



# NONPARAMETRIC ESTIMATION OF AN ADDITIVE MODEL WITH A LINK FUNCTION

---

*Joel Horowitz and Enno Mammen*

THE INSTITUTE FOR FISCAL STUDIES  
DEPARTMENT OF ECONOMICS, UCL  
**cemmap** working paper CWP19/02

# NONPARAMETRIC ESTIMATION OF AN ADDITIVE MODEL WITH A LINK FUNCTION

by

Joel L. Horowitz  
Department of Economics  
Northwestern University  
Evanston, IL 60208-2600  
U.S.A.  
joel-horowitz@northwestern.edu

and

Enno Mammen  
Institute for Applied Mathematics  
Ruprecht-Karls-Universität Heidelberg  
D 69120 Heidelberg  
GERMANY  
mammen@statlab.uni-heidelberg.de

July 2002

## ABSTRACT

This paper describes an estimator of the additive components of a nonparametric additive model with a known link function. When the additive components are twice continuously differentiable, the estimator is asymptotically normally distributed with a rate of convergence in probability of  $n^{-2/5}$ . This is true regardless of the (finite) dimension of the explanatory variable. Thus, in contrast to the existing asymptotically normal estimator, the new estimator has no curse of dimensionality. Moreover, the asymptotic distribution of each additive component is the same as it would be if the other components were known with certainty.

MSC: AMS 2000 subject classifications. Primary 62G08; secondary 62G20

Keywords and phrases: Additive models, multivariate curve estimation, nonparametric regression, kernel estimates, orthogonal series estimator.

Short title: Additive regression model with a link function

---

The research of Joel L. Horowitz was supported in part by NSF Grant SES-9910925, the Alexander von Humboldt Foundation, and Deutsche Forschungsgemeinschaft Sonderforschungsbereich 373, "Quantifikation und Simulation Ökonomischer Prozesse.". The research of Enno Mammen was supported in part by Deutsche Forschungsgemeinschaft MA 1026/6-2.

# NONPARAMETRIC ESTIMATION OF AN ADDITIVE MODEL WITH A LINK FUNCTION

## 1. Introduction

This paper is concerned with nonparametric estimation of the functions  $m_1, \dots, m_d$  in the model

$$(1.1) \quad Y = F[\mu + m_1(X^1) + \dots + m_d(X^d)] + U,$$

where  $X^j$  ( $j = 1, \dots, d$ ) is the  $j$ 'th component of the random vector  $X \in \mathbb{R}^d$  for some finite  $d \geq 2$ ,  $F$  is a known function,  $\mu$  is an unknown constant,  $m_1, \dots, m_d$  are unknown functions, and  $U$  is an unobserved random variable satisfying  $E(U | X = x) = 0$  for almost every  $x$ . Estimation is based on an *iid* random sample  $\{Y_i, X_i : i = 1, \dots, n\}$  of  $(Y, X)$ . We describe an estimator of the additive components  $m_1, \dots, m_d$  that converges in probability pointwise at the rate  $n^{-2/5}$  when  $F$  and the  $m_j$ 's are twice continuously differentiable and the second derivative of  $F$  is sufficiently smooth. In contrast to previous estimators, only two derivatives are needed regardless of the dimension of  $X$ , so asymptotically there is no curse of dimensionality. The centered, scaled estimator of each additive component is asymptotically normally distributed with the same mean and variance that it would have if the other components were known.

Linton and Härdle (1996) (hereinafter LH) developed an estimator of the additive components of (1.1) that is based on marginal integration. The marginal integration method is discussed in more detail below. The estimator of LH converges at the rate  $n^{-2/5}$  and is asymptotically normally distributed, but it requires the  $m_j$ 's to have an increasing number of derivatives as the dimension of  $X$  increases. Thus, it suffers from the curse of dimensionality. Our estimator avoids this problem.

There is a large body of research on estimation of (1.1) when  $F$  is the identity function so that  $Y = \mu + m_1(X^1) + \dots + m_d(X^d) + U$ . Stone (1985, 1986) showed that  $n^{-2/5}$  is the optimal  $L_2$  rate of convergence of an estimator of the  $m_j$ 's when they are twice continuously differentiable. Stone (1994) and Newey (1997) describe spline estimators whose  $L_2$  rate of convergence is  $n^{-2/5}$ , but the pointwise rates of convergence and asymptotic distributions of spline and other series estimators remain unknown. Breiman and Friedman (1985); Buja, Hastie, and Tibshirani (1989); Hastie and Tibshirani (1990); Opsomer and Ruppert (1997); Mammen, Linton and Nielsen (1999); and Opsomer (2000) have investigated the properties of backfitting

procedures. Mammen, Linton and Nielsen (1999) give conditions under which a backfitting estimator of the  $m_j$ 's converges at the pointwise rate  $n^{-2/5}$  when these functions are twice continuously differentiable. The estimator is asymptotically normally distributed and avoids the curse of dimensionality, but extending it to models in which  $F$  is not the identity function appears to be quite difficult. Horowitz, Klemelä, and Mammen (2002) discuss optimality properties of a variety of estimators for nonparametric additive models.

Tjøstheim and Auestad (1994), Linton and Nielsen (1995); Chen, Härdle, Linton, and Severance-Lossin (1996); and Fan, Härdle, and Mammen (1998) have investigated the properties of marginal integration estimators for the case in which  $F$  is the identity function. These estimators are based on the observation that when  $F$  is the identity function, then  $m_1(x^1)$ , say, is given up to an additive constant by

$$(1.2) \quad \int \mathbf{E}(Y | X = x) w(x^2, \dots, x^d) dx^2 \dots dx^d,$$

where  $w$  is a non-negative function satisfying

$$\int w(x^2, \dots, x^d) dx^2 \dots dx^d = 1.$$

Therefore,  $m_1(x^1)$  can be estimated up to an additive constant by replacing  $\mathbf{E}(Y | X = x)$  in (1.2) with a nonparametric estimator. Linton and Nielsen (1995); Chen, Härdle, Linton, and Severance-Lossin (1996); and Fan, Härdle, and Mammen (1998) have given conditions under which a variety of estimators based on the marginal integration idea converge at rate  $n^{-2/5}$  and are asymptotically normal. LH extend marginal integration to the case in which  $F$  is not the identity function. However, marginal integration estimators have a curse of dimensionality: the smoothness of the  $m_j$ 's must increase as the dimension of  $X$  increases to achieve  $n^{-2/5}$  convergence. The reason for this is that estimating  $\mathbf{E}(Y | X = x)$  requires carrying out a  $d$ -dimensional nonparametric regression. If  $d$  is large and the  $m_j$ 's are only twice differentiable, then the bias of the resulting estimator of  $\mathbf{E}(Y | X = x)$  converges to zero too slowly as  $n \rightarrow \infty$  to estimate the  $m_j$ 's with a  $n^{-2/5}$  rate. For example the estimator of Fan, Härdle, and Mammen (1998), which imposes the weakest smoothness conditions of any existing marginal integration estimator, requires more than two derivatives if  $d \geq 5$ .

This paper describes a two-stage estimation procedure that does not require a  $d$ -dimensional nonparametric regression and, thereby, avoids the curse of dimensionality. In the first stage, nonlinear least squares is used to obtain a series approximation to each  $m_j$ . The first-

stage procedure imposes the additive structure of (1.1) and yields estimates of the  $m_j$ 's that have smaller asymptotic biases than do estimators based on marginal integration or other procedures that require  $d$ -dimensional nonparametric estimation. The first-stage estimates are inputs to the second stage. The second-stage estimate of, say,  $m_1$  is obtained by taking one Newton step from the first-stage estimate toward a local linear estimate. In large samples, the second-stage estimator has a structure similar to that of a local linear estimator, so deriving its pointwise rate of convergence and asymptotic distribution is relatively easy. The main results of this paper can also be obtained by using a local constant estimate in the second stage, and the results of Monte Carlo experiments described in Section 5 show that a local constant estimator has better finite-sample performance under some conditions. However, a local linear estimator has better boundary behavior and better ability to adapt to non-uniform designs, among other desirable properties (Fan and Gijbels 1996).

The remainder of this paper is organized as follows. Section 2 provides an informal description of the two-stage estimator. The main results are presented in Section 3. Section 4 discusses the selection of bandwidths. Section 5 presents the results of a small simulation study, and Section 6 presents concluding comments. The proofs of theorems are in Section 7. Throughout the paper, subscripts index observations and superscripts denote components of vectors. Thus,  $X_i$  is the  $i$ 'th observation of  $X$ ,  $X^j$  is the  $j$ 'th component of  $X$ , and  $X_i^j$  is the  $i$ 'th observation of the  $j$ 'th component.

## 2. Informal Description of the Estimator

Assume that the support of  $X$  is  $\mathcal{X} \equiv [-1, 1]^d$ , and normalize  $m_1, \dots, m_d$  so that

$$\int_{-1}^1 m_j(v) dv = 0; \quad j = 1, \dots, d.$$

For any  $x \in \mathbb{R}^d$  define  $m(x) = m_1(x^1) + \dots + m_d(x^d)$ , where  $x^j$  is the  $j$ 'th component of  $x$ . Let  $\{p_k : k = 1, 2, \dots\}$  denote a basis for smooth functions on  $[-1, 1]$ . A precise definition of "smooth" and conditions that the basis functions must satisfy are given in Section 3. These conditions include:

$$(2.1) \quad \int_{-1}^1 p_k(v) dv = 0;$$

$$(2.2) \quad \int_{-1}^1 p_j(v) p_k(v) dv = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{otherwise;} \end{cases}$$

and

$$(2.3) \quad m_j(x^j) = \sum_{k=1}^{\infty} \theta_{jk} p_k(x^j)$$

for each  $j=1, \dots, d$ , each  $x^j \in [0,1]$ , and suitable coefficients  $\{\theta_{jk}\}$ . For any positive integer  $\kappa$ , define

$$P_{\kappa}(x) = [1, p_1(x^1), \dots, p_{\kappa}(x^1), p_1(x^2), \dots, p_{\kappa}(x^2), \dots, p_1(x^d), \dots, p_{\kappa}(x^d)]'$$

Then for  $\theta_{\kappa} \in \mathbb{R}^{\kappa d+1}$ ,  $P_{\kappa}(x)' \theta_{\kappa}$  is a series approximation to  $\mu + m(x)$ . Section 3 gives conditions that  $\kappa$  must satisfy. These require that  $\kappa \rightarrow \infty$  at an appropriate rate as  $n \rightarrow \infty$ .

To obtain the first-stage estimators of the  $m_j$ 's, let  $\{Y_i, X_i : i=1, \dots, n\}$  be a random sample of  $(Y, X)$ . Let  $\hat{\theta}_{n\kappa}$  be a solution to

$$\text{minimize: } S_{n\kappa}(\theta) \equiv n^{-1} \sum_{i=1}^n \{Y_i - F[P_{\kappa}(X_i)' \theta]\}^2,$$

where  $\Theta_{\kappa} \subset \mathbb{R}^{\kappa d+1}$  is a compact parameter set. The series estimator of  $\mu + m(x)$  is

$$\tilde{\mu} + \tilde{m}(x) = P_{\kappa}(x)' \hat{\theta}_{n\kappa},$$

where  $\tilde{\mu}$  is the first component of  $\hat{\theta}_{n\kappa}$ . The estimator of  $m_j(x^j)$  for any  $j=1, \dots, d$  and any  $x^j \in [0,1]$  is the product of  $[p_1(x^j), \dots, p_{\kappa}(x^j)]$  with the appropriate components of  $\hat{\theta}_{n\kappa}$ .

To obtain the second-stage estimator of (say)  $m_1(x^1)$ , let  $\tilde{X}_i$  denote the  $i$ 'th observation of  $\tilde{X} \equiv (X^2, \dots, X^d)$ . Define  $\tilde{m}_{-1}(\tilde{X}_i) = \tilde{m}_2(X_i^2) + \dots + \tilde{m}_d(X_i^d)$ , where  $X_i^j$  is the  $i$ 'th observation of the  $j$ 'th component of  $X$  and  $\tilde{m}_j$  is the series estimator of  $m_j$ . Let  $K$  be a probability density function on  $[-1,1]$ , and define  $K_h(v) = K(v/h)$  for any real, positive constant  $h$ . Conditions that  $K$  and  $h$  must satisfy are given in Section 3. These include  $h \rightarrow 0$  at an appropriate rate as  $n \rightarrow \infty$ . Define

$$S'_{nj1}(x^1, \tilde{m}) =$$

$$-2 \sum_{i=1}^n \{Y_i - F[\tilde{\mu} + \tilde{m}_1(x^1) + \tilde{m}_{-1}(\tilde{X}_i)]\} F'[\tilde{\mu} + \tilde{m}_1(x^1) + \tilde{m}_{-1}(\tilde{X}_i)] (X_i^1 - x^1)^j K_h(x^1 - X_i^1)$$

for  $j=0,1$  and

$$S''_{nj1}(x^1, \tilde{m}) = 2 \sum_{i=1}^n F'[\tilde{\mu} + \tilde{m}_1(x^1) + \tilde{m}_{-1}(\tilde{X}_i)]^2 (X_i^1 - x^1)^j K_h(x^1 - X_i^1) \\ - 2 \sum_{i=1}^n \{Y_i - F[\tilde{\mu} + \tilde{m}_1(x^1) + \tilde{m}_{-1}(\tilde{X}_i)]\} F''[\tilde{\mu} + \tilde{m}_1(x^1) + \tilde{m}_{-1}(\tilde{X}_i)] (X_i^1 - x^1)^j K_h(x^1 - X_i^1)$$

for  $j=0,1,2$ . The second-stage estimator of  $m_1(x^1)$  is

$$(2.4) \quad \hat{m}_1(x^1) = \tilde{m}_1(x^1) - \frac{S''_{n21}(x^1, \tilde{m})S'_{n01}(x^1, \tilde{m}) - S''_{n11}(x^1, \tilde{m})S'_{n11}(x^1, \tilde{m})}{S''_{n01}(x^1, \tilde{m})S''_{n21}(x^1, \tilde{m}) - S''_{n11}(x^1, \tilde{m})^2}.$$

The second stage estimators of  $m_2(x^2), \dots, m_d(x^d)$  are obtained similarly.

The estimator (2.4) can be understood intuitively as follows. If  $\tilde{\mu}$  and  $\tilde{m}_{-1}$  were the true values of  $\mu$  and  $m_{-1}$ , the local linear estimator of  $m_1(x^1)$  would minimize

$$(2.5) \quad S_{n1}(x^1, b_0, b_1) = \sum_{i=1}^n \{Y_i - F[\tilde{\mu} + b_0 + b_1(X_i^1 - x^1) + \tilde{m}_{-1}(\tilde{X}_i)]\}^2.$$

Moreover,  $S'_{nj1}(x^1, \tilde{m}) = \partial S_{n1}(x^1, b_0, b_1) / \partial b_j$  ( $j=0,1$ ) evaluated at  $b_0 = \tilde{m}_1(x^1)$  and  $b_1 = 0$ .  $S''_{nj1}(x^1, \tilde{m})$  gives the second derivatives of  $S_{n1}(x^1, b_0, b_1)$  evaluated at the same point. The estimator (2.4) is the result of taking one Newton step from the starting values  $b_0 = \tilde{m}_1(x^1)$ ,  $b_1 = 0$  toward the minimum of the right-hand side of (2.5).

Section 3 gives conditions under which  $\hat{m}_1(x^1) - m_1(x^1) = O_p(n^{-2/5})$  and  $n^{2/5}[\hat{m}_1(x^1) - m_1(x^1)]$  is asymptotically normally distributed for any finite  $d$  when  $F$  and the  $m_j$ 's are twice continuously differentiable.

### 3. Main Results

This section has two parts. Section 3.1 states the assumptions that are used to prove the main results. Section 3.2 states the results. The main results are the  $n^{-2/5}$ -consistency and asymptotic normality of the  $m_j$ 's.

The following additional notation is used. For any matrix  $A$ , define the norm  $\|A\| = [\text{trace}(A'A)]^{1/2}$ . Define  $U = Y - F[\mu + m(X)]$ ,  $V(x) = \text{Var}(U | X = x)$ ,  $Q_\kappa = \mathbf{E}\{F'[\mu + m(X)]^2 P_\kappa(X)P_\kappa(X)'\}$ , and  $\Psi_\kappa = Q_\kappa^{-1} \mathbf{E}\{F'[\mu + m(X)]^2 V(X)P_\kappa(X)P_\kappa(X)'\} Q_\kappa^{-1}$

whenever the latter quantity exists.  $Q_\kappa$  and  $\Psi_\kappa$  are  $d(\kappa) \times d(\kappa)$  positive semidefinite matrices, where  $d(\kappa) = \kappa d + 1$ . Let  $\lambda_{\kappa, \min}$  denote the smallest eigenvalue of  $Q_\kappa$ . Let  $Q_{\kappa, ij}$  denote the  $(i, j)$  element of  $Q_\kappa$ . Define  $\zeta_\kappa = \sup_{x \in \mathcal{X}} \|P_\kappa(x)\|$ . Let  $\{\theta_{jk}\}$  be the coefficients of the series expansion (2.3). For each  $\kappa$  define

$$\theta_\kappa = (\mu, \theta_{11}, \dots, \theta_{1\kappa}, \theta_{21}, \dots, \theta_{2\kappa}, \dots, \theta_{d1}, \dots, \theta_{d\kappa})'.$$

### 3.1 Assumptions

The main results are obtained under the following assumptions.

A1: The data,  $\{(Y_i, X_i) : i = 1, \dots, n\}$ , are an *iid* random sample from the distribution of  $(Y, X)$ , and  $E(Y | X = x) = F[\mu + m(x)]$  for almost every  $x \in \mathcal{X} \equiv [-1, 1]^d$ .

A2: (i) The support of  $X$  is  $\mathcal{X}$ . (ii) The distribution of  $X$  is absolutely continuous with respect to Lebesgue measure. (iii) The probability density function of  $X$  is bounded, bounded away from zero, and twice continuously differentiable on  $\mathcal{X}$ . (iv) There are constants  $c_V > 0$  and  $C_V < \infty$  such that  $c_V \leq \text{Var}(U | X = x) \leq C_V$  for all  $x \in \mathcal{X}$ . (v) There is a constant  $C_U < \infty$  such that  $E|U|^j \leq C_U^{j-2} j! E(U^2) < \infty$  for all  $j \geq 2$ .

A3: (i) There is a constant  $C_m < \infty$  such that  $|m_j(v)| \leq C_m$  for each  $j = 1, \dots, d$  and all  $v \in [-1, 1]$ . (ii) Each function  $m_j$  is twice continuously differentiable on  $[-1, 1]$ . (iii) There are constants  $C_{F1} < \infty$ ,  $c_{F2} > 0$ , and  $C_{F2} < \infty$  such that  $F(v) \leq C_{F1}$  and  $c_{F1} \leq F'(v) \leq C_{F2}$  for all  $v \in [\mu - C_m d, \mu + C_m d]$ . (iv)  $F$  is twice continuously differentiable on  $[\mu - C_m d, \mu + C_m d]$ . (v) There are constants  $C_{F3} < \infty$  and  $s > 5/7$  such that  $|F''(v_2) - F''(v_1)| \leq C_{F3} |v_2 - v_1|^s$  for all  $v_2, v_1 \in [\mu - C_m d, \mu + C_m d]$ .

A4: (i) There are constants  $C_Q < \infty$  and  $c_\lambda > 0$  such that  $|Q_{\kappa, ij}| \leq C_Q$  and  $\lambda_{\kappa, \min} > c_\lambda$  for all  $\kappa$  and all  $i, j = 1, \dots, d(\kappa)$ . (ii) The largest eigenvalue of  $\Psi_\kappa$  is bounded for all  $\kappa$ .

A5: (i) The functions  $\{p_k\}$  satisfy (2.1) and (2.2). (ii) There is a constant  $c_\kappa > 0$  such that  $\zeta_\kappa \geq c_\kappa$  for all sufficiently large  $\kappa$ . (iii)  $\zeta_\kappa = O(\kappa^{1/2})$  as  $\kappa \rightarrow \infty$ . (iv) There are a constant  $C_\theta < \infty$  and vectors  $\theta_{\kappa 0} \in \Theta_\kappa \equiv [-C_\theta, C_\theta]^{d(\kappa)}$  such that  $\sup_{x \in \mathcal{X}} |\mu + m(x) - P_\kappa(x)' \theta_{\kappa 0}| = O(\kappa^{-2})$  as  $\kappa \rightarrow \infty$ . (v) For each  $\kappa$ ,  $\theta_\kappa$  is an interior point of  $\Theta_\kappa$ .



A6: (i)  $\kappa = C_\kappa n^{4/15+\nu}$  for some constant  $C_\kappa$  satisfying  $0 < C_\kappa < \infty$  and some  $\nu$  satisfying  $0 < \nu < \min\{1/30, (7s-5)/[30(1+s)]\}$ . (ii)  $h = C_h n^{-1/5}$  for some constant  $C_h$  satisfying  $0 < C_h < \infty$ .

A7: The function  $K$  is a bounded, continuous probability density function on  $[-1,1]$  and is symmetrical about 0.

Assumption A2(v) restricts the thickness of the tails of the distribution of  $U$  and is used to prove consistency of the first-stage estimator. Assumption A3 defines the sense in which  $F$  and the  $m_j$ 's must be smooth. A3(iii) is needed for identification. A4 insures the existence and non-singularity of the covariance matrix of the asymptotic form of the first-stage estimator. Assumption A4(i) implies A4(ii) if  $U$  is homoskedastic. Assumptions A5(iii) and A5(iv) bound the magnitudes of the basis functions and insure that the errors in the series approximations to the  $m_j$ 's converge to zero sufficiently rapidly as  $\kappa \rightarrow \infty$ . These assumptions are satisfied by spline and (for periodic functions) Fourier bases. Assumption A6 states the rates at which  $\kappa \rightarrow \infty$  and  $h \rightarrow 0$  as  $n \rightarrow \infty$ . The assumed rate of convergence of  $h$  is well known to be asymptotically optimal for one-dimensional kernel mean-regression when the conditional mean function is twice continuously differentiable. The required rate for  $\kappa$  insures that the asymptotic bias and variance of the first-stage estimator are sufficiently small to achieve an  $n^{-2/5}$  rate of convergence in the second stage. The  $L_2$  rate of convergence of a series estimator of  $m_j$  is maximized by setting  $\kappa \propto n^{1/5}$ , which is slower than the rates permitted by A6(i) (Newey (1997)). Thus, A6(i) requires the first-stage estimator to be undersmoothed. Undersmoothing is needed to insure sufficiently rapid convergence of the bias of the first-stage estimator.

### 3.2 Theorems

This section states two theorems that give the main results of the paper. Theorem 1 gives the asymptotic behavior of the first-stage series estimator under assumptions A1-A6(i). Theorem 2 gives the properties of the second-stage estimator. For  $i=1, \dots, n$ , define  $U_i = Y_i - F[\mu + m(X_i)]$  and  $b_{\kappa 0}(x) = \mu + m(x) - P_\kappa(x)' \theta_{\kappa 0}$ . Let  $\|v\|$  denote the Euclidean norm of any finite-dimensional vector  $v$ .

Theorem 1: Let A1-A6(i) hold. Then

$$(a) \quad \lim_{n \rightarrow \infty} \left\| \hat{\theta}_{n\kappa} - \theta_{\kappa 0} \right\| = 0$$

almost surely,

$$(b) \quad \hat{\theta}_{n\kappa} - \theta_{\kappa 0} = O_p(\kappa^{1/2}/n^{1/2} + \kappa^{-2}),$$

and

$$(c) \quad \sup_{x \in \mathcal{X}} |\tilde{m}(x) - m(x)| = O_p(\kappa/n^{1/2} + \kappa^{-3/2}).$$

In addition,

$$\hat{\theta}_{n\kappa} - \theta_{\kappa 0} = n^{-1} Q_\kappa^{-1} \sum_{i=1}^n F'[\mu + m(X_i)] P_\kappa(X_i) U_i$$

(d)

$$+ n^{-1} Q_\kappa^{-1} \sum_{i=1}^n F'[\mu + m(X_i)]^2 P_\kappa(X_i) b_\kappa(X_i) + R_n,$$

where  $\|R_n\| = O_p(\kappa^{3/2}/n + n^{-1/2})$ . ■

Now let  $f_X$  denote the probability density function of  $X$ . For  $j = 0, 1$ , define

$$S'_{nj1}(x^1, m) =$$

$$- 2 \sum_{i=1}^n \{Y_i - F[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)]\} F'[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)] (X_i^1 - x^1)^j K_h(x^1 - X_i^1).$$

Also define

$$D_0(x^1) = 2 \int F'[\mu + m_1(x^1) + m_{-1}(\tilde{x})]^2 f_X(x^1, \tilde{x}) d\tilde{x},$$

$$D_1(x^1) = 2 \int F'[\mu + m_1(x^1) + m_{-1}(\tilde{x})]^2 [\partial f_X(x^1, \tilde{x}) / \partial x^1] d\tilde{x},$$

$$A_K = \int_{-1}^1 v^2 K(v) dv,$$

$$B_K = \int_{-1}^1 K(v)^2 dv,$$

$$g(x^1, \tilde{x}) = F''[\mu + m_1(x^1) + m_{-1}(\tilde{x})] m_1'(x^1) + F'[\mu + m_1(x^1) + m_{-1}(\tilde{x})] m_1''(x^1),$$

$$\beta_1(x^1) = 2C_h^2 A_K D_0(x^1)^{-1} \int g(x^1, \tilde{x}) F'[\mu + m_1(x^1) + m_{-1}(\tilde{x})] f_X(x^1, \tilde{x}) d\tilde{x},$$

and

$$V_1(x^1) = B_K C_h^{-1} D(x^1)^{-2} \int \text{Var}(U | x^1, \tilde{x}) F'[\mu + m_1(x^1) + m_{-1}(\tilde{x})]^2 f_X(x^1, \tilde{x}) d\tilde{x}.$$

The next theorem gives the asymptotic properties of the second-stage estimator.

**Theorem 2:** Let A1-A6 hold. Then

(a)  $\hat{m}_1(x^1) - m_1(x^1) =$

$$[nhD_0(x^1)]^{-1}\{-S'_{n01}(x^1, m) + [D_1(x^1)/D_0(x^1)]S'_{n11}(x^1, m)\} + o_p(n^{-2/5})$$

uniformly over  $|x^1| \leq 1-h$  and  $\hat{m}_1(x^1) - m_1(x^1) = O_p[(\log n)^{1/2}n^{-2/5}]$  uniformly over  $|x^1| \leq 1$ .

(b)  $n^{2/5}[\hat{m}_1(x^1) - m_1(x^1)] \rightarrow^d N[\beta_1(x^1), V_1(x^1)].$

(c) If  $j \neq 1$ , then  $n^{2/5}[\hat{m}_1(x^1) - m_1(x^1)]$  and  $n^{2/5}[\hat{m}_j(x^j) - m_j(x^j)]$  are asymptotically independently normally distributed. ■

Theorem 2(a) implies that asymptotically,  $n^{2/5}[\hat{m}_1(x^1) - m_1(x^1)]$  is not affected by random sampling errors in the first stage estimator. In fact, the second-stage estimator of  $m_1(x^1)$  has the same asymptotic distribution that it would have if  $m_2, \dots, m_d$  were known and local linear estimation were used to estimate  $m_1(x^1)$  directly. In this sense, our estimator achieves an oracle bound. Parts (b) and (c) of Theorem 2 imply that the estimators of  $m_1(x^1), \dots, m_d(x^d)$  are asymptotically independently distributed.

It is also possible to use a local constant estimator in the second stage. The resulting second-stage estimator is

$$\hat{m}_{1,LC}(x^1) = \tilde{m}_1(x^1) - S'_{n01}(x^1, \tilde{m}) / S''_{n01}(x^1, \tilde{m}).$$

The following modification of Theorem 2, which we state without proof, gives the asymptotic properties of the local constant second-stage estimator. Define

$$g_{LC}(x^1, \tilde{x}) = (\partial^2 / \partial \zeta^2) \{F[m_1(\zeta + x^1) + m_{-1}(\tilde{x})] - F[m_1(x^1) + m_{-1}(\tilde{x})]\} f_X(\zeta + x^1, \tilde{x}) \Big|_{\zeta=0}$$

and

$$\beta_{1,LC}(x^1) = 2C_h^2 A_K D_0(x^1)^{-1} \int g_{LC}(x^1, \tilde{x}) F'[\mu + m_1(x^1) + m_{-1}(\tilde{x})] f_X(x^1, \tilde{x}) d\tilde{x}.$$

**Theorem 3:** Let A1-A6 hold. Then

(a)  $\hat{m}_{1,LC}(x^1) - m_1(x^1) = -[nhD_0(x^1)]^{-1} S'_{n01}(x^1, m) + o_p(n^{-2/5}),$

uniformly over  $|x^1| \leq 1-h$  and  $\hat{m}_1(x^1) - m_1(x^1) = O_p[(\log n)^{1/2}n^{-2/5}]$  uniformly over  $|x^1| \leq 1$ .

(b)  $n^{2/5}[\hat{m}_{1,LC}(x^1) - m_1(x^1)] \rightarrow^d N[\beta_{1,LC}(x^1), V_1(x^1)].$

(c) If  $j \neq 1$ , then  $n^{2/5}[\hat{m}_{1,LC}(x^1) - m_1(x^1)]$  and  $n^{2/5}[\hat{m}_{j,LC}(x^j) - m_j(x^j)]$  are asymptotically independently normally distributed. ■

$V_1(x^1)$  and  $\beta_1(x^1)$  and  $\beta_{1,LC}(x^1)$  can be estimated consistently by replacing unknown population parameters with consistent estimators. Section 4 gives a method for estimating the derivatives of  $m_1$  that are in the expressions for  $\beta_1(x^1)$  and  $\beta_{1,LC}(x^1)$ . As is usual in nonparametric estimation, reasonably precise bias estimation is possible only by making assumptions that amount to undersmoothing. One way of doing this is to assume that the second derivative of  $m_1$  satisfies a Lipschitz condition. Alternatively, one can set  $h = C_h n^{-\gamma}$  for  $1/5 < \gamma < 1$ . Then  $n^{(1-\gamma)/2}[\hat{m}_1(x^1) - m_1(x^1)] \rightarrow^d N[0, V_1(x^1)]$ , and  $n^{(1-\gamma)/2}[\hat{m}_{1,LC}(x^1) - m_1(x^1)] \rightarrow^d N[0, V_1(x^1)]$ .

#### 4. Bandwidth Selection

This section presents a plug-in and a penalized least squares (PLS) method for choosing  $h$  in applications. We begin with a description of the plug-in method. This method estimates the value of  $h$  that minimizes the asymptotic integrated mean-square error (AIMSE) of  $n^{2/5}[\hat{m}_1(x^1) - m_1(x^1)]$  for  $j = 1, \dots, d$ . We discuss only local linear estimation, but similar results hold for local constant estimation. The AIMSE of  $n^{2/5}(\hat{m}_1 - m_1)$  is defined as

$$AIMSE_1 = n^{4/5} \int_{-1}^1 w(x^1) [\beta_1(x^1)^2 + V_1(x^1)] dx^1$$

where  $w(\cdot)$  is a non-negative weight function that integrates to one. We also define the integrated squared error as (ISE)

$$ISE_1 = n^{4/5} \int_{-1}^1 w(x^1) [\hat{m}_1(x^1) - m_1(x^1)]^2 dx^1.$$

We define the asymptotically optimal bandwidth for estimating  $m_1$  as  $C_{h1} n^{-1/5}$ , where  $C_{h1}$  minimizes  $AIMSE_1$ . Let  $\tilde{\beta}_1(x^1) = \beta_1(x^1) / C_h^2$  and  $\tilde{V}_1(x^1) = C_h V_1(x^1)$ . Then

$$(4.1) \quad C_{h1} = \left[ \frac{\int_{-1}^1 w(x^1) \tilde{V}_1(x^1) dx^1}{\int_{-1}^1 w(x^1) \tilde{\beta}_1(x^1)^2 dx^1} \right]^{1/5}.$$

The results for the plug-in method rely on the following two theorems. Theorem 4 shows that the difference between the ISE and AIMSE is asymptotically negligible. Theorem 5 gives a method for estimating the first and second derivatives of  $m_j$ . Let  $G^{(\ell)}$  denote the  $\ell$ 'th derivative of any  $\ell$ -times differentiable function  $G$ .

**Theorem 4:** Let A1-A6 hold. Then for a continuous weight function  $w(\cdot)$  and as  $n \rightarrow \infty$ ,  $AIMSE_1 = ISE_1 + o_p(1)$ . ■

**Theorem 5:** Let A1-A6 hold. Let  $L$  be a twice differentiable probability density function on  $[-1,1]$ , and let  $\{g_n : n=1,2,\dots\}$  be a sequence of strictly positive real numbers satisfying  $g_n \rightarrow 0$  and  $g_n^2 n^{4/5} (\log n)^{-1} \rightarrow \infty$  as  $n \rightarrow \infty$ . For  $\ell=1,2$  define

$$\hat{m}_1^{(\ell)}(x^1) = g_n^{-1-\ell} \int_{-1}^1 L^{(\ell)}[(x^1 - v)/g_n] \hat{m}_1(v) dv.$$

Then as  $n \rightarrow \infty$  and for  $\ell=1,2$

$$\sup_{|x^1| \leq 1} |\hat{m}_1^{(\ell)}(x^1) - m_1^{(\ell)}(x^1)| = o_p(1). \quad \blacksquare$$

A plug-in estimator of  $C_{h1}$  can now be obtained by replacing unknown population quantities on the right-hand side of (4.1) with consistent estimators. Theorem 5 provides consistent estimators of the required derivatives of  $m_1$ . Estimators of the conditional variance of  $U$  and of  $f_X$  can be obtained by using standard kernel methods.

We now describe the PLS method. This method simultaneously estimates the bandwidths for second-stage estimation of all of the functions  $m_j$  ( $j=1,\dots,d$ ). Let  $h_j = C_{hj} n^{-1/5}$  be the bandwidth for  $\hat{m}_j$ . Then the PLS method selects the  $C_{hj}$ 's that minimize an estimate of the average squared error

$$ASE(\bar{h}) = n^{-1} \sum_{i=1}^n \{F[\tilde{\mu} + \hat{m}(X_i)] - F[\mu + m(X_i)]\}^2,$$

where  $\bar{h} = (C_{h1} n^{-1/5}, \dots, C_{hd} n^{-1/5})$ . Specifically, the PLS method selects the  $C_{hj}$ 's to

$$(4.2) \quad \underset{C_{h1}, \dots, C_{hd}}{\text{minimize}}: PLS(\bar{h}) = n^{-1} \sum_{i=1}^n \{Y_i - F[\tilde{\mu} + \hat{m}(X_i)]\}^2 \\ + 2K(0)n^{-1} \sum_{i=1}^n \{F'[\mu + m(X_i)]^2 \hat{V}(X_i)\} \sum_{j=1}^d [n^{4/5} C_{hj} \hat{D}_j(X_i^j)]^{-1},$$

where the  $C_{hj}$ 's are restricted to a compact, positive interval that excludes 0,

$$\hat{D}_j(x^j) = \frac{1}{nh_j} \sum_{i=1}^n K_{h_j}(X_i^j - x^j) F'[\tilde{\mu} + \hat{m}(X_i)]^2,$$

and

$$\hat{V}(x) = \left[ \sum_{i=1}^n K_{h_1}(X_i^1 - x^1) \dots K_{h_d}(X_i^d - x^d) \right]^{-1} \\ \times \sum_{i=1}^n K_{h_1}(X_i^1 - x^1) \dots K_{h_d}(X_i^d - x^d) \{Y_i - F[\tilde{\mu} + \hat{m}(X_i)]\}^2.$$

The bandwidths used for  $\hat{V}$  may be different from those used for  $\hat{m}$  because  $\hat{V}$  is a full dimensional nonparametric estimator. We now argue that the difference

$$n^{-1} \sum_{i=1}^n U_i^2 + ASE(\bar{h}) - PLS(\bar{h})$$

is asymptotically negligible and, therefore, that the solution to (4.2) estimates the bandwidths that minimize  $ASE$ . A proof of this result only requires additional smoothness conditions on  $F$  and more restrictive assumptions on  $\kappa$ . The proof can be carried out by making arguments similar to those used in the proof of Theorem 2 but with a higher-order stochastic expansion for  $\hat{m} - m$ . Here, we provide only a heuristic outline. For this purpose, note that

$$n^{-1} \sum_{i=1}^n U_i^2 + ASE(\bar{h}) - PLS(\bar{h}) = 2n^{-1} \sum_{i=1}^n \{F[\tilde{\mu} + \hat{m}(X_i)] - F[\mu + m(X_i)]\} U_i \\ - 2K(0)n^{-1} \sum_{i=1}^n F'[\mu + m(X_i)]^2 \hat{V}(X_i) \sum_{j=1}^d [n^{4/5} C_{h_j} \hat{D}_j(X_i^j)]^{-1}.$$

We now approximate  $F[\tilde{\mu} + \hat{m}(X_i)] - F[\mu + m(X_i)]$  by a linear expansion in  $\hat{m} - m$  and replace  $\hat{m} - m$  with the stochastic approximation of Theorem 2(a). (A rigorous argument would require a higher-order expansion of  $\hat{m} - m$ .) Thus,  $F[\tilde{\mu} + \hat{m}(X_i)] - F[\mu + m(X_i)]$  is approximated by a linear form in  $U_i$ . Dropping higher-order terms leads to an approximation of

$$\frac{2}{n} \sum_{i=1}^n \{F[\tilde{\mu} + \hat{m}(X_i)] - F[\mu + m(X_i)]\} U_i$$

that is a  $U$  statistic in  $U_i$ . The off-diagonal terms of the  $U$  statistic can be shown to be of higher order and, therefore, asymptotically negligible. Thus, we get

$$\frac{2}{n} \sum_{i=1}^n \{F[\tilde{\mu} + \hat{m}(X_i)] - F[\mu + m(X_i)]\} U_i \approx \frac{2}{n} \sum_{i=1}^n F'[\mu + m(X_i)]^2 \text{Var}(U_i | X_i) \\ \times \sum_{j=1}^d [n^{4/5} C_{hj} D_{0j}(X_i^j)] h^{-1} K(0),$$

where

$$D_{0j}(x^j) = 2E\{F'[\mu + m(X_i)]^2 | X_i^j = x^j\} f_{X^j}(x^j)$$

and  $f_{X^j}$  is the probability density function of  $X^j$ . Now by standard kernel smoothing arguments,  $D_{0j}(x^j) \approx \hat{D}_j(x^j)$ . In addition, it is clear that  $\hat{V}(X_i) \approx V(U_i | X_i)$ , which establishes the desired result.

## 5. Monte Carlo Experiments

This section presents the results of a small set of Monte Carlo experiments that compare the finite-sample performances of the two-stage estimator and the estimator of LH. Experiments were carried out with  $d = 2$  and  $d = 5$ . The sample size is  $n = 500$ . The experiments with  $d = 2$  consist of estimating  $f_1$  and  $f_2$  in the binary logit model

$$P(Y = 1 | X = x) = L[f_1(x^1) + f_2(x^2)],$$

where  $L$  is the cumulative logistic distribution function:

$$L(v) = e^v / [1 + e^v]; \quad -\infty < v < \infty.$$

The experiments with  $d = 5$  consist of estimating  $f_1$  and  $f_2$  in the binary logit model

$$P(Y = 1 | X = x) = L \left[ f_1(x^1) + f_2(x^2) + \sum_{j=3}^5 x^j \right].$$

In all of the experiments,  $f_1(x) = \sin(\pi x)$  and  $f_2(x) = \Phi(3x)$ , where  $\Phi$  is the standard normal distribution function. The components of  $X$  are independently distributed as  $U[-1,1]$ . Estimation is carried out under the assumption that the additive components have two (but not necessarily more) continuous derivatives. Under this assumption, the two-stage estimator has the rate of convergence  $n^{-2/5}$ . The LH estimator has this rate of convergence if  $d = 2$  but not if  $d = 5$ .

$B$ -splines were used for the first-stage of the two-stage estimator. The kernel used for the second stage and for the LH estimator is

$$K(v) = (15/16)(1-v^2)^2 I(|v| \leq 1).$$

Experiments were carried out using both local-constant and local-linear estimators in the second stage of the two-stage method. There were 1000 Monte Carlo replications per experiment with the two-stage estimator but only 500 replications with the LH estimator because of the very long computing times it entails. The experiments were carried out in GAUSS using GAUSS random number generators.

The results of the experiments are summarized in Table 1, which shows the empirical integrated mean-square errors (EIMSE's) of the estimators at the values of the tuning parameters that minimize the EIMSE's. Lengthy computing times precluded using data-based methods for selecting tuning parameters in the experiments. The EIMSE's of the local-constant and local linear two-stage estimates of  $f_1$  are considerably smaller than the EIMSE's of the LH estimator. The EIMSE's of the local-constant and LH estimators of  $f_2$  are approximately equal same, whereas the local-linear estimator of  $f_2$  has a larger EIMSE.

## 6. Conclusions

This paper has described an estimator of the additive components of a nonparametric additive model with a known link function. The estimator is asymptotically normally distributed and has a pointwise rate of convergence in probability of  $n^{-2/5}$  when the unknown functions are twice continuously differentiable, regardless of the dimension of the explanatory variable  $X$ . In contrast, achieving the rate of convergence  $n^{-2/5}$  with the only other currently available estimator for this model requires the additive components to have an increasing number of derivatives as the dimension of  $X$  increases.

## 7. Appendix: Proofs of Theorems

Assumptions A1-A7 hold throughout this section.

a. Theorem 1

This section begins with lemmas that are used to prove Theorem 1.

Lemma 1: There are constants  $a > 0$  and  $C < \infty$  such that

$$P \left[ \sup_{\theta \in \Theta_\kappa} |S_{nk}(\theta) - ES_{nk}(\theta)| > \varepsilon \right] \leq C \exp(-na\varepsilon^2)$$

for any sufficiently small  $\varepsilon > 0$  and all sufficiently large  $n$ .

Proof: Write



$$S_{n\kappa}(\theta) = n^{-1} \sum_{i=1}^n Y_i^2 - 2S_{n\kappa 1}(\theta) + S_{n\kappa 2}(\theta),$$

where

$$S_{n\kappa 1}(\theta) = n^{-1} \sum_{i=1}^n Y_i F[P_\kappa(X_i)' \theta]$$

and

$$S_{n\kappa 2}(\theta) = n^{-1} \sum_{i=1}^n F[P_\kappa(X_i)' \theta]^2.$$

It suffices to prove that

$$\mathbf{P} \left[ \sup_{\theta \in \Theta_\kappa} |S_{n\kappa j}(\theta) - \mathbf{E}S_{n\kappa j}(\theta)| > \varepsilon \right] \leq \tilde{C} \exp(-na\varepsilon^2) \quad (j=1,2)$$

for any  $\varepsilon > 0$ , some  $\tilde{C} < \infty$ , and all sufficiently large  $n$ . The proof is given only for  $j=1$ . Similar arguments apply when  $j=2$ .

Define  $\tilde{S}_{n\kappa 1}(\theta) = S_{n\kappa 1}(\theta) - \mathbf{E}S_{n\kappa 1}(\theta)$ . Divide  $\Theta_\kappa$  into hypercubes of edge-length  $\ell$ . Let  $\Theta_\kappa^{(1)}, \dots, \Theta_\kappa^{(M)}$  denote the  $M = (2C_\theta / \ell)^{d(\kappa)}$  cubes thus created. Let  $\theta_{\kappa j}$  be the point the center of  $\Theta_\kappa^{(j)}$ . The maximum distance between  $\theta_{\kappa j}$  and any other point in  $\Theta_\kappa^{(j)}$  is  $r = d(\kappa)^{1/2} \ell / 2$ , and  $M = \exp\{d(\kappa)[\log(C_\theta / r) + (1/2)\log d(\kappa)]\}$ . Now

$$\left[ \sup_{\theta \in \Theta_\kappa} |\tilde{S}_{n\kappa 1}(\theta)| > \varepsilon \right] \subset \bigcup_{j=1}^M \left[ \sup_{\theta \in \Theta_\kappa^{(j)}} |\tilde{S}_{n\kappa 1}(\theta)| > \varepsilon \right].$$

Therefore,

$$\mathbf{P}_n \equiv \mathbf{P} \left[ \sup_{\theta \in \Theta_\kappa} |\tilde{S}_{n\kappa 1}(\theta)| > \varepsilon \right] \leq \sum_{j=1}^M \mathbf{P} \left[ \sup_{\theta \in \Theta_\kappa^{(j)}} |\tilde{S}_{n\kappa 1}(\theta)| > \varepsilon \right].$$

Now for  $\theta \in \Theta_\kappa^{(j)}$ ,

$$\begin{aligned} |\tilde{S}_{n\kappa 1}(\theta)| &\leq |\tilde{S}_{n\kappa 1}(\theta_{\kappa j})| + |\tilde{S}_{n\kappa 1}(\theta) - \tilde{S}_{n\kappa 1}(\theta_{\kappa j})| \\ &\leq |\tilde{S}_{n\kappa 1}(\theta_{\kappa j})| + [C_{F1}(\ell + \zeta_\kappa r)] n^{-1} \sum_{i=1}^n |Y_i| \\ &\leq |\tilde{S}_{n\kappa 1}(\theta_{\kappa j})| + 2C_{F1} \zeta_\kappa r n^{-1} \sum_{i=1}^n |Y_i| \end{aligned}$$

for all sufficiently large  $\kappa$  and, therefore,  $n$ . Therefore, for all sufficiently large  $n$ ,

$$\mathbf{P} \left[ \sup_{\theta \in \Theta_{\kappa}^{(j)}} |\tilde{S}_{n\kappa 1}(\theta)| > \varepsilon \right] \leq \mathbf{P} [|\tilde{S}_{n\kappa 1}(\theta_{\kappa j})| > \varepsilon/2] + \mathbf{P} \left[ 2C_{F1} \zeta_{\kappa} r n^{-1} \sum_{i=1}^n |Y_i| > \varepsilon/2 \right].$$

Choose  $r = \zeta_{\kappa}^2$ . Then  $\varepsilon/2 - 2C_{F1} \zeta_{\kappa} r \mathbf{E}(|Y|) > \varepsilon/4$  for all sufficiently large  $\kappa$ . Moreover,

$$\begin{aligned} \mathbf{P} \left[ 2C_{F1} \zeta_{\kappa} r n^{-1} \sum_{i=1}^n |Y_i| > \varepsilon/2 \right] &\leq \mathbf{P} \left[ 2C_{F1} \zeta_{\kappa} r n^{-1} \sum_{i=1}^n (|Y_i| - \mathbf{E}|Y|) > \varepsilon/4 \right] \\ &\leq 2 \exp(-a_1 n \varepsilon^2 \zeta_{\kappa}^2) \end{aligned}$$

for some constant  $a_1 > 0$  and all sufficiently large  $\kappa$  by Bernstein's inequality (Bosq 1998, p. 22). Also by Bernstein's inequality, there is a constant  $a_2 > 0$  such that

$$\mathbf{P} [|\tilde{S}_{n\kappa 1}(\theta_{\kappa j})| > \varepsilon/2] \leq 2 \exp(-a_2 n \varepsilon^2)$$

for all  $n$ ,  $\kappa$ , and  $j$ . Therefore,

$$\begin{aligned} \mathbf{P}_n &\leq 2[M \exp(-a_2 n \varepsilon^2) + \exp(-a_1 n \varepsilon^2)] \\ &\leq 2 \exp\{-a_2 n \varepsilon^2 \zeta_{\kappa}^2 + 2dC_{\kappa} n^{\gamma} [\log(C_{\theta}/r) + (1/2) \log(2C_{\kappa} d) + (1/2) \gamma \log n]\} \\ &\quad + 2 \exp(-a_1 n \varepsilon^2), \end{aligned}$$

where  $\gamma = 4/15 + \nu$ . It follows that  $\mathbf{P}_n \leq 4 \exp(-an \varepsilon^2)$  for a suitable  $a > 0$  and all sufficiently large  $n$ . Q.E.D.

$$\text{Define } S_{\kappa}(\theta) = \mathbf{E}[S_{n\kappa}(\theta)]$$

and

$$\tilde{\theta}_{\kappa} = \arg \min_{\theta \in \Theta_{\kappa}} S_{\kappa}(\theta).$$

**Lemma 2:** For any  $\eta > 0$ ,  $S_{\kappa}(\hat{\theta}_{n\kappa}) - S_{\kappa}(\tilde{\theta}_{\kappa}) < \eta$  almost surely for all sufficiently large  $n$ .

**Proof:** For each  $\kappa$ , let  $\mathcal{N}_{\kappa} \subset \mathbb{R}^{d(\kappa)}$  be an open set containing  $\tilde{\theta}_{\kappa}$ . Let  $\bar{\mathcal{N}}_{\kappa}$  denote the complement of  $\mathcal{N}_{\kappa}$  in  $\Theta_{\kappa}$ . Define  $T_{\kappa} = \bar{\mathcal{N}}_{\kappa} \cap \tilde{\Theta}_{\kappa}$ . Then  $T_{\kappa} \subset \mathbb{R}^{d(\kappa)}$  is compact. Define

$$\eta = \min_{\theta \in T_{\kappa}} S_{\kappa}(\theta) - S_{\kappa}(\tilde{\theta}_{\kappa}).$$

Let  $A_n$  be the event  $|S_{n\kappa}(\theta) - S_{\kappa}(\theta)| < \eta/2$  for all  $\theta \in \Theta_{\kappa}$ . Then

$$A_n \Rightarrow S_{\kappa}(\hat{\theta}_{n\kappa}) < S_{n\kappa}(\hat{\theta}_{n\kappa}) + \eta/2$$

and

$$A_n \Rightarrow S_{n\kappa}(\tilde{\theta}_\kappa) < S_\kappa(\tilde{\theta}_\kappa) + \eta/2.$$

But  $S_{n\kappa}(\hat{\theta}_{n\kappa}) \leq S_{n\kappa}(\tilde{\theta}_\kappa)$  by definition, so

$$A_n \Rightarrow S_\kappa(\hat{\theta}_{n\kappa}) < S_{n\kappa}(\tilde{\theta}_\kappa) + \eta/2.$$

Therefore,

$$A_n \Rightarrow S_\kappa(\hat{\theta}_{n\kappa}) < S_\kappa(\tilde{\theta}_\kappa) + \eta \Rightarrow S_\kappa(\hat{\theta}_{n\kappa}) - S_\kappa(\tilde{\theta}_\kappa) < \eta.$$

So  $A_n \Rightarrow \hat{\theta}_{n\kappa} \in \mathcal{N}_\kappa$ . Since  $\mathcal{N}_\kappa$  is arbitrary, the result follows from Lemma 1 and Theorem 1.3.4 of Serfling (1980, p. 10). Q.E.D.

Define  $b_\kappa(x) = \mu + m(x) - P_\kappa(x)' \theta_\kappa$  and

$$S_{\kappa 0}(\theta) = \mathbf{E}\{Y - F[P_\kappa(X)' \theta + b_\kappa(X)]\}^2.$$

Then

$$\theta_\kappa = \arg \min_{\theta \in \Theta_\kappa} S_{\kappa 0}(\theta).$$

**Lemma 3:** For any  $\eta > 0$ ,  $S_{\kappa 0}(\tilde{\theta}_\kappa) - S_{\kappa 0}(\theta_\kappa) < \eta$  for all sufficiently large  $n$ .

**Proof:** Observe that  $|S_\kappa(\theta) - S_{\kappa 0}(\theta)| \rightarrow 0$  as  $n \rightarrow \infty$  uniformly over  $\theta \in \Theta_\kappa$  because  $b_\kappa(x) \rightarrow 0$  for almost every  $x \in \mathcal{X}$ . For each  $\kappa$ , let  $\mathcal{N}_\kappa \subset \mathbb{R}^{d(\kappa)}$  be an open set containing  $\theta_{\kappa 0}$ . Define  $T_\kappa = \bar{\mathcal{N}}_\kappa \cap \Theta_\kappa$ . Then  $T_\kappa \subset \mathbb{R}^{d(\kappa)}$  is compact. Define

$$\eta = \min_{\theta \in T_\kappa} S_{\kappa 0}(\theta) - S_{\kappa 0}(\theta_{\kappa 0}).$$

Choose  $n$  and, therefore,  $\kappa$  large enough that  $|S_\kappa(\theta) - S_{\kappa 0}(\theta)| < \eta/2$  for all  $\theta \in \Theta$ . Now proceed as in the proof of Lemma 2. Q.E.D.

Define  $Z_{\kappa i} = F'[\mu + m(X_i)]P_\kappa(X_i)$  and  $\hat{Q}_\kappa = n^{-1} \sum_{i=1}^n Z_{\kappa i} Z_{\kappa i}'$ . Then  $Q_\kappa = \mathbf{E} \hat{Q}_\kappa$ . Let  $Z_{\kappa i}^k$  ( $k=1, \dots, d(\kappa)$ ) denote the  $k$ 'th component of  $Z_{\kappa i}$ . Let  $Z_\kappa$  denote the  $n \times d(\kappa)$  matrix whose  $(i, k)$  element is  $Z_{\kappa i}^k$ .

**Lemma 4:**  $\|\hat{Q}_\kappa - Q_\kappa\|^2 = O_p(\kappa^2/n)$ .

**Proof:** Let  $Q_{ij}$  denote the  $(i, j)$  element of  $Q_\kappa$ . Then

$$\mathbf{E} \|\hat{Q}_\kappa - Q_\kappa\|^2 = \sum_{k=1}^{d(\kappa)} \sum_{j=1}^{d(\kappa)} \mathbf{E} \left( n^{-1} \sum_{i=1}^n Z_{\kappa i}^k Z_{\kappa i}^j - Q_{kj} \right)^2$$

$$\begin{aligned}
&= \sum_{k=1}^{d(\kappa)} \sum_{j=1}^{d(\kappa)} \left( \mathbf{E} n^{-2} \sum_{i=1}^n \sum_{\ell=1}^n Z_{\kappa i}^k Z_{\kappa i}^j Z_{\kappa \ell}^k Z_{\kappa \ell}^j - Q_{kj}^2 \right) \\
&= \sum_{k=1}^{d(\kappa)} \sum_{j=1}^{d(\kappa)} \mathbf{E} n^{-2} \sum_{i=1}^n \left( Z_{\kappa i}^k \right)^2 \left( Z_{\kappa i}^j \right)^2 - n^{-1} \sum_{k=1}^{d(\kappa)} \sum_{j=1}^{d(\kappa)} Q_{kj}^2 \\
&\leq n^{-1} \mathbf{E} \left[ \sum_{k=1}^{d(\kappa)} \left( Z_{\kappa i}^k \right)^2 \sum_{j=1}^{d(\kappa)} \left( Z_{\kappa i}^j \right)^2 \right] = O(\kappa/n).
\end{aligned}$$

The lemma now follows from Markov's inequality. Q.E.D.

Define  $\gamma_n = I(\lambda_{\kappa, \min} \geq c_\lambda/2)$ , where  $I$  is the indicator function. Let  $\bar{U} = (U_1, \dots, U_n)'$ .

**Lemma 5:**  $\gamma_n \left\| \hat{Q}_\kappa^{-1} Z'_\kappa \bar{U} / n \right\| = O_p(\kappa^{1/2}/n^{1/2})$  as  $n \rightarrow \infty$ .

**Proof:** For any  $x \in \mathcal{X}$ ,

$$\begin{aligned}
n^{-2} \mathbf{E} \left( \gamma_n \left\| \hat{Q}_\kappa^{-1/2} Z'_\kappa \bar{U} \right\|^2 \middle| X = x \right) &= n^{-2} \gamma_n \mathbf{E} \left( \bar{U}' Z_\kappa \hat{Q}_\kappa^{-1} Z_\kappa \bar{U} \middle| X = x \right) \\
&= n^{-2} \mathbf{E} \left[ \text{Trace} \left( Z_\kappa \hat{Q}_\kappa^{-1} Z'_\kappa \bar{U} \bar{U}' \right) \middle| X = x \right] \\
&\leq n^{-2} \gamma_n C_V \text{Trace} \left( \hat{Q}_\kappa^{-1} Z_\kappa Z'_\kappa \right) \\
&= n^{-1} C_V \gamma_n d(\kappa) \leq C \kappa / n
\end{aligned}$$

for some constant  $C < \infty$ . Therefore,  $\gamma_n \left\| \hat{Q}_\kappa^{-1/2} Z'_\kappa \bar{U} / n \right\| = O_p(\kappa^{1/2}/n^{1/2})$  by Markov's inequality. Now

$$\gamma_n \left\| \hat{Q}_\kappa^{-1} Z'_\kappa \bar{U} / n \right\| = \gamma_n \left[ (\bar{U}' Z_\kappa / n) \hat{Q}_\kappa^{-1/2} \hat{Q}_\kappa^{-1} \hat{Q}_\kappa^{-1/2} (Z'_\kappa \bar{U} / n) \right]^{1/2}.$$

Define  $\xi = \hat{Q}_\kappa^{-1/2} Z'_\kappa \bar{U} / n$ . Let  $\eta_1, \dots, \eta_{d(\kappa)}$  and  $q_1, \dots, q_{d(\kappa)}$  denote the eigenvalues and eigenvectors of  $\hat{Q}_\kappa^{-1}$ . Let  $\eta_{\max} = \max(\eta_1, \dots, \eta_{d(\kappa)})$ . The spectral decomposition of  $\hat{Q}_\kappa^{-1}$  gives

$$\hat{Q}_\kappa^{-1} = \sum_{\ell=1}^{d(\kappa)} \eta_\ell q_\ell q_\ell', \text{ so}$$

$$\begin{aligned} \gamma_n \left\| \hat{Q}_\kappa^{-1} Z'_\kappa \bar{U} / n \right\|^2 &= \gamma_n \sum_{\ell=1}^{d(\kappa)} \eta_\ell \xi' q_\ell q'_\ell \xi \\ &\leq \gamma_n \eta_{\max} \sum_{\ell=1}^{d(\kappa)} \xi' q_\ell q'_\ell \xi \leq \gamma_n \eta_{\max} \xi' \xi = O_p(\kappa/n). \end{aligned}$$

Q.E.D.

Define

$$B_n = \hat{Q}_\kappa^{-1} n^{-1} \sum_{i=1}^n F'[\mu + m(X_i)] Z_{\kappa i} b_{\kappa 0}(X_i).$$

**Lemma 6:**  $\|B_n\| = O(\kappa^{-2})$  with probability approaching 1 as  $n \rightarrow \infty$ .

**Proof:** Let  $\xi$  be the  $n \times 1$  vector whose  $i$ 'th component is  $F'[\mu + m(X_i)] b_{\kappa 0}(X_i)$ . Then  $B_n = \hat{Q}_\kappa^{-1} Z'_\kappa \xi / n$ , and  $\gamma_n \|B_n\|^2 = n^{-2} \gamma_n \xi' Z'_\kappa \hat{Q}_\kappa^{-2} Z'_\kappa \xi$ . Therefore, by the same arguments used to prove Lemma 5,  $\gamma_n \|B_n\|^2 \leq C n^{-1} \gamma_n \xi' \xi = \gamma_n O(\kappa^{-4})$ . The lemma follows from the fact that  $P(\gamma_n = 1) \rightarrow 1$  as  $n \rightarrow \infty$ . Q.E.D.

**Proof of Theorem 1:** To prove Part (a), write

$$(7.1) \quad S_{\kappa 0}(\hat{\theta}_{n\kappa}) - S_{\kappa 0}(\theta_\kappa) =$$

$$[S_{\kappa 0}(\hat{\theta}_{n\kappa}) - S_\kappa(\hat{\theta}_{n\kappa})] + [S_\kappa(\hat{\theta}_{n\kappa}) - S_\kappa(\tilde{\theta}_\kappa)] + [S_\kappa(\tilde{\theta}_\kappa) - S_{\kappa 0}(\tilde{\theta}_\kappa)] + [S_{\kappa 0}(\tilde{\theta}_\kappa) - S_{\kappa 0}(\theta_\kappa)].$$

Given any  $h > 0$ , it follows from Lemmas 2-3 and uniform convergence of  $S_\kappa$  to  $S_{\kappa 0}$  that each term on the right-hand side of (4.1) is less than  $\eta/4$  almost surely for all sufficiently large  $n$ . Therefore  $S_{\kappa 0}(\hat{\theta}_{n\kappa}) - S_{\kappa 0}(\theta_\kappa) < \eta$  almost surely for all sufficiently large  $n$ . It follows that  $\|\hat{\theta}_{n\kappa} - \theta_\kappa\| \rightarrow 0$  almost surely as  $n \rightarrow \infty$  because  $\theta_\kappa$  uniquely minimizes  $S_\kappa$ . Part (a) follows because uniqueness of the series representation of each function  $m_j$  implies that  $\|\theta_\kappa - \theta_{\kappa 0}\| \rightarrow 0$  as  $n \rightarrow \infty$ .

To prove the remaining parts of the theorem, observe that  $\hat{\theta}_{n\kappa}$  satisfies the first-order condition  $\partial S_{n\kappa}(\hat{\theta}_{n\kappa}) / \partial \theta = 0$  almost surely for all sufficiently large  $n$ . Define  $M_i = \mu + m(X_i)$  and  $\Delta M_i = P_\kappa(X_i)' \hat{\theta}_{n\kappa} - M_i = P_\kappa(X_i)' (\hat{\theta}_{n\kappa} - \theta_{\kappa 0}) - b_{\kappa 0}(X_i)$ . Then a Taylor series expansion yields

$$n^{-1} \sum_{i=1}^n Z_{\kappa i} U_i - (\hat{Q}_\kappa + R_{n1})(\hat{\theta}_{n\kappa} - \theta_{\kappa 0}) + n^{-1} \sum_{i=1}^n F'(M_i) Z_{\kappa i} b_{\kappa 0}(X_i) + R_{n2} = 0,$$

almost surely for all sufficiently large  $n$ .  $R_{n1}$  is defined by

$$\begin{aligned} R_{n1} = & n^{-1} \sum_{i=1}^n \{-U_i F''(M_i) + U_i [F''(\tilde{M}_i) - F''(M_i)] + [(3/2)F''(\tilde{M}_i)F'(M_i) \\ & + (1/2)F''(\tilde{M}_i)F''(\tilde{M}_i)\Delta M_i + (1/2)F''(\tilde{M}_i)F''(\tilde{M}_i)(\Delta M_i)^2] \Delta M_i - \\ & [F''(\tilde{M}_i)F'(M_i) - (1/2)F''(\tilde{M}_i)F'(M_i) + F''(\tilde{M}_i)F''(\tilde{M}_i)b_{\kappa 0}(X_i)]b_{\kappa 0}(X_i)\} P_{\kappa}(X_i)P_{\kappa}(X_i)', \end{aligned}$$

where  $\tilde{M}_i$  and  $\tilde{M}_i$  are points between  $P_{\kappa}(X_i)'\hat{\theta}_{n\kappa}$  and  $M_i$ .  $R_{n2}$  is defined by

$$\begin{aligned} R_{n2} = & -n^{-1} \sum_{i=1}^n \{U_i F''(\tilde{M}_i) + U_i [F''(\tilde{M}_i) - F''(M_i)] \\ & + [F''(\tilde{M}_i)F'(M_i) - (1/2)F''(\tilde{M}_i)F'(M_i)]b_{\kappa 0}(X_i) - (1/2)F''(\tilde{M}_i)F''(\tilde{M}_i)b_{\kappa 0}(X_i)^2\} b_{\kappa 0}(X_i). \end{aligned}$$

Now let  $\xi$  be any  $d(\kappa) \times 1$  vector. Then

$$\begin{aligned} \gamma_n \left\| [(\hat{Q}_{\kappa} + R_{n1})^{-1} - \hat{Q}_{\kappa}^{-1}] \hat{Q}_{\kappa} \xi \right\|^2 &= \gamma_n \left\| (\hat{Q}_{\kappa} + R_{n1})^{-1} R_{n1} \xi \right\|^2 \\ &= \text{Trace} \{ \gamma_n [\xi' R_{n1} (\hat{Q}_{\kappa} + R_{n1})^{-2} R_{n1} \xi] \} \\ &\leq O_p(\kappa) (\xi' \xi) \|R_{n1}\|^2 \\ &= O_p(\kappa) (\xi' \xi) O \left[ \sup_{x \in \mathcal{X}} |P_{\kappa}(x)'(\hat{\theta}_{n\kappa} - \theta_{\kappa 0}) - b_{\kappa 0}(x)|^2 + \sup_{x \in \mathcal{X}} |b_{\kappa 0}(x)|^2 \right] \\ &= (\xi' \xi) O_p \left( \kappa^2 \|\hat{\theta}_{\kappa} - \theta\|^2 + \kappa^{-3} \right). \end{aligned}$$

Setting  $\xi = \hat{Q}_{\kappa}^{-1} Z_{\kappa}' \bar{U} / n$  and applying Lemma 5 yields  $\left\| [(\hat{Q}_{\kappa} + R_{n1})^{-1} - \hat{Q}_{\kappa}^{-1}] Z_{\kappa}' \bar{U} / n \right\|^2 = O_p \left[ (\kappa^3 / n) \|\hat{\theta}_{n\kappa} - \theta_{\kappa 0}\|^2 + 1 / (n\kappa^2) \right]$ . Setting  $\xi = \hat{Q}_{\kappa}^{-1} \left[ n^{-1} \sum_{i=1}^n F'(M_i)^2 P_{\kappa}(X_i) b_{\kappa 0}(X_i) + R_{n2} \right]$ ,

applying Lemma 6, and using the result that  $\|\hat{Q}_{\kappa}^{-1} R_{n2}\| = o_p(\kappa^{-2})$  yields

$$\left\| [(\hat{Q}_{\kappa} + R_{n1})^{-1} - \hat{Q}_{\kappa}^{-1}] \left[ n^{-1} \sum_{i=1}^n F'(M_i) Z_{\kappa i}' b_{\kappa 0}(X_i) + R_{n2} \right] \right\|^2 = O_p \left( \|\hat{\theta}_{n\kappa} - \theta_{\kappa 0}\|^2 / \kappa + 1 / \kappa^5 \right)$$

It follows from these results that

$$\begin{aligned}\hat{\theta}_{n\kappa} - \theta_{\kappa 0} &= n^{-1} \hat{Q}_\kappa^{-1} \sum_{i=1}^n F'[\mu + m(X_i)] P_\kappa(X_i) U_i \\ &\quad + n^{-1} \hat{Q}_\kappa^{-1} \sum_{i=1}^n F'[\mu + m(X_i)]^2 P_\kappa(X_i) b_{\kappa 0}(X_i) + R_n,\end{aligned}$$

where  $\|R_n\| = O_p(\kappa^2/n + n^{-1/2})$ . Part (d) of the theorem now follows from Lemma 4. Part (b) follows by applying Lemmas 5 and 6 to Part (d). Part (c) follows from Part (b) and Assumption A5(iii). Q.E.D.

### b. Theorem 2

This section begins with lemmas that are used to prove Theorem 2. For any  $\tilde{x} \equiv (x^2, \dots, x^d) \in [-1, 1]^{d-1}$ , set  $m_{-1}(\tilde{x}) = m_2(x^2) + \dots + m_d(x^d)$ , and  $\bar{b}_{\kappa 0}(\tilde{x}) = \mu + m_{-1}(\tilde{x}) - \bar{P}_\kappa(\tilde{x})\bar{\theta}_{\kappa 0}$ , where

$$\bar{P}_\kappa(x) = [1, 0, \dots, 0, p_1(x^2), \dots, p_\kappa(x^2), \dots, p_1(x^d), \dots, p_\kappa(x^d)]',$$

and

$$\bar{\theta}_{\kappa 0} = (\mu, 0, \dots, 0, \theta_{21}, \dots, \theta_{2\kappa}, \dots, \theta_{d1}, \dots, \theta_{d\kappa})'.$$

In other words,  $\bar{P}$  and  $\bar{\theta}_{\kappa 0}$  are obtained by replacing  $p_1(x^1), \dots, p_\kappa(x^d)$  with zeros in  $P_\kappa$  and  $\theta_{11}, \dots, \theta_{1\kappa}$  with zeros in  $\theta_{\kappa 0}$ . Also define

$$\delta_{n1}(\tilde{x}) = n^{-1} \bar{P}_\kappa(\tilde{x})' Q_\kappa^{-1} \sum_{j=1}^n F'[\mu + m(X_j)] P_\kappa(X_j) U_j,$$

and

$$\delta_{n2}(\tilde{x}) = n^{-1} \bar{P}_\kappa(\tilde{x})' Q_\kappa^{-1} \sum_{j=1}^n F'[\mu + m(X_j)]^2 P_\kappa(X_j) b_{\kappa 0}(X_j).$$

For  $x^1 \in [-1, 1]$  and  $j = 0, 1$  define

$$H_{nj1}(x^1) = (nh)^{-1/2} \sum_{i=1}^n F'[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)]^2 (X_i^1 - x^1)^j K_h(x^1 - X_i^1) \delta_{n1}(\tilde{X}_i),$$

$$H_{nj2}(x^1) = (nh)^{-1/2} \sum_{i=1}^n F'[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)]^2 (X_i^1 - x^1)^j K_h(x^1 - X_i^1) \delta_{n2}(\tilde{X}_i),$$

and

$$H_{nj3}(x^1) = -(nh)^{-1/2} \sum_{i=1}^n F'[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)]^2 (X_i^1 - x^1)^j K_h(x^1 - X_i^1) \bar{b}_{\kappa 0}(\tilde{X}_i).$$

Let  $V(x) = \text{Var}(U | X = x)$ .

**Lemma 7:** For  $j=0,1$  and  $k=1,2,3$ ,  $H_{nj}(x^1) = o_p(1)$  as  $n \rightarrow \infty$  uniformly over  $x^1 \in [-1,1]$ .

**Proof:** The proof is given only for  $j=0$ . Similar arguments apply with  $j=1$ . First consider  $H_{n01}(x^1)$ . We can write

$$H_{n01}(x^1) = \sum_{j=1}^n a_j(x^1) U_j,$$

where

$$a_j(x^1) =$$

$$\begin{aligned} & n^{-3/2} h^{-1/2} \sum_{i=1}^n F'[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)]^2 K_h(x^1 - X_i^1) \bar{P}_\kappa(\tilde{X}_i)' Q_\kappa^{-1} F'[\mu + m(X_j)] P_\kappa(X_j) \\ & \equiv n^{-3/2} h^{-1/2} \sum_{i=1}^n K_h(x^1 - X_i^1) A_{ij}(x^1). \end{aligned}$$

Using Lipschitz continuity of  $a_j$ , the moment conditions A2(v) on  $U_j$ , and arguments similar to those used to prove Lemma 1, the conclusion of the lemma for  $j=0$ ,  $k=1$  follows from

$$(7.2) \quad \sup_{|x^1| \leq 1} \sum_{j=1}^n a_j(x^1)^2 = o_p(1).$$

To prove (7.2), it suffices to show that

$$(7.3) \quad \sup_{|x^1| \leq 1, 1 \leq j \leq n} |a_{jk}(x^1)| = o_p(n^{-1/2}) \quad (k=1,2),$$

where

$$a_{j1}(x^1) = n^{-3/2} h^{-1/2} K_h(x^1 - X_j^1) A_{jj}(x^1)$$

and

$$a_{j2}(x^1) = n^{-3/2} h^{-1/2} \sum_{\substack{i=1 \\ i \neq j}}^n K_h(x^1 - X_j^1) A_{ij}(x^1).$$

To prove (7.3), note first that by A3(iii), A5(iii), and A6(i),

$$\sup_{|x^1| \leq 1, 1 \leq i, j \leq n} |A_{ij}(x^1)| = O(\kappa) = O(n^{3/10}).$$

This immediately implies (7.3) for  $k=1$ . To prove (7.3) for  $k=2$ , note that for some constant  $C > 0$ ,



$$\sup_{\substack{|x^1| \leq 1 \\ 1 \leq i, j \leq n \\ i \neq j}} |\mathbf{E}[A_{ij}(x^1) | X_i^1]| \leq C$$

and

$$\sup_{\substack{|x^1| \leq 1 \\ 1 \leq i, j \leq n \\ i \neq j}} |\mathbf{E}[A_{ij}(x^1)^2 | X_i^1]| \leq C\kappa.$$

For a proof of these bounds, note that the density of  $X$  is bounded from above and below so that

$$\mathbf{E}\{F'[\mu + m(X_j)]P_\kappa(X_j) | X_i^1\} = \mathbf{E}\{F'[\mu + m(X_j)]P_\kappa(X_j)\}$$

are the Fourier coefficients of a bounded function. Therefore, their sum of squares converges.

Thus, it remains to show that

$$(7.4) \quad \sup_{|x^1| \leq 1, 1 \leq j \leq n} n^{-1}h^{-1/2} \sum_{\substack{i=1 \\ i \neq j}}^n K_h(x^1 - X_i^1) B_{ij}(x^1) = o_p(1),$$

where  $B_{ij}(x^1) = A_{ij}(x^1) - \mathbf{E}[A_{ij}(x^1) | X_i^1]$ . To prove (7.4), note that for  $\delta > 0$ ,

$$\begin{aligned} & \mathbf{P} \left[ n^{-1}h^{-1/2} \sum_{\substack{i=1 \\ i \neq j}}^n K_h(x^1 - X_i^1) B_{ij}(x^1) > \delta \right] \\ & \leq \mathbf{E} \exp \left[ n^{-3/4}h^{-1/2} \sum_{\substack{i=1 \\ i \neq j}}^n K_h(x^1 - X_i^1) B_{ij}(x^1) \right] \exp(-n^{1/4}\delta) \\ & \leq \{1 + n^{-3/2}h^{-1} \mathbf{E} K_h^2(x^1 - X_1^1) B_{12}^2(x^1) \exp[O(n^{-3/4}h^{-1/2}n^{3/10})]\}^n \exp(-n^{1/4}\delta) \\ & \leq \exp[O(n^{-1/2}n^{3/10})] \exp(-n^{1/4}\delta) \\ & = O(1) \exp(-n^{1/4}\delta). \end{aligned}$$

We now prove the lemma for  $j = 0$ ,  $k = 2$ . We can write

$$H_{n02}(x^1) = (nh)^{-1/2} \sum_{i=1}^n F'[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)]^2 K_h(x^1 - X_i^1) \bar{P}_\kappa(\tilde{X}_i)' B_n,$$

where

$$B_n = n^{-1} Q_\kappa^{-1} \sum_{j=1}^n F'[\mu + m(X_j)]^2 P_\kappa(X_j) b_{\kappa 0}(X_j).$$

Arguments like those used to prove Lemma 6 show that  $\mathbf{E} \|B_n\|^2 = O(\kappa^{-4})$ . Therefore,

$$\begin{aligned} \sup_{|x^1| \leq 1} |H_{n02}(x^1)| &= \sup_{|x^1| \leq 1} (nh)^{-1/2} \sum_{i=1}^n K_h(x^1 - X_i^1) \cdot O_p(\kappa^{-3/2}) \\ &= O_p(n^{1/2} h^{1/2} \kappa^{-3/2}) \\ &= o_p(1). \end{aligned}$$

For the proof with  $j=0$ ,  $k=3$ , note that

$$\begin{aligned} \sup_{|x^1| \leq 1} |H_{n03}(x^1)| &= \sup_{|x^1| \leq 1} (nh)^{-1/2} \sum_{i=1}^n K_h(x^1 - X_i^1) \cdot O_p(\kappa^{-2}) \\ &= o_p(1). \end{aligned}$$

Q.E.D.

Let  $P_\kappa(x^1, \tilde{X}) = [1, p_1(x^1), \dots, p_\kappa(x^1), p_1(X^2), \dots, p_\kappa(X^2), \dots, p_1(X^d), \dots, p_\kappa(X^d)]'$  and  $b_\kappa(x^1, \tilde{X}_i) = \mu + m_1(x^1) + m_{-1}(\tilde{X}_i) - P_\kappa(x^1, \tilde{X}_i) \theta_{\kappa 0}$ . Define

$$\delta_{n3}(x^1, \tilde{X}_i) = n^{-1} P_\kappa(x^1, \tilde{X}_i)' Q_\kappa^{-1} \sum_{j=1}^n F'[\mu + m(X_j)] P_\kappa(X_j) U_j$$

and

$$\delta_{n4}(x^1, \tilde{X}_i) = n^{-1} P_\kappa(x^1, \tilde{X}_i)' Q_\kappa^{-1} \sum_{j=1}^n F'[\mu + m(X_j)]^2 P_\kappa(X_j) b_{\kappa 0}(X_j).$$

Also, for  $j=0,1$  define

$$\begin{aligned} L_{nj1}(x^1) &= (nh)^{-1/2} \sum_{i=1}^n U_i F''[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)] (X_i^1 - x^1)^j K_h(x^1 - X_i^1) \delta_{n3}(x^1, \tilde{X}_i), \\ L_{nj2}(x^1) &= (nh)^{-1/2} \sum_{i=1}^n U_i F''[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)] (X_i^1 - x^1)^j K_h(x^1 - X_i^1) \delta_{n4}(x^1, \tilde{X}_i), \\ L_{nj3}(x^1) &= -(nh)^{-1/2} \sum_{i=1}^n U_i F''[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)] (X_i^1 - x^1)^j K_h(x^1 - X_i^1) b_{\kappa 0}(x^1, \tilde{X}_i), \\ L_{nj4}(x^1) &= (nh)^{-1/2} \sum_{i=1}^n \{F[\mu + m_1(X_i^1) + m_{-1}(X_i)] - F[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)]\} \\ &\quad \cdot F''[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)] (X_i^1 - x^1)^j K_h(x^1 - X_i^1) \delta_{n3}(x^1, \tilde{X}_i), \end{aligned}$$

$$L_{nj5}(x^1) = (nh)^{-1/2} \sum_{i=1}^n \{F[\mu + m_1(X_i^1) + m_{-1}(X_i)] - F[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)]\}$$

$$\cdot F''[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)](X_i^1 - x^1)^j K_h(x^1 - X_i^1) \delta_{n4}(x^1, \tilde{X}_i),$$

and

$$L_{nj6}(x^1) = -(nh)^{-1/2} \sum_{i=1}^n \{F[\mu + m_1(X_i^1) + m_{-1}(X_i)] - F[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)]\}$$

$$\cdot F''[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)](X_i^1 - x^1)^j K_h(x^1 - X_i^1) b_{\kappa 0}(x^1, \tilde{X}_i).$$

**Lemma 8:** As  $n \rightarrow \infty$ ,  $L_{nj6}(x^1) = o_p(1)$  uniformly over  $x^1 \in [-1, 1]$  for each  $j = 0, 1$ ;

$k = 1, \dots, 6$ .

Proof: The proof is given only for  $j = 0$ . The arguments are similar for  $j = 1$ . By Theorem 1,  $\delta_{n4}(x^1, \tilde{X}_i)$  is the asymptotic bias component of the stochastic expansion of  $P_\kappa(x^1, \tilde{X}_i)(\hat{\theta}_{n\kappa} - \theta_{\kappa 0})$  and is  $O_p(\kappa^{-3/2})$  uniformly over  $(x^1, \tilde{X}_i) \in [-1, 1]^d$ . This fact and standard bounds on

$$\sup_{|x^1| \leq 1} \sum_{i=1}^n |U_i| K_h(x^1 - X_i^1)$$

and

$$\sup_{|x^1| \leq 1} \sum_{i=1}^n K_h(x^1 - X_i^1)$$

establish the conclusion of the lemma for  $j = 0$ ,  $k = 2, 5$ . For  $j = 0$ ,  $k = 3, 6$ , proceed similarly using

$$\sup_{|x^1| \leq 1} |b_{\kappa 0}(x^1)| = O(\kappa^{-2}).$$

For  $j = 0$ ,  $k = 5$ , one can use arguments similar to those made for  $H_{n01}(x^1)$  in the proof of

Lemma 7. It remains to consider  $L_{n01}(x^1) = D_n(x^1)B_n$ , where

$$D_n(x^1) = (nh)^{-1/2} \sum_{i=1}^n U_i F''[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)] K_h(x^1 - X_i^1) P_\kappa(x^1, \tilde{X}_i)$$

and

$$B_n = n^{-1} Q_\kappa^{-1} \sum_{j=1}^n F'[\mu + m(X_j)]^2 P_\kappa(X_j) U_j.$$

Now,  $\mathbf{E}\|B_n\|^2 = O(\kappa n^{-1})$ , and  $D_n(x^1)$  contains elements of the form

$$(nh)^{-1/2} p_r(x^1) \sum_{i=1}^n U_i F''[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)] K_h(x^1 - X_i^1)$$

and

$$(nh)^{-1/2} \sum_{i=1}^n U_i p_r(X_i^\ell) F''[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)] K_h(x^1 - X_i^1)$$

for  $0 \leq r \leq \kappa$ ,  $2 \leq \ell \leq d$ . These expressions can be bounded uniformly over  $|x^1| \leq 1$  by terms that are  $O_p[(\log n)^{1/2} |p_r(x^1)|]$  and  $O_p[(\log n)^{1/2}]$ , respectively. This gives

$$\sup_{|x^1| \leq 1} \|D_n(x^1)\|^2 = O_p(\kappa \log n).$$

Therefore,

$$\sup_{|x^1| \leq 1} |L_{n01}(x^1)|^2 \leq \sup_{|x^1| \leq 1} \|D_n(x^1)\|^2 \|B_n\|^2 = o_p(1).$$

Q.E.D.

**Lemma 9:** The following hold uniformly over  $|x^1| \leq 1 - h$ :

$$(nh)^{-1} S''_{n01}(x^1, \tilde{m}) = D_0(x^1) + o_p(1),$$

$$(nh)^{-1} S''_{n21}(x^1, \tilde{m}) = A_K h^2 D_0(x^1) [1 + o_p(1)],$$

and

$$(nh)^{-1} S''_{n11}(x^1, \tilde{m}) = h^2 A_K D_1(x^1) [1 + o_p(1)].$$

**Proof:** This follows from the Theorem 1(c) and standard bounds on

$$\sup_{|x^1| \leq 1} \sum_{i=1}^n U_i^r (X_i^1 - x^1)^s K_h(x^1 - X_i^1)$$

for  $r = 0, 1$ ;  $s = 0, 1, 2$ . Q.E.D.

**Lemma 10:** The following hold uniformly over  $|x^1| \leq 1 - h$ :

$$(a) \quad (nh)^{-1/2} S'_{n01}(x^1, \tilde{m}) = (nh)^{-1/2} S'_{n01}(x^1, m) + (nh)^{1/2} D_0(x^1) \Delta m_1(x^1) + o_p(1),$$

and

$$(b) \quad (nh)^{-1/2} S'_{n11}(x^1, \tilde{m}) = (nh)^{-1/2} S'_{n11}(x^1, m) + o_p(1).$$

**Proof:** Only (a) is proved. The proof of (b) is similar. For each  $i = 1, \dots, n$ , let  $m^*(x^1, \tilde{X}_i)$  and  $m^{**}(x^1, \tilde{X}_i)$  denote quantities that are between  $\tilde{\mu} + \tilde{m}_1(x^1) + \tilde{m}_{-1}(\tilde{X}_i)$  and

$\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)$ . The values of  $m^*(x^1, \tilde{X}_i)$  and  $m^{**}(x^1, \tilde{X}_i)$  may be different in different uses. A Taylor series expansion and Theorem 1(c) give

$$\begin{aligned} (nh)^{-1/2} S'_{n01}(x^1, \tilde{m}) &= (nh)^{-1/2} S'_{n01}(x^1, m) + \sum_{j=1}^4 J_{nj}(x^1) + n(nh)^{-1/2} O \left[ \sup_{(x^1, \tilde{x}) \in \mathcal{X}} \|\Delta m(x^1, \tilde{x})\|^3 \right] \\ &= (nh)^{-1/2} S'_{n01}(x^1, m) + \sum_{j=1}^4 J_{nj}(x^1) + o_p(1) \end{aligned}$$

uniformly over  $|x^1| \leq 1-h$ , where

$$\begin{aligned} J_{n1}(x^1) &= 2(nh)^{-1/2} \sum_{i=1}^n F'[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)]^2 K_h(x^1 - X_i^1) \Delta m_1(x^1), \\ J_{n2}(x^1) &= 2(nh)^{-1/2} \sum_{i=1}^n F'[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)]^2 K_h(x^1 - X_i^1) \Delta m_{-1}(\tilde{X}_i), \\ J_{n3}(x^1) &= \\ &\quad - 2(nh)^{-1/2} \sum_{i=1}^n \{Y_i - F[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)]\} F''[m^*(x^1, \tilde{X}_i)] K_h(x^1 - X_i^1) \Delta m(x^1, \tilde{X}_i), \\ J_{n4}(x^1) &= 2(nh)^{-1/2} \sum_{i=1}^n F'[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)] \{F''[m^*(x^1, \tilde{X}_i)] + 2F''[m^{**}(x^1, \tilde{X}_i)]\} \\ &\quad \cdot K_h(x^1 - X_i^1) [\Delta m(x^1, \tilde{X}_i)]^2. \end{aligned}$$

It follows from Theorem 1(d) and Lemma 7 that  $J_{n2}(x^1) = \sum_{k=1}^3 H_{n0k}(x^1) + o_p(1) = o_p(1)$

uniformly over  $|x^1| \leq 1-h$ . In addition, it follows from Theorem 1(c) that for some constant  $C < \infty$

$$\begin{aligned} J_{n4}(x^1) &< C(nh)^{-1/2} \sum_{i=1}^n K_h(x^1 - X_i^1) \left[ \sup_{(x^1, \tilde{x}) \in \mathcal{X}} \|\Delta m(x^1, \tilde{x})\|^2 \right] \\ &= O_p \left[ (nh)^{1/2} \sup_{(x^1, \tilde{x}) \in \mathcal{X}} \|\Delta m(x^1, \tilde{x})\|^2 \right] = o_p(1) \end{aligned}$$

uniformly over  $|x^1| \leq 1-h$ . Now consider  $J_{n3}(x^1)$ . It follows from Assumption A3(v) that

$$\begin{aligned}
J_{n3}(x^1) &= \\
&- 2(nh)^{-1/2} \sum_{i=1}^n \{Y_i - F[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)]\} F''[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)] K_h(x^1 - X_i^1) \Delta m(x^1, \tilde{X}_i) \\
&+ O_p \left[ (nh)^{1/2} \sup_{x \in \mathcal{X}} |\Delta m(x^1, \tilde{x})|^{1+s} \right] \\
&= \sum_{k=1}^6 L_{n0k}(x^1) + O_p \left[ (nh)^{1/2} \sup_{x \in \mathcal{X}} |\Delta m(x^1, \tilde{x})|^{1+s} \right] + o_p(1)
\end{aligned}$$

uniformly over  $|x^1| \leq 1-h$ . Therefore,  $J_{n3}(x^1) = o_p(1)$  uniformly by Lemma 8 and Theorem 1(c), and

$$(nh)^{-1/2} S'_{n01}(x^1, \tilde{m}) = (nh)^{-1/2} S'_{n01}(x^1, m) + J_{n1}(x^1) + o_p(1)$$

uniformly over  $|x^1| \leq 1-h$ .

Now consider  $J_{n1}(x^1)$ . Set

$$\tilde{J}_{n1}(x^1) = 2(nh)^{-1/2} \sum_{i=1}^n F'[\mu + m_1(x^1) + m_{-1}(\tilde{X}_i)]^2 K_h(x^1 - X_i^1).$$

It follows from Theorem 2.37 of Pollard (1984) that  $\tilde{J}_{n1}(x^1) - E[\tilde{J}_{n1}(x^1)] = o(\log n)$  almost surely as  $n \rightarrow \infty$ . In addition,  $E[(nh)^{-1/2} \tilde{J}_{n1}(x^1)] = D(x^1) + O(h^2)$ . Therefore,

$$\begin{aligned}
J_{n1}(x^1) &= (nh)^{1/2} D(x^1) \Delta m_1(x^1) + O[\log n \Delta m_1(x^1)] \\
&= (nh)^{1/2} D(x^1) \Delta m_1(x^1) + o_p(1)
\end{aligned}$$

uniformly over  $|x^1| \leq 1-h$ . Q.E.D.

Proof of Theorem 2: By the definition of  $\hat{m}_1(x^1)$ ,

$$(4.3) \quad \hat{m}_1(x^1) - m_1(x^1) = \tilde{m}_1(x^1) - m_1(x^1) - \frac{S''_{n21}(x^1, \tilde{m}) S'_{n01}(x^1, \tilde{m}) - S''_{n11}(x^1, \tilde{m}) S'_{n11}(x^1, \tilde{m})}{S''_{n01}(x^1, \tilde{m}) S''_{n21}(x^1, \tilde{m}) - S''_{n11}(x^1, \tilde{m})^2}.$$

Part (a) follows by applying Lemmas 9 and 10 to the right-hand side of (4.3). Define

$$w = [nhD_0(x^1)]^{-1} \{-S'_{n01}(x^1, m) + [D_1(x^1)/D_0(x^1)] S'_{n11}(x^1, m)\}.$$

Methods identical to those used to establish asymptotic normality of local linear estimators show that  $\mathbf{E}(n^{2/5}w) = \beta_1 + o(1)$ ,  $\text{Var}(n^{2/5}w) = V_1(x^1) + o(1)$ , and  $n^{2/5}[\hat{m}_1(x^1) - m_1(x^1)]$  is asymptotically normal, which proves Part (b). Q.E.D.

Proof of Theorem 4: It follows from Theorem 2(a) that

$$n^{4/5} \int_{1-h \leq |x^1| \leq 1} w(x^1) [\hat{m}_1(x^1) - m_1(x^1)]^2 dx^1 = o_p(1).$$

Now consider

$$n^{4/5} \int_{|x^1| \leq 1-h} w(x^1) [\hat{m}_1(x^1) - m_1(x^1)]^2 dx^1.$$

By replacing the integrand with the expansion of Theorem 2(a), one obtains a  $U$ -statistic in  $U_i$  conditional on  $X_1, \dots, X_n$ . This  $U$  statistic has vanishing conditional variance. Q.E.D.

Proof of Theorem 5: Use Theorem 2(a) to replace  $\hat{m}_1$  with  $m_1$  in the expression for  $\hat{m}_1^{(\ell)}$ . The result now follows from standard methods for bounding kernel estimators. Q.E.D.

## REFERENCES

- Breiman, L. and Friedman, J.H. (1985). Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80, 580-598.
- Buja, A., Hastie, T. and Tibshirani, R.J. (1989). Linear smoothers and additive models. *Annals of Statistics*, 17, 453-510.
- Chen, R., Härdle, W., Linton, O.B., and Severance-Lossin, E. (1996). Estimation in additive nonparametric regression, in *Proceedings of the COMPSTAT Conference Semmering*, ed. by W. Härdle and M. Schimek, Heidelberg: Physika Verlag.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, London: Chapman & Hall.
- Fan, J., Härdle, W., and Mammen, E. (1998). Direct estimation of low-dimensional components in additive models. *Annals of Statistics*, 26, 943-971.
- Hastie, T.J. and Tibshirani, R.J. (1990). *Generalized Additive Models*, London: Chapman & Hall.
- Horowitz, J.L., Klemelä, J. and Mammen, E. (2002). Optimal estimation in additive regression models, working paper, Institute für Angewandte Mathematik, Ruprecht-Karls-Universität Heidelberg, Germany.
- Linton, O.B. and Härdle, W. (1996). Estimating additive regression with known links. *Biometrika*, 83, 529-540.
- Linton, O.B. and Nielsen, J.P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, 82, 93-100.
- Mammen, E., Linton, O.B., and Nielsen, J.P. (1999). The existence and asymptotic properties of backfitting projection algorithm under weak conditions. *Annals of Statistics*, 27, 1443-1490.
- Newey, W.K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79, 147-168.
- Opsomer, J.D. (2000). Asymptotic properties of backfitting estimators. *Journal of Multivariate Analysis*, 73, 166-179.
- Opsomer, J.D. and Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *Annals of Statistics*, 25, 186-211.
- Stone, C.J. (1985). Additive regression and other nonparametric models. *Annals of Statistics*, 13, 689-705.
- Stone, C.J. (1986). The dimensionality reduction principle for generalized additive models. *Annals of Statistics*, 14, 590-606.



Tjøstheim, D. and Auestad, B.H. (1994). Nonparametric identification of nonlinear time series: projections. *Journal of the American Statistical Association*, 89, 1398-1409.

TABLE 1: Results of Monte Carlo Experiments<sup>1</sup>

Estimator	$\kappa_1$	$\kappa_2$	$h_1$	$h_2$	<u>Empirical IMSE</u>	
					$f_1$	$f_2$
$d = 2$						
LH			.9	.9	.116	.015
2-Stage with Local Constant Smoothing	2	2	.4	.9	.052	.015
2-Stage with Local Linear Smoothing	4	2	.5	1.4	.052	.023
$d = 5$						
LH			1.0	1.0	.145	.019
2-Stage with Local Constant Smoothing	2	2	.4	.9	.060	.018
2-Stage with Local Linear Smoothing	2	2	.6	1.3	.057	.029

<sup>1</sup> In the two-stage estimator,  $\kappa_j$  and  $h_j$  ( $j=1,2$ ) are the series length and bandwidth used to estimate  $f_j$ . In the LH estimator,  $h_j$  ( $j=1,2$ ) is the bandwidth used to estimate  $f_j$ . The values of  $\kappa_1, \kappa_2, h_1$ , and  $h_2$  minimize the IMSE's of the estimates.