K,$$

where C_K is the compact support of $K(\cdot)$. By Lemma 1 and (4.1),

$$\left|\sum_{j=1}^{n} S(\mathbf{X}_j, \mathbf{X}_i)[u(\mathbf{X}_j) - u^*(\mathbf{X}_j; \mathbf{X}_i)]\right| = O_p(\|\mathbf{H}\|^2).$$

Thus,

$$\frac{1}{n}\sum_{i=1}^{n}[\mathbf{S}'(\mathbf{X}_i)(\mathbf{U}-\mathbf{U}^*(\mathbf{X}_i))]^2 = O_p(\|\mathbf{H}\|^4). \tag{4.8}$$

It follows from (4.6)–(4.8) and Condition 4 that

$$\frac{1}{n}\sum_{i=1}^{n}[\hat{u}(\mathbf{X}_i)-u(\mathbf{X}_i)]^2 = O_p(n^{-1}\,|\mathbf{H}|^{-1}) + O_p(\|\mathbf{H}\|^4) = o_p(1). \tag{4.9}$$

By the Cauchy–Schwarz inequality, (4.5) and (4.9),

$$\begin{aligned}&\left|\frac{1}{n}\sum_{i=1}^{n}[B_i-u(\mathbf{X}_i)][\hat{u}(\mathbf{X}_i)-u(\mathbf{X}_i)]\right|\\&\quad\leqslant\left|\frac{1}{n}\sum_{i=1}^{n}[B_i-u(\mathbf{X}_i)]^2\right|^{1/2}\left|\frac{1}{n}\sum_{i=1}^{n}[\hat{u}(\mathbf{X}_i)-u(\mathbf{X}_i)]^2\right|^{1/2}\xrightarrow{p}0.\end{aligned} \tag{4.10}$$

The desired result follows from (4.4), (4.5), (4.9), and (4.10). ∎

By applying a Taylor expansion to $m(\cdot)$ in a manner similar to (4.2), the next result follows easily from Lemma 1 and (4.1) [see the argument for (4.8)].

LEMMA 5. *Suppose Conditions* 1, 2 *and* 4 *hold. Then*

$$\|(\mathbf{I}-\mathbf{S})\,\mathbf{M}\| = O_p(\|\mathbf{H}\|^2).$$

4.1. *Proof of Theorem* 1

Recall that $\hat{\alpha} = (\tilde{\mathbf{B}}'\tilde{\mathbf{B}})^{-1}\,\tilde{\mathbf{B}}'\tilde{\mathbf{Y}}$. Thus

$$E_n(\hat{\alpha}) = (\tilde{\mathbf{B}}'\tilde{\mathbf{B}})^{-1}\,\tilde{\mathbf{B}}'(\mathbf{I}-\mathbf{S})(\alpha\mathbf{B}+\mathbf{M}) = \alpha + (\tilde{\mathbf{B}}'\tilde{\mathbf{B}})^{-1}\,\tilde{\mathbf{B}}'(\mathbf{I}-\mathbf{S})\,M,$$

and

$$\mathrm{var}_n(\hat{\alpha}) = (\tilde{\mathbf{B}}'\tilde{\mathbf{B}})^{-1}\,\tilde{\mathbf{B}}'(\mathbf{I}-\mathbf{S})\,\mathbf{V}(\mathbf{I}-\mathbf{S})'\,\tilde{\mathbf{B}}(\tilde{\mathbf{B}}'\tilde{\mathbf{B}})^{-1}.$$

By Lemmas 4 and 5,

$$\begin{aligned}|\mathrm{bias}_n(\hat{\alpha})| &= |(\tilde{\mathbf{B}}'\tilde{\mathbf{B}})^{-1}\,\tilde{\mathbf{B}}'(\mathbf{I}-\mathbf{S})\,\mathbf{M}| \leqslant (\tilde{\mathbf{B}}'\tilde{\mathbf{B}})^{-1}\,\|\tilde{\mathbf{B}}\|\cdot\|(\mathbf{I}-\mathbf{S})\,\mathbf{M}\|\\&= O_p(n^{-1/2}\,\|\mathbf{H}\|^2).\end{aligned}$$

According to Condition 3 and Lemma 4, there is a positive constant c_8 such that

$$\begin{aligned}\mathrm{var}_n(\hat{\alpha}) &= (\tilde{\mathbf{B}}'\tilde{\mathbf{B}})^{-1}\,\tilde{\mathbf{B}}'(\mathbf{I}-\mathbf{S})\,\mathbf{V}(\mathbf{I}-\mathbf{S})'\,\tilde{\mathbf{B}}(\tilde{\mathbf{B}}'\tilde{\mathbf{B}})^{-1}\\&\leqslant c_8 n^{-2}\,\|(\mathbf{I}-\mathbf{S})'\tilde{\mathbf{B}}\|^2.\end{aligned} \tag{4.11}$$

Note that $\|(\mathbf{I}-\mathbf{S})'\,\tilde{\mathbf{B}}\|^2 \leqslant 2\,\|\tilde{\mathbf{B}}\|^2 + 2\,\|\mathbf{S}'\tilde{\mathbf{B}}\|^2$. By Lemma 4, $\|\tilde{\mathbf{B}}\|^2 = O_p(n)$. It follows from

$$\|\mathbf{S}'\tilde{\mathbf{B}}\| = \|\mathbf{S}'(\mathbf{I}-\mathbf{S})\,\mathbf{B}\| \leqslant \|\mathbf{S}(\mathbf{B}-\mathbf{U})\| + \|\mathbf{S}\| \cdot \|\mathbf{U}-\mathbf{S}\mathbf{B}\|,$$

Lemmas 2, 3 and (4.9) that

$$\|\mathbf{S}'\tilde{\mathbf{B}}\|^2 = O_p(|\mathbf{H}|^{-1} + |\mathbf{H}|^{-1}\,\{|\mathbf{H}|^{-1} + n\,\|\mathbf{H}\|^4\}).$$

Hence,

$$n^{-2}\,\|(\mathbf{I}-\mathbf{S})'\,\tilde{\mathbf{B}}\|^2 = O_p(n^{-1}) + O_p(n^{-1}) \times (n^{-1}\,|\mathbf{H}|^{-1} + n^{-1}\,|\mathbf{H}|^{-2} + |\mathbf{H}|^{-1}\,\|\mathbf{H}\|^4). \quad (4.12)$$

The conclusion of Theorem 1 follows from (4.11), (4.12) and Condition 4. ▮

4.2. *Proof of Theorem* 2

Write

$$\hat{m}(\mathbf{X}_i) = \mathbf{S}'(\mathbf{X}_i)(\mathbf{Y}-\hat{\alpha}\mathbf{B}) = \mathbf{S}'(\mathbf{X}_i)[(\mathbf{Y}-\alpha\mathbf{B}) + (\hat{\alpha}-\alpha)\,\mathbf{B}].$$

Then

$$\begin{aligned} E_n[\hat{m}(\mathbf{X}_i) - m(\mathbf{X}_i)] &= E_n[\mathbf{S}'(\mathbf{X}_i)(\mathbf{Y}-\alpha\mathbf{B}) - m(\mathbf{X}_i)] + \mathbf{S}'(\mathbf{X}_i)\,\mathbf{B}E_n(\hat{\alpha}-\alpha) \\ &= \mathbf{S}'(\mathbf{X}_i)\,\mathbf{M} - m(\mathbf{X}_i) + \mathbf{S}'(\mathbf{X}_i)\,\mathbf{B}E_n(\hat{\alpha}-\alpha). \end{aligned}$$

According to (4.9) and Conditions 1 and 2,

$$|\mathbf{S}'(\mathbf{X}_i)\,\mathbf{B}| \leqslant |\hat{u}(\mathbf{X}_i) - u(\mathbf{X}_i)| + |u(\mathbf{X}_i)| = O_p(n^{-1/2}\,|\mathbf{H}|^{-1/2} + \|\mathbf{H}\|^2) + O(1).$$

It follows from Lemma 5 and Theorem 1 that

$$|E_n[\hat{m}(\mathbf{X}_i) - m(\mathbf{X}_i)]| = O_p(\|\mathbf{H}\|^2) + O_p(n^{-1/2}\,\|\mathbf{H}\|^2).$$

Next we consider the variance term. Observe that

$$\begin{aligned} &\mathrm{var}_n\{\mathbf{S}'(\mathbf{X}_i)[(\mathbf{Y}-\alpha\mathbf{B}) + (\hat{\alpha}-\alpha)\,\mathbf{B}]\} \\ &\quad = \mathrm{var}_n[\mathbf{S}'(\mathbf{X}_i)(\mathbf{Y}-\alpha\mathbf{B})] + \mathrm{var}_n[\mathbf{S}'(\mathbf{X}_i)(\hat{\alpha}-\alpha)\,\mathbf{B}] \\ &\quad\quad + \mathrm{cov}_n[\mathbf{S}'(\mathbf{X}_i)(\mathbf{Y}-\alpha\mathbf{B}), \mathbf{S}'(\mathbf{X}_i)(\hat{\alpha}-\alpha)\,\mathbf{B}]. \end{aligned} \quad (4.13)$$

By Condition 3 and Lemma 2,

$$\operatorname{var}_n[\mathbf{S}'(\mathbf{X}_i)(\mathbf{Y}-\alpha\mathbf{B})]=\mathbf{S}'(\mathbf{X}_i)\operatorname{var}_n(\mathbf{Y}-\alpha\mathbf{B})\,\mathbf{S}(\mathbf{X}_i)=O_p(n^{-1}\,|\mathbf{H}|^{-1}). \quad (4.14)$$

Similarly, by Conditions 2, 3, Lemma 2 and Theorem 1,

$$\operatorname{var}_n[\mathbf{S}'(\mathbf{X}_i)(\hat{\alpha}-\alpha)\,\mathbf{B}]=o_p(n^{-1}\,|\mathbf{H}|^{-1}). \quad (4.15)$$

By the Cauchy–Schwarz inequality,

$$|\operatorname{cov}_n[\mathbf{S}'(\mathbf{X}_i)(\mathbf{Y}-\alpha\mathbf{B}),\mathbf{S}'(\mathbf{X}_i)(\hat{\alpha}-\alpha)\,\mathbf{B}]|=o_p(n^{-1}\,|\mathbf{H}|^{-1}). \quad (4.16)$$

The conclusion of Theorem 2 follows from (4.13)–(4.16). ∎

4.3. *Proof of Theorem* 3

We have

$$\mathrm{SD}_n^2(\hat{\alpha})=\operatorname{var}_n(\hat{\alpha})=(\tilde{\mathbf{B}}'\tilde{\mathbf{B}})^{-1}\,\tilde{\mathbf{B}}'(\mathbf{I}-\mathbf{S})\,\mathbf{V}(\mathbf{I}-\mathbf{S})'\,\tilde{\mathbf{B}}(\tilde{\mathbf{B}}'\tilde{\mathbf{B}})^{-1}.$$

According to (4.12),

$$\|(\mathbf{I}-\mathbf{S})'\,\tilde{\mathbf{B}}\|^2\geqslant\tfrac{1}{2}\|\tilde{\mathbf{B}}\|^2-\|\mathbf{S}'\tilde{\mathbf{B}}\|^2=\tfrac{1}{2}\|\tilde{\mathbf{B}}\|^2\,[1+o_p(1)].$$

It follows from (3.1) that there is a positive constant c_9 such that

$$\lim_n P(\operatorname{var}_n(\hat{\alpha})\geqslant c_9(\tilde{\mathbf{B}}'\tilde{\mathbf{B}})^{-1})=1. \quad (4.17)$$

Observe that $\hat{\alpha}-\alpha=(\tilde{\mathbf{B}}'\tilde{\mathbf{B}})^{-1}\,\tilde{\mathbf{B}}(\mathbf{I}-\mathbf{S})(\mathbf{Y}-\alpha\mathbf{B}-\mathbf{M})+(\tilde{\mathbf{B}}'\tilde{\mathbf{B}})^{-1}\,\tilde{\mathbf{B}}'(\mathbf{I}-\mathbf{S})$ $\mathbf{M}$. By (4.17), Condition 4, Lemmas 4 and 5,

$$\frac{|(\tilde{\mathbf{B}}'\tilde{\mathbf{B}})^{-1}\,\tilde{\mathbf{B}}'(\mathbf{I}-\mathbf{S})\,\mathbf{M}|}{\mathrm{SD}_n(\hat{\alpha})}=O_p(\|\mathbf{H}\|^2)=o_p(1).$$

Thus the conditional asymptotic normality of $\hat{\alpha}-\alpha$ follows from

$$\begin{aligned}P_n\left(\frac{(\tilde{\mathbf{B}}'\tilde{\mathbf{B}})^{-1}\,\tilde{\mathbf{B}}'(\mathbf{I}-\mathbf{S})(\mathbf{Y}-\alpha\mathbf{B}-\mathbf{M})}{\mathrm{SD}_n(\hat{\alpha})}\leqslant t\right)&=P_n\left(\frac{\mathbf{b}'\mathbf{Z}}{\sqrt{\operatorname{var}_n(\mathbf{b}'\mathbf{Z})}}\leqslant t\right)\\&=\Phi(t)+o_p(1),\end{aligned} \quad (4.18)$$

where $\mathbf{b}'=(b_1,...,b_n)=\tilde{\mathbf{B}}'(\mathbf{I}-\mathbf{S})$ and $\mathbf{Z}=(Z_1,...,Z_n)'=\mathbf{Y}-\alpha\mathbf{B}-\mathbf{M}$.

Set $s_n^2=\operatorname{var}_n(\mathbf{b}'\mathbf{Z})=\sum b_i^2\sigma^2(B_i,\mathbf{X}_i)$. Since $E_n(\mathbf{b}'\mathbf{Z})=0$, then (4.18) follows from the conditional Lindeberg's condition:

$$\frac{1}{s_n^2}\sum_{i=1}^n E_n(b_i^2 Z_i^2 1_{\{|b_i Z_i|\geqslant c s_n\}})=o_p(1), \qquad c>0. \quad (4.19)$$

To verify (4.19), we note that from (3.1) that there is a positive constant c_{10} such that $s_n^2 \geqslant c_{10} \sum_i b_i^2$. Thus

$$\frac{1}{s_n^2} \sum_{i=1}^{n} E_n(b_i^2 Z_i^2 1_{\{|b_i Z_i| \geqslant c s_n\}}) \leqslant c_{10}^{-1} E_n(Z_1^2 1_{\{Z_1^2 \geqslant c^2 c_{10} \sum_i b_i^2/\max_i b_i^2\}}). \quad (4.20)$$

Set $\Omega_n = \{\max_i b_i^2/\sum_1^n b_i^2 > \eta\}$ for $\eta > 0$. Then

$$\lim_n P(\Omega_n) = 0, \qquad \eta > 0. \quad (4.21)$$

[Proof of (4.21) will be given shortly.] By Markov's inequality,

$$\begin{aligned} P(E_n[Z_1^2 1_{\{Z_1^2 \geqslant c^2 c_{10} \sum_i b_i^2/\max_i b_i^2\}}] > \varepsilon) &\leqslant P(E_n[Z_1^2 1_{\{Z_1^2 \geqslant c^2 c_{10}/\eta\}}] > \varepsilon) + P(\Omega_n) \\ &\leqslant \frac{1}{\varepsilon} E[Z_1^2 1_{\{Z_1^2 \geqslant c^2 c_{10}/\eta\}}] + P(\Omega_n), \qquad \eta, \varepsilon > 0. \end{aligned}$$

According to (3.3), $E[Z_1^2 1_{\{Z_1^2 > c^2 c_{10}/\eta\}}] \to 0$ as $\eta \to 0$. Hence, by (4.21)

$$E_n(Z_1^2 1_{\{Z_1^2 > c^2 c_{10} \sum_i b_i^2/\max_i b_i^2\}}) = o_p(1), \qquad c > 0. \quad (4.22)$$

It follows from (4.20) and (4.22) that (4.19) holds.

We now prove (4.21). By Lemma 4,

$$n^{-1}\mathbf{b}'\mathbf{b} = n^{-1} \|(\mathbf{I} - \mathbf{S})' \tilde{\mathbf{B}}\|^2 \geqslant 2n^{-1} \|\tilde{\mathbf{B}}\|^2 \xrightarrow{p} 2^{-1} \int \sigma_B^2(\mathbf{x}) f(\mathbf{x})\, d\mathbf{x}.$$

Thus (4.21) follows from $\max_{1 \leqslant i \leqslant n} b_i^2 = o_p(n)$. For a vector $\mathbf{v} = (v_i)$ and a matrix $\mathbf{A} = (a_{ij})$, define

$$\|\mathbf{v}\|_\infty = \max_i |v_i| \qquad \text{and} \qquad \|\mathbf{A}\|_\infty = \max_i \sum_j |a_{ij}|.$$

Then

$$\max_i |b_i| = \|\mathbf{b}\|_\infty \leqslant (1 + \|\mathbf{S}\|_\infty)(1 + \|\mathbf{S}\|_\infty) \|\mathbf{B}\|_\infty. \quad (4.23)$$

It follows from

$$\sum_{i=1}^{n} \mathbf{S}_i(\mathbf{x}) = \mathbf{a}'[\mathbf{X}'(\mathbf{x})\, \mathbf{W}(\mathbf{x})\, \mathbf{X}(\mathbf{x})]^{-1} \begin{pmatrix} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) \\ \sum_{i=1}^{n} \mathbf{H}(\mathbf{X}_i - \mathbf{x})\, K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) \end{pmatrix},$$

Lemma 1, (4.1) and (4.2) that

$$\|\mathbf{S}\|_{\infty} = O_p(1). \tag{4.24}$$

According to (3.2) and Markov's inequality,

$$\|\mathbf{B}-\mathbf{U}\|_{\infty} = O_p(n^{1/2}/\log n).$$

Since $\|\mathbf{U}\|_{\infty} = O(1)$ by Conditions 1 and 2, we have

$$\|\mathbf{B}\|_{\infty} \leqslant \|\mathbf{B}-\mathbf{U}\|_{\infty} + \|\mathbf{U}\|_{\infty} = O_p(n^{1/2}/\log n). \tag{4.25}$$

It follows from (4.23)–(4.25) that $\max_i b_i^2 = o_p(n)$. Hence (4.21) is valid.

The unconditional asymptotic normality follows from the dominated convergence theorem. This completes the proof of Theorem 3. ∎

4.4. *Proof of Theorem* 4

Write

$$\begin{aligned}\hat{m}(\mathbf{x}) - m(\mathbf{x}) = \mathbf{S}'(\mathbf{x})(\mathbf{Y}-\alpha\mathbf{B}-\mathbf{M}) + \mathbf{S}'(\mathbf{x})\,\mathbf{M} - m(\mathbf{x})\\ -\mathbf{S}'(\mathbf{x})\,\mathbf{B}(\hat{\alpha}-\alpha).\end{aligned} \tag{4.26}$$

It follows from Condition 3 and Lemma 2 that there is a positive constant c_{11} such that

$$\operatorname{var}_n[\mathbf{S}'(\mathbf{x})(\mathbf{Y}-\alpha\mathbf{B}-\mathbf{M})] \leqslant c_{11} n^{-1}\,|\mathbf{H}|^{-1} \qquad \text{on} \quad \Psi_n.$$

Thus, by Chebyshev's inequality and Lemma 2,

$$|\mathbf{S}'(\mathbf{x})(\mathbf{Y}-\alpha\mathbf{B}-\mathbf{M})| = O_p(n^{-1/2}\,|\mathbf{H}|^{-1/2}). \tag{4.27}$$

By Lemma 5,

$$|\mathbf{S}'(\mathbf{x})\,\mathbf{M} - m(\mathbf{x})| = O_p(\|\mathbf{H}\|^2). \tag{4.28}$$

It now follows from (4.26)–(4.28) and Theorem 3 that (3.4) holds.

To prove (3.5), set

$$\hat{m}_0(\mathbf{x}) = \mathbf{a}'[\mathbf{X}'(\mathbf{x})\,\mathbf{W}(\mathbf{x})\,\mathbf{X}(\mathbf{x})]^{-1}\,\mathbf{X}'(\mathbf{x})\,\mathbf{W}(\mathbf{x})(\mathbf{Y}-\alpha\mathbf{B}) = \mathbf{S}'(\mathbf{x})(\mathbf{Y}-\alpha\mathbf{B}).$$

That is, $\hat{m}_0(\mathbf{x})$ is the local linear estimate of $m(\cdot)$ if α were known. Then from Theorems 1 and 2, we have

$$\begin{aligned}|\hat{m}_0(\mathbf{x}) - \hat{m}(\mathbf{x})| &= |(\hat{\alpha}-\alpha)\,\mathbf{S}'(\mathbf{x})\,\mathbf{B}| = |(\hat{\alpha}-\alpha)\,\hat{u}(\mathbf{x})|\\ &= O_p(n^{-1/2}).\end{aligned} \tag{4.29}$$

$$|E_n[\hat{m}(\mathbf{x})] - E_n[\hat{m}_0(\mathbf{x})]| = O_p(n^{-1/2}), \tag{4.30}$$

and

$$|\mathrm{var}_n[\hat{m}(\mathbf{x})] - \mathrm{var}_n[\hat{m}_0(\mathbf{x})]| = O_p(n^{-1}\,|\mathbf{H}|^{-1}). \tag{4.31}$$

By an argument similar to showing (4.1) and (4.2), there is a positive constant c_{12} such that

$$\lim_n P(\mathbf{S}'(\mathbf{x})\,\mathbf{S}(\mathbf{x}) > c_{12} n^{-1}\,|\mathbf{H}|^{-1}) = 1. \tag{4.32}$$

It follows from (3.1), (4.31) and (4.32) that there is a positive constant c_{13} such that

$$\lim_n P(\mathrm{var}_n[\hat{m}(\mathbf{x})] > c_{13} n^{-1}\,|\mathbf{H}|^{-1}) = 1. \tag{4.33}$$

Thus from (4.29), (4.33) and Condition 4,

$$\frac{|\hat{m}(\mathbf{x}) - \hat{m}_0(\mathbf{x})|}{\mathrm{SD}_n[\hat{m}(\mathbf{x})]} = O_p(|\mathbf{H}|^{1/2}) = o_p(1).$$

Similarly, by (4.30), (4.33) and Condition 4,

$$\frac{|E_n[\hat{m}(\mathbf{x})] - E_n[\hat{m}_0(\mathbf{x})]|}{\mathrm{SD}_n[\hat{m}(\mathbf{x})]} = o_p(1).$$

Hence, in view of (4.31), it is sufficient to show that

$$P_n\left(\frac{\hat{m}_0(\mathbf{x}) - E_n\hat{m}_0(\mathbf{x})}{\mathrm{SD}_n[\hat{m}_0(\mathbf{x})]} \leqslant t\right) = \Phi(t) + o_p(1). \tag{4.34}$$

This will be proved by verifying the conditional Lyapounov's condition. Note that $\hat{m}_0(\mathbf{x}) - E_n\hat{m}_0(\mathbf{x}) = \mathbf{S}'(\mathbf{x})(\mathbf{Y} - \alpha\mathbf{B} - \mathbf{M})$. By Lemma 2 and (3.3),

$$\begin{aligned}\sum_i |S_i(\mathbf{x})|^{2+\delta}\, E[\,|\,Y_i - \alpha B_i - m(\mathbf{X}_i)|^{2+\delta}\,|\,B_i, \mathbf{X}_i] &= O\left(\max_{1\leqslant i\leqslant n} |S(\mathbf{X}_i, \mathbf{x})|^{\delta} \sum_i S_i^2(\mathbf{x})\right)\\ &= O_p(n^{-\delta}\,|\mathbf{H}|^{-\delta})\,\mathbf{S}'(\mathbf{x})\,\mathbf{S}(\mathbf{x}).\end{aligned} \tag{4.35}$$

By (3.1), there is a positive constant c_{14} such that

$$\mathrm{var}_n[\mathbf{S}'(\mathbf{x})(\mathbf{Y} - \alpha\mathbf{B} - \mathbf{M})] \geqslant c_{14}\mathbf{S}'(\mathbf{x})\,\mathbf{S}(\mathbf{x}). \tag{4.36}$$

It follows from (4.32), (4.35), (4.36) and Condition 4 that

$$\frac{\sum_i |S_i(\mathbf{x})|^{2+\delta}\, E[\,|Y_i-\alpha B_i-m(\mathbf{X}_i)|^{2+\delta}\,|\,B_i, \mathbf{X}_i]}{\{\mathrm{var}_n[\mathbf{S}'(\mathbf{x})(\mathbf{Y}-\alpha\mathbf{B}-\mathbf{M})]\}^{(2+\delta)/2}}$$
$$= O_p(\{n^{-1}\,|\mathbf{H}|^{-1}\}^{\delta/2}) = o_p(1).$$

Thus the conditional Lyapounov's condition holds. Hence, (4.34) is valid.

The unconditional result follows from the dominated convergence theorem. This completes the proof of Theorem 4. ∎

ACKNOWLEDGMENTS

Professor Stephen J. Marron's thoughtful advice is greatly appreciated. We also thank the referee for helpful suggestions that improve the readability of the paper. Scott A. Hamilton's research was supported in part by NIH Grant 5T32HD07237-10. Young K. Truong's research was supported in part by National Science Foundation Grant DMS-9403800 and National Institutes of Health Grant CA61937.

REFERENCES

[1] Chen, H. (1988). Convergence rates for parametric components in a partly linear model. *Ann. Statist.* **16** 136–146.

[2] Chen, H., and Shiau, J.-J. H. (1991). A two-stage spline smoothing method for partially linear models. *J. Statist. Planning Inference* **27** 187–202.

[3] Chen, H., and Shiau, J.-J. H. (1994). Data-driven efficient estimation for a partially linear model. *Ann. Statist.* **22** 211–237.

[4] Denby, L. (1986). Smooth regression function. Statistical Research Report 26. AT&T Bell Laboratories, Princeton, New Jersey.

[5] Engle, R., Granger, C., Rice, J., and Weiss, A. (1986). Nonparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc.* **81** 310–320.

[6] Eubank, R. L., and Whitney, P. (1989). Convergence rates for estimation in certain partially linear models. *J. Statist. Planning Inference* **23** 33–43.

[7] Eubank, R. L., Hart, J. D., and Speckman, P. (1990). Trigonometric series regression estimators with an application to partially linear models. *J. Multivariate Anal.* **32** 70–83.

[8] Fan, J. (1992). Design adaptive nonparametric function estimation: A unified approach, *J. Amer. Statist. Assoc.* **87** 998–1004.

[9] Fan, J. (1993). Local linear regression smoothers and their minimax efficiency. *Ann. Statist.* **21** 196–216.

[10] Fan, J., and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *Ann. Statist.* **20** 2008–2036.

[11] Fan, J., Hu, T., and Truong, Y. K. (1994). Robust non-parametric function estimation. *Scand. J. Statist.* **21** 433–446.

[12] Green, P., Jennison, C., and Seheult, A. (1985). Analysis of field experiments by least squares smoothing. *J. Royal Statist. Soc. Ser. B.* **47** 299–315.

[13] Heckman, N. E. (1986). Spline smoothing in a partly linear model. *J. Royal Statist. Soc. Ser. B.* **48** 244–248.

[14] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

[15] Rice, J. (1986). Convergence rates for partially splined models. *Statist. Probabil. Lett.* **4** 203–208.

[16] Ruppert, D., and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.* **22** 1346–1370.

[17] Scheffé, H. (1959). *The Analysis of Variance*. Wiley, New York.

[18] Shiau, J., Wahba, G., and Johnson, D. R. (1986). Partial spline models for the inclusion of tropopause and frontal boundary information in otherwise smooth two and three dimensional objective analysis. *J. Atmospheric Oceanic Technol.* **3** 714–725.

[19] Speckman, P. (1988). Kernel smoothing in partial linear models. *J. Royal Statist. Soc. Ser. B* **50** 413–436.

[20] Stone, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–645.

[21] Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

[22] Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.

[23] Stone, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation (with discussion). *Ann. Statist.* **22** 118–184.

[24] Truong, Y. K., and Stone, C. J. (1994). Semiparametric time series regression. *J. Time Series Anal.* **15** 405–428.

[25] Wahba, G. (1984). Partial spline models for the semiparametric estimation of functions of several variables. In *Statistical Analysis of Time Series*, *Japan–US Joint Sem.*, pp. 312–329. Institute of Statistical Mathematics. Tokyo.

[26] Wand, W. P., and Jones, M. C. (1993). Comparison of smoothing parameterizations in bivariate kernel density estimation. *J. Amer. Statist. Assoc.* **88** 520–528.