Nonparametric Estimation of an Additive Model with Cross-Terms and a Known Link Function

Maria Ponomareva*
Northwestern University

December 7, 2009

Abstract

In this paper I extend the two-step estimator of the additive nonparametric model with a known link function proposed in Horowitz and Mammen (2004) to cover the additive models with multiplicative interaction terms. I find the same rate of convergence ($n^{2/5}$) for estimators of both the univariate additive part and the multiplicative interaction part. I show that this convergence rate does not depend on the dimension of the vector of covariates.

1 Introduction

Additive models with multiplicative interaction terms provide the nonparametric extension to the simple linear regression model that includes interactions between the covariates. These models are useful in economic analysis since they allow to capture the cross-factor effects that are common in e.g. the production function analysis or demand analysis. However, the majority of the literature that studies those types of nonparametric models focuses on pure additive models that do not include any interaction terms (see e.g. Linton and Nielsen (1995), Linton (2000), or Horowitz and Mammen (2004)). In this paper I fill this gap and provide a nonparametric estimator of the additive univariate terms and two-factor interaction terms in the additive model with a

*I am grateful to Joel Horowitz for his continuous encouragement and support. I thank Maria Goltsman and Eugene Orlov for valuable comments and suggestions. All remaining errors are mine.

known link function. I extend two-step estimator of Horowitz and Mammen (2004) to cover additive models with bi-factor interaction terms and show that it possesses similar properties. When both additive univariate components and cross-term univariate components are twice continuously differentiable, the estimator is asymptotically normally distributed with the convergence rate $n^{2/5}$ regardless of the dimension of the vector of covariates. The estimator discussed in this paper can be easily adjusted to cover the case where the multiplicative cross-terms are composed of more than two univariate factors while preserving the convergence rate. In particular, adding factors to the multiplicative terms affects the the rate of the second-order term of the asymptotic expansion of the estimator discussed in this paper, but it does not affect the rate of the first-order term.

The rest of the paper is organized as follows. In section 2 I formally state the problem, provide the conditions for identification and describe the two-step estimator of the univariate components of the additive model with bi-factor cross-terms. In section I present theoretical results that supports the validity of this two-step procedure. I illustrate the behavior of the estimator with two small simulation exercises in section 5. Finally, section 6 concludes. All proofs are collected in the Appendix.

2 The Model

Consider the following model:

$$Y_i = F\left(\mu + \sum_{j=1}^{d} m_j(X_i^j) + \sum_{k,l=1,k<l}^{d} h_{1k}(X_i^k)h_{2l}(X_i^l)\right) + U_i, \qquad (1)$$

where $E[U_i|X_i] = 0$, $X_i \in \mathbb{R}^d$ is a continuous random vector with density $f(x)$, and F is a known link function. Throughout the paper I assume that density function $f(\cdot)$ has a compact support. The goal is to estimate unknown additive and multiplicative components: μ, $\{m_j, j = 1, \ldots, d\}$, and $\{h_{1k}, h_{2l}, k, l = 1, \ldots, d, k < l\}$ based on the sample $\{(Y_i, X_i), i = 1, \ldots, n\}$. For the ease of the presentation of the results, in this paper we will deal with a simplification of the model in (1) that includes only a single cross-term component:

$$Y_i = F\left(\mu + \sum_{j=1}^{d} m_j(X_i^j) + h_1(X_i^1)h_2(X_i^2)\right) + U_i. \tag{2}$$

There are several nonparametric estimators for this type of models available in the literature. In general, one can ignore the structure of the model and use a series estimator developed by Newey (1997), or a local linear estimator proposed by Stone (1977) and Cleveland (1979), or finally a Nadaraya-Watson kernel estimator. However, all these estimators suffer from the curse of dimensionality, so that the rate of convergence of these estimators depend on the dimension of the vector of regressors. In the context of the additive model without cross-term component, Stone (1985) shows the rate of convergence estimator based on splines does not depend on the number of covariates and therefore does not suffer from the curse of dimensionality.

For pure additive models, Linton and Nielsen (1995), Linton and Härdle (1996), and Chen, Hardle, Linton and Severance-Lossin (1996) propose various estimation procedure that based on marginal integration. Linton and Nielsen (1995) also consider a pure multiplicative submodel with two components show that it also can be estimated by marginal integration procedure. In the case when the dimension of the problem is $d = \dim(X_i) \leq 2$, their marginal integration estimator has a convergence rate $n^{2/5}$. However, when $d \geq 3$, Linton and Nielsen (1995) estimator requires the number of continuous derivatives of additive or multiplicative terms to go up with the size of the problem.

The optimal convergence rate for the estimator based on the marginal integration in this case is $n^{\frac{r}{2r+1}}$, where r is the number of continuous derivatives. To avoid this "curse of dimensionality", in the case of the pure additive model without the cross terms, Horowitz and Mammen (2004) propose a two-step method that achieves a $n^{2/5}$ convergence rate regardless the dimension of X under the minimal assumption that second order derivatives are continuous. Another two-step procedure that also attains the convergence rate of $n^{2/5}$ is proposed in Linton (2000). In this paper I show how the approach in Horowitz and Mammen (2004) can be extended to the case when the model contains bothadditive and multiplicative terms. Next subsection addresses identification issues and describes the estimator.

2.1 Two-Step Estimator

Identification. For the identification purposes, since μ, $\{m_j, 1 \le j \le d\}$ and $\{h_1, h_2\}$ in (2) are identified only up to a location, we need to fix the location. Let the support of each component of the vector of regressor X_i be $[-1, 1]$, so that $suppX_i^j = [-1, 1]$. Then we pin down the location by assuming that functions $\{m_j, 1 \le j \le d\}$ and $\{h_1, h_2\}$ satisfy

$$\int_{-1}^{1} m_j(x)dx = 0, \text{ and } \int_{-1}^{1} h_l(x)dx = 0$$

for any $j = 1, \ldots, d$ and $l = 1, 2$. Also, observe that cross-term $h_1(\cdot)h_2(\cdot)$ identifies functions h_1 and h_2 only up to scale. Therefore, we assume that $h_1(-1) = \overline{h}_1 \neq 0$ to pin down the scale. Finally, if e.g. $h_1(x) = 0$ for any $x \in [-1, 1]$, then $h_2(\cdot)$ is not identified, and vice versa. Therefore, we assume that $h_1(x) \neq 0$ and $h_2(x) \neq 0$ for any $x \in [-1, 1]$.

Two-Step Estimator. The procedure proposed in this paper is based on the two-step estimator introduced in Horowitz and Mammen (2004). The first step estimator is a series estimator that estimates all the components of the model in (2) simultaneously. The second step estimator is a local linear estimator that treats each component individually, while employing the first step estimator for the remaining parameters. The series estimator uses the following notation: let $\{p_k, k = 1, 2, \ldots\}$ be an orthonormal basis for twice continuously differentiable functions on $[-1, 1]$, that satisfy the following conditions:

Condition 1. *(a)* $\int_{-1}^{1} p_k(x)dx = 0$ *for any* k*;*

(b) $\int_{-1}^{1} p_k(x)p_j(x)dx = 0$ *for any* k, j *such that* $k \neq j$ *and* $\int_{-1}^{1} p_k^2(x)dx = 1$ *for any* k*;*

(c) For any $1 \le j \le d$*,* $l = 1, 2$ *and any* $x \in [-1, 1]^d$*, there exist* $\theta_k^{(m_j)}$ *and* $\theta_k^{(h_l)}$ *such that*

$$m_j(x) = \sum_{k=1}^{\infty} \theta_k^{(m_j)} p_k(x) \text{ and } h_l(x) = \sum_{k=1}^{\infty} \theta_k^{(h_l)} p_k(x). \tag{3}$$

These are standard conditions imposed on the basis functions that justify the use of series estimator. The two-step procedure for estimation of cross-term components can be summarized as follows.

Step 1: Series Estimator. Let $p^\kappa(v) = (p_1(v), \ldots, p_\kappa(v))'$ be the vector of first κ basis functions. For any $x \in [-1,1]^d$, define

$$g(x) \equiv \mu + \sum_{j=1}^{d} m_j(x^j) + h_1(x^1)h_2(x^2) \tag{4}$$

and let

$$\phi_\kappa(\theta; x) = \mu + p^\kappa(x^1)'\theta^{(m_1)} + \ldots + p^\kappa(x^d)'\theta^{(m_d)} + \left[p^\kappa(x^1)'\theta^{(h_1)}\right]\left[p^\kappa(x^2)'\theta^{(h_2)}\right].$$

Let $\{(Y_i, X_i), i = 1, \ldots, n\}$ be a random sample from the distribution of (Y, X). Suppose that $\widehat{\theta}_{n\kappa}$ is a solution to the minimization problem:

$$\min_{\theta \in \Theta_\kappa} S_{n\kappa}(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n} \{Y_i - F(\phi_\kappa(\theta; X_i))\}^2, \tag{5}$$

where $\Theta_\kappa \subset \mathbb{R}^{1+\kappa(d+2)}$ is a compact set. Then the first-stage series estimator of $g(x)$ is $\phi_\kappa(\widehat{\theta}_{n\kappa}; x)$. The corresponding first-stage series estimators for $m_j(x)$ and $a_l(x)$ are $\widetilde{m}_j(x^j) \equiv p^\kappa(x^j)'\widehat{\theta}_{n\kappa}^{(m_j)}$ and $\widetilde{h}_l(x^l) = p^\kappa(x^l)'\widehat{\theta}_{n\kappa}^{(h_l)}$ respectively, for any $j = 1, \ldots, d$ and $l = 1, 2$.

Step 2: Local Linear Estimator.

Define $\widetilde{m}(x) = \widetilde{\mu} + \sum_{l=1}^{d} \widetilde{m}_l(x^l)$ and $\widetilde{g}(x) = \widetilde{m}(x) + \widetilde{h}_1(x^1)\widetilde{h}_2(x^2)$. Let K be a kernel function defined on $[-1,1]$ and let $K_h(v) \equiv K\left(\frac{v}{h}\right)$ for $h > 0$. Consider the following optimization problem:

$$\begin{gathered}\min_{b_0, b_1} S_n(b_0, b_1; x^1) \equiv \\ \equiv \tfrac{1}{n}\sum_{i=1}^{n} \{Y_i - F(\widetilde{m}(X_i) + (b_0 + b_1(X_i^1 - x^1))\widetilde{h}_2(X_i^2))\}^2 K_h(x^1 - X_i^1)\end{gathered} \tag{6}$$

Define the set of first and second partial derivatives of $S_n(b_0, b_1; x^1)$ with respect to b_0 and b_1, evaluated at the point $(b_0^* = \widetilde{h}_1(x^1), b_1^* = 0)$:

$S'_{nb_0}(x^1, \widetilde{g}) \equiv \left.\frac{\partial S_n(b_0, b_1; x^1)}{\partial b_0}\right|_{(b_0^*, b_1^*)}$;

$S'_{nb_1}(x^1, \widetilde{g}) \equiv \left.\frac{\partial S_n(b_0, b_1; x^1)}{\partial b_1}\right|_{(b_0^*, b_1^*)}$;

$S''_{nb_0b_0}(x^1, \widetilde{g}) \equiv \left.\frac{\partial^2 S_n(b_0, b_1; x^1)}{\partial b_0^2}\right|_{(b_0^*, b_1^*)}$;

$$S''_{nb_0b_1}(x^1,\widetilde{g}) \equiv \left.\frac{\partial^2 S_n(b_0,b_1;x^1)}{\partial b_0 \partial b_1}\right|_{(b_0^*,b_1^*)};$$
$$S''_{nb_1b_1}(x^1,\widetilde{g}) \equiv \left.\frac{\partial^2 S_n(b_0,b_1;x^1)}{\partial b_1^2}\right|_{(b_0^*,b_1^*)}.$$

The second-stage estimator of a cross-term $h_1(x^1)$ is given by:

$$\widehat{h}_1(x^1) = \widetilde{h}_1(x^1) - \frac{S''_{nb_1b_1}(x^1,\widetilde{g})S'_{nb_0}(x^1,\widetilde{g}) - S''_{nb_0b_1}(x^1,\widetilde{g})S'_{nb_1}(x^1,\widetilde{g})}{S''_{nb_0b_0}(x^1,\widetilde{g})S''_{nb_1b_1}(x^1,\widetilde{g}) - \left[S''_{nb_0b_1}(x^1,\widetilde{g})\right]^2} \tag{7}$$

This is a one Newton step descend from the point $(\widetilde{h}_1(x^1), 0)$ towards the minimum of (6). Pure additive components $m_j(\cdot)$ can be estimated analogously with the appropriate choice of the objective function in (6). Next section provides a set of sufficient conditions for consistency and asymptotic normality of this two-step estimator.

3 Results

First, I introduce some additional notation. For any matrix A define the norm $\|A\| \equiv (trace(AA'))^{1/2}$. For any $k \in \mathbb{N}$ let I_k be $k \times k$ identity matrix. For any two $\alpha_1, \alpha_2 \in \mathbb{R}$ define the following two matrices:

$$\gamma(\alpha_1,\alpha_2) = \begin{pmatrix} I_k & 0 & 0 \\ 0 & \alpha_1 & 0 \\ 0 & 0 & \alpha_2 \end{pmatrix}$$

and

$$\Gamma_\kappa(\alpha_1,\alpha_2) = \begin{pmatrix} 1 & 0 \\ 0 & \gamma(\alpha_1,\alpha_2) \otimes I_\kappa \end{pmatrix}.$$

Both $\gamma(\alpha_1,\alpha_2)$ and $\Gamma_\kappa(\alpha_1,\alpha_2)$ are diagonal matrices. Using this above notation, I define matrices

$$Q_\kappa = E[F'(g(X))^2 \left(\Gamma_\kappa(h_2(X^2),h_1(X^1))P_\kappa(X)\right)\left(\Gamma_\kappa(h_2(X^2),h_1(X^1))P_\kappa(X)\right)']$$

and

$$\Psi_\kappa = Q_\kappa^{-1} E[F'(g(X))^2 V(X) \left(\Gamma_\kappa(h_2(X^2), h_1(X^1)) P_\kappa(X)\right) \times \left(\Gamma_\kappa(h_2(X^2), h_1(X^1)) P_\kappa(X)\right)'] Q_\kappa^{-1},$$

where $V(x) = Var(U|X = x)$. Note that Q_κ and Ψ_κ are positive semidefinite $d(\kappa) \times d(\kappa)$ matrices, where $d(\kappa) = 1 + \kappa(d + 2)$. Finally, let $\lambda_{\kappa,\min}$ be the smallest eigenvalue of Q_κ and $\zeta_\kappa = \sup_{x \in [-1,1]^d} \|P_\kappa(x)\|$. Below is the set of conditions that are sufficient for consistency and asymptotic normality of the two-step estimator. These conditions closely follow assumptions A1-A7 from Horowitz and Mammen (2004), so that the condition that we need for consistency and asymptotic normality of the two-step estimator of multiplicative components are not more restrictive than the conditions we need for consistency and asymptotic normality of the two-step estimator of pure additive model.

Assumption 1. *The data* $\{(Y_i, X_i), i = 1, \ldots, n\}$ *are i.i.d. sample from* $(\Omega, \mathcal{F}, P)$, *and* $E(Y|X = x) = F(\mu + \sum_{j=1}^{d} m_j(x^j) + h_1(x^1)h_2(x^2))$ *for almost any* $x \in [-1, 1]^d$.

This is a standard random sampling assumption. Also, this assumption specifies the structural form defined in (2) for the conditional expectation of Y conditional on X.

Assumption 2. *(i) The support of* X *is* $\mathcal{X} = [-1, 1]^d$.

(ii) The distribution of X *is absolutely continuous with respect to Lebesgue measure.*

(iii) The probability density function of X *is a bounded twice continuously differentiable function on* $[-1, 1]^d$ *and is bounded away from zero.*

(iv) There are constants $c_V > 0$ *and* $C_V < \infty$ *such that* $c_V < Var(U|X = x) < C_V$ *for all* $x \in [-1, 1]^d$.

(v) $E|U|^2 < \infty$ *for all* $j \geqslant 3$. *Also, there is a constant* $C_U < \infty$ *such that* $E|U|^j < C_U^{j-2} j!$.

Assumption 3. *(i) There is a constant* $C_{ac} < \infty$ *such that* $|m_j(v)|, |a_l(v)| < C_{ac}$, *for each* $j = 1, \ldots, d$, $l = 1, 2$ *and any* $v \in [-1, 1]$.

(ii) *Each function* $\{m_j, 1 \leq j \leq d\}$*,* h_1 *and* h_2 *is twice continuously differentiable on* $[-1, 1]$.

(iii) *There are constants* $C_{F_1} < \infty$*,* $c_{F_2} > 0, C_{F_2} < \infty$ *such that* $|F(v_1) - F(v_2)| < C_{F_1} |v_1 - v_2|$ *and* $c_{F_2} < F'(v) < C_{F_2}$ *for all* $v_1, v_2, v \in V = [\min_{x\in[-1,1]^d} g(x), \max_{x\in[-1,1]^d} g(x)]$*. Here function* g *is defined in (4).*

(iv) F *is twice continuously differentiable on* V*.*

(v) *There are constants* $C_{F_3} < \infty$ *and* $S_1 > 5/7$ *such that* $|F''(v_1) - F''(v_2)| < C_{F_3} |v_1 - v_2|^{S_1}$ *for all* $v_1, v_2 \in V$*.*

Assumptions 2 and 3 impose some regularity conditions on the parameters of the model (functions $\{m_j\}$, h_1, h_2, and F) and distributions of vector of covariates X and error term U. In particular, we require known link function F to be Lipschitz continuous, so that applying function F to the series expansion of g preserves the convergence rate of this expansion.

Assumption 4. (i) *There are constants* $C_Q < \infty$ *and* $c_\lambda > 0$ *such that* $|Q_{ij}| \leqslant C_Q$ *and* $\lambda_{\kappa,\min} \geqslant c_\lambda$ *for all* κ *and all* $i, j = 1, \ldots, d(\kappa)$*.*

(ii) *The largest eigenvalue of* Ψ_κ *is bounded for all* κ*.*

Assumption 5. (i) *The basis functions* $\{p_k, k = 1, 2, \ldots\}$ *satisfy conditions 1 (a), (b) and (c).*

(ii) *There is a constant* $c_\kappa > 0$ *such that* $\zeta_\kappa > c_\kappa$ *for all* κ *sufficiently large.*

(iii) $\zeta_\kappa = O(\kappa^{1/2})$ *as* $\kappa \to \infty$*.*

(iv) *There are constant* $C_\theta < \infty$ *and vector* $\theta_{\kappa 0} \in [-C_\theta, C_\theta]^{d(\kappa)}$ *such that* $\sup_{x\in[-1,1]^d} |g(x) - \phi_\kappa(\theta_{\kappa 0}; x)| = O(\kappa^{-2})$ *as* $\kappa \to \infty$*.*

(v) *For each* κ*,* $\theta_{\kappa 0}$ *is an interior point of* $[-C_\theta, C_\theta]^{d(\kappa)}$*.*

Assumptions 4 and 5 provide regularity condition that are sufficient for the consistency of a first step series estimator.

Assumption 6. (i) $\kappa = C_\kappa n^{4/15+v}$ *for some* $0 < C_\kappa < \infty$ *and* $0 < v < \min\{1/30, (7S_1 - 5)/[30(1 + S_1)]\}$*.*

(ii) $h = C_h n^{-1/5}$ *for some* $0 < C_h < \infty$.

Assumption 7. *The function* K *is a bounded, continuous probability density function on* $[-1, 1]$ *and is symmetric about* 0.

Here Assumptions 6 and 7 specify the rate at which the number of basis functions κ used in a first step estimation goes to infinity with the sample size, rate at which the bandwidth h used in the second step shrinks to zero, and restricts the choice of kernel functions to the family of symmetric probability density functions.

Under those assumptions, theorem below shows that the first step series estimator is consistent and characterizes its asymptotic expansion and convergence rate.

Theorem 2. *Suppose that Assumptions 1 through 7 hold. Then*

(a) $\left\|\widehat{\theta}_{n\kappa} - \theta_{\kappa 0}\right\| \to 0$ *almost surely as* $n \to \infty$,

(b) $\widehat{\theta}_{n\kappa} - \theta_{\kappa 0} = O_p\left(\frac{\kappa^{1/2}}{n^{1/2}} + \frac{1}{\kappa^2}\right)$,

(c) $\sup_{x \in \mathcal{X}} \left|\phi_\kappa(\widehat{\theta}_{n\kappa}; x) - g(x)\right| = O_p\left(\frac{\kappa}{n^{1/2}} + \frac{1}{\kappa^{3/2}}\right)$,

and finally,

(d)

$$\begin{aligned}\widehat{\theta}_{n\kappa} - \theta_{\kappa 0} =& Q_\kappa^{-1} \frac{1}{n} \sum_{i=1}^n U_i F'(g(X_i)) \Gamma_\kappa(h_2(X_i^2), h_1(X_i^1)) P_\kappa(X_i) \\ &+ Q_\kappa^{-1} \frac{1}{n} \sum_{i=1}^n \left[F'(g(X_i))\right]^2 \left(\Gamma_\kappa(h_2(X_i^2), h_1(X_i^1)) P_\kappa(X_i)\right) \\ &\times \left(\Gamma_\kappa(h_2(X_i^2), h_1(X_i^1)) P_\kappa(X_i)\right)' b_{\kappa 0}(X_i) + R_n,\end{aligned}$$

where $\|R_n\| = O_p\left(\frac{1}{n^{1/2}\kappa^{1/2}} + \frac{\kappa^2}{n}\right)$.

Define the following set of auxiliary variables:

$D_0(x^1) = 2 \int \left[F'(g(x^1, \widetilde{x}_{-1}))\right]^2 \left[h_2(\widetilde{x}^2)\right]^2 h_1'(x^1) f_X(x^1, \widetilde{x}_{-1}) d\widetilde{x}_{-1}$;
$D_1(x^1) = 2 \int \left[F'(g(x^1, \widetilde{x}_{-1}))\right]^2 \left[h_2(\widetilde{x}^2)\right]^2 h_1'(x^1) \frac{\partial f_X(x^1, \widetilde{x}_{-1})}{\partial x^1} d\widetilde{x}_{-1}$;
$A_K = \int_{-1}^{1} \nu^2 K(\nu) d\nu$;

$B_K = \int_{-1}^{1} [K(\nu)]^2 \, d\nu;$
$\beta_1(x^1) = C_h^2 \frac{2A_K}{D_0(x^1)} \int q(x^1, \widetilde{x}_{-1}) F'(g(x^1, \widetilde{x}_{-1}))[h_2(\widetilde{x}^2)]^2 f_X(x^1, \widetilde{x}_{-1}) d\widetilde{x}_{-1};$
$V_1(x^1) = C_h^{-1} \frac{B_K}{D_0(x^1)} \int Var(U|x^1, \widetilde{x}_{-1}) \left[F'(g(x^1, \widetilde{x}_{-1}))\right]^2 [h_2(\widetilde{x}^2)]^2 f_X(x^1, \widetilde{x}_{-1}) d\widetilde{x}_{-1};$
$q(x^1, \widetilde{x}_{-1}) = F''(g(x^1, \widetilde{x}_{-1})) h_1'(x^1) \left\{ 4 \frac{\partial^2 g((x^1, \widetilde{x}_{-1}))}{\partial (x^1)^2} - 3h_1'(x^1) h_2(\widetilde{x}^2) \right\}$
$\quad + F'(g(x^1, \widetilde{x}_{-1})) h_1''(x^1).$

Theorem 3 (Asymptotic normality for multiplicative components)**.** *Suppose that Assumptions 1 through 7 hold, and let* $\widehat{h}_1(x^1)$ *be a second stage estimator of* $h_1(x^1)$ *defined in (7). Then*

(a) $\widehat{h}_1(x^1) - h_1(x^1) = \frac{-S'_{1b_0}(x^1, g) + \left[D_1(x^1)/D_0(x^1)\right] S'_{1b_1}(x^1, g)}{nhD_0(x^1)} + o_p(n^{-2/5})$ *uniformly over* $|x^1| \leq 1 - h$,

(b) $n^{2/5} \left[\widehat{h}_1(x^1) - h_1(x^1)\right] \xrightarrow{d} N(\beta_1(x^1), V_1(x^1))$,

(c) $n^{2/5}(\widehat{h}_1(x^1) - h_1(x^1))$ *and* $n^{2/5}(\widehat{h}_2(x^2) - h_2(x^2))$ *are asymptotically independently distributed. Moreover,* $n^{2/5}(\widehat{h}_1(x^1) - h_1(x^1))$ *and* $n^{2/5}(\widehat{m}_j(x^j) - m_j(x^j))$ *are asymptotically independently distributed, for any* $j \neq 1$.

Theorem 3 shows that the second step estimator has pointwise convergence rate equal to $n^{2/5}$ and is asymptotically normally distributed. Moreover, estimators of both univariate additive and cross-term components in the model (2) evaluated at different points are asymptotically independent. Next section describes the optimal choice of the bandwidth in empirical applications.

4 Bandwidth selection

The general rule of the bandwidth selection is to choose bandwidth that minimizes Mean Integrated Square Error (*MISE*) or Asymptotic *MISE* (*AMISE*). In our case, for each h_1 and any given weight function $w(\cdot)$, the *AMISE* for the bandwidth $h = C_h n^{-1/5}$ is given by

$$AMISE_{h_1}(C_h) = n^{\frac{4}{5}} \int_{-1}^{1} w(x^1)[(\beta_1(x^1))^2 + V_1(x^1)] dx^1.$$

Recall that $\beta_1(x^1)$ and $V_1(x^1)$ in the expression for $AMISE_{h_1}$ depend on the choice of C_h. The asymptotically optimal bandwidth sets C_h in $h = C_h n^{-1/5}$ as to minimize the $AMISE_{h_1}$. Define $\beta_1(x^1) = C_h^2 \beta_{01}(x^1)$ and $V_1(x^1) = C_h^{-1} V_{01}(x^1)$. Then

$$AMISE_{h_1}(C_h) = n^{\frac{4}{5}} \int_{-1}^{1} w(x^1)[C_h^4(\beta_{01}(x^1))^2 + C_h^{-1} V_{01}(x^1)]dx^1.$$

Therefore, the asymptotically optimal C_h^* for the estimate of h_1 is given by

$$C_h^* = \frac{1}{4}\left[\frac{\int_{-1}^{1} w(x^1)V_{01}(x^1)dx^1}{\int_{-1}^{1} w(x^1)(\beta_{01}(x^1))^2 dx^1}\right]^{\frac{1}{5}}. \tag{8}$$

A plug-in estimator of C_h^* can be obtained by substitution of $\beta_{01}(x^1)$ and $V_{01}(x^1)$ in (8) with their consistent estimates (e.g. kernel estimators). The same method can be used to obtain plug-in estimators of the asymptotically optimal choice of C_h for other multiplicative cross-terms (h_2 in our case) and additive univariate components $\{m_j, j = 1, \ldots, d\}$.

To estimate asymptotically optimal bandwidths simultaneously for all component of the additive model, one can use the Penalized Least Squares method proposed in Horowitz and Mammen (2004).

5 Simulation

In this section, I consider two simple examples with $d = 2$. Both models have a logistic link function $F(v) = \frac{e^v}{1+e^v}$.

Design 1: Let

$$P(Y = 1|X = x) = F\left(f_1(x^1) + f_2(x^2) + x^1 x^2\right),$$

where

$$f_1(x^1) = \sin(\pi x^1) \text{ and } f_2(x^2) = \Phi(3x^2).$$

The covariates x^1 and x^2 are independent and uniformly distributed over $[-1; 1]$. Sample

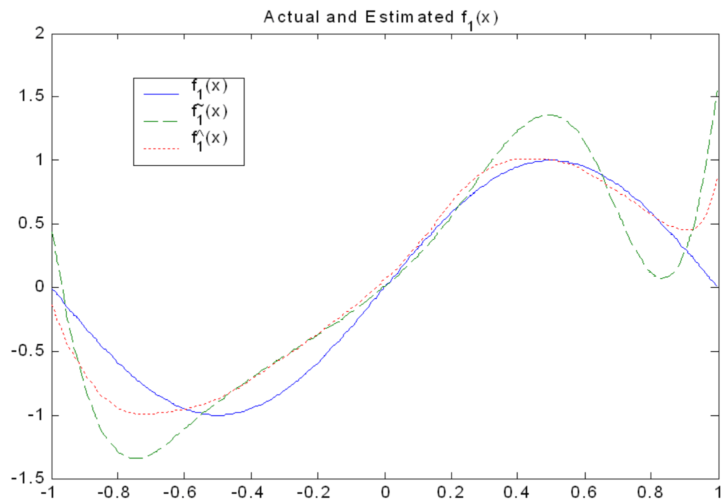

Figure 1: Design 1: First and second step estimators of f_1

size is $n = 500$. The first step estimator uses B-splines and $\kappa_1 = 4$ for f_1 and $\kappa_2 = 2$ for f_2; Tchebychev polynomials are used to obtain a first-step estimator of interaction terms. The second-stage estimator sets $h_1 = 0.5$ and $h_2 = 1.5$ and uses a kernel function $K(v) = \frac{15}{16}\left(1 - v^2\right)^2 I\left(|v| \leqslant 1\right)$. Figures 1 and 2 plot the results of the two-step estimation procedure: first step series estimators (dashed lines) and second step estimators (dotted line). Observe that the second step estimator has smaller integrated squared error than series estimator.

Design 2: Let

$$P(Y = 1|X = x) = F\left(x^1 + x^2 + f_1(x^1)(f_2(x^2) - 1/2)\right),$$

were $f_1(\cdot)$ and $f_2(\cdot)$ are the same as in previous case. B-splines and Tchebychev polynomials are used to obtain a first step estimator, with the parameters specified as in the Design 1. The second step estimation uses $h_1 = 0.5$ and $h_2 = 1.5$ and a kernel function $K(v) = \frac{15}{16}\left(1 - v^2\right)^2 I\left(|v| \leqslant 1\right)$. Figures 3 and 4 show the results of the estimation procedure: dashed lines for the first step series estimator and dotted line for the second step estimators. Again, we observe that the second step estimator behaves better than the series estimator in the case of function $f_2(\cdot)$, but for this particular sample it turns out to be inferior in the integrated squared error sense to series estimator for $f_1(\cdot)$.

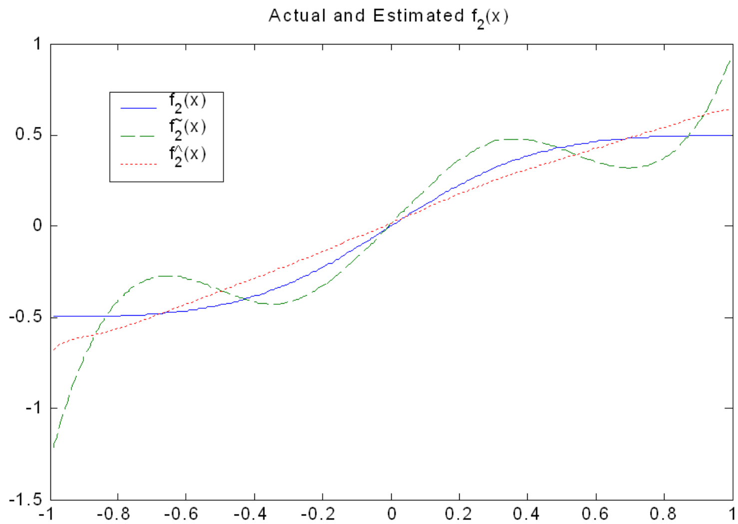

Figure 2: Design 1: First and second step estimators of f_2

Overall, the two-step procedure performs satisfactory in these examples.

6 Conclusion

In this paper I show how the technique developed in Horowitz and Mammen (2004) can be used to estimate both univariate additive terms and multiplicative two-factor terms in the non-parametric additive model with a known link function. I show that both estimators of univariate components and cross-term components are asymptotically normally independently distributed with the rate of convergence $n^{2/5}$. This is equal to the rate of convergence obtained by Horowitz and Mammen for the additive model without multiplicative cross-terms. The method proposed in this paper can be extended to cover the estimation of additive models where cross-terms are composed of more than two univariate factors.

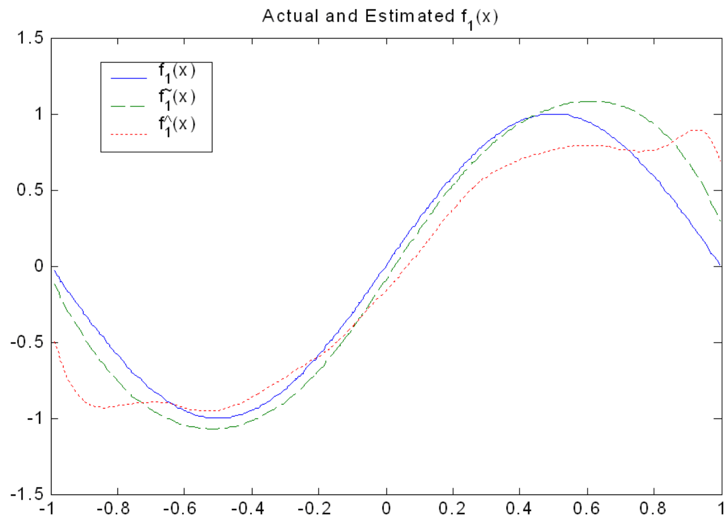

Figure 3: Design 2: First and second step estimators of f_1

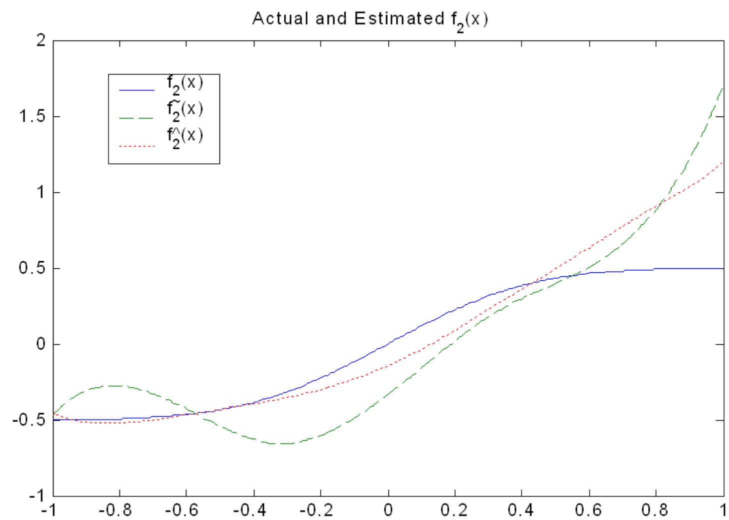

Figure 4: Design 2: First and second step estimators of f_2

7 Appendix

The following notation is used throughout the appendix:

$g(x) = \mu + \sum_{j=1}^{d} m_j(x^j) + h_1(x^1)h_2(x^2),.$

$m(x) = \mu + \sum_{j=1}^{d} m_j(x^j),$

$\phi_\kappa(\theta; x) = \mu + p^\kappa(x^1)'\theta^{(m_1)} + \ldots + p^\kappa(x^d)'\theta^{(m_d)} + \left[p^\kappa(x^1)'\theta^{(h_1)}\right]\left[p^\kappa(x^2)'\theta^{(h_2)}\right],$

I_k is $k \times k$ identity matrix,

$$\gamma(\alpha_1, \alpha_2) = \begin{pmatrix} I_k & 0 & 0 \\ 0 & \alpha_1 & 0 \\ 0 & 0 & \alpha_2 \end{pmatrix}$$

$$\Gamma_\kappa(\alpha_1, \alpha_2) = \begin{pmatrix} 1 & 0 \\ 0 & \gamma(\alpha_1, \alpha_2) \otimes I_\kappa \end{pmatrix}.$$

$b_{\kappa 0}(x) = g(x) - \phi_\kappa(\theta_{\kappa 0}; x),$

$\alpha_{1\kappa 0}(x) = h_1(x^1) - p^\kappa(x^1)'\theta_{\kappa 0}^{(h_1)},$

$\alpha_{2\kappa 0}(x) = h_2(x^2) - p^\kappa(x^2)'\theta_{\kappa 0}^{(h_2)},$

$\widehat{\Gamma}_\kappa(\widehat{\theta}_{n\kappa}; x) = \Gamma_\kappa(p^\kappa(x^2)'\widehat{\theta}_{n\kappa}^{(h_2)}, p^\kappa(x^1)'\widehat{\theta}_{n\kappa}^{(h_1)}),$

$\Delta\widehat{\Gamma}_\kappa(\widehat{\theta}_{n\kappa}; x) = \widehat{\Gamma}_\kappa(\widehat{\theta}_{n\kappa}; x) - \Gamma_\kappa(0,0),$

$\Delta\Gamma_{\kappa 0}(x) = \Gamma_\kappa(\alpha_{2\kappa 0}(x^2), \alpha_{1\kappa 0}(x^1)) - \Gamma_\kappa(0,0),$

$$\Lambda_\kappa(x) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & p^\kappa(x^1)p^\kappa(x^2)' \\ 0 & p^\kappa(x^2)p^\kappa(x^1)' & 0 \end{pmatrix}.$$

Proof of Theorem 2: I begin with six auxiliary lemmas that help prove Theorem 2.

Lemma 4. *There is a constant $a > 0$ and $C < \infty$ such that*

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

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

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

$$S_{n\kappa 1}(\theta) = n^{-1}\sum_{i=1}^{n} Y_i F(\phi_\kappa(\theta; X_i))$$

and

$$S_{n\kappa 2}(\theta) = n^{-1}\sum_{i=1}^{n}\left[F(\phi_\kappa(\theta; X_i))\right]^2.$$

Define $\widetilde{S}_{n\kappa 1}(\theta) = S_{n\kappa 1}(\theta) - ES_{n\kappa 1}(\theta)$. Divide Θ_κ into hypercubes with edge-length equal to some $l > 0$. Let $\Theta_\kappa^{(j)}$ be the j'th cube and define $\theta_{\kappa j}$ to be the center of the cube, for every $j = 1, \ldots, J \equiv (C_\theta/l)^{d(\kappa)}$.

For any $\theta \in \Theta_\kappa^{(j)}$, we have

$$\left|\widetilde{S}_{n\kappa 1}(\theta)\right| \leqslant \left|\widetilde{S}_{n\kappa 1}(\theta_{\kappa j})\right| + \left|\widetilde{S}_{n\kappa 1}(\theta) - \widetilde{S}_{n\kappa 1}(\theta_{\kappa j})\right| \leqslant \left|\widetilde{S}_{n\kappa 1}(\theta_{\kappa j})\right| + 2C_{F_1}\zeta_\kappa^2 r n^{-1}\sum_{i=1}^{n}|Y_i|.$$

Choose $r = [\zeta_\kappa^2]^{-c}$, $c > 1$, so that $2C_{F_1}\zeta_\kappa^2 rE|Y| < \varepsilon/4$ for all sufficiently large κ. By Bernstein's inequality,

$$P\left[2C_{F_1}\zeta_\kappa^2 r n^{-1}\sum_{i=1}^{n}|Y_i| > \varepsilon/2\right] \leqslant 2\exp\left[-a_5 n\varepsilon^2\zeta_\kappa^{4c-4}\right]$$

and

$$P\left[\left|\widetilde{S}_{n\kappa 1}(\theta_{\kappa j})\right| > \varepsilon/2\right] = P\left[n^{-1}\left|\sum_{i=1}^{n}\{Y_iF(\phi_\kappa(\theta_{\kappa j}; X_i)) - E\left[Y_iF(\phi_\kappa(\theta_{\kappa j}; X_i)\right]\}\right| > \varepsilon/2\right]$$
$$\leq 2\exp\left[-b_5 n\varepsilon^2\right].$$

Therefore,

$$P\left[\sup_{\theta\in\Theta_\kappa}\left|\widetilde{S}_{n\kappa 1}(\theta)\right| > \varepsilon\right] \leq 2M\left[\exp\left[-b_5 n\varepsilon^2\right] + \exp\left[-a_5 n\varepsilon^2\zeta_\kappa^{4c-4}\right]\right]$$
$$\leq \exp\left[d(\kappa)\left(\ln\left(C_\theta/r\right) + (1/2)\ln\left(d(\kappa)\right)\right)\right]\left[\exp\left[-b_5 n\varepsilon^2\right] + \exp\left[-a_5 n\varepsilon^2\right]\right]$$

Observe that by the choice of r, $d(\kappa)\left(\ln\left(C_\theta/r\right) + (1/2)\ln\left(d(\kappa)\right)\right) < 0$ when n is sufficiently large. The result of the lemma follows immediately. □

Define $S_\kappa(\theta) = ES_{n\kappa}(\theta)$ and let $\widetilde{\theta}_\kappa = \arg\min_{\theta\in\Theta_\kappa} S_\kappa(\theta)$. Also define $S_{\kappa 0}(\theta) =$

$E\left[Y-F\left(\phi_\kappa(\theta;X)+b_{\kappa 0}(X)\right)\right]^2$ and let $\theta_{\kappa 0}=\arg\min_{\theta\in\Theta_\kappa} S_{\kappa 0}(\theta)$.

Lemma 5. *For any $\eta>0$, $S_\kappa(\widehat{\theta}_{n\kappa})-S_\kappa(\widetilde{\theta}_\kappa)<\eta$ almost surely for all n sufficiently large.*

Lemma 6. *For any $\eta>0$, $S_{\kappa 0}(\widetilde{\theta}_\kappa)-S_{\kappa 0}(\theta_{\kappa 0})<\eta$ for all sufficiently large n.*

Proof: The proof of Lemmas 5 and 6 closely follows the proof of Lemmas 2 and 3 in Horowitz and Mammen (2004) and therefore is omitted here.

Define $G_i=g(X_i)$, $Z_{\kappa i}=F'(G_i)\Gamma_\kappa(h_2(X_i^2),h_1(X_i^1))P_\kappa(X_i)$. Also define $\widehat{Q}_\kappa = n^{-1}\sum_{i=1}^n Z_{\kappa i}Z'_{\kappa i}$ and $B_n=\widehat{Q}_\kappa^{-1}\frac{1}{n}\sum_{i=1}^n F'(G_i)Z_{\kappa i}b_{\kappa 0}(X_i)$. Let $\overline{U}=(U_1,\ldots,U_n)'$ be the vector of residuals. Under Assumptions 1 through 7 the following results hold:

Lemma 7. $\left\|\widehat{Q}_\kappa-Q_\kappa\right\|^2=O_p(\kappa^2/n)$.

Lemma 8. $\gamma_n\left\|\frac{1}{n}\widehat{Q}_\kappa^{-1}Z'_\kappa\overline{U}\right\|^2=O_p(\kappa/n)$.

Lemma 9. $\|B_n\|^2=O_p(\kappa^{-4})$.

Proof: The proof of Lemmas 7, 8 and 9 follows the proof for corresponding lemmas in Horowitz and Mammen (2004) and therefore is omitted here.

Now we are ready to prove Theorem 2. To prove part (a) of Theorem 2, note that

$$\begin{aligned}S_{\kappa 0}(\widehat{\theta}_{n\kappa})-S_{\kappa 0}(\theta_{\kappa 0}) &= \left[S_{\kappa 0}(\widehat{\theta}_{n\kappa})-S_\kappa(\widehat{\theta}_{n\kappa})\right]+\left[S_\kappa(\widehat{\theta}_{n\kappa})-S_\kappa(\widetilde{\theta}_\kappa)\right]\\ &\quad+\left[S_\kappa(\widetilde{\theta}_\kappa)-S_{\kappa 0}(\widetilde{\theta}_\kappa)\right]+\left[S_{\kappa 0}(\widetilde{\theta}_\kappa)-S_{\kappa 0}(\theta_{\kappa 0})\right].\end{aligned}$$

From Lemmas 5 and 6 and uniform convergence of S_κ to $S_{\kappa 0}$ it follows that for any given $h>0$ and $\eta>0$ each of the four terms in the expression for $S_{\kappa 0}(\widehat{\theta}_{n\kappa})-S_{\kappa 0}(\theta_{\kappa 0})$ is less than $\eta/4$ almost surely for all sufficiently large n. Therefore $S_{\kappa 0}(\widehat{\theta}_{n\kappa})-S_{\kappa 0}(\theta_{\kappa 0})<\eta$ almost surely for all sufficiently large n. Part (a) follows because $\theta_{\kappa 0}$ uniquely minimizes $S_{\kappa 0}(\cdot)$.

To prove part (d) of Theorem 2, define $\Delta G_i=\phi_\kappa(\widehat{\theta}_{\kappa n};X_i)-G_i=\phi_\kappa(\widehat{\theta}_{\kappa n};X_i)-\phi_\kappa(\theta_{\kappa 0};X_i)-b_{\kappa 0}(X_i)$. From Taylor series expansion for $\partial S_{n\kappa}(\widehat{\theta}_{n\kappa})/\partial\theta=0$ it follows that:

$$\frac{1}{n}\sum_{i=1}^n Z_{\kappa i}U_i-\left[\widehat{Q}_\kappa+R_{n1}\right](\widehat{\theta}_{n\kappa}-\theta_{\kappa 0})+\frac{1}{n}\sum_{i=1}^n F'(G_i)\,Z_{\kappa i}b_{\kappa 0}(X_i)+R_{n2}=0. \tag{9}$$

The residual terms have the form:

$$
\begin{aligned}
R_{n1} =& \frac{1}{n}\sum_{i=1}^{n}\{-U_i F''(\widehat{G}_i) - F(G_i)F''(\widehat{G}_i) + (2F'(G_i)F''(\widetilde{G}_i) + F'(G_i)F''(\widehat{G}_i))\Delta G_i \\
&+ ((F''(\widetilde{G}_i))^2 + \frac{1}{2}F''(\widetilde{\widetilde{G}}_i)F''(\widehat{G}_i))\,(\Delta G_i)^2\}(\widehat{\Gamma}_\kappa(\widehat{\theta}_{n\kappa}; X_i)P_\kappa(X_i))(\widehat{\Gamma}_\kappa(\widehat{\theta}_{n\kappa}; X_i)P_\kappa(X_i))' \\
&+ \frac{1}{n}\sum_{i=1}^{n}(F'(G_i))^2(\Delta\widehat{\Gamma}_\kappa(\widehat{\theta}_{n\kappa}; X_i)P_\kappa(X_i))(\widehat{\Gamma}_\kappa(\widehat{\theta}_{n\kappa}; X_i)P_\kappa(X_i))' \\
&+ \frac{1}{n}\sum_{i=1}^{n}(F'(G_i))^2(\widehat{\Gamma}_\kappa(\widehat{\theta}_{n\kappa}; X_i)P_\kappa(X_i))(\Delta\widehat{\Gamma}_\kappa(\widehat{\theta}_{n\kappa}; X_i))P_\kappa(X_i))' \\
&+ \frac{1}{n}\sum_{i=1}^{n}\{-U_i F'(G_i) + F(G_i)F'(G_i) + (-U_i F''(\widetilde{G}_i) + F(G_i)F''(\widetilde{G}_i))\Delta G_i \\
&+ \frac{1}{2}F'(G_i)F''(\widetilde{\widetilde{G}}_i)\,(\Delta G_i)^2 + \frac{1}{2}F''(\widetilde{G}_i)F''(\widetilde{\widetilde{G}}_i)\,(\Delta G_i)^3\}\Lambda_\kappa(X_i),
\end{aligned}
$$

and

$$
\begin{aligned}
R_{n2} =& -\frac{1}{n}\sum_{i=1}^{n}\{U_i(F''(\overline{\overline{G}}_i)\widehat{\Gamma}_\kappa(\theta_{\kappa 0}; X_i) + F'(G_i)\Delta\Gamma_{\kappa 0}(X_i)) \\
&+ (\frac{1}{2}F'(G_i)F''(\overline{G}_i) + F'(G_i)F''(\overline{\overline{G}}_i)b_{\kappa 0}(X_i) \\
&+ \frac{1}{2}F''(\overline{G}_i)F''(\overline{\overline{G}}_i)\,(b_{\kappa 0}(X_i))^2)\widehat{\Gamma}_\kappa(\theta_{\kappa 0}; X_i) \\
&+ (F'(G_i))^2\,\Delta\Gamma_{\kappa 0}(X_i)\}P_\kappa(X_i)b_{\kappa 0}(X_i),
\end{aligned}
$$

where $\widetilde{G}_i$ and $\widetilde{\widetilde{G}}_i$ are the points between $\widehat{G}_i = \phi_\kappa(\widehat{\theta}_{n\kappa}; X_i)$ and G_i, $\overline{G}_i$, and $\overline{\overline{G}}_i$ are the points between $\phi_\kappa(\theta_{\kappa 0}; X_i)$ and G_i.

Following Horowitz and Mammen (2004), one can show that

$$
||[(\widehat{Q}_\kappa + R_{n1})^{-1} - \widehat{Q}_\kappa^{-1}]Z_\kappa'\overline{U}/n||^2 = O_p\left(\frac{\kappa^3}{n}||\widehat{\theta}_{n\kappa} - \theta_{\kappa 0}||^2 + \frac{1}{\kappa n}\right)
$$

and that

$$||((\widehat{Q}_\kappa+R_{n1})^{-1}-\widehat{Q}_\kappa^{-1})(\frac{1}{n}\sum_{i=1}^{n}F'\left(G_i\right)Z_{\kappa i}b_{\kappa 0}\left(X_i\right)+R_{n2})||^2 = O_p\left(||\widehat{\theta}_{n\kappa}-\theta_{\kappa 0}||^2\frac{1}{\kappa^2}+\kappa^{-6}\right).$$

From the above results, it follows that

$$\widehat{\theta}_{n\kappa}-\theta_{\kappa 0} = \frac{1}{n}\widehat{Q}_\kappa^{-1}Z_\kappa'\overline{U}+\widehat{Q}_\kappa^{-1}\frac{1}{n}\sum_{i=1}^{n}F'(g(X_i))Z_{\kappa i}b_{\kappa 0}(X_i)+R_n, \tag{10}$$

where $\|R_n\| = O_p\left(\frac{1}{n^{1/2}\kappa^{1/2}}+\frac{\kappa^2}{n}\right)$.

Part (d) then follows from applying the result of Lemma 7 to the right-hand side of (9).

Part (b) of Theorem 2 follows immediately from part (d). To prove part (c), observe that $\phi_\kappa(\widehat{\theta}_{n\kappa};x)-\phi_\kappa(\theta_{\kappa 0};x) = \widehat{\Gamma}_\kappa(\widetilde{\theta}_{n\kappa};x)P_\kappa'(x)(\widehat{\theta}_{n\kappa}-\theta_{\kappa 0})$, where $\widetilde{\theta}_{n\kappa}$ lies between $\widehat{\theta}_{n\kappa}$ and $\theta_{\kappa 0}$. Therefore, $\sup_{x\in\mathcal{X}}\left|\phi_\kappa(\widehat{\theta}_{n\kappa};x)-\phi_\kappa(\theta_{\kappa 0};x)\right| = O_p\left(\frac{\kappa}{n^{1/2}}+\frac{1}{\kappa^{3/2}}\right)$.

Part (c) follows from this result and Assumption 5(iii). □

Proof of Theorem 3

I begin by showing several auxiliary lemmas that will be used in the proof of Theorem 3:

Lemma 10. *Let*

$$\begin{aligned}\delta_{n1}(x) =& n^{-1}\left(\Gamma_\kappa(h_2(x^2),h_1(x^1))P_\kappa(x)\right)'Q_\kappa^{-1}\\ &\times\sum_{j=1}^{n}F'\left(g(X_j)\right)\left(\Gamma_\kappa(h_2(X_j^2),h_1(X_j^1))P_\kappa(X_j)\right)U_j\end{aligned}$$

and

$$\begin{aligned}\delta_{n2}(x) =& n^{-1}\left(\Gamma_\kappa(h_2(x^2),h_1(x^1))P_\kappa(x)\right)'Q_\kappa^{-1}\\ &\times\sum_{j=1}^{n}\left[F'(g(X_j))\right]^2\left(\Gamma_\kappa(h_2(X_j^2),h_1(X_j^1))P_\kappa(X_j)\right)b_{\kappa 0}(X_j).\end{aligned}$$

Define

$$
\begin{aligned}
H_{k1}(x^1) &= (nh)^{-\frac{1}{2}} \sum_{i=1}^{n} \left[F'(g(x^1, X_i))\right]^2 h_2(X_i^2)(X_i^1 - x^1)^k K_h(x^1 - X_i^1)\delta_{n1}(X_i), \\
H_{k2}(x^1) &= (nh)^{-\frac{1}{2}} \sum_{i=1}^{n} \left[F'(g(x^1, X_i))\right]^2 h_2(X_i^2)(X_i^1 - x^1)^k K_h(x^1 - X_i^1)\delta_{n2}(X_i), \\
H_{k3}(x^1) &= (nh)^{-\frac{1}{2}} \sum_{i=1}^{n} \left[F'(g(x^1, X_i))\right]^2 h_2(X_i^2)(X_i^1 - x^1)^k K_h(x^1 - X_i^1)b_{\kappa 0}(X_i).
\end{aligned}
$$

Then $H_{kl}(x^1) = o_p(1)$*, where* $k = 0, 1$ *and* $l = 1, 2, 3$*.*

Proof: The proof is given for the case when $k = 0$. Similar arguments apply to the case when $k = 1$.

Write $H_{01}(x^1) = \sum_{j=1}^{n} a_j(x^1)U_j$, where

$$
\begin{aligned}
a_j(x^1) =& n^{-1} (nh)^{-\frac{1}{2}} \sum_{i=1}^{n} \left[F'(g(x^1, X_i))\right]^2 h_2(X_i^2)K_h(x^1 - X_i^1) \\
&\times \left(\Gamma_\kappa(h_2(X_i^2), h_1(X_i^1))P_\kappa(X_i)\right)' \\
&\times Q_\kappa^{-1} F'\left(g(X_j)\right) \left(\Gamma_\kappa(h_2(X_j^2), h_1(X_j^1))P_\kappa(X_j)\right).
\end{aligned}
$$

Rewrite $a_j(x^1) \equiv n^{-\frac{3}{2}} h^{-\frac{1}{2}} \sum_{i=1}^{n} K_h(x^1 - X_i^1)A_{ij}(x^1)$ and define

$$
a_{j1}(x^1) = n^{-\frac{3}{2}} h^{-\frac{1}{2}} K_h(x^1 - X_i^1)A_{jj}(x^1)
$$

and

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

To proof the claim for $H_{01}(x^1)$ it suffice to show that $\sup_{|x^1| \leqslant 1, 1 \leqslant i,j \leqslant n} |a_{js}(x^1)| = o_p(n^{-\frac{1}{2}})$ for $s = 1, 2$. From the Assumptions 3(iii), 5(iii) and 6(i) it follows that $\sup_{|x^1| \leqslant 1, 1 \leqslant i,j \leqslant n} |A_{ij}(x^1)| = O(\kappa)$. This in turn implies that $\sup_{|x^1| \leqslant 1, 1 \leqslant i,j \leqslant n} |a_{j1}(x^1)|$ $= n^{-\frac{3}{2}} h^{-\frac{1}{2}} O(\kappa) = o_p(n^{-\frac{1}{2}})$.

The proof of the result for $a_{j2}(x^1)$ is identical to the proof of the same result of Lemma 7 in Horowitz and Mammen (2004). Therefore, $H_{01}(x^1) = o_p(1)$.

To prove the result for $H_{02}(x^1)$, write

$$H_{02}(x^1) = (nh)^{-\frac{1}{2}} \sum_{i=1}^{n} \left[F'(g(x^1, X_i))\right]^2 h_2(X_i^2) K_h(x^1 - X_i^1)$$
$$\times \left[\left(\Gamma_\kappa(h_2(X_i^2), h_1(X_i^1)) P_\kappa(X_i)\right)\right]' B_n,$$

where $B_n = n^{-1} Q_\kappa^{-1} \sum_{j=1}^{n} [F'(g(X_j))]^2 \left(\Gamma_\kappa(h_2(X_j^2), h_1(X_j^1)) P_\kappa(X_j)\right) b_{\kappa 0}(X_j)$.
Arguments similar to those used to prove Lemma 9 imply that $E\|B_n\|^2 = O_p(\kappa^{-4})$. So, $H_{02}(x^1) = O_p([\kappa\kappa^{-4}]^{1/2} (nh)^{1/2}) = o_p(1)$.

Finally, to prove the claim for $H_{03}(x^1)$, recall that
$H_{03}(x^1) = (nh)^{-\frac{1}{2}} \sum_{i=1}^{n} [F'(g(x^1, X_i))]^2 h_2(X_i^2)(X_i^1 - x^1)^j K_h(x^1 - X_i^1) b_{\kappa 0}(X_i)$.

Then the result follows directly from Assumptions 3 and 1(i). □

Define $\widetilde{g}(x) = \phi_\kappa(\widehat{\theta}_{n\kappa}; x)$; that is, $\widetilde{g}(x)$ is the first-step series estimator of $g(x)$.

Lemma 11. *The following holds uniformly over $|x^1| \leqslant 1$:*

(a) $(nh)^{-1} S''_{nb_0b_0}(x^1, \widetilde{g}) = D_0(x^1) + o_p(1)$;

(b) $(nh)^{-1} S''_{nb_0b_1}(x^1, \widetilde{g}) = h^2 A_K D_1(x^1)(1 + o_p(1))$;

(c) $(nh)^{-1} S''_{nb_1b_1}(x^1, \widetilde{g}) = h^2 A_K D_0(x^1) + o_p(1)$.

Proof: The result follows from Theorem 2(c) and bounds on $\sup_{|x^1|\leqslant 1} \sum_{i=1}^{n} U_i^r (X_i^1 - x^1)^s K_h(x^1 - X_i^1)$ for $r = 0, 1$ and $s = 0, 1, 2$. □.

Define $\Delta h_l(x^l) = h_l(x^l) - \widetilde{h}_l(x^l)$, $l = 1, 2$, and $\Delta m(x) = m(x) - \widetilde{m}(x)$.

Lemma 12. *The following holds uniformly over $|x^1| \leqslant 1$:*

(a) $(nh)^{-\frac{1}{2}} S'_{nb_0}(x^1, \widetilde{g}) = (nh)^{-\frac{1}{2}} S'_{nb_0}(x^1, g) + (nh)^{\frac{1}{2}} D_0(x^1) \Delta h_1(x^1) + o_p(1)$;

(b) $(nh)^{-\frac{1}{2}} S'_{nb_1}(x^1, \widetilde{g}) = (nh)^{-\frac{1}{2}} S'_{nb_1}(x^1, g) + o_p(1)$.

Proof: To prove part (a), first observe that Taylor series expansion for $S'_{nb_0}(x^1, \widetilde{g})$ yields:

$$S'_{nb_0}(x^1, \widetilde{g}) = S'_{nb_0}(x^1, g) + \sum_{k=1}^{6} J_k(x^1),$$

where

$$
\begin{aligned}
J_1(x^1) =& 2\sum_{i=1}^{n} \left[F'(g(x^1, X_i))\right]^2 [\Delta m(X_i) + \Delta h_1(x^1) h_2(X_i^2) + \widetilde{h}_1(x^1)\Delta h_2(X_i^2)] \\
& \times \left[h_2(X_i^2) + \Delta h_2(X_i^2)\right] K_h(x^1 - X_i^1); \\
J_2(x^1) =& -2\sum_{i=1}^{n} \left[Y_i - F(g(x^1, X_i))\right] F'(g(x^1, X_i))\Delta h_2(X_i^2) K_h(x^1 - X_i^1); \\
J_3(x^1) =& -2\sum_{i=1}^{n} \left[Y_i - F(g(x^1, X_i))\right] F''(g^{**}(x^1, X_i))[\Delta m(X_i) + \Delta h_1(x^1) h_2(X_i^2) \\
& + \widetilde{h}_1(x^1)\Delta h_2(X_i^2)] \left[h_2(X_i^2) + \Delta h_2(X_i^2)\right] K_h(x^1 - X_i^1); \\
J_4(x^1) =& 2\sum_{i=1}^{n} F'(g(x^1, X_i)) F''(g^{**}(x^1, X_i))[\Delta m(X_i) + \Delta h_1(x^1) h_2(X_i^2) \\
& + \widetilde{h}_1(x^1)\Delta h_2(X_i^2)]^2 \left[h_2(X_i^2) + \Delta h_2(X_i^2)\right] K_h(x^1 - X_i^1); \\
J_5(x^1) =& \sum_{i=1}^{n} F'(g(x^1, X_i)) F''(g^{*}(x^1, X_i))[\Delta m(X_i) + \Delta h_1(x^1) h_2(X_i^2) \\
& + \widetilde{h}_1(x^1)\Delta h_2(X_i^2)]^2 \left[h_2(X_i^2) + \Delta h_2(X_i^2)\right] K_h(x^1 - X_i^1); \\
J_6(x^1) =& \sum_{i=1}^{n} F''(g^{*}(x^1, X_i)) F''(g^{**}(x^1, X_i))[\Delta m(X_i) + \Delta h_1(x^1) h_2(X_i^2) \\
& + \widetilde{h}_1(x^1)\Delta h_2(X_i^2)]^3 \left[h_2(X_i^2) + \Delta h_2(X_i^2)\right] K_h(x^1 - X_i^1).
\end{aligned}
$$

and where $g^*(x^1, X_i)$, $g^{**}(x^1, X_i)$ are points between $g(x^1, X_i)$ and $\widetilde{g}(x^1, X_i)$.

From Theorem 2(c) and Lemma 10, it follows that

$$
(nh)^{-\frac{1}{2}}\, J_1 = \left[D_0(x^1) + O_p(h^2)\right] \Delta h_1(x^1) + o_p(1).
$$

Theorem 2(c), Lemma 10 and standard bounds on $\sup_{|x^1|\leqslant 1} \sum_{i=1}^{n} (X_i^1 - x^1)^s K_h(x^1 - X_i^1)$ for $s = 0, 1, 2$ imply that

$$
\begin{aligned}
&(nh)^{-1/2}\, J_2 = o_p(1), \\
&(nh)^{-1/2}\, J_3 = o_p(1), \\
&(nh)^{-1/2}\, J_4 = o_p(1), \\
&(nh)^{-1/2}\, J_5 = o_p(1),
\end{aligned}
$$

$(nh)^{-1/2}\, J_6 = O_p((nh)^{1/2}\left[\frac{\kappa}{n^{1/2}} + \frac{1}{\kappa^{3/2}}\right]^3) = o_p(1).$

This concludes the proof of part (a) of Theorem 3.

To prove part (b) of Theorem 3, observe that the Taylor series expansion for $S'_{nb_1}(x^1, \widetilde{g})$ gives:

$$S'_{nb_1}(x^1, \widetilde{g}) = S'_{nb_1}(x^1, g) + \sum_{k=1}^{6} L_k(x^1), \tag{11}$$

where

$$\begin{aligned} L_1(x^1) =& 2\sum_{i=1}^{n} \left[F'(g(x^1, X_i))\right]^2 [\Delta m(X_i) + \Delta h_1(x^1)h_2(X_i) + \widetilde{h}_1(x^1)\Delta h_2(X_i^2)] \\ & \times \left[h_2(X_i^2) + \Delta h_2(X_i^2)\right](X_i^1 - x^1)K_h(x^1 - X_i^1); \end{aligned}$$

$$L_2(x^1) = -2\sum_{i=1}^{n} \left[Y_i - F(g(x^1, X_i))\right] F'(g(x^1, X_i))\Delta h_2(X_i^2)(X_i^1 - x^1)K_h\left(x^1 - X_i^1\right);$$

$$\begin{aligned} L_3(x^1) =& -2\sum_{i=1}^{n} \left[Y_i - F(g(x^1, X_i))\right] F''(g^{**}(x^1, X_i))[\Delta m(X_i) + \Delta h_1(x^1)h_2(X_i^2) \\ & + \widetilde{h}_1(x^1)\Delta h_2(X_i^2)]\left[h_2(X_i^2) + \Delta h_2(X_i^2)\right](X_i^1 - x^1)K_h(x^1 - X_i^1); \end{aligned}$$

$$\begin{aligned} L_4(x^1) =& 2\sum_{i=1}^{n} F'(g(x^1, X_i))F''(g^{**}(x^1, X_i))[\Delta m(X_i) + \Delta h_1(x^1)h_2(X_i^2) \\ & + \widetilde{h}_1(x^1)\Delta h_2(X_i^2)]^2\left[h_2(X_i^2) + \Delta h_2(X_i^2)\right](X_i^1 - x^1)K_h(x^1 - X_i^1); \end{aligned}$$

$$\begin{aligned} L_5(x^1) =& \sum_{i=1}^{n} F'(g(x^1, X_i))F''(g^{*}(x^1, X_i))[\Delta m(X_i) + \Delta h_1(x^1)h_2(X_i^2) \\ & + \widetilde{h}_1(x^1)\Delta h_2(X_i^2)]^2\left[h_2(X_i^2) + \Delta h_2(X_i^2)\right](X_i^1 - x^1)K_h(x^1 - X_i^1); \end{aligned}$$

$$\begin{aligned} L_6(x^1) =& \sum_{i=1}^{n} F''(g^{*}(x^1, X_i))F''(g^{**}(x^1, X_i))[\Delta m(X_i) + \Delta h_1(x^1)h_2(X_i^2) \\ & + \widetilde{h}_1(x^1)\Delta h_2(X_i^2)]^3\left[h_2(X_i^2) + \Delta h_2(X_i^2)\right](X_i^1 - x^1)K_h(x^1 - X_i^1). \end{aligned}$$

Using the result in Theorem 2(c), Lemma 10 and standard bounds $\sup_{|x^1|\leqslant 1}\sum_{i=1}^{n}(X_i^1 - x^1)^s K_h(x^1 - X_i^1)$ for $s = 0, 1, 2$, one can show that $(nh)^{-\frac{1}{2}}\, J_m = o_p(1)$ for $m = 1, \ldots, 6$. Therefore, $(nh)^{-\frac{1}{2}}\, S'_{nb_1}(x^1, \widetilde{g}) = (nh)^{-\frac{1}{2}}\, S'_{nb_1}(x^1, g) + o_p(1)$, which concludes the proof of the lemma. □

Now we are ready to prove Theorem 3. By definition, $\widehat{h}_1(x^1) - h_1(x^1) = \widetilde{h}_1(x^1) - h_1(x^1) - \frac{S''_{nb_1b_1}(x^1,\widetilde{g})S'_{nb_0}(x^1,\widetilde{g}) - S''_{nb_0b_1}(x^1,\widetilde{g})S'_{nb_1}(x^1,\widetilde{g})}{S''_{nb_0b_0}(x^1,\widetilde{g})S''_{nb_1b_1}(x^1,\widetilde{g}) - \left[S''_{nb_0b_1}(x^1,\widetilde{g})\right]^2}$. Part (a) then follows from the results of Lemmas 11 and 12 and Assumption 6(ii).

Part (b): Define $\eta = \frac{-S'_{nb_0}(x^1,g) + \left[D_1(x^1)/D_0(x^1)\right]S'_{nb_1}(x^1,g)}{nhD_0(x^1)}$.

Arguments like those used to prove asymptotic normality of a local linear estimator imply that $E(n^{2/5}\eta) = \beta_1(x^1) + o(1)$ and $Var(n^{2/5}\eta) = V_1(x^1) + o(1)$, and that $n^{2/5}\eta \xrightarrow{d} N(\beta_1(x^1), V_1(x^1))$. This proves the result in Theorem 3(b).

Part (c): One can show that $Cov(n^{2/5}(\widehat{h}_1(x^1) - h_1(x^1)), n^{2/5}(\widehat{h}_2(x^2) - h_2(x^2))) = o(1)$. Therefore $n^{2/5}(\widehat{h}_1(x^1) - h_1(x^1))$ and $n^{2/5}(\widehat{h}_2(x^2) - h_2(x^2))$ have zero asymptotic covariance. Part (c) then follows from this result and part (b) of Theorem 3. $\square$

References

[1] Chen, R., W. Hardle, O.B. Linton, and E. Severance-Lossin (1996). "Nonparametric Estimation of Additive Separable Regression Models," *Statistical Theory and Computational Aspects of Smoothing*, Physica Verlag, 247-265.

[2] Cleveland, W.S. (1979). "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 74(368), 829-836.

[3] Horowitz, J.L., and E. Mammen (2004). "Nonparametric Estimation of an Additive Model with a Link Function," *Annals of Statistics*, 32(6), 2412-2443.

[4] Linton, O.B. (2000). "Efficient Estimation of Generalized Additive Nonparametric Regression Models," *Econometric Theory*, 16, 502-523.

[5] Linton, O.B., and W. Härdle (1996). "Estimation of Additive Regression Models with Known Links," *Biometrika*, 83(3), 529-540.

[6] Linton, O.B., and J.P. Nielsen (1995). "A Kernel Method of Estimating Structural Nonparametric Regression Based on Marginal Integration,"

[7] Newey, W.K. (1997). "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79(1), 147-168.

[8] Stone, C.J. (1977). "Consistent Nonparametric Regression," *Annals of Statistics*, 5(4), 595-645.

[9] Stone, C.J. (1985). "Additive Regression and Other Nonparametric Models," *Annals of Statistics*, 13(2), 689-705.