

Nonparametric Estimation and Testing of Interaction in Additive Models

Stefan Sperlich

Institut für Statistik und Ökonometrie
Humboldt Universität zu Berlin
Spandauer Str.1
10178 Berlin
Germany

Dag Tjøstheim

Department of Mathematics
University of Bergen
5007 Bergen
Norway

Lijian Yang

Department of Statistics and Probability
Michigan State University
East Lansing, Michigan 48824
U. S. A.

February 16, 1998

Abstract

We consider an additive model with second order interaction terms. It is shown how the components of this model can be estimated using marginal integration, and the asymptotic distribution of the estimators is derived. Moreover, two test statistics for testing the presence of interactions are proposed. Asymptotics for the test functions are obtained, but in this case the asymptotics produce inaccurate results unless the number of observations is very large. For small or moderate sample sizes a bootstrap procedure is suggested and is shown to work well on a simulated example. Finally, our methods are illustrated on a five-dimensional production function for a set of Wisconsin farm data. In particular, the separability hypothesis for the production function is discussed. ¹

Keywords: Testing Additivity, Derivative Estimation, Nonparametric Regression, Additive Models.

¹This research was financially supported by Sonderforschungsbereich 373 "Quantifikation und Simulation Ökonomischer Prozesse" Deutsche Forschungsgemeinschaft, at Humboldt Universität zu Berlin. We thank Stefan Profit and Oliver Linton for inspiration and helpful discussion.

1 Introduction

Linearity is often used as a simplifying device in econometric modeling. If a linearity assumption cannot be entertained, even as a rough approximation, a very large class of nonlinear models is subsumed under the general regression model

$$(1) \quad Y = m(X) + \sigma(X)\varepsilon \quad ,$$

where $X = (X_1, \dots, X_d)$ is a vector of explanatory variables, and where ε is independent of X with $E(\varepsilon) = 0$ and $Var(\varepsilon) = 1$. Although in principle this model can be estimated using nonparametric methods, in practice the curse of dimensionality would in general render such a task impractical.

A viable middle alternative in modelling complexity is to let m be additive, i. e.

$$(2) \quad m(x) = c + \sum_{\alpha=1}^d f_{\alpha}(x_{\alpha}) \quad ,$$

where the functions f_{α} are unknown. The additivity assumption has been employed in several areas of economic theory, for example in connection with separability hypotheses for production functions. Traditionally, additive models have been estimated using backfitting (Hastie and Tibshirani 1990), but recently the method of marginal integration (Auestad and Tjøstheim (1991), Linton and Nielsen (1995), Newey (1994), Tjøstheim and Auestad (1994)) has attracted a fair amount of attention, an advantage being that an explicit asymptotic theory can be constructed. It should be remarked that important progress has been made recently (Linton, Mammen and Nielsen (1997), Opsomer and Ruppert (1997)) in the asymptotic theory of backfitting.

A weakness of the purely additive model is that interactions between the explanatory variables are completely ignored, and in certain econometric contexts - production function modeling being one of them - the absence of interaction terms has been criticized. In this paper we therefore allow for second order pairwise interactions resulting in a model

$$(3) \quad m(x) = c + \sum_{\alpha=1}^d f_{\alpha}(x_{\alpha}) + \sum_{1 \leq \alpha < \beta \leq d} f_{\alpha\beta}(x_{\alpha}, x_{\beta}) \quad .$$

Such models have been mentioned in Hastie and Tibshirani (1990) and briefly in Tjøstheim and Auestad (1994) and a hierarchy of increasing order of interactions has been discussed. The main objective of this paper is to consider estimation and testing of the interaction terms using marginal integration theory. Again the latter makes it possible to construct a precise asymptotic theory. An application to testing for the presence of interaction terms in a production function will be given in Section 5.3.

Not surprisingly, the problems of testing and estimation are intimately connected. The estimation theory is used in constructing tests, and once the test has been performed, estimation theory can again be used to construct confidence intervals for the model resulting from the testing procedure.

We construct two basic functionals for testing of the presence of interaction between a pair of variables (x_{α}, x_{β}) . The most obvious one is to estimate $f_{\alpha\beta}$ and then use a test functional

$$(4) \quad \int \hat{f}_{\alpha\beta}^2(x_{\alpha}, x_{\beta})\pi(x_{\alpha}, x_{\beta})dx_{\alpha}dx_{\beta} \quad ,$$

where π is an appropriate non-negative weight function. The other functional is based on the fact that $\partial^2 m / \partial x_\alpha \partial x_\beta$ is zero iff there is no interaction between x_α and x_β . By marginal integration techniques this can be achieved without estimating $f_{\alpha\beta}$ itself and its derivative, but it does require the estimation of a second order mixed partial derivative of the marginal regressor in the direction (x_α, x_β) .

It is well known that the asymptotic distribution of test functionals of the above type is not a very accurate description of the finite sample properties unless the sample size n is fairly large, see e.g. Hjellvik, Yao and Tjøstheim (1998). As a consequence for a moderate sample size we have adopted a wild bootstrap scheme for constructing the null distribution of the test functional.

Our test is in effect a test of additivity with the added bonus that the alternative is formulated in terms of interactions between pairs of variables. Thus as an outcome of the testing procedure we should be capable of indicating which pairs (if any) of variables should be included to describe the interaction. We again refer to the example of Section 5.3.

Other tests of additivity or interaction terms have been proposed. The one coming closest to ours is a test by Gozalo and Linton (1997), which is based on the differences in modelling m by a purely additive model as in equation (2) opposed to using the general model (1). The curse of dimensionality may of course lead to bias - as pointed out by the authors themselves. Also, this test is less specific in indicating what should be done if the additivity hypothesis is rejected. A rather different approach to additivity testing (in a time series context) is taken by Chen, Liu and Tsay (1995). Still another methodology is considered by Eubank, Hart, Simpson and Stefanski (1995).

2 Some Simple Properties of the Model

In general, $X = (X_1, X_2, \dots, X_d)$ represents a sequence of independent identically distributed (i.i.d.) vectors of explanatory variables, ε refers to a sequence of i.i.d. random variables independent of X , and such that $E(\varepsilon) = 0$ and $Var(\varepsilon) = 1$. In the expression (3) for $m(x)$, c is a constant, $\{f_\alpha(\cdot)\}_{\alpha=1}^d$ and $\{f_{\alpha\beta}(\cdot)\}_{1 \leq \alpha < \beta \leq d}$ are real-valued unknown functions, and where for $\alpha = 1, 2, \dots, d$ one has

$$(5) \quad E f_\alpha(X_\alpha) = \int f_\alpha(x_\alpha) \varphi_\alpha(x_\alpha) dx_\alpha = 0,$$

and for all $1 \leq \alpha < \beta \leq d$,

$$(6) \quad \int f_{\alpha\beta}(x_\alpha, x_\beta) \varphi_\alpha(x_\alpha) dx_\alpha = \int f_{\alpha\beta}(x_\alpha, x_\beta) \varphi_\beta(x_\beta) dx_\beta = 0,$$

with $\{\varphi_\alpha(\cdot)\}_{\alpha=1}^d$ being marginal densities (assumed to exist) of the X_α 's. Equations (5) and (6) are identifiability conditions. If one starts with a function of the form given in (3) not necessarily satisfying (5) and (6), the following steps could be taken:

1. Replace all $\{f_{\alpha\beta}(x_\alpha, x_\beta)\}_{1 \leq \alpha < \beta \leq d}$ by $\{f_{\alpha\beta}(x_\alpha, x_\beta) - f_{\alpha,\alpha\beta}(x_\alpha) - f_{\beta,\alpha\beta}(x_\beta) + c_{0,\alpha\beta}\}_{1 \leq \alpha < \beta \leq d}$, where

$$f_{\alpha,\alpha\beta}(x_\alpha) = \int f_{\alpha\beta}(x_\alpha, u) \varphi_\beta(u) du$$

$$f_{\beta,\alpha\beta}(x_\beta) = \int f_{\alpha\beta}(u, x_\beta)\varphi_\alpha(u)du$$

$$c_{0,\alpha\beta} = \int f_{\alpha\beta}(u, v)\varphi_\alpha(u)\varphi_\beta(v)dudv$$

and adjust the $\{f_\beta(x_\beta)\}_{\beta=1}^d$'s and the constant term c accordingly so as to keep $m()$ the same function;

2. Replace all $\{f_\beta(x_\beta)\}_{\beta=1}^d$ by $\{f_\beta(x_\beta) - c_{0,\beta}\}_{\beta=1}^d$, where $c_{0,\beta} = \int f_\beta(u)\varphi_\beta(u)du$, and adjust the constant term c accordingly so as to keep $m()$ the same function.

Let X_α be the $(d-1)$ -dimensional random variable obtained by removing X_α from $X = (X_1, \dots, X_d)$, and let $X_{\alpha\beta}$ be defined analogously. With some abuse of notation we write $X = (X_\alpha, X_\beta, X_{\alpha\beta})$ to highlight the directions in d -space represented by the α and β coordinates. We denote the marginal density of X_α , that of $X_{\alpha\beta}$ and of X by $\varphi_\alpha(x_\alpha)$, $\varphi_{\alpha\beta}(x_{\alpha\beta})$, and $\varphi(x)$, respectively.

We now define by marginal integration

$$(7) \quad F_\alpha(x_\alpha) = \int m(x_\alpha, x_\alpha)\varphi_\alpha(x_\alpha)dx_\alpha,$$

for every $1 \leq \alpha \leq d$ and

$$(8) \quad F_{\alpha\beta}(x_\alpha, x_\beta) = \int m(x_\alpha, x_\beta, x_{\alpha\beta})\varphi_{\alpha\beta}(x_{\alpha\beta})dx_{\alpha\beta},$$

for every pair $1 \leq \alpha < \beta \leq d$. Denote by D_α and $D_{\alpha,\beta}$ the subset of $\{1, 2, \dots, d\}$ with α , respectively α and β , removed. Moreover, let

$$D_{\alpha\alpha} = \{(\gamma, \delta) \mid 1 \leq \gamma < \delta \leq d, \gamma \in D_\alpha, \delta \in D_\alpha\},$$

$$D_{\alpha\beta} = \{(\gamma, \delta) \mid 1 \leq \gamma < \delta \leq d, \gamma \in D_\alpha \cap D_\beta, \delta \in D_\alpha \cap D_\beta\}$$

and

$$c_{\alpha\beta} = \int f_{\alpha\beta}(u, v)\varphi_{\alpha\beta}(u, v)dudv$$

for every pair $1 \leq \alpha < \beta \leq d$. Then (5) and (6) entail the following lemma.

Lemma 1

- 1)
$$F_\alpha(x_\alpha) = f_\alpha(x_\alpha) + c + \sum_{(\gamma,\delta) \in D_{\alpha\alpha}} c_{\delta\gamma}$$

$$F_{\alpha\beta}(x_\alpha, x_\beta) = f_{\alpha\beta}(x_\alpha, x_\beta) + f_\alpha(x_\alpha) + f_\beta(x_\beta) + c + \sum_{(\gamma,\delta) \in D_{\alpha\beta}} c_{\delta\gamma}$$
- 2)
$$F_{\alpha\beta}(x_\alpha, x_\beta) - F_\alpha(x_\alpha) - F_\beta(x_\beta) + \int m(x)\varphi(x)dx = f_{\alpha\beta}(x_\alpha, x_\beta) + c_{\alpha\beta}$$
- 3)
$$c_{\alpha\beta} = \int \{F_{\alpha\beta}(u, x_\beta) - F_\alpha(u)\} \varphi_\alpha(u)du - F_\beta(x_\beta) + \int m(x)\varphi(x)dx$$

$$f_{\alpha\beta}(x_\alpha, x_\beta) = F_{\alpha\beta}(x_\alpha, x_\beta) - F_\alpha(x_\alpha) - \int \{F_{\alpha\beta}(u, x_\beta) - F_\alpha(u)\} \varphi_\alpha(u)du$$

Proof.

- 1) Both formulas follow from the definitions of $D_{\alpha\alpha}$, $D_{\alpha\beta}$, $c_{\alpha\beta}$ and equations (7) and (8).
- 2) Note first that the population mean is simply

$$\int m(x)\varphi(x)dx = c + \sum_{(1 \leq \gamma < \delta \leq d)} c_{\delta\gamma}.$$

Using this and the formulas in 1), one arrives at

$$\begin{aligned} F_{\alpha\beta}(x_\alpha, x_\beta) - F_\alpha(x_\alpha) - F_\beta(x_\beta) + \int m(x)\varphi(x)dx = \\ f_{\alpha\beta}(x_\alpha, x_\beta) + \sum_{1 \leq \gamma < \delta \leq d} c_{\delta\gamma} + \sum_{(\gamma, \delta) \in D_{\alpha\beta}} c_{\delta\gamma} - \sum_{(\gamma, \delta) \in D_{\alpha\alpha}} c_{\delta\gamma} - \sum_{(\gamma, \delta) \in D_{\beta\beta}} c_{\delta\gamma} \\ = f_{\alpha\beta}(x_\alpha, x_\beta) + c_{\alpha\beta}. \end{aligned}$$

- 3) We only need to integrate both sides of the equation in 2) and note that the right hand side comes out as $c_{\alpha\beta}$ because of the identifiability condition (5). The rest follows by the equation in 2). Q.E.D.

We define another auxiliary function

$$\tilde{f}_{\alpha\beta}(x_\alpha, x_\beta) := F_{\alpha\beta}(x_\alpha, x_\beta) - F_\alpha(x_\alpha) - F_\beta(x_\beta) + \int m(x)\varphi(x)dx = f_{\alpha\beta}(x_\alpha, x_\beta) + c_{\alpha\beta},$$

which is a convenient substitute for $f_{\alpha\beta}(x_\alpha, x_\beta)$ as shown in the following.

Corollary 1

$$\tilde{f}_{\alpha\beta}(x_\alpha, x_\beta) \equiv 0 \iff f_{\alpha\beta}(x_\alpha, x_\beta) \equiv 0$$

Proof.

\implies : by the previous lemma, $F_{\alpha\beta}(x_\alpha, x_\beta) - F_\alpha(x_\alpha) - F_\beta(x_\beta) + \int m(x)\varphi(x)dx \equiv 0$ implies $f_{\alpha\beta}(x_\alpha, x_\beta) + c_{\alpha\beta} \equiv 0$, or $f_{\alpha\beta}(x_\alpha, x_\beta) \equiv -c_{\alpha\beta}$, which by (6) gives

$$0 = \int f_{\alpha\beta}(x_\alpha, x_\beta)\varphi_\alpha(x_\alpha)dx_\alpha = - \int c_{\alpha\beta}\varphi_\beta(x_\beta)dx_\beta = -c_{\alpha\beta}.$$

and therefore $f_{\alpha\beta}(x_\alpha, x_\beta) \equiv 0$;

\impliedby : first by the definition of $c_{\alpha\beta}$, $f_{\alpha\beta}(x_\alpha, x_\beta) \equiv 0$ gives $c_{\alpha\beta} = 0$, and thus $f_{\alpha\beta}(x_\alpha, x_\beta) + c_{\alpha\beta} \equiv 0$. Q.E.D.

The corollary provides a marginal integration tool for testing the presence of the interaction term $f_{\alpha\beta}(x_\alpha, x_\beta)$:

$$\int \tilde{f}_{\alpha\beta}^2(x_\alpha, x_\beta)\pi(x_\alpha, x_\beta)dx_\alpha dx_\beta \neq 0$$

where $\pi(x_\alpha, x_\beta)$ is any weight function. This observation suggests the use of the following statistic for testing of additivity of the α -th and β -th directions:

$$(9) \quad \int \widehat{f}_{\alpha\beta}^2(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta$$

where

$$(10) \quad \widehat{f}_{\alpha\beta}(x_\alpha, x_\beta) = \widehat{F}_{\alpha\beta}(x_\alpha, x_\beta) - \widehat{F}_\alpha(x_\alpha) - \widehat{F}_\beta(x_\beta) + \frac{1}{n} \sum_{j=1}^n Y_j.$$

with $\widehat{F}_{\alpha,\beta}$, \widehat{F}_α and \widehat{F}_β being defined in the next section, and where it follows from the strong law of large numbers that

$$\frac{1}{n} \sum_{j=1}^n Y_j \xrightarrow{\text{a.s.}} \int m(x) \varphi(x) dx .$$

As an alternative it is also possible to consider the mixed derivative of $f_{\alpha\beta}$. We will use the notation $f_{\alpha\beta}^{(r,s)}$ to denote the derivative $\frac{\partial^{r+s}}{\partial x_\alpha^r \partial x_\beta^s} f_{\alpha\beta}$ and analogously $F_{\alpha\beta}^{(r,s)}$ for $\frac{\partial^{r+s}}{\partial x_\alpha^r \partial x_\beta^s} F_{\alpha\beta}$. We only have to check whether

$$\int \left\{ F_{\alpha\beta}^{(1,1)}(x_\alpha, x_\beta) \right\}^2 \pi(x_\alpha, x_\beta) dx_\alpha dx_\beta$$

is zero which is equivalent to $f_{\alpha\beta} \equiv 0$.

3 The Estimators

3.1 Estimation of interaction

To use the marginal integration type statistic (9), estimators of the interaction terms must be prescribed. Imagine the X -variables to be scaled so that we can choose the same bandwidth h for the directions represented by α , β and g for $\underline{\alpha\beta}$. Further, let K and L be kernel functions and define $K_h(\cdot) = \frac{1}{h} K(\cdot/h)$ and $L_g(\cdot) = \frac{1}{g} L(\cdot/g)$. We will give more detailed descriptions of the kernels K and L and the bandwidths h and g in the next sections. For ease of notation we use the same letters K and L (and later K^*) to denote kernel functions of varying dimensions. It will be clear from the context what the dimensions are in each specific case.

Following the ideas of Linton and Nielsen (1995) and Tjøstheim and Auestad (1994) we estimate the marginal influence of x_α , x_β and (x_α, x_β) by the integration estimator as follows

$$\widehat{F}_{\alpha\beta}(x_\alpha, x_\beta) = \frac{1}{n} \sum_{l=1}^n \widehat{m}(x_\alpha, x_\beta, X_{l\underline{\alpha\beta}}) \quad , \quad \widehat{F}_\alpha(x_\alpha) = \frac{1}{n} \sum_{l=1}^n \widehat{m}(x_\alpha, X_{l\underline{\alpha}}),$$

where $X_{l\underline{\alpha\beta}}$ ($X_{l\underline{\alpha}}$) is the l^{th} observation of X with X_α and X_β (X_α) removed.

The estimator $\widehat{m}(x_\alpha, x_\beta, X_{l\underline{\alpha\beta}})$ will be called the pre-estimator in the following. To compute it we make use of a special kind of multidimensional local linear kernel estimation; see Ruppert and Wand (1994) for the general case. We consider the problem of minimizing

$$(11) \quad \sum_{i=1}^n \{Y_i - a_0 - a_1(X_{i\underline{\alpha}} - x_\alpha) - a_2(X_{i\underline{\beta}} - x_\beta)\}^2 K_h(X_{i\underline{\alpha}} - x_\alpha, X_{i\underline{\beta}} - x_\beta) L_g(X_{i\underline{\alpha\beta}} - X_{l\underline{\alpha\beta}})$$

for each l fixed. Accordingly we define

$$\widehat{m}(x_\alpha, x_\beta, X_{\underline{\alpha\beta}}) = e_1(Z_{\alpha\beta}^T W_{l,\alpha\beta} Z_{\alpha\beta})^{-1} Z_{\alpha\beta}^T W_{l,\alpha\beta} Y$$

in which

$$Y = (Y_1, \dots, Y_n)^T,$$

$$W_{l,\alpha\beta} = \text{diag} \left\{ \frac{1}{n} K_h(X_{i\alpha} - x_\alpha, X_{i\beta} - x_\beta) L_g(X_{i\alpha\beta} - X_{l\alpha\beta}) \right\}_{i=1}^n,$$

$$Z_{\alpha\beta} = \begin{pmatrix} 1 & X_{1\alpha} - x_\alpha & X_{1\beta} - x_\beta \\ \vdots & \vdots & \vdots \\ 1 & X_{n\alpha} - x_\alpha & X_{n\beta} - x_\beta \end{pmatrix},$$

and $e_1 = (1, 0, 0)$. It should be noted that this is a local linear estimator in the directions α , β and a local constant one for the nuisance directions $\underline{\alpha\beta}$.

Similarly, to produce the pre-estimator $\widehat{m}(x_\alpha, X_{\underline{\alpha}})$, with $e_1 = (1, 0)$, we define

$$\widehat{m}(x_\alpha, X_{\underline{\alpha}}) = e_1(Z_\alpha^T W_{l,\alpha} Z_\alpha)^{-1} Z_\alpha^T W_{l,\alpha} Y$$

in which

$$W_{l,\alpha} = \text{diag} \left\{ \frac{1}{n} K_h(X_{i\alpha} - x_\alpha) L_g(X_{i\alpha} - X_{l\alpha}) \right\}_{i=1}^n,$$

$$Z_\alpha = \begin{pmatrix} 1 & X_{1\alpha} - x_\alpha \\ \vdots & \vdots \\ 1 & X_{n\alpha} - x_\alpha \end{pmatrix}.$$

This estimator results from minimizing

$$\sum_{i=1}^n \{Y_i - a_0 - a_1(X_{i\alpha} - x_\alpha)\}^2 K_h(X_{i\alpha} - x_\alpha) L_g(X_{i\alpha} - X_{l\alpha}),$$

which gives a local linear smoother for the direction α and a local constant one for the other directions.

In order to derive the asymptotics of these estimators we make use of the idea of equivalent kernels; see Ruppert and Wand (1994) and Fan et al. (1993). The main idea is that the local polynomial smoother of degree p is asymptotically equivalent to, i. e. it has the same leading term as, a kernel estimator with "higher order kernel" given by

$$(12) \quad K_\nu^*(u) := \sum_{t=0}^p s_{\nu t} u^t K(u)$$

for the one dimensional case, where $S = (\int u^{t+s} K(u) du)_{0 \leq t, s \leq p}$ and $S^{-1} = (s_{\nu t})_{0 \leq \nu, t \leq p}$. Estimates of derivatives of m can then be obtained by taking different rows of S^{-1} .

For the one dimensional case and $p = 1$ we have $S^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \mu_2^{-1}(K) \end{pmatrix}$, with $\mu_2(K) = \int u^2 K(u) du$.

For the two dimensional case with $p = 1$ the equivalent kernel is

$$(13) \quad K_\nu^*(u, v) := K(u, v) s_\nu(1, u, v)^T, \text{ with } s_\nu \text{ being the } (\nu + 1)^{\text{th}} \text{ row of}$$

$$S^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \mu_2^{-1} & 0 \\ 0 & 0 & \mu_2^{-1} \end{pmatrix}, \quad 0 \leq \nu \leq 2.$$

To estimate the function m itself ($\nu = 0$) using a local linear smoother ($p = 1$) we have simply $K_0^*(u) = K(u)$ and $K_0^*(u, v) = K(u, v)$, but K_ν^* becomes increasingly important when we estimate derivatives. We will come back to this point in Section 3.2.

We are interested in the asymptotics of the estimator $\widehat{f}_{\alpha\beta}(x_\alpha, x_\beta)$ given in (10). We need the following assumptions.

A1: *The kernels $K(\cdot)$ and $L(\cdot)$ are positive, bounded, symmetric, compactly supported and Lipschitz continuous. The bivariate kernel K is a product kernel such that (with some abuse of notation) $K(u, v) = K(u)K(v)$, where $K(u)$ and $K(v)$ are identical functions with $\int K(u) du = 1$. The $d - 1$, respectively $d - 2$ dimensional kernel $L(\cdot)$ is also a product of univariate kernels $L(u)$ of order $q \geq 2$, i.e.*

$$\int u^r L(u) du = \begin{cases} 1 & \text{for } r = 0 \\ 0 & \text{for } 0 < r < q \\ c_r \in \mathbb{R} & \text{for } r \geq q \end{cases}.$$

A2: *Bandwidths satisfy $\frac{nh^2 q^{2(d-2)}}{\ln^2(n)} \rightarrow \infty$, $\frac{q^q}{h^2} \rightarrow 0$ and $h = h_0 n^{-\frac{1}{6}}$.*

A3: *The functions f_α , $f_{\alpha\beta}$ have bounded Lipschitz continuous $(p + 1)^{\text{th}}$ derivatives..*

A4: *The variance function, $\sigma^2(\cdot)$, is bounded and Lipschitz continuous.*

A5: *The d -dimensional density φ has compact support S_X with $\inf_{x \in S_X} \varphi(x) > 0$ and is Lipschitz continuous.*

Remark to assumption (A1): Product kernels are chosen here for the ease of notation, especially in the proofs. The theorems work also for other multivariate kernels. In the following we will use the notation $\|K\|_2^2 := \int K^2(x) dx$ for a kernel K (or K^*) of any dimension.

Theorem 1 *Let (x_α, x_β) be in the interior of the support of $\varphi_{\alpha\beta}(\cdot)$. Then under conditions (A1)-(A5),*

$$(14) \quad \sqrt{nh^2} \{ \widehat{f}_{\alpha\beta}(x_\alpha, x_\beta) - \widetilde{f}_{\alpha\beta}(x_\alpha, x_\beta) - h^2 B_1(x_\alpha, x_\beta) \} \xrightarrow{\mathcal{L}} N \{0, V_1(x_\alpha, x_\beta)\},$$

where $\widetilde{f}_{\alpha\beta}$ is given by (10) and

$$V_1(x_\alpha, x_\beta) = \|K^*\|_2^2 \int \sigma^2(x) \frac{\varphi_{\alpha\beta}^2(x_{\alpha\beta})}{\varphi(x)} dx_{\alpha\beta}$$

and

$$B_1(x_\alpha, x_\beta) = \mu_2(K) \frac{1}{2} \left\{ f_{\alpha\beta}^{(2,0)}(x_\alpha, x_\beta) - \int f_{\alpha\beta}^{(2,0)}(x_\alpha, u_\beta) \varphi_{\underline{\beta}}(u_{\underline{\beta}}) du \right. \\ \left. + f_{\alpha\beta}^{(0,2)}(x_\alpha, x_\beta) - \int f_{\alpha\beta}^{(0,2)}(u_\alpha, x_\beta) \varphi_{\underline{\alpha}}(u_{\underline{\alpha}}) du \right\}$$

The proof of Theorem 1 is given in the appendix.

3.2 Estimation of derivatives

Since the estimation of derivatives for additive separable models has already been introduced in the paper of Severance-Lossin and Sperlich (1997), we concentrate in this section on estimating the mixed derivatives of the function $F_{\alpha\beta}$. Although the derivative estimation for multivariate functions with the aid of multidimensional local polynomials is worked out in Ruppert and Wand (1994) and other papers, the case of mixed derivatives has not been explicitly considered yet. Our interest in this estimator is motivated by testing the hypothesis of additivity with no second order interaction. Since $F_{\alpha\beta}^{(1,1)} = f_{\alpha\beta}^{(1,1)}$, to test for $F_{\alpha\beta}^{(1,1)} \equiv 0$ is equivalent to testing the hypothesis that $f_{\alpha\beta}$ is zero.

Following the ideas of the previous section (or see also Severance-Lossin and Sperlich (1997)) at the pre-estimator point $(x_\alpha, x_\beta, X_{i\underline{\alpha}\underline{\beta}})$ we implement a special version of the local polynomial estimator. For our purpose it is enough to use a bivariate local polynomial estimator. We want to minimize

$$\sum_{i=1}^n \{Y_i - a_0 - a_1(X_{i\alpha} - x_\alpha) - a_2(X_{i\beta} - x_\beta) - a_3(X_{i\alpha} - x_\alpha)(X_{i\beta} - x_\beta) - a_4(X_{i\alpha} - x_\alpha)^2 - a_5(X_{i\beta} - x_\beta)^2\}^2 K_h(X_{i\alpha} - x_\alpha) K_h(X_{i\beta} - x_\beta) L_g(X_{i\underline{\alpha}\underline{\beta}} - X_{l\underline{\alpha}\underline{\beta}})$$

and accordingly define our estimator by

$$(15) \quad \hat{F}_{\alpha\beta}^{(1,1)}(x_\alpha, x_\beta) = \frac{1}{n} \sum_{i=1}^n e_4 (Z_{\alpha\beta}^T W_{i,\alpha\beta} Z_{\alpha\beta})^{-1} Z_{\alpha\beta}^T W_{i,\alpha\beta} Y$$

where $Y, W_{i,\alpha\beta}$ are defined as above in Section 3.1 and $e_4 = (0, 0, 0, 1, 0, 0)$.

Thus in equation (15) above, $Z_{\alpha\beta}$ is

$$Z_{\alpha\beta} = \begin{pmatrix} 1 & X_{1\alpha} - x_\alpha & X_{1\beta} - x_\beta & (X_{1\alpha} - x_\alpha)(X_{1\beta} - x_\beta) & (X_{1\alpha} - x_\alpha)^2 & (X_{1\beta} - x_\beta)^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n\alpha} - x_\alpha & X_{n\beta} - x_\beta & (X_{n\alpha} - x_\alpha)(X_{n\beta} - x_\beta) & (X_{n\alpha} - x_\alpha)^2 & (X_{n\beta} - x_\beta)^2 \end{pmatrix}.$$

This estimator is bivariate locally quadratic for the directions α and β and locally constant else. Certainly it is also possible to use polynomials of higher degree but for ease of presentation we restrict ourself to local quadratic polynomials.

Recalling the ideas of the preceding section we can now put the equivalent kernel K^* to effective use. Using a local quadratic smoother ($p = 2$) we have for the two dimensional case

$$K_\nu^*(u, v) := K(u, v) s_\nu(1, u, v, uv, u^2, v^2)^T$$

where s_ν is the $(\nu + 1)^{th}$, $0 \leq \nu \leq 5$, row of

$$\mathcal{S}^{-1} = \begin{pmatrix} \frac{\mu_4 + \mu_2^2}{\mu_4 - \mu_2^2} & 0 & 0 & 0 & \frac{1 - \mu_2}{\mu_4 - \mu_2^2} & \frac{-\mu_2}{\mu_4 - \mu_2^2} \\ 0 & \mu_2^{-1} & 0 & 0 & 0 & 0 \\ 0 & 0 & \mu_2^{-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu_2^{-2} & 0 & 0 \\ \frac{-\mu_2}{\mu_4 - \mu_2^2} & 0 & 0 & 0 & (\mu_4 - \mu_2^2)^{-1} & 0 \\ \frac{\mu_4 + \mu_2^2}{\mu_4 - \mu_2^2} & 0 & 0 & 0 & 0 & (\mu_4 - \mu_2^2)^{-1} \end{pmatrix}$$

where $\mu_j = \mu_j(K) = \int u^j K(u) du$. The relationship between \mathcal{S}^{-1} and $(Z_{\alpha\beta}^T W_{1,\alpha\beta} Z_{\alpha\beta})^{-1}$ is given in Lemma A2 of the appendix.

If we want to estimate the mixed derivative, we use $K_3^*(u, v) = K(u, v)uv\mu_2^{-2}(K)$ where

$$\begin{aligned} \int uv K_3^*(u, v) dudv &= 1 \\ \int u^q K_3^*(u, v) dudv &= \int v^q K_3^*(u, v) dudv = 0 \quad \text{for } q = 0, 1, 2, 3, \dots \end{aligned}$$

To state the asymptotics for the joint derivative estimator we need bandwidth conditions that differ slightly from (A2):

$$\text{A6: Bandwidths satisfy } \frac{nh^2 g^{2(d-2)}}{\ln^2(n)} \rightarrow \infty, \frac{g^q}{h^2} \rightarrow 0 \text{ and } h = h_0 n^{-\frac{1}{10}}.$$

Then we have

Theorem 2 Under conditions (A1), (A3)-(A6),

$$\sqrt{nh^6} \{ \widehat{F}_{\alpha\beta}^{(1,1)}(x_\alpha, x_\beta) - F_{\alpha\beta}^{(1,1)}(x_\alpha, x_\beta) - h^2 B_2(x_\alpha, x_\beta) \} \xrightarrow{\mathcal{L}} N \{0, V_2(x_\alpha, x_\beta)\},$$

where

$$V_2(x_\alpha, x_\beta) = \|K_3^*\|_2^2 \int \sigma^2(x) \frac{\varphi_{\alpha\beta}^2(x_{\alpha\beta})}{\varphi(x)} dx_{\alpha\beta}$$

and

$$\begin{aligned} B_2(x_\alpha, x_\beta) &= \mu_4(K)\mu_2^{-1}(K) \left[\frac{1}{2} \{ f_{\alpha\beta}^{(2,1)}(x_\alpha, x_\beta) \int \varrho_\beta + f_{\alpha\beta}^{(1,2)}(x_\alpha, x_\beta) \int \varrho_\alpha \} \right. \\ &\quad + \frac{1}{3!} \{ f_{\alpha\beta}^{(3,1)}(x_\alpha, x_\beta) + f_{\alpha\beta}^{(1,3)}(x_\alpha, x_\beta) + f_{\alpha\beta}^{(3,0)}(x_\alpha, x_\beta) \int \varrho_\beta + f_{\alpha\beta}^{(0,3)}(x_\alpha, x_\beta) \int \varrho_\alpha \\ &\quad + f_\alpha^{(3)}(x_\alpha) \int \varrho_\beta + f_\beta^{(3)}(x_\beta) \int \varrho_\alpha + \sum_{\gamma \in D_{\alpha\beta}} \int f_{\alpha\gamma}^{(3,0)}(x_\alpha, x_\gamma) \varrho_\beta + \\ &\quad \left. \sum_{\gamma \in D_{\alpha\beta}} \int f_{\alpha\gamma}^{(0,3)}(x_\alpha, x_\gamma) \varrho_\alpha \} \right] \end{aligned}$$

with

$$\varrho_\alpha = \frac{\varphi_{\alpha\beta}(x_{\alpha\beta})}{\varphi(x)} \frac{\partial \varphi(x)}{\partial x_\alpha} dx_{\alpha\beta}$$

and ϱ_β defined analogously.

4 Testing for Interaction

We are now in a position to state the problem of testing for second order interaction. As mentioned in Sections 1 and 2 for the model (3) we consider the null hypothesis $H_0 : f_{\alpha\beta} \equiv 0$, i.e. there is no interaction between X_α and X_β for a fixed pair (α, β) . Applying this test to any pair of different directions $X_\gamma, X_\delta, 1 \leq \gamma < \delta \leq d$ this can be regarded as a test for additivity in the regression model.

In Section 2 we pointed out that for this purpose it is equivalent to consider $\tilde{f}_{\alpha\beta}$ instead of $f_{\alpha\beta}$. We propose two procedures; the first one is focused on $\tilde{f}_{\alpha\beta}$ directly, the second one on the mixed derivative $\tilde{f}_{\alpha\beta}^{(1,1)}$.

4.1 Considering the interaction function

We will briefly sketch the idea as to how the test statistic can be analysed and then state the theorem giving the asymptotics. The detailed proof is postponed to the appendix.

We consider $\int \tilde{f}_{\alpha\beta}^2(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta$. To study this test statistic, note first that by Theorem 1, equation (14) and some tedious calculations we get the following decomposition

$$\int \tilde{f}_{\alpha\beta}^2(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta = 2 \sum_{1 \leq i < j \leq n} H(X_i, \varepsilon_i, X_j, \varepsilon_j) + \sum_{i=1}^n H(X_i, \varepsilon_i, X_i, \varepsilon_i) + \int \tilde{f}_{\alpha\beta}^2(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta + 2h^2 \int \tilde{f}_{\alpha\beta}(x_\alpha, x_\beta) B_1(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta + o_p(h^2)$$

where

$$H(X_i, \varepsilon_i, X_j, \varepsilon_j) = \varepsilon_i \varepsilon_j \int \frac{1}{n^2} (w_{i\alpha\beta} - w_{i\alpha} - w_{i\beta})(w_{j\alpha\beta} - w_{j\alpha} - w_{j\beta}) \sigma(X_i) \sigma(X_j) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta$$

with weights $w_{i\alpha}, w_{i\beta}$ and $w_{i\alpha\beta}$ defined in the appendix, equation (20) and (23).

We then calculate the asymptotics of $H(X_i, \varepsilon_i, X_i, \varepsilon_i)$ and $H(X_i, \varepsilon_i, X_j, \varepsilon_j)$, put the results together and obtain (cf. Lemma A3 of the appendix)

Theorem 3 *Under assumptions (A1) to (A5), as $h \rightarrow 0$ and $nh^2 \rightarrow \infty$*

$$\begin{aligned} & nh \int \tilde{f}_{\alpha\beta}^2(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta - \frac{\{2K^{(2)}(0)\}^2}{h} \int \frac{\varphi_{\alpha\beta}^2(z_{\alpha\beta})}{\varphi(z)} \sigma^2(z) dz \\ & - nh \int \tilde{f}_{\alpha\beta}^2(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta - 2nh^3 \int \tilde{f}_{\alpha\beta}(x_\alpha, x_\beta) B_1(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta \\ & \xrightarrow{\mathcal{L}} N \left\{ 0, 8 \left\| K^{(2)} \right\|_2^4 \int \frac{\varphi_{\alpha\beta}(z_{1\alpha\beta})^2}{\varphi(z_1)} \frac{\varphi_{\alpha\beta}(z_{2\alpha\beta})^2}{\varphi(z_{1\alpha}, z_{1\beta}, z_{2\alpha\beta})} \sigma^2(z_1) \sigma^2(z_{1\alpha}, z_{1\beta}, z_{2\alpha\beta}) dz_1 dz_{2\alpha\beta} \right\}, \end{aligned}$$

where $K^{(2)}$ is the 2-fold convolution of the kernel K , and where B_1 is defined in the formulation of Theorem 1.

4.2 Considering the mixed derivative of the joint influence

In contrast to the preceding method one can test for interaction without estimating the function of interaction $f_{\alpha\beta}$ explicitly but looking at the mixed derivative of the function $F_{\alpha\beta}$. Our test statistic is $\int \hat{F}_{\alpha\beta}^{(1,1)2} \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta$ which certainly for our purpose is the same as $\int \hat{f}_{\alpha\beta}^{(1,1)2} \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta$.

As can be seen from the proofs of Theorems 1 to 3, the asymptotics for this test statistic are the same as in Theorem 3 with the only difference that we now have to deal with K_3^* and end up with asymptotic formulas containing K_1^* instead of K ; see the definition in Section 3.1. Thus we state the following theorem without an explicit proof.

Theorem 4 *Under assumptions (A1) and (A3)-(A6), as $h \rightarrow 0$ and $nh^6 \rightarrow \infty$*

$$\begin{aligned} & nh^5 \int \hat{F}_{\alpha\beta}^{(1,1)2}(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta - \frac{2 \left(K_1^{*(2)}(0) \right)^2}{h} \int \frac{\varphi_{\alpha\beta}^2(z_{\alpha\beta})}{\varphi(z)} \sigma^2(z) dz \\ & - nh^5 \int f_{\alpha\beta}^{(1,1)2}(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta - 2nh^7 \int f_{\alpha\beta}^{(1,1)}(x_\alpha, x_\beta) B_2(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta \\ & \xrightarrow{\mathcal{L}} N \left\{ 0, 8 \left\| K_1^{*(2)} \right\|_2^4 \int \frac{\varphi_{\alpha\beta}(z_{1\alpha\beta})^2}{\varphi(z_1)} \frac{\varphi_{\alpha\beta}(z_{2\alpha\beta})^2}{\varphi(z_{1\alpha}, z_{1\beta}, z_{2\alpha\beta})} \sigma^2(z_1) \sigma^2(z_{1\alpha}, z_{1\beta}, z_{2\alpha\beta}) dz_1 dz_{2\alpha\beta} \right\}, \end{aligned}$$

where B_2 is defined in the formulation of Theorem 2.

5 An empirical investigation of the test procedures

In nonparametric statistics for small and moderate sample sizes one has to be careful when using the asymptotic distribution in practice. In our case we have the additional problem of having unknown expressions in the bias and variance of the test statistics, and we are dealing with a type of nonparametric test functional which has been known (Hjellvik, Yao and Tjøstheim (1998)) to possess a low degree of accuracy in its asymptotic distribution. It is therefore not unexpected when a simulation experiment (Section 5.2, Figure 2) for $n = 150$ observations reveals a very bad approximation for the asymptotics, and we must look for alternative ways to proceed for low and moderate sample sizes.

5.1 Using the wild bootstrap

One possible alternative is to use the bootstrap or the wild bootstrap, the latter being first introduced by Wu (1986) and Liu (1988). Härdle and Mammen (1993) set it into the context of nonparametric hypothesis testing as it will be used here.

The basic idea is to resample from residuals estimated under the null hypothesis by drawing each bootstrap residual from a two-point (a, b) distribution $G_{(a,b),i}$ which has mean zero, variance equal

to the square of the residual and third moment equal to the cube of the residual for all $i = 1, 2, \dots, n$. Thus, through the use of one single observation one attempts to reconstruct the distribution for each residual separately up to the third moment. For this we do not need additional assumptions on ε or $\sigma(\cdot)$.

Let T_n be the test statistic described in Theorem 3 or 4 and let n^* be the number of bootstrap replications. The whole procedure for the test using the wild bootstrap then consists of the following steps:

1. Estimate the regression function $m_0 = m_{0,\alpha\beta}$ under the hypothesis $H_{0,\alpha\beta}$ that $f_{\alpha\beta} = \hat{f}_{\alpha\beta} = 0$ in model (3) for a fixed pair (α, β) , $1 \leq \alpha < \beta \leq d$ and construct the residuals $\tilde{u}_i = \tilde{u}_{i,\alpha\beta} = Y_i - \widehat{m}_0(X_i)$, for $i = 1, 2, \dots, n$
2. For each X_i , randomly draw a bootstrap residual u_i^* from the distribution $G_{(a,b),i}$ such that for $U \sim G_{(a,b),i}$,

$$\begin{aligned} E_{G_{(a,b),i}}(U) &= 0 & , & & E_{G_{(a,b),i}}(U^2) &= \tilde{u}_i^2 \\ \text{and} & & & & E_{G_{(a,b),i}}(U^3) &= \tilde{u}_i^3 . \end{aligned}$$

3. Generate a sample $\{(Y_i^*, X_i)\}_{i=1}^n$ with $Y_i^* = \widehat{m}_0 + u_i^*$. Here, for the estimation of m_0 it is recommended to use slightly oversmoothing bandwidths, see Härdle and Mammen (1993).
4. Calculate the bootstrap test statistic T_n^* using the sample $\{(Y_i^*, X_i)\}_{i=1}^n$ in the same way as the original T_n is calculated.
5. Repeat steps 2-4 n^* times and use the n^* different T_n^* to determine the quantiles of the test statistic under the null hypothesis and subsequently the critical value for the rejection boundaries.

For the two-point distribution $G_{(a,b),i}$ we have used the so called *golden cut* method, setting $G_{(a,b),i} = q\delta_a + (1-q)\delta_b$ where δ_a, δ_b denote point measures at $a = \tilde{u}_i(1 - \sqrt{5})/2$, $b = \tilde{u}_i(1 + \sqrt{5})/2$ with q has to be $(5 + \sqrt{5})/10$.

For the marginal integration estimator Dalelane (1998) recently proved that the wild bootstrap works for the case of i.i.d. observations. In the setting of times series some work on this has been done by Achmus (1998). Dalelane showed via strong approximation that it holds in supremum norm whereas Achmus proved that the wild bootstrap holds at least locally for time series. There is still some work needed to establish a theory of the wild bootstrap for the test statistic we are using. But this is beyond the scope of the paper.

5.2 The simulation study

To compare our two test procedures, to investigate and demonstrate their behavior empirically as well as to present how well the estimation works we did a simulation study for a small sample size

of $n = 150$ observations. The data have been generated from the model

$$m(x) = E(Y|X = x) = c + \sum_{j=1}^3 f_j(x_j) + f_{1,2}(x_1, x_2)$$

where

$$(16) \quad \begin{aligned} f_1(u) &= 2u \\ f_2(u) &= 1.5 \sin(-1.5u) \\ f_3(u) &= -u^2 + E(u^2) \end{aligned}$$

and

$$(17) \quad f_{1,2}(u, v) = auv$$

with $a = 0$ under the null hypothesis and $a = 1$ under the alternative. The input variables X_j , $j = 1, 2, 3$ are i. i. d. uniform on $[-2, 2]$. To generate the response Y we added normally distributed error terms with standard deviation $\sigma_\epsilon = 0.5$ to the regression function $m(x)$.

For calculating the test statistic we used the quartic kernel $\frac{15}{16}(1 - u^2)^2 \mathbb{1}\{|u| \leq 1\}$ for $K(u)$ and $L(u)$ and product kernels for higher dimensions. We chose different bandwidths depending on the testing procedure and whether the direction was of interest or not (in the previous sections we distinguished them by denoting them h and g).

When we considered the test statistic based on the estimation of $\tilde{f}_{1,2}$ (*direct test*) we used $h = 0.9$, $g = 1.1$ and for the pre-estimation to do the wild bootstrap $h = 1.0$ and $g = 1.2$. To calculate the test statistic based on the joint derivative $f_{1,2}^{(1,1)}$ (*testing derivatives*) we selected $h = 1.5$, $g = 1.6$ and $h = 1.4$, $g = 1.5$, respectively.

In Figure 1 we depict the performance of the estimation procedure for $a = 1.0$ in (17) using the local linear smoother. The estimates of f_1 , f_2 and f_3 with their corresponding true functions are displayed in the upper part of Figure 1; the corresponding estimate $f_{1,2}$ in the lower part. The estimation procedure is working quite well.

Turning to testing, we consider the null hypothesis $H_0 : f_{1,2}(u) \equiv 0$. First we take a look at the asymptotics. In Figure 2 we have plotted kernel estimates of the standardized densities of the test procedures compared to the standard normal distribution. The densities of the test statistics have been estimated with a quartic kernel and bandwidth 0.2. To make the densities comparable we also smoothed the normal densities using the same kernel. We see clearly that the test statistics we introduced in the previous sections look more like a χ^2 distributed random variable than a normal one. Thus, even if we could estimate bias and variance of the test statistics well, the asymptotic distribution of them is hardly usable for testing for such a moderate sample of observations.

This conclusion is consistent with the results of Hjellvik, Yao and Tjøstheim (1998) for a similar type of functional designed for testing of linearity. For that functional roughly 100000 observations were needed to obtain a good approximation. The reason is that for a functional of the type $\int \tilde{f}_{\alpha\beta}^2(x_\alpha, x_\beta) \pi(x_\alpha, x_\beta) dx_\alpha dx_\beta$ several of the leading terms of the Edgeworth expansion are nearly of the same magnitude, so that very many observations are needed for the dominance of the first order term yielding normality. We refer to Hjellvik, Yao and Tjøstheim (1998) for more details.

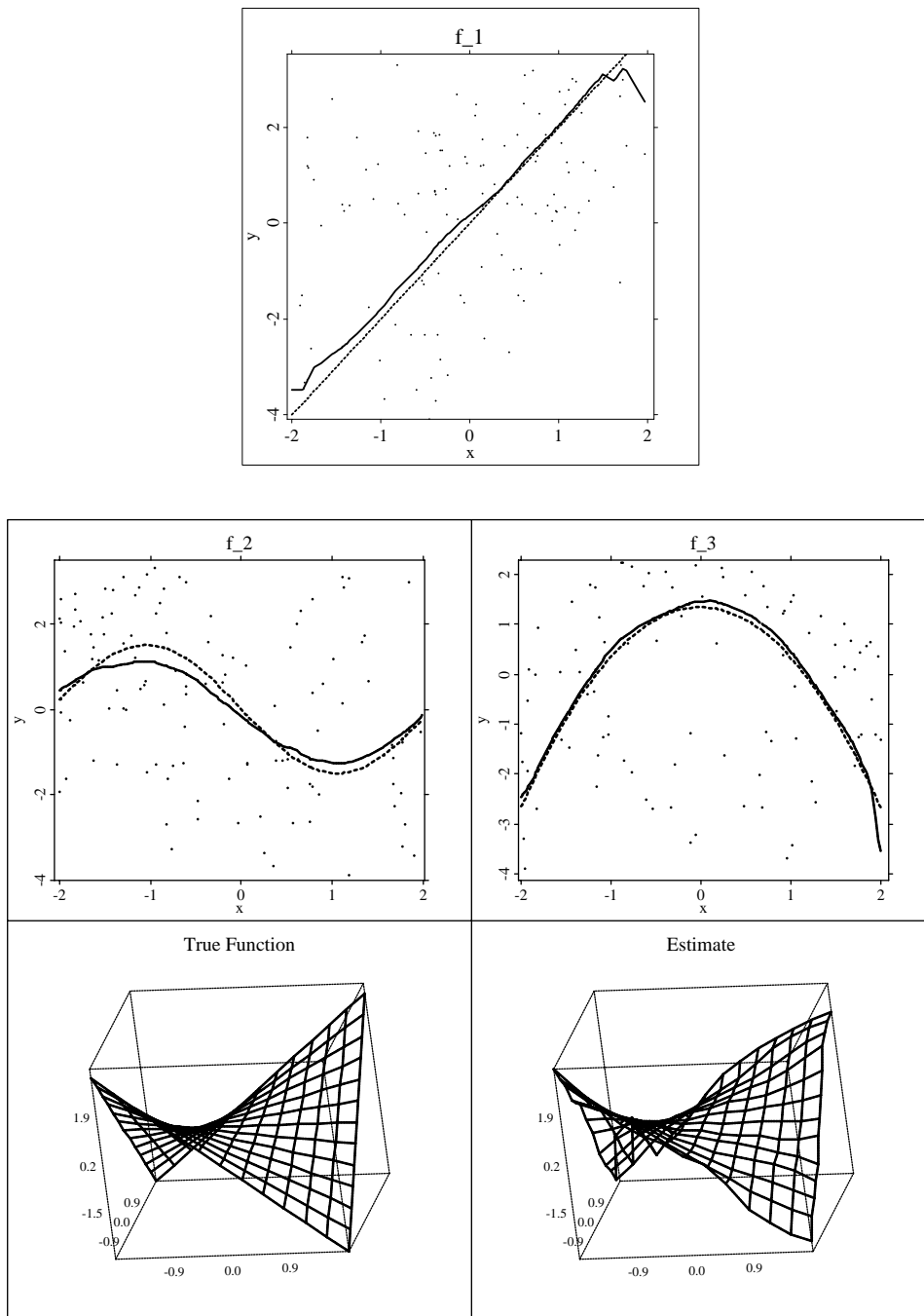


Figure 1: Additive components (dashed) and their estimates (solid), f_1 (top), f_2 (upper left), f_3 (upper right) and $\tilde{f}_{1,2}$ (lower).

To get the results of Table 1 and Figure 3, describing the bootstrap version of the tests, we did 249

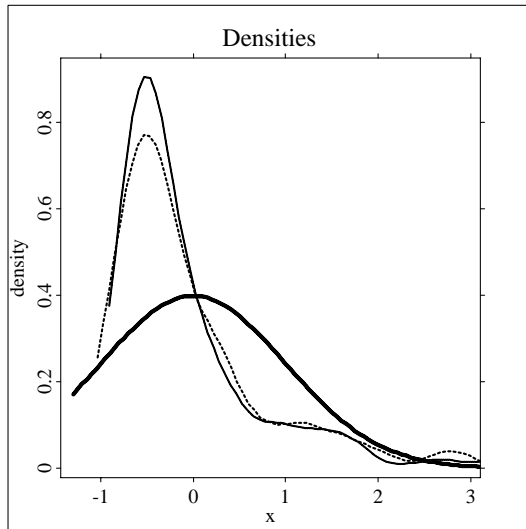


Figure 2: *Densities of the test statistics; direct method (solid), testing derivatives (dashed) and normal density (thick, solid).*

bootstrap replications and considered the test statistics

$$\frac{1}{n} \sum_{i=1}^n \widehat{f}_{1,2}(X_1, X_2) \mathbb{1}\{|X_k| \leq 1.6 \text{ for } k = 1, 2\}$$

and

$$\frac{1}{n} \sum_{i=1}^n \widehat{f}_{1,2}^{(1,1)}(X_1, X_2) \mathbb{1}\{|X_k| \leq 1.6 \text{ for } k = 1, 2\}$$

respectively, i.e. we used a weight function for the test statistic to correct for boundary effects caused by the estimation.

Table 1 is presenting the error of the first kind for both methods and at different significance levels.

TABLE 1: Percentage of rejection under H_0

significance level in %	1	5	10	15	20
direct method	3.0	6.0	12.7	17.3	22.3
testing derivatives	0.5	4.5	11.4	14.4	18.2

For both test procedures obtaining an accurate error of the first kind with the aid of wild bootstrap depends on a proper choice of bandwidth. Thus, Table 1 does not give decisive information whether the *direct testing* procedure is superior to the other one. We just see that wild bootstrap obviously works quite well and can be used for this test problem. For a comparison of the *direct method* against the *derivative approach* and to be able to judge the tests more generally we have to look at

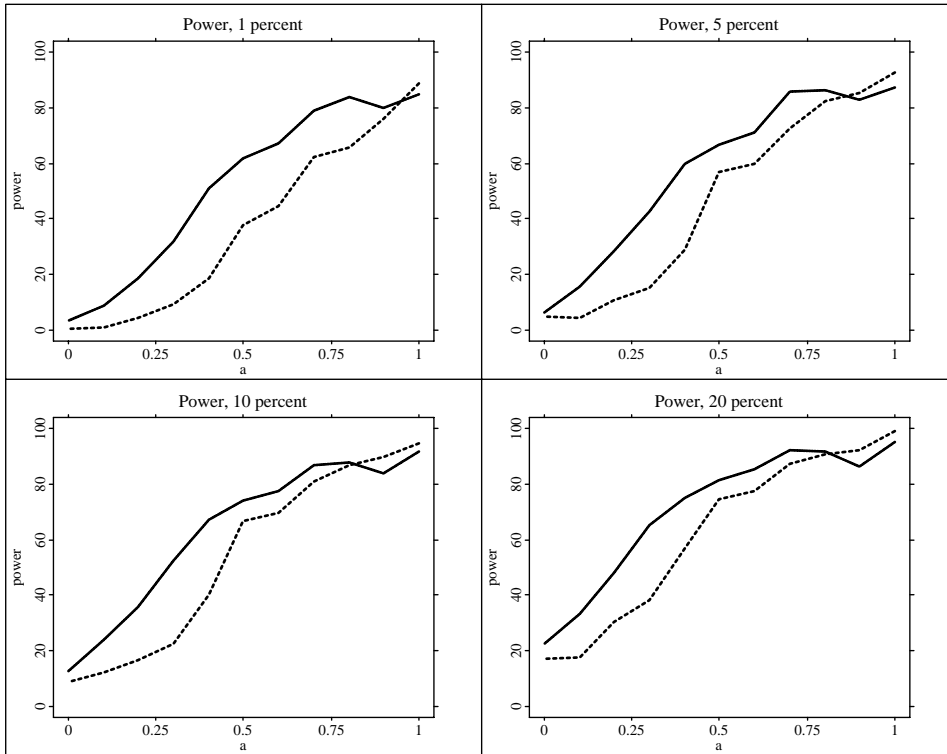


Figure 3: Power functions at the 1, 5, 10 and 20% significance levels for both procedures: direct method (solid) and testing derivatives (dashed).

the power at a wide range of examples. The power as a function of a in (17) is displayed for both methods and different levels in Figure 3. Both procedures are working well. For this particular model the power function of the *direct method* is steeper, but it is quite likely that the comparative advantages of the two methods depend on the particular model or design.

5.3 An Application to Production Function Estimation

In this section we use our estimation and testing procedures for a five dimensional production function.

Separability and additivity of production functions have been discussed since the early paper by Leontief (1947). These conditions yield many important economic results, e.g. they allow the aggregation of inputs or decentralization in decision-making. But there has been much discussion in the past whether production functions can be taken to be additive (strongly separable ²) for a particular data set. This discussion goes back at least to Denny and Fuss (1977), Fuss, McFadden

²The expression "strong separability" is equivalently used for "additivity" or "generalized additivity", see Berndt and Christensen (1973).

and Mundlak (1978), Deaton and Muellbauer (1980, pp.117-165). Our test procedure is an adequate tool to investigate the hypothesis of additivity.

We consider the example and data of Severance-Lossin and Sperlich (1997) and look at the estimation of a production function for livestock in Wisconsin. In that paper strong separability (additivity) among the input factors was assumed and the additive components and their derivatives were estimated using the marginal integration estimator. Whereas their interest was focused mainly on the return to scale and thus on the derivative estimation, we are more interested in examining the validity of the assumption of additivity by looking at the second order interaction terms. We use a subset of $n = 250$ observations of an original data set of more than 1000 Wisconsin farms collected by the Farm Credit Service of St. Paul, Minnesota in 1987. Severance-Lossin and Sperlich removed outliers and incomplete records and selected farms which only produced animal outputs. The data consists of farm level inputs and outputs measured in dollars. The output Y in this analysis is livestock, the input variables are family labor X_1 , hired labor X_2 , miscellaneous inputs (e.g. repairs, rent, custom hiring, supplies, insurance, gas, oil, and utilities) X_3 , animal inputs (purchased feed, breeding, and veterinary services) X_4 , and intermediate run assets (assets with a useful life of one to ten years) X_5 .

The underlying additive model is of the form

$$(18) \quad \ln(y) = c + \sum_{\alpha=1}^d f_{\alpha} \{\ln(x_{\alpha})\} .$$

This model can be viewed as a generalization of the Cobb-Douglas production technology. In the Cobb-Douglas model we would have $f_{\alpha} \{\ln(x_{\alpha})\} = \beta_{\alpha} \ln(x_{\alpha})$.

We have extended this model by including interaction terms $f_{\alpha\beta}$ to obtain

$$(19) \quad \ln(y) = c + \sum_{\alpha=1}^d f_{\alpha} \{\ln(x_{\alpha})\} + \sum_{1 \leq \alpha < \beta \leq d} f_{\alpha\beta} \{\ln(x_{\alpha}), \ln(x_{\beta})\}$$

and the assumed strong separability (additivity) can be checked by testing the null hypothesis $H_0 : f_{\alpha\beta} \equiv 0$ for all α, β .

First we estimated all functions f_{α} and $f_{\alpha\beta}$. The estimation results are given in Figures 4 to 6. For the estimation we used the quartic kernel for K and L . The data were divided by their standard deviations so that we could choose the same bandwidths for each direction. We tried different bandwidths and $h = 1.7$ and $g = 3.3$ yield reasonable smooth estimates. However, we know by experience that the integration estimator is quite robust against different choices of bandwidths. For a detailed discussion of the bandwidth choice and robustness, compare Sperlich, Linton and Härdle (1997).

In Figure 4 the univariate function estimates (not centered to zero) are displayed together with a kind of partial residuals $\hat{r}_{i\alpha} := y_i - \sum_{j \neq \alpha} f_j(X_{ij}) = f_{\alpha}(X_{i\alpha}) + \varepsilon_i$. To see clearly the shape of the estimates we display the main part of the point clouds including the function estimates. As mentioned already in Severance-Lossin and Sperlich, the graphs in Figure 4 give some indication of nonlinearity in family labor, hired labor and intermediate run assets. They even seem to indicate

that the elasticities for these inputs increase and finally could lead to increasing returns to scale. An obvious inference from the economic point of view would be that larger farms are more productive.

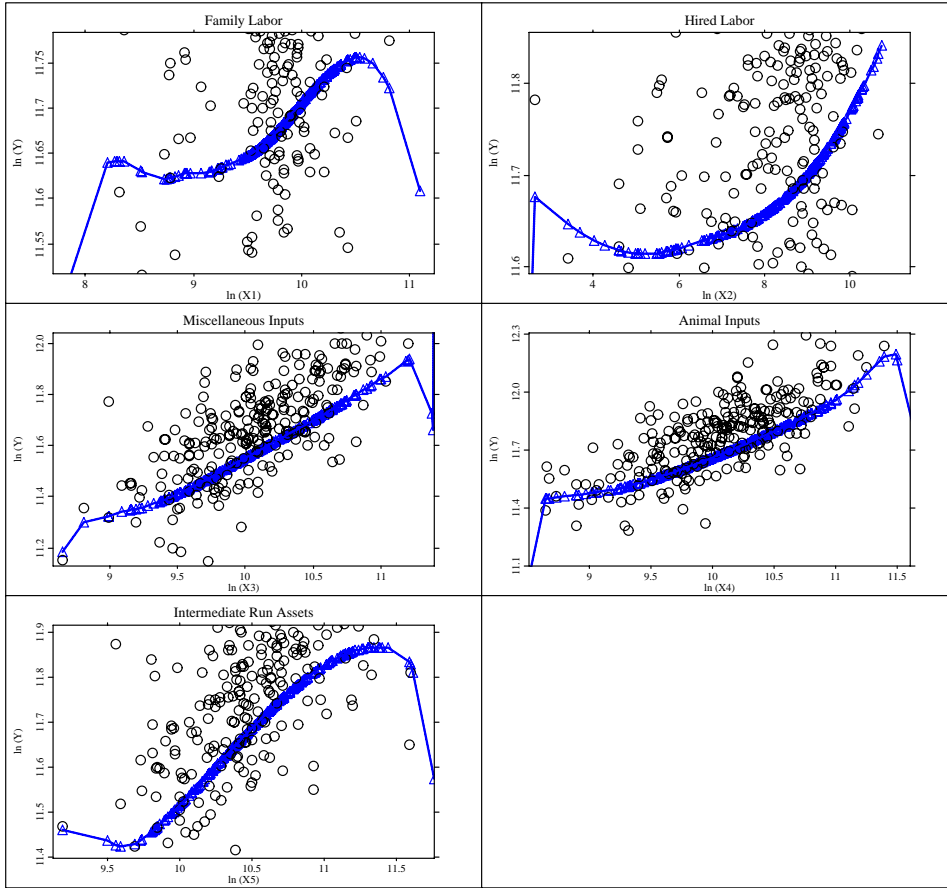


Figure 4: *Function estimates for the univariate additive components and partial residuals.*

In Figures 5 and 6 we have shown the estimates of the bivariate interaction terms $f_{\alpha\beta}$. For their estimation and presentation we trimmed the data by removing 2% of the most extreme data points. The interaction terms seem not to provide an (economic) interpretation.

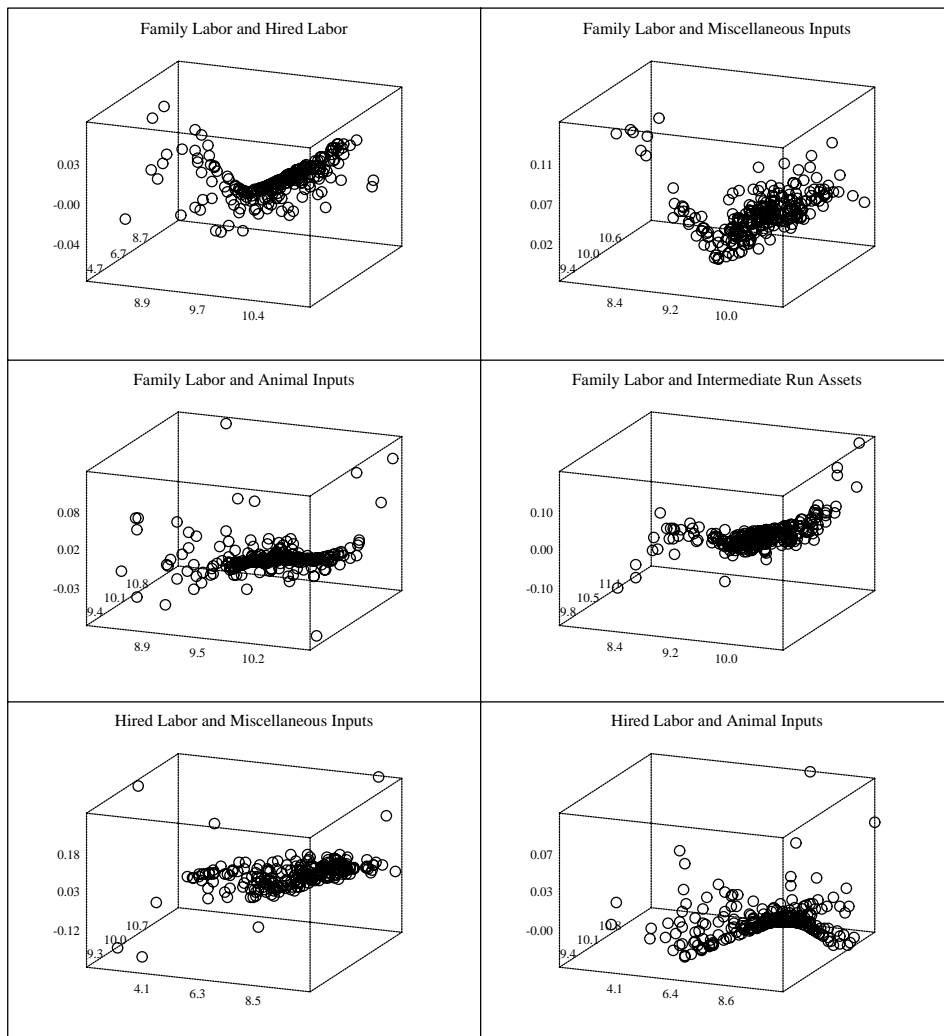


Figure 5: *Estimates of the interaction terms.*

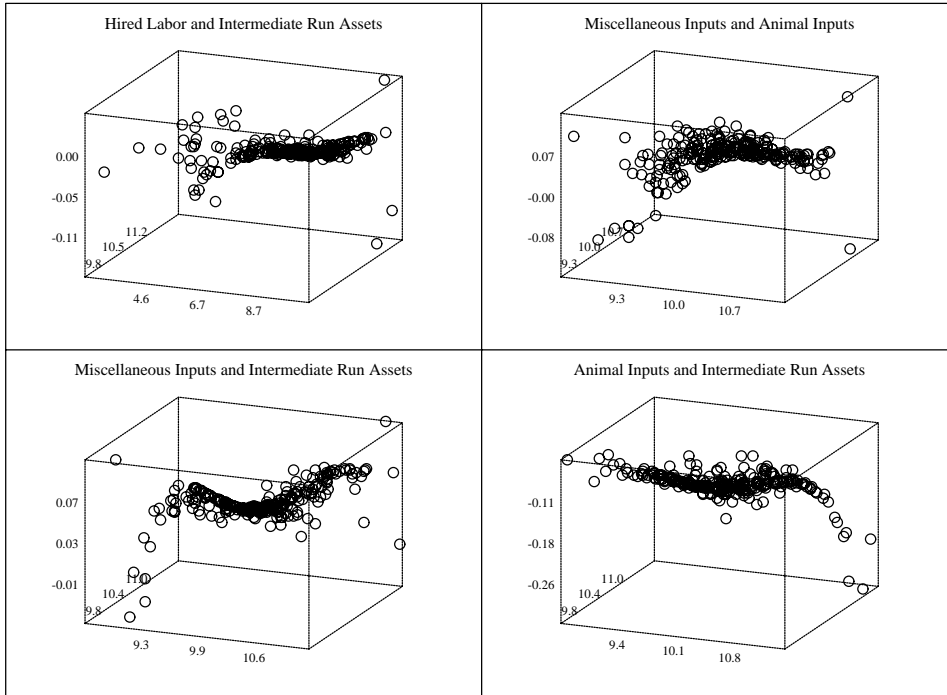


Figure 6: *Estimates of the interaction terms.*

For the testing we again used the quartic kernel and trimmed the data by removing 2% at the tails. We did 249 bootstrap replications.

Since we know about the sensitivity of the test procedures against the choice of bandwidths, we applied the procedures for a wide range of different bandwidths. For the first method, which employs the estimate of the interaction term directly, we used $h = 1.3$ to 2.1 , $g = 2.9$ to 3.7 for the pre-estimation to get estimates for the bootstrap and $h = 1.6$ to 2.4 , $g = 3.1$ to 3.9 to calculate the test statistics. For the second method which involves the mixed derivatives of the interaction term, we used $h = 1.6$ to 2.4 , $g = 3.1$ to 3.9 for the pre-estimation to get estimates for the bootstrap and $h = 2.1$ to 2.9 , $g = 3.1$ to 3.9 to calculate the test statistics.

To test the different interaction terms for significance, we used the following iterative model selection procedure. First we calculated the p-values for each interaction term $f_{\alpha\beta}$ including all the other functions in the model (19). Then we removed the function $f_{\alpha\beta}$ with the highest p-value, and again determined the p-values for the remaining interaction terms as above. Stepwise dropping out the interaction terms with the highest p-value, we end up with the most significant ones.

This procedure was applied for both testing methods. For large bandwidths the interactions are smoothed out and we never rejected the null hypothesis of no interaction for any of the pairwise terms, but for small bandwidths some of the interactions terms turned out to be significant. For the first method, where we consider the interaction terms directly, the term $f_{1,3}$ (family labor and miscellaneous inputs) was significant at a 5% level with a p-value of about 2%. Of the other terms $f_{1,3}$ and $f_{1,5}$ came closest to being significant.

For the second method, considering the derivatives, $f_{1,5}$ (family labor and intermediate run assets) and $f_{3,5}$ (miscellaneous inputs and intermediate run assets) had the lowest p-values, $f_{1,5}$ having a p-value of less than 1%.

Even though the two procedures are not entirely consistent in their selection of relevant interaction terms, both procedures indicate that a weak form of interaction is present, and that the variable *family labor* plays a significant role in the interaction. There are fairly clear indications from Figure 5 that $f_{1,3}$ and $f_{1,5}$ are not multiplicative in their input factors. This would make it difficult for a parametric test to detect the interaction.

A Appendix

A.1 Proof of Theorem 1

The proof of Theorem 1 makes use of the following lemmas, whose proofs are not difficult. They can be found in Severance-Lossin and Sperlich (1997).

Lemma A1 *Let D_n, B_n and A be matrices, possibly having random variables as their entries. Further, let $D_n = A + B_n$ where A^{-1} exists and $B_n = (b_{ij})_{1 \leq i, j \leq d}$ where $b_{ij} = O_p(\delta_n)$ with d fixed, independent of n . Then $D_n^{-1} = A^{-1}(I + C_n)$ where $C_n = (c_{ij})_{1 \leq i, j \leq d}$ and $c_{ij} = O_p(\delta_n)$. Here δ_n denotes a function of n , going to zero with increasing n .*

Lemma A2 *Let $W_{l,\alpha}, W_{l,\alpha\beta}, Z_\alpha, Z_{\alpha\beta}$ and S be defined as in Section 3.1 and $H = \text{diag}(h^{i-1})_{i=1, \dots, p+1}$. Then*

$$\text{a) } (H^{-1} Z_\alpha^T W_{l,\alpha} Z_\alpha H^{-1})^{-1} = \frac{1}{\varphi(x_\alpha, X_{l\alpha})} S^{-1} \left\{ I + O_p \left(h + \frac{\ln n}{\sqrt{nhg^{d-1}}} \right) \right\}$$

and

$$\text{b) } (H^{-1} Z_{\alpha\beta}^T W_{l,\alpha\beta} Z_{\alpha\beta} H^{-1})^{-1} = \frac{1}{\varphi(x_\alpha, x_\beta, X_{l\alpha\beta})} S^{-1} \left\{ I + O_p \left(h + \frac{\ln n}{\sqrt{nh^2g^{d-2}}} \right) \right\}.$$

Define $E_i[\cdot] = E[\cdot | X_{i1}, \dots, X_{id}]$ and $E_*[\cdot] = E[\cdot | \mathbf{X}]$, where \mathbf{X} is the design matrix $\{X_{i\alpha}\}_{i,\alpha=1,1}^{n,d}$. The proof of Theorem 1 can now be divided into two parts:

I) We start by considering the univariate estimator \hat{F}_α . This is one component of the estimator $\tilde{f}_{\alpha\beta}$ of interest in Theorem 1. First we will separate the difference between the estimator and the true function into a bias and a variance part.

Defining the vector

$$F_i = \begin{pmatrix} c + f_\alpha(x_\alpha) + \sum_{\gamma \in D_\alpha} f_{\alpha\gamma}(x_\alpha, X_{i\gamma}) + \sum_{\gamma \in D_\alpha} f_\gamma(X_{i\gamma}) + \sum_{(\gamma, \delta) \in D_{\alpha\alpha}} f_{\gamma\delta}(X_{i\gamma}, X_{i\delta}) \\ f_\alpha^{(1)}(x_\alpha) + \sum_{\gamma \in D_\alpha} f_{\alpha\gamma}^{(1,0)}(x_\alpha, X_{i\gamma}) \end{pmatrix}$$

and applying Lemma A2 **a**), we have

$$\begin{aligned} & \hat{F}_\alpha(x_\alpha) - F_\alpha(x_\alpha) \\ &= \frac{1}{n} \sum_{i=1}^n e_1 (Z_\alpha^T W_{i,\alpha} Z_\alpha)^{-1} Z_\alpha^T W_{i,\alpha} Y - F_\alpha(x_\alpha) \\ &= \frac{1}{n} \sum_{i=1}^n e_1 (Z_\alpha^T W_{i,\alpha} Z_\alpha)^{-1} Z_\alpha^T W_{i,\alpha} (Y - Z_\alpha F_i) + O_p(n^{-1/2}) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{\varphi(x_\alpha, X_{i\alpha})} e_1 S^{-1} \left\{ I + O_p \left(h + \frac{\ln n}{\sqrt{nhg^{d-1}}} \right) \right\} H^{-1} Z_\alpha^T W_{i,\alpha} (Y - Z_\alpha F_i) + O_p(n^{-\frac{1}{2}}) \end{aligned}$$

When we compute the matrix product and use for $Y_l = \sigma(X_l)\varepsilon_l + m(X_l)$ the Taylor expansion of $m(X_l)$ around $(x_\alpha, X_{l\alpha})$ we get

$$\begin{aligned} & \hat{F}_\alpha(x_\alpha) - F_\alpha(x_\alpha) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{\varphi(x_\alpha, X_{i\alpha})} \frac{1}{n} \sum_{l=1}^n K_h(X_{l\alpha} - x_\alpha) L_g(X_{l\alpha} - X_{i\alpha}) \left\{ 1 + O_p \left(h + \frac{\ln n}{\sqrt{nhg^{d-1}}} \right) \right\} \times \\ & \quad \left[\frac{(X_{l\alpha} - x_\alpha)^2}{2} \{ f_\alpha^{(2)}(x_\alpha) + \sum_{\gamma \in D_\alpha} f_{\alpha\gamma}^{(2,0)}(x_\alpha, X_{l\gamma}) \} + \sum_{\gamma \in D_\alpha} \{ f_\gamma(X_{l\gamma}) - f_\gamma(X_{i\gamma}) \} + \right. \\ & \quad \left. \sum_{(\gamma, \delta) \in D_{\alpha\alpha}} \{ f_{\gamma\delta}(X_{l\gamma}, X_{l\delta}) - f_{\gamma\delta}(X_{i\gamma}, X_{i\delta}) \} + O_p\{(X_{l\alpha} - x_\alpha)^3\} + \sigma(X_l)\varepsilon_l \right] + O_p(n^{-\frac{1}{2}}) \end{aligned}$$

Now we can separate this expression into a systematic "bias" and a stochastic "variance". Then we have

$$\hat{F}_\alpha(x_\alpha) - F_\alpha(x_\alpha) = \frac{1}{n} \sum_{i=1}^n \frac{E_i(\hat{a}_i)}{\varphi(x_\alpha, X_{i\alpha})} + \frac{1}{n} \sum_{i=1}^n \frac{\hat{a}_i - E_i(\hat{a}_i)}{\varphi(x_\alpha, X_{i\alpha})} + O_p \left(\frac{h}{\sqrt{nhg^{d-1}}} + \frac{\ln n}{nhg^{d-1}} \right)$$

where,

$$\hat{a}_i = \frac{1}{n} \sum_{l=1}^n K_h(X_{l\alpha} - x_\alpha) L_g(X_{l\alpha} - X_{i\alpha}) \times [\dots]$$

and the expression in the brackets $[\dots]$ is as in the formula above. Notice that \hat{a}_i is $O_p\{(nhg^{d-1})^{-1/2}\}$. It remains to work with the first order approximations.

Let

$$T_{1n} = \frac{1}{n} \sum_{i=1}^n \frac{E_i(\hat{a}_i)}{\varphi(x_\alpha, X_{i\alpha})} \quad ; \quad T_{2n} = \frac{1}{n} \sum_{i=1}^n \frac{\hat{a}_i - E_i(\hat{a}_i)}{\varphi(x_\alpha, X_{i\alpha})}.$$

For the bias part we prove that

$$T_{1n} = h^2 \mu_2(K) \frac{1}{2} \{f_\alpha^{(2)}(x_\alpha) + \sum_{\gamma \in D_\alpha} \frac{1}{n} \sum_{i=1}^n f_{\alpha\gamma}^{(2,0)}(x_\alpha, X_{i\gamma})\} + o_p(h^2).$$

Consider $\varphi(x_\alpha, X_{i\alpha})^{-1} E_i(\hat{a}_i)$, which is, in fact, an approximation of the (conditional) bias of the Nadaraya-Watson estimator at $(x_\alpha, X_{i\alpha})$. This is, by assumptions (A1), (A2), (A3) and (A5)

$$\begin{aligned} & \frac{E_i(\hat{a}_i)}{\varphi(x_\alpha, X_{i\alpha})} \\ &= \frac{1}{\varphi(x_\alpha, X_{i\alpha})} E_i \left[\frac{1}{n} \sum_{l=1}^n K_h(X_{l\alpha} - x_\alpha) L_g(X_{l\alpha} - X_{i\alpha}) \left[\frac{(X_{l\alpha} - x_\alpha)^2}{2} \right. \right. \\ & \quad \left. \left. \left\{ f_\alpha^{(2)}(x_\alpha) + \sum_{\gamma \in D_\alpha} f_{\alpha\gamma}^{(2,0)}(x_\alpha, X_{l\gamma}) \right\} + \sum_{\gamma \in D_\alpha} \{f_\gamma(X_{l\gamma}) - f_\gamma(X_{i\gamma})\} + \right. \right. \\ & \quad \left. \left. \sum_{(\gamma, \delta) \in D_{\alpha\alpha}} \{f_{\gamma\delta}(X_{l\gamma}, X_{l\delta}) - f_{\gamma\delta}(X_{i\gamma}, X_{i\delta})\} + O_p\{(X_{l\alpha} - x_\alpha)^3\} \right] \right] \\ &= \frac{1}{\varphi(x_\alpha, X_{i\alpha})} \int K_h(z - x_\alpha) L_g(w - X_{i\alpha}) \varphi(z, w) \times \left[\frac{(z - x_\alpha)^2}{2} \right. \\ & \quad \left. \left\{ f_\alpha^{(2)}(x_\alpha) + \sum_{\gamma \in D_\alpha} f_{\alpha\gamma}^{(2,0)}(x_\alpha, X_{l\gamma}) \right\} + \sum_{\gamma \in D_\alpha} \{f_\gamma(w_\gamma) - f_\gamma(X_{i\gamma})\} \right. \\ & \quad \left. + \sum_{(\gamma, \delta) \in D_{\alpha\alpha}} \{f_{\gamma\delta}(w_\gamma, w_\delta) - f_{\gamma\delta}(X_{i\gamma}, X_{i\delta})\} + O_p\{(z - x_\alpha)^3\} \right] dw dz + o_p(1) \\ &= \frac{1}{\varphi(x_\alpha, X_{i\alpha})} \int K(u) L(v) \varphi(x_\alpha + uh, X_{i\alpha} + vg) \times \left[\frac{(uh)^2}{2} \left\{ f_\alpha^{(2)}(x_\alpha) + \right. \right. \\ & \quad \left. \left. \sum_{\gamma \in D_\alpha} f_{\alpha\gamma}^{(2,0)}(x_\alpha, X_{l\gamma}) \right\} + O_p\{(uh)^3\} + \sum_{\gamma \in D_\alpha} \{f_\gamma(X_{i\gamma} + gv_\gamma) - f_\gamma(X_{i\gamma})\} \right. \\ & \quad \left. + \sum_{(\gamma, \delta) \in D_{\alpha\alpha}} \{f_{\gamma\delta}(X_{i\gamma} + gv_\gamma, X_{i\delta} + v_\delta g) - f_{\gamma\delta}(X_{i\gamma}, X_{i\delta})\} \right] dv du + o_p(1) \\ &= h^2 \mu_2(K) \frac{1}{2} \left\{ f_\alpha^{(2)}(x_\alpha) + \sum_{\gamma \in D_\alpha} f_{\alpha\gamma}^{(2,0)}(x_\alpha, X_{l\gamma}) \right\} + o_p(h^2) + O_p(g^q) \end{aligned}$$

since $E_*(\varepsilon_i) = 0$, respectively $E_i(\varepsilon_l) = 0$ for all i and l . We have used here the substitutions $u = \frac{z - x_\alpha}{h}$ and $v = \frac{w - X_{i\alpha}}{g}$, where v and w are $(d-1)$ -dimensional vectors with γ^{th} component v_γ , respectively w_γ .

Since the random variables $\varphi(x_\alpha, X_{i\alpha})^{-1} E_i(\hat{a}_i)$ are bounded we have using (A2)

$$T_{1n} = h^2 \mu_2(K) \frac{1}{2} \{f_\alpha^{(2)}(x_\alpha) + \sum_{\gamma \in D_\alpha} \frac{1}{n} \sum_{i=1}^n f_{\alpha\gamma}^{(2,0)}(x_\alpha, X_{i\gamma})\} + o_p(h^2)$$

For the stochastic term we use the same technique as in Severance-Lossin and Sperlich (1997) to prove that

$$(20) \quad w_{i\alpha} = \frac{1}{n} K_h(x_\alpha - X_{i\alpha}) \frac{\varphi_\alpha(X_{i\alpha})}{\varphi(x_\alpha, X_{i\alpha})},$$

$$(21) \quad T_{2n} = \sum_{i=1}^n w_{i\alpha} \sigma(X_i) \varepsilon_i + o_p\{(nh)^{-1/2}\}.$$

with a rate of \sqrt{nh} .

II) Analogous to the univariate case of \hat{F}_α , we proceed for the bivariate case considering $\hat{F}_{\alpha\beta}$:

We need the following definition

$$F_i = \begin{pmatrix} c + f_\alpha(x_\alpha) + f_\beta(x_\beta) + f_{\alpha\beta}(x_\alpha, x_\beta) + \sum_{\gamma \in D_{\alpha\beta}} \{\dots\} \\ f_\alpha^{(1)}(x_\alpha) + \sum_{\gamma \in D_{\alpha,\beta}} f_{\alpha\gamma}^{(1,0)}(x_\alpha, X_{i\gamma}) + f_{\alpha\beta}^{(1,0)}(x_\alpha, x_\beta) \\ f_\beta^{(1)}(x_\beta) + \sum_{\gamma \in D_{\alpha,\beta}} f_{\alpha\gamma}^{(1,0)}(x_\beta, X_{i\gamma}) + f_{\alpha\beta}^{(0,1)}(x_\alpha, x_\beta) \end{pmatrix}$$

where $\{\dots\}$ is

$$\left\{ f_{\alpha\gamma}(x_\alpha, X_{i\gamma}) + f_{\beta\gamma}(x_\beta, X_{i\gamma}) + f_\gamma(X_{i\gamma}) + \sum_{\gamma, \delta \in D_{\alpha\beta}} f_{\gamma\delta}(X_{i\gamma}, X_{i\delta}) \right\}.$$

Applying Lemma A2 **b)** we have for the estimator

$$\begin{aligned} & \hat{F}_{\alpha\beta}(x_\alpha, x_\beta) - F_{\alpha\beta}(x_\alpha, x_\beta) \\ &= \frac{1}{n} \sum_{i=1}^n e_1 (Z_{\alpha\beta}^T W_{i,\alpha\beta} Z_{\alpha\beta})^{-1} Z_{\alpha\beta}^T W_{i,\alpha\beta} Y - F_{\alpha\beta}(x_\alpha, x_\beta) \\ &= \frac{1}{n} \sum_{i=1}^n e_1 (Z_{\alpha\beta}^T W_{i,\alpha\beta} Z_{\alpha\beta})^{-1} Z_{\alpha\beta}^T W_{i,\alpha\beta} (Y - Z_{\alpha\beta} F_i) + O_p(n^{-\frac{1}{2}}) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{\varphi(x_\alpha, x_\beta, X_{i\alpha\beta})} e_1 \mathcal{S}^{-1} \left\{ I + O_p \left(h + \frac{\ln n}{\sqrt{nh^2 g^{d-2}}} \right) \right\} \\ & \quad \times H^{-1} Z_{\alpha\beta}^T W_{i,\alpha\beta} (Y - Z_{\alpha\beta} F_i) + O_p(n^{-\frac{1}{2}}). \end{aligned}$$

As above in I) we do the matrix calculation, replace Y_l by $Y_l = \sigma(X_l) \varepsilon_l + m(X_l)$ and use the Taylor expansion of m around $(x_\alpha, x_\beta, X_{l\alpha\beta})$. Then we get

$$(22) \quad \begin{aligned} & \hat{F}_{\alpha\beta}(x_\alpha, x_\beta) - F_{\alpha\beta}(x_\alpha, x_\beta) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{\varphi(x_\alpha, x_\beta, X_{i\alpha\beta})} \frac{1}{n} \sum_{l=1}^n K_h(X_{l\alpha} - x_\alpha) K_h(X_{l\beta} - x_\beta) L_g(X_{l\alpha\beta} - X_{i\alpha\beta}) \times \\ & \quad \left\{ I + O_p \left(h + \frac{\ln n}{\sqrt{nh^2 g^{d-2}}} \right) \right\} \left[\frac{(X_{l\alpha} - x_\alpha)^2}{2} \{f_\alpha^{(2)}(x_\alpha) + \sum_{\gamma \in D_{\alpha,\beta}} f_{\alpha\gamma}^{(2,0)}(x_\alpha, X_{l\gamma}) \right. \end{aligned}$$

$$\begin{aligned}
& + f_{\alpha\beta}^{(2,0)}(x_\alpha, x_\beta) \Big\} + \frac{(X_{l\beta} - x_\beta)^2}{2} \left\{ f_\beta^{(2)}(x_\beta) + \sum_{\gamma \in D_{\alpha,\beta}} f_{\beta\gamma}^{(2,0)}(x_\beta, X_{l\gamma}) \right. \\
& + f_{\alpha\beta}^{(0,2)}(x_\alpha, x_\beta) \Big\} + \sum_{\gamma \in D_{\alpha,\beta}} \{f_\gamma(X_{l\gamma}) - f_\gamma(X_{i\gamma})\} + \sum_{(\gamma,\delta) \in D_{\alpha\beta}} \{f_{\gamma\delta}(X_{l\gamma}, X_{l\delta}) \\
& - f_{\gamma\delta}(X_{i\gamma}, X_{i\delta})\} + (X_{l\alpha} - x_\alpha)(X_{l\beta} - x_\beta) f_{\alpha\beta}^{(1,1)}(x_\alpha, x_\beta) + O_p\{(X_{l\alpha} - x_\alpha)^3\} \\
& + O_p\{(X_{l\beta} - x_\beta)^3\} + o_p\{(X_{l\alpha} - x_\alpha)(X_{l\beta} - x_\beta)\} + \sigma(X_l)\varepsilon_l \Big] + O_p(n^{-\frac{1}{2}})
\end{aligned}$$

We pursue the same steps as for the one dimensional case and separate this expression into a systematic "bias" and a stochastic "variance":

$$\frac{1}{n} \sum_{i=1}^n \frac{E_i(\hat{a}_i)}{\varphi(x_\alpha, x_\beta, X_{i\underline{\alpha\beta}})} + \frac{1}{n} \sum_{i=1}^n \frac{\hat{a}_i - E_i(\hat{a}_i)}{\varphi(x_\alpha, x_\beta, X_{i\underline{\alpha\beta}})} + O_p\left(\frac{h}{\sqrt{nh^2g^{d-2}}} + \frac{\ln n}{nh^2g^{d-2}}\right)$$

where,

$$\hat{a}_i = \frac{1}{n} \sum_{l=1}^n K_h(X_{l\alpha} - x_\alpha) K_h(X_{l\beta} - x_\beta) L_g(X_{l\underline{\alpha\beta}} - X_{i\underline{\alpha\beta}}) \times [\dots]$$

and $[\dots]$ is the expression in the same angular brackets of equation (22). Notice that \hat{a}_i is $O_p\{(nh^2g^{d-2})^{-1/2}\}$.

Again we only have to work with the first order approximations.

Let

$$\mathcal{T}_{1n} = \frac{1}{n} \sum_{i=1}^n \frac{E_i(\hat{a}_i)}{\varphi(x_\alpha, x_\beta, X_{i\underline{\alpha\beta}})} \quad ; \quad \mathcal{T}_{2n} = \frac{1}{n} \sum_{i=1}^n \frac{\hat{a}_i - E_i(\hat{a}_i)}{\varphi(x_\alpha, x_\beta, X_{i\underline{\alpha\beta}})}.$$

We prove:

$$\begin{aligned}
\mathcal{T}_{1n} &= h^2 \mu_2(K) \frac{1}{2} \left\{ f_\alpha^{(2)}(x_\alpha) + \sum_{\gamma \in D_{\alpha,\beta}} \frac{1}{n} \sum_{i=1}^n f_{\alpha\gamma}^{(2,0)}(x_\alpha, X_{i\gamma}) + f_\beta^{(2)}(x_\beta) + \right. \\
& \left. \sum_{\gamma \in D_{\alpha,\beta}} \frac{1}{n} \sum_{i=1}^n f_{\beta\gamma}^{(2,0)}(x_\beta, X_{i\gamma}) + f_{\alpha\beta}^{(2,0)}(x_\alpha, x_\beta) + f_{\alpha\beta}^{(0,2)}(x_\alpha, x_\beta) \right\} + o_p(h^2)
\end{aligned}$$

Consider $\varphi(x_\alpha, x_\beta, X_{i\underline{\alpha\beta}})^{-1} E_i(\hat{a}_i)$, which is again an approximation of the (conditional) bias of the Nadaraya-Watson estimator at $(x_\alpha, x_\beta, X_{i\underline{\alpha\beta}})$. By assumptions (A1), (A2), (A3) and (A5) we have

$$\begin{aligned}
& \frac{E_i(\hat{a}_i)}{\varphi(x_\alpha, x_\beta, X_{i\underline{\alpha\beta}})} \\
&= \frac{1}{\varphi(x_\alpha, x_\beta, X_{i\underline{\alpha\beta}})} \int K_h(z_\alpha - x_\alpha) K_h(z_\beta - x_\beta) L_g(w - X_{i\underline{\alpha\beta}}) \varphi(z, w) \\
& \quad \left[\frac{(z_\alpha - x_\alpha)^2}{2} \left\{ f_\alpha^{(2)}(x_\alpha) + \sum_{\gamma \in D_{\alpha,\beta}} f_{\alpha\gamma}^{(2,0)}(x_\alpha, X_{l\gamma}) + f_{\alpha\beta}^{(2,0)}(x_\alpha, x_\beta) \right\} + \right.
\end{aligned}$$

$$\begin{aligned}
& \frac{(z_\beta - x_\beta)^2}{2} \left\{ f_\beta^{(2)}(x_\beta) + \sum_{\gamma \in D_{\alpha,\beta}} f_{\beta\gamma}^{(2,0)}(x_\beta, X_{l\gamma}) + f_{\alpha\beta}^{(0,2)}(x_\alpha, x_\beta) \right\} + \\
& \sum_{\gamma \in D_{\alpha,\beta}} \{f_\gamma(w_\gamma) - f_\gamma(X_{i\gamma})\} + \sum_{\gamma, \delta \in D_{\alpha\beta}} \{f_{\gamma\delta}(w_\gamma, w_\delta) - f_{\gamma\delta}(X_{i\gamma}, X_{i\delta})\} + \\
& (z_\alpha - x_\alpha)(z_\beta - x_\beta) f_{\alpha\beta}^{(1,1)}(x_\alpha, x_\beta) + O_p\{(z_\alpha - x_\alpha)^3\} + O_p\{(z_\beta - x_\beta)^3\} \\
& + o_p\{(z_\alpha - x_\alpha)(z_\beta - x_\beta)\} \Big] dw dz + o_p(1) \\
= & \frac{1}{\varphi(x_\alpha, x_\beta, X_{i\alpha\beta})} \int K(u_\alpha) K(u_\beta) L(w) \varphi(x_\alpha + u_\alpha h, x_\beta + u_\beta h, X_{i\alpha\beta} + v g) \times \\
& \left[\frac{(hu_\alpha)^2}{2} \{f_\alpha^{(2)}(x_\alpha) + \sum_{\gamma \in D_{\alpha,\beta}} f_{\alpha\gamma}^{(2,0)}(x_\alpha, X_{l\gamma}) + f_{\alpha\beta}^{(2,0)}(x_\alpha, x_\beta)\} + \frac{(hu_\beta)^2}{2} \right. \\
& \left. \{f_\beta^{(2)}(x_\beta) + \sum_{\gamma \in D_{\alpha,\beta}} f_{\beta\gamma}^{(2,0)}(x_\beta, X_{l\gamma}) + f_{\alpha\beta}^{(0,2)}(x_\alpha, x_\beta)\} + \sum_{\gamma \in D_{\alpha,\beta}} \{f_\gamma(X_{i\gamma} + gv_\gamma) \right. \\
& \left. - f_\gamma(X_{i\gamma})\} + \sum_{(\gamma, \delta) \in D_{\alpha\beta}} \{f_{\gamma\delta}(X_{i\gamma} + gv_\gamma, X_{i\delta} + gv_\delta) - f_{\gamma\delta}(X_{i\gamma}, X_{i\delta})\} + \right. \\
& \left. (hu_\alpha)(hu_\beta) f_{\alpha\beta}^{(1,1)}(x_\alpha, x_\beta) + O_p(h^3) \right] dv du + o_p(1) \\
= & h^2 \mu_2(K) \frac{1}{2} \{f_\alpha^{(2)}(x_\alpha) + \sum_{\gamma \in D_{\alpha,\beta}} f_{\alpha\gamma}^{(2,0)}(x_\alpha, X_{i\gamma}) + f_\beta^{(2)}(x_\beta) + \\
& \sum_{\gamma \in D_{\alpha,\beta}} f_{\beta\gamma}^{(2,0)}(x_\beta, X_{i\gamma}) + f_{\alpha\beta}^{(2,0)}(x_\alpha, x_\beta) + f_{\alpha\beta}^{(0,2)}(x_\alpha, x_\beta)\} + o_p(h^2) + O_p(g^q)
\end{aligned}$$

since $E_*[\varepsilon_i] = 0$. We have used here the substitutions $u = \frac{z - (x_\alpha, x_\beta)^T}{h}$ and $v = \frac{w - X_{i\alpha\beta}}{g}$, where v, w are $(d-2)$ -dimensional vectors with γ^{th} component v_γ, w_γ .

Since the $\varphi(x_\alpha, x_\beta, X_{i\alpha\beta})^{-1} E_i(\hat{a}_i)$ are independent and bounded we have

$$\begin{aligned}
\mathcal{T}_{1n} &= h^2 \mu_2(K) \frac{1}{2} \{f_\alpha^{(2)}(x_\alpha) + \sum_{\gamma \in D_{\alpha,\beta}} \frac{1}{n} \sum_{i=1}^n f_{\alpha\gamma}^{(2,0)}(x_\alpha, X_{i\gamma}) + f_\beta^{(2)}(x_\beta) + \\
& \sum_{\gamma \in D_{\alpha,\beta}} \frac{1}{n} \sum_{i=1}^n f_{\beta\gamma}^{(2,0)}(x_\beta, X_{i\gamma}) + f_{\alpha\beta}^{(2,0)}(x_\alpha, x_\beta) + f_{\alpha\beta}^{(0,2)}(x_\alpha, x_\beta)\} + o_p(h^2)
\end{aligned}$$

Thus, combining with the bias formulas obtained for $\hat{F}_\alpha(x_\alpha)$ and $\hat{F}_\beta(x_\beta)$, the bias of $\hat{F}_{\alpha\beta}(x_\alpha, x_\beta) - \hat{F}_\alpha(x_\alpha) - \hat{F}_\beta(x_\beta)$ is as claimed in Theorem 1

$$\begin{aligned}
h^2 B_1 &= h^2 \mu_2(K) \frac{1}{2} \left\{ f_{\alpha\beta}^{(2,0)}(x_\alpha, x_\beta) - \int f_{\alpha\beta}^{(2,0)}(x_\alpha, u_\beta) \varphi_{\underline{\alpha}}(u) du \right. \\
& \left. + f_{\alpha\beta}^{(0,2)}(x_\alpha, x_\beta) - \int f_{\alpha\beta}^{(0,2)}(u_\alpha, x_\beta) \varphi_{\underline{\beta}}(u) du \right\} + o_p(h^2).
\end{aligned}$$

We now turn to the variance part \mathcal{T}_{2n} . In Fan, Härdle, Mammen (1997) it is shown that for

$$(23) \quad w_{i\alpha\beta} = \frac{1}{n} K_h(x_\alpha - X_{i\alpha}, x_\beta - X_{i\beta}) \frac{\varphi_{\alpha\beta}(X_{i\alpha\beta})}{\varphi(x_\alpha, x_\beta, X_{i\alpha\beta})}$$

$$(24) \quad \mathcal{T}_{2n} = \sum_{i=1}^n w_{i\alpha\beta} \sigma(X_i) \varepsilon_i + o_p\{(nh^2)^{-1/2}\}$$

with a rate of $\sqrt{nh^2}$ and it also obeys a Central Limit Theorem.

Finally we want to calculate the variance of the combined estimator $\hat{F}_{\alpha\beta}(x_\alpha, x_\beta) - \hat{F}_\alpha(x_\alpha) - \hat{F}_\beta(x_\beta)$. Because of the faster rate of the stochastic term in I than the one in II, it is enough to consider II, i.e. $\sum_{i=1}^n w_{i\alpha\beta} \sigma(X_i) \varepsilon_i + o_p\{(nh^2)^{-1/2}\}$. It is easy to show that the variance is then

$$\|K_0^*\|_2^2 \int \sigma^2(x) \frac{\varphi_{\alpha\beta}^2(x_{\alpha\beta})}{\varphi(x)} dx_{\alpha\beta}$$

QED.

A.2 Proof of Theorem 2

This proof is analogous to that of Theorem 1 for the two dimensional terms. The main difference is that at the beginning the kernel $K(\cdot)$ has to be replaced by $K^*(\cdot)$, i. e. $K^*(u, v) = K(u, v)uv\mu_2^{-2}(K)$ and the weights are

$$(25) \quad w_{i\alpha\beta} = \frac{1}{nh^3} K_{3,h}^*(x_\alpha - X_{i\alpha}, x_\beta - X_{i\alpha}) \frac{\varphi_{\alpha\beta}(X_{i\alpha\beta})}{\varphi(x_\alpha, x_\beta, X_{i\alpha\beta})},$$

where $K_{3,h}^*(\cdot, \cdot) = \frac{1}{h^2} K_3^*(\frac{\cdot}{h}, \frac{\cdot}{h})$.

QED.

A.3 Proof of Theorem 3

Consider the decomposition

$$\int \hat{f}_{\alpha\beta}^2(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta = 2 \sum_{1 \leq i < j \leq n} H(X_i, \varepsilon_i, X_j, \varepsilon_j) + \sum_{i=1}^n H(X_i, \varepsilon_i, X_i, \varepsilon_i) +$$

$$\int \tilde{f}_{\alpha\beta}^2(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta + 2h^2 \int \tilde{f}_{\alpha\beta}(x_\alpha, x_\beta) B_1(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta + o_p(h^2)$$

in which

$$H(X_i, \varepsilon_i, X_j, \varepsilon_j) = \varepsilon_i \varepsilon_j \int \frac{1}{n^2} (w_{i\alpha\beta} - w_{i\alpha} - w_{i\beta})(w_{j\alpha\beta} - w_{j\alpha} - w_{j\beta}) \sigma(X_i) \sigma(X_j) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta$$

with $w_{i\alpha}$, $w_{i\beta}$ and $w_{i\alpha\beta}$ as in equation (20) and (23).

We first simplify $H(X_i, \varepsilon_i, X_j, \varepsilon_j)$ by substituting alternatively $u = (x_\alpha - X_{i\alpha})/h$, $v = (x_\beta - X_{i\beta})/h$

$$\begin{aligned} H(X_i, \varepsilon_i, X_j, \varepsilon_j) &= \frac{\varepsilon_i \varepsilon_j}{n^2} \int \left\{ K(u)K(v) \frac{\varphi_{\alpha\beta}(X_{i\alpha\beta})}{\varphi(X_i)h} - K(u) \frac{\varphi_\alpha(X_{i\alpha})}{\varphi(X_i)} - K(v) \frac{\varphi_{\alpha\beta}(X_{i\beta})}{\varphi(X_i)} \right\} \\ &\quad \times \left\{ K\left(u + \frac{X_{i\alpha} - X_{j\alpha}}{h}\right) K\left(v + \frac{X_{i\beta} - X_{j\beta}}{h}\right) \frac{\varphi_{\alpha\beta}(X_{j\alpha\beta})}{\varphi(X_{i\alpha}, X_{i\beta}, X_{j\alpha\beta})h} \right. \\ &\quad \left. - K\left(u + \frac{X_{i\alpha} - X_{j\alpha}}{h}\right) \frac{\varphi_\alpha(X_{j\alpha})}{\varphi(X_{i\alpha}, X_{j\alpha})} - K\left(v + \frac{X_{i\beta} - X_{j\beta}}{h}\right) \frac{\varphi_{\alpha\beta}(X_{j\beta})}{\varphi(X_{i\beta}, X_{j\beta})} \right\} \\ &\quad \times \sigma(X_i)\sigma(X_j)\varphi_{\alpha\beta}(X_{i\alpha}, X_{i\beta})dudv \{1 + o_p(1)\} \end{aligned}$$

Denoting by $K^{(r)}$ the r -fold convolution of the kernel K , one obtains

$$\sum_{1 \leq i < j \leq n} H(X_i, \varepsilon_i, X_j, \varepsilon_j) = \sum_{1 \leq i < j \leq n} (H_1 + H_2 + H_3 + H_4 + H_5) \{1 + o_p(1)\}$$

where

$$\begin{aligned} H_1 &= \frac{\varepsilon_i \varepsilon_j \sigma(X_i)\sigma(X_j)}{n^2 h^2} K^{(2)}\left(\frac{X_{i\alpha} - X_{j\alpha}}{h}\right) K^{(2)}\left(\frac{X_{i\beta} - X_{j\beta}}{h}\right) \varphi_{\alpha\beta}(X_{i\alpha\beta})\varphi_{\alpha\beta}(X_{j\alpha\beta}) \\ &\quad \times \left\{ \frac{1}{\varphi(X_i)\varphi(X_{i\alpha}, X_{i\beta}, X_{j\alpha\beta})} + \frac{1}{\varphi(X_j)\varphi(X_{j\alpha}, X_{j\beta}, X_{i\alpha\beta})} \right\} \\ H_2 &= -\frac{\varepsilon_i \varepsilon_j \sigma(X_i)\sigma(X_j)}{n^2 h} \frac{\varphi_{\alpha\beta}(X_{j\alpha\beta})}{\varphi(X_{i\alpha}, X_{i\beta}, X_{j\alpha\beta})} \left\{ K^{(2)}\left(\frac{X_{i\alpha} - X_{j\alpha}}{h}\right) \frac{\varphi_\alpha(X_{i\alpha})}{\varphi(X_i)} + K^{(2)}\left(\frac{X_{i\beta} - X_{j\beta}}{h}\right) \frac{\varphi_\beta(X_{i\beta})}{\varphi(X_i)} \right\} \\ &\quad - \frac{\varepsilon_i \varepsilon_j \sigma(X_i)\sigma(X_j)}{n^2 h} \frac{\varphi_{\alpha\beta}(X_{i\alpha\beta})}{\varphi(X_{j\alpha}, X_{j\beta}, X_{i\alpha\beta})} \left\{ K^{(2)}\left(\frac{X_{j\alpha} - X_{i\alpha}}{h}\right) \frac{\varphi_\alpha(X_{j\alpha})}{\varphi(X_j)} + K^{(2)}\left(\frac{X_{j\beta} - X_{i\beta}}{h}\right) \frac{\varphi_\beta(X_{j\beta})}{\varphi(X_j)} \right\} \\ H_3 &= -\frac{\varepsilon_i \varepsilon_j \sigma(X_i)\sigma(X_j)}{n^2 h} \frac{\varphi_{\alpha\beta}(X_{i\alpha\beta})}{\varphi(X_i)} \left\{ K^{(2)}\left(\frac{X_{i\alpha} - X_{j\alpha}}{h}\right) \frac{\varphi_\alpha(X_{j\alpha})}{\varphi(X_{i\alpha}, X_{j\alpha})} + K^{(2)}\left(\frac{X_{i\beta} - X_{j\beta}}{h}\right) \frac{\varphi_\beta(X_{j\beta})}{\varphi(X_{i\beta}, X_{j\beta})} \right\} \\ &\quad - \frac{\varepsilon_i \varepsilon_j \sigma(X_i)\sigma(X_j)}{n^2 h} \frac{\varphi_{\alpha\beta}(X_{j\alpha\beta})}{\varphi(X_j)} \left\{ K^{(2)}\left(\frac{X_{j\alpha} - X_{i\alpha}}{h}\right) \frac{\varphi_\alpha(X_{i\alpha})}{\varphi(X_{j\alpha}, X_{i\alpha})} + K^{(2)}\left(\frac{X_{j\beta} - X_{i\beta}}{h}\right) \frac{\varphi_\beta(X_{i\beta})}{\varphi(X_{j\beta}, X_{i\beta})} \right\} \\ H_4 &= \frac{\varepsilon_i \varepsilon_j \sigma(X_i)\sigma(X_j)}{n^2} \left\{ K^{(2)}\left(\frac{X_{i\alpha} - X_{j\alpha}}{h}\right) \frac{\varphi_\alpha(X_{i\alpha})}{\varphi(X_i)} \frac{\varphi_\alpha(X_{j\alpha})}{\varphi(X_{i\alpha}, X_{j\alpha})} + K^{(2)}\left(\frac{X_{i\beta} - X_{j\beta}}{h}\right) \frac{\varphi_\beta(X_{i\beta})}{\varphi(X_i)} \frac{\varphi_\beta(X_{j\beta})}{\varphi(X_{i\beta}, X_{j\beta})} \right\} \\ &\quad + \frac{\varepsilon_i \varepsilon_j \sigma(X_i)\sigma(X_j)}{n^2} \left\{ K^{(2)}\left(\frac{X_{j\alpha} - X_{i\alpha}}{h}\right) \frac{\varphi_\alpha(X_{j\alpha})}{\varphi(X_j)} \frac{\varphi_\alpha(X_{i\alpha})}{\varphi(X_{j\alpha}, X_{i\alpha})} + K^{(2)}\left(\frac{X_{j\beta} - X_{i\beta}}{h}\right) \frac{\varphi_\beta(X_{j\beta})}{\varphi(X_j)} \frac{\varphi_\beta(X_{i\beta})}{\varphi(X_{j\beta}, X_{i\beta})} \right\} \\ H_5 &= \frac{\varepsilon_i \varepsilon_j \sigma(X_i)\sigma(X_j)}{n^2} \left\{ \frac{\varphi_\alpha(X_{i\alpha})}{\varphi(X_i)} \frac{\varphi_\beta(X_{j\beta})}{\varphi(X_{i\beta}, X_{j\beta})} + \frac{\varphi_{\alpha\beta}(X_{i\beta})}{\varphi(X_i)} \frac{\varphi_\alpha(X_{j\alpha})}{\varphi(X_{i\alpha}, X_{j\alpha})} + \frac{\varphi_\alpha(X_{j\alpha})}{\varphi(X_j)} \frac{\varphi_\beta(X_{i\beta})}{\varphi(X_{j\beta}, X_{i\beta})} \right. \\ &\quad \left. + \frac{\varphi_{\alpha\beta}(X_{j\beta})}{\varphi(X_j)} \frac{\varphi_\alpha(X_{i\alpha})}{\varphi(X_{j\alpha}, X_{i\alpha})} \right\} \end{aligned}$$

All of these are symmetric and nondegenerate U -Statistics. We will derive the asymptotic variance of H_1 and one will be able to see in the process that all the other H_i 's are of higher order and thus negligible. Now we calculate

$$E \{ H_1^2(X_1, \varepsilon_1, X_2, \varepsilon_2) \} = \frac{1}{n^4 h^4} \int K^{(2)2} \left(\frac{z_{1\alpha} - z_{2\alpha}}{h} \right) K^{(2)2} \left(\frac{z_{1\beta} - z_{2\beta}}{h} \right) \varphi_{\underline{\alpha\beta}}^2(z_{1\underline{\alpha\beta}}) \varphi_{\underline{\alpha\beta}}^2(z_{2\underline{\alpha\beta}}) \\ \times \left\{ \frac{1}{\varphi(z_1) \varphi(z_{1\alpha}, z_{1\beta}, z_{2\underline{\alpha\beta}})} + \frac{1}{\varphi(z_2) \varphi(z_{2\alpha}, z_{2\beta}, z_{1\underline{\alpha\beta}})} \right\}^2 \sigma^2(z_1) \sigma^2(z_2) \varphi(z_1) \varphi(z_2) dz_1 dz_2.$$

Introducing the change of variable

$$z_{2\alpha} = z_{1\alpha} - hu, \quad z_{2\beta} = z_{1\beta} - hv$$

we obtain

$$E \{ H_1^2(X_1, \varepsilon_1, X_2, \varepsilon_2) \} = \frac{1}{n^4 h^2} \int K^{(2)2}(u) K^{(2)2}(v) \varphi_{\underline{\alpha\beta}}^2(z_{1\underline{\alpha\beta}}) \varphi_{\underline{\alpha\beta}}^2(z_{2\underline{\alpha\beta}}) \sigma^2(z_1) \sigma^2(z_{1\alpha}, z_{1\beta}, z_{2\underline{\alpha\beta}}) \\ \left\{ \frac{1}{\varphi(z_1) \varphi(z_{1\alpha}, z_{1\beta}, z_{2\underline{\alpha\beta}})} + \frac{1}{\varphi(z_{1\alpha}, z_{1\beta}, z_{2\underline{\alpha\beta}}) \varphi(z_{1\alpha}, z_{1\beta}, z_{1\underline{\alpha\beta}})} \right\}^2 \varphi(z_1) \varphi(z_{1\alpha}, z_{1\beta}, z_{2\underline{\alpha\beta}}) dz_1 du dv dz_{2\underline{\alpha\beta}} \{1 + o(1)\},$$

or

$$E \{ H_1^2(X_1, \varepsilon_1, X_2, \varepsilon_2) \} = \frac{4}{n^4 h^2} \left\| K^{(2)} \right\|_2^4 \int \frac{\varphi_{\underline{\alpha\beta}}^2(z_{1\underline{\alpha\beta}})}{\varphi(z_1)} \frac{\varphi_{\underline{\alpha\beta}}^2(z_{2\underline{\alpha\beta}})}{\varphi(z_{1\alpha}, z_{1\beta}, z_{2\underline{\alpha\beta}})} \times \\ \sigma^2(z_1) \sigma^2(z_{1\alpha}, z_{1\beta}, z_{2\underline{\alpha\beta}}) dz_1 dz_{2\underline{\alpha\beta}} \{1 + o(1)\}.$$

To prove that $\sum_{i < j} H_1(X_i, \varepsilon_i, X_j, \varepsilon_j)$ is asymptotically normal, one needs to show that

$$E \{ G_1^2(X_1, \varepsilon_1, X_2, \varepsilon_2) \} + n^{-1} E \{ H_1^4(X_1, \varepsilon_1, X_2, \varepsilon_2) \} = o_p \left[\{ E H_1^2(X_1, \varepsilon_1, X_2, \varepsilon_2) \}^2 \right]$$

where

$$(26) \quad G_1(x, \varepsilon, y, \delta) = E \{ H_1(X_1, \varepsilon_1, x, \varepsilon) H_1(X_1, \varepsilon_1, y, \delta) \},$$

see Hall (1984).

Lemma A3 As $h \rightarrow 0$ and $nh^2 \rightarrow \infty$

$$n^{-1} E \{ H_1^4(X_1, \varepsilon_1, X_2, \varepsilon_2) \} = O_p(n^{-9} h^{-6}) = o_p \left[\{ E H_1^2(X_1, \varepsilon_1, X_2, \varepsilon_2) \}^2 \right].$$

Proof. As in the case of the second moment, the fourth moment can be calculated as

$$E \{ H_1^4(X_1, \varepsilon_1, X_2, \varepsilon_2) \} = \frac{h^2}{n^8 h^8} \int K^{(2)4}(u) K^{(2)4}(v) \varphi_{\underline{\alpha\beta}}^4(z_{1\underline{\alpha\beta}}) \varphi_{\underline{\alpha\beta}}^4(z_{2\underline{\alpha\beta}}) \sigma^4(z_1) \sigma^4(z_{1\alpha}, z_{1\beta}, z_{2\underline{\alpha\beta}}) \\ \left\{ \frac{1}{\varphi(z_1) \varphi(z_{1\alpha}, z_{1\beta}, z_{2\underline{\alpha\beta}})} + \frac{1}{\varphi(z_{1\alpha}, z_{1\beta}, z_{2\underline{\alpha\beta}}) \varphi(z_{1\alpha}, z_{1\beta}, z_{1\underline{\alpha\beta}})} \right\}^4 \varphi(z_1) \varphi(z_{1\alpha}, z_{1\beta}, z_{2\underline{\alpha\beta}}) dz_1 du dv dz_{2\underline{\alpha\beta}} \{1 + o_p(1)\}$$

which implies that

$$n^{-1} E \{ H_1^4(X_1, \varepsilon_1, X_2, \varepsilon_2) \} = O_p(n^{-9} h^{-6}) = \{ E H_1^2(X_1, \varepsilon_1, X_2, \varepsilon_2) \}^2 O_p(n^{-1} h^{-2})$$

which proves the lemma as $n^{-1} h^{-2} \rightarrow 0$.

Lemma A4 As $h \rightarrow 0$ and $nh^2 \rightarrow \infty$

$$G_1(x, \varepsilon, y, \delta) = \frac{2\varepsilon\delta\varphi_{\underline{\alpha\beta}}(x_{\underline{\alpha\beta}})\varphi_{\underline{\alpha\beta}}(y_{\underline{\alpha\beta}})\sigma(x)\sigma(y)}{n^4h^2\varphi(x)}K^{(4)}\left(\frac{x_\alpha - y_\alpha}{h}\right)K^{(4)}\left(\frac{x_\beta - y_\beta}{h}\right) \times \\ \int \left\{ \frac{1}{\varphi(y)\varphi(y_\alpha, y_\beta, z_{\underline{\alpha\beta}})} + \frac{1}{\varphi(x_\alpha, x_\beta, z_{\underline{\alpha\beta}})\varphi(x_\alpha, x_\beta, y_{\underline{\alpha\beta}})} \right\} \varphi_{\underline{\alpha\beta}}^2(z_{\underline{\alpha\beta}})\sigma^2(x_\alpha, x_\beta, z_{\underline{\alpha\beta}})dz_{\underline{\alpha\beta}} \{1 + o(1)\}$$

Proof. According to the definition of G_1

$$G_1(x, \varepsilon, y, \delta) = E \{ H_1(X_1, \varepsilon_1, x, \varepsilon) H_1(X_1, \varepsilon_1, y, \delta) \} = \frac{\varepsilon\delta}{n^4h^4} \times \\ E \left[K^{(2)}\left(\frac{X_{1\alpha} - x_\alpha}{h}\right) K^{(2)}\left(\frac{X_{1\beta} - x_\beta}{h}\right) \left\{ \frac{1}{\varphi(x_1)\varphi(x_{1\alpha}, x_{1\beta}, X_{\underline{\alpha\beta}})} + \frac{1}{\varphi(X_1)\varphi(X_{1\alpha}, X_{1\beta}, x_{\underline{\alpha\beta}})} \right\} \right. \\ \left. \times \varphi_{\underline{\alpha\beta}}(X_{1\underline{\alpha\beta}})\varphi_{\underline{\alpha\beta}}(x_{\underline{\alpha\beta}})\sigma(X_1)\sigma(x) \right. \\ \left. \times K^{(2)}\left(\frac{X_{1\alpha} - y_\alpha}{h}\right) K^{(2)}\left(\frac{X_{1\beta} - y_\beta}{h}\right) \left\{ \frac{1}{\varphi(y_1)\varphi(y_{1\alpha}, y_{1\beta}, X_{\underline{\alpha\beta}})} + \frac{1}{\varphi(X_1)\varphi(X_{1\alpha}, X_{1\beta}, y_{\underline{\alpha\beta}})} \right\} \right. \\ \left. \times \varphi_{\underline{\alpha\beta}}(X_{1\underline{\alpha\beta}})\varphi_{\underline{\alpha\beta}}(y_{\underline{\alpha\beta}})\sigma(X_1)\sigma(y) \right]$$

or

$$G_1(x, \varepsilon, y, \delta) = \frac{\varepsilon\delta\varphi_{\underline{\alpha\beta}}(x_{\underline{\alpha\beta}})\varphi_{\underline{\alpha\beta}}(y_{\underline{\alpha\beta}})\sigma(x)\sigma(y)}{n^4h^4} \int \varphi_{\underline{\alpha\beta}}^2(z_{\underline{\alpha\beta}})\sigma^2(z) \\ \times K^{(2)}\left(\frac{z_\alpha - x_\alpha}{h}\right) K^{(2)}\left(\frac{z_\beta - x_\beta}{h}\right) \left\{ \frac{1}{\varphi(x)\varphi(x_\alpha, x_\beta, z_{\underline{\alpha\beta}})} + \frac{1}{\varphi(z)\varphi(z_\alpha, z_\beta, x_{\underline{\alpha\beta}})} \right\} \\ \times K^{(2)}\left(\frac{z_\alpha - y_\alpha}{h}\right) K^{(2)}\left(\frac{z_\beta - y_\beta}{h}\right) \left\{ \frac{1}{\varphi(y)\varphi(y_\alpha, y_\beta, z_{\underline{\alpha\beta}})} + \frac{1}{\varphi(z)\varphi(z_\alpha, z_\beta, y_{\underline{\alpha\beta}})} \right\} \varphi(z)dz.$$

Introducing the change of variable

$$z_\alpha = x_\alpha + hu, z_\beta = x_\beta + hv$$

we obtain

$$G_1(x, \varepsilon, y, \delta) = \frac{\varepsilon\delta\varphi_{\underline{\alpha\beta}}(x_{\underline{\alpha\beta}})\varphi_{\underline{\alpha\beta}}(y_{\underline{\alpha\beta}})\sigma(x)\sigma(y)}{n^4h^4} \int \varphi_{\underline{\alpha\beta}}^2(z_{\underline{\alpha\beta}})\sigma^2(x_\alpha, x_\beta, z_{\underline{\alpha\beta}}) \\ \times K^{(2)}(u) K^{(2)}(v) \left\{ \frac{1}{\varphi(x)\varphi(x_\alpha, x_\beta, z_{\underline{\alpha\beta}})} + \frac{1}{\varphi(x_\alpha, x_\beta, z_{\underline{\alpha\beta}})\varphi(x_\alpha, x_\beta, x_{\underline{\alpha\beta}})} \right\} K^{(2)}\left(u + \frac{x_\alpha - y_\alpha}{h}\right) \\ \times K^{(2)}\left(v + \frac{x_\beta - y_\beta}{h}\right) \left\{ \frac{1}{\varphi(y)\varphi(y_\alpha, y_\beta, z_{\underline{\alpha\beta}})} + \frac{1}{\varphi(x_\alpha, x_\beta, z_{\underline{\alpha\beta}})\varphi(x_\alpha, x_\beta, y_{\underline{\alpha\beta}})} \right\} \\ \times \varphi(x_\alpha, x_\beta, z_{\underline{\alpha\beta}})h^2dudvdz_{\underline{\alpha\beta}} \{1 + o(1)\}.$$

Using convolution notation, one has

$$G_1(x, \varepsilon, y, \delta) = \frac{\varepsilon\delta\varphi_{\underline{\alpha\beta}}(x_{\underline{\alpha\beta}})\varphi_{\underline{\alpha\beta}}(y_{\underline{\alpha\beta}})\sigma(x)\sigma(y)}{n^4h^2}K^{(4)}\left(\frac{x_\alpha - y_\alpha}{h}\right)K^{(4)}\left(\frac{x_\beta - y_\beta}{h}\right) \times$$

$$\int \left\{ \frac{1}{\varphi(x)\varphi(x_\alpha, x_\beta, z_{\underline{\alpha\beta}})} + \frac{1}{\varphi(x_\alpha, x_\beta, z_{\underline{\alpha\beta}})\varphi(x_\alpha, x_\beta, x_{\underline{\alpha\beta}})} \right\} \left\{ \frac{1}{\varphi(y)\varphi(y_\alpha, y_\beta, z_{\underline{\alpha\beta}})} + \frac{1}{\varphi(x_\alpha, x_\beta, z_{\underline{\alpha\beta}})\varphi(x_\alpha, x_\beta, y_{\underline{\alpha\beta}})} \right\}$$

$$\times \varphi_{\underline{\alpha\beta}}^2(z_{\underline{\alpha\beta}})\sigma^2(x_\alpha, x_\beta, z_{\underline{\alpha\beta}})\varphi(x_\alpha, x_\beta, z_{\underline{\alpha\beta}})dz_{\underline{\alpha\beta}} \{1 + o(1)\}$$

or

$$G_1(x, \varepsilon, y, \delta) = \frac{2\varepsilon\delta\varphi_{\underline{\alpha\beta}}(x_{\underline{\alpha\beta}})\varphi_{\underline{\alpha\beta}}(y_{\underline{\alpha\beta}})\sigma(x)\sigma(y)}{n^4h^2\varphi(x)}K^{(4)}\left(\frac{x_\alpha - y_\alpha}{h}\right)K^{(4)}\left(\frac{x_\beta - y_\beta}{h}\right)$$

$$\times \int \left\{ \frac{1}{\varphi(y)\varphi(y_\alpha, y_\beta, z_{\underline{\alpha\beta}})} + \frac{1}{\varphi(x_\alpha, x_\beta, z_{\underline{\alpha\beta}})\varphi(x_\alpha, x_\beta, y_{\underline{\alpha\beta}})} \right\} \varphi_{\underline{\alpha\beta}}^2(z_{\underline{\alpha\beta}})\sigma^2(x_\alpha, x_\beta, z_{\underline{\alpha\beta}})dz_{\underline{\alpha\beta}} \{1 + o(1)\}$$

which is what we set out to prove. By techniques used in the two previous lemmas, one has

Lemma A5 *As $h \rightarrow 0$ and $nh^2 \rightarrow \infty$*

$$E \{G_1(X_1, \varepsilon_1, X_2, \varepsilon_2)^2\} = O(n^{-8}h^{-2}) = o \left[\{EH_1(X_1, \varepsilon_1, X_2, \varepsilon_2)^2\}^2 \right].$$

Lemmas A3 and A5, plus the Martingale Central Limit Theorem of Hall (1984) implies

Proposition A1 *As $h \rightarrow 0$ and $nh^2 \rightarrow \infty$*

$$nh \sum_{1 \leq i < j \leq n} H(X_i, \varepsilon_i, X_j, \varepsilon_j) \xrightarrow{\mathcal{L}} N \left\{ 0, 2 \left\| K^{(2)} \right\|_2^4 \int \frac{\varphi_{\underline{\alpha\beta}}^2(z_{1\underline{\alpha\beta}})}{\varphi(z_1)} \frac{\varphi_{\underline{\alpha\beta}}^2(z_{2\underline{\alpha\beta}})}{\varphi(z_{1\underline{\alpha}}, z_{1\underline{\beta}}, z_{2\underline{\alpha\beta}})} \sigma^2(z_1)\sigma^2(z_{1\underline{\alpha}}, z_{1\underline{\beta}}, z_{2\underline{\alpha\beta}})dz_1dz_{2\underline{\alpha\beta}} \right\}.$$

The “diagonal” term $\sum_{i=1}^n H(X_i, \varepsilon_i, X_i, \varepsilon_i)$ has the following property

Proposition A2 *As $h \rightarrow 0$ and $nh^2 \rightarrow \infty$*

$$\sum_{i=1}^n H(X_i, \varepsilon_i, X_i, \varepsilon_i) = \frac{2\{K^{(2)}(0)\}^2}{nh^2} \int \frac{\varphi_{\underline{\alpha\beta}}^2(z_{\underline{\alpha\beta}})}{\varphi(z)} \sigma^2(z)dz + O_p \left(\frac{1}{\sqrt{nnh^2}} \right)$$

Proof. This follows by simply calculating the mean and variance of $H(X_1, \varepsilon_1, X_1, \varepsilon_1)$.

Putting these results together, Theorem 3 is proved.

QED.

References

- ACHMUS, S. (1998) Nichtparametrische additive Modelle. *Thesis at Technische Universität Braunschweig, Germany*
- AUESTAD, B. H. AND TJØSTHEIM, D. (1991) Functional identification in nonlinear time series. In *G.G. Roussas (ed.), Nonparametric Functional Estimation and Related Topics, Amsterdam: Kluwer*, 483-507.
- BERNDT, E.R. AND CHRISTENSEN, L.R. (1973) The internal structure of functional relationships: separability, substitution, and aggregation. *Review of Economic Studies*, **40**: 403-410.
- BUJA, A, HASTIE, T.J. AND TIBSHIRANI, R.J. (1989) Linear smoothers and additive models. *Ann. Statist.*, **17**: 453-510.
- CHEN, R. , LIU, J. S. AND TSAY, R. S. (1995) Additivity tests for nonlinear autoregression. *Biometrika*, **82** : 369-383.
- DALELANE, C. (1998) Bootstrap Confidence Bands for the Integration Estimator in Additive Models. *Manuscript for Diploma, Humboldt-Universität zu Berlin, Germany*
- DEATON, A. AND MUELLBAUER, J. (1980) Economics and Consumer Behavior. *Cambridge University Press, New York*.
- DENNY, M. AND FUSS, M. (1977) The use of approximation analysis to test for separability and the existence of consistent aggregates. *The American Economic Review*, **67** : 404-418.
- EUBANK, R. L. , HART, J. D. , SIMPSON, D. G. AND STEFANSKI, L. A. (1995) Testing for additivity in nonparametric regression. *Ann. Statist.*, **23** : 1896-1920.
- FAN, J. , HÄRDLE, W. AND MAMMEN, E. (1997) Direct estimation of low dimensional components in additive models. To appear in the *Ann. Statist.*,
- FAN, J., GASSER, T., GIJBELS, I., BROCKMANN, M. AND ENGEL, J. (1993) Local polynomial fitting: A standard for nonparametric regression. *CORE Discussion Paper 9315*.
- FUSS, M., MCFADDEN, D. AND MUNDLAK, Y. (1978) A survey of functional forms in the economic analysis of production. In *Fuss, M. and McFadden, D. (eds.), Production Economics: A dual approach to theory and applications*, **1**: 219-268.
- GOZALO, P.L. AND LINTON, O.B. (1997) Testing additivity in generalized nonparametric regression models. *Preprint, Yale University*
- HÄRDLE, W. AND MAMMEN, E. (1993) Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, **21** : 1926-1947.
- HALL, P. (1984) Central limit theorem for integrated square error of multivariate nonparametric density estimators, *Journal of Multivariate Analysis*, **14** : 1-16.

- HASTIE, T. J. AND R. J. TIBSHIRANI. (1990) Generalized Additive Models. *Chapman and Hall: London*
- HJELLVIK, V. , YAO, Q. AND TJØSTHEIM, D. (1998) Linearity testing using local polynomial approximation. to appear in the *Journal of Statistical Planning and Inference*,
- LEONTIEF, W. (1947) Introduction to a theory of the internal structure of functional relationships. *Econometrika*, **15** : 361-373.
- LINTON, O.B AND NIELSEN, J.P. (1995) A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, **82** : 93-101.
- LINTON, O.B., MAMMEN, E. AND NIELSEN, J.P. (1997) The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Preprint, Yale University*
- LIU, R. (1988) Bootstrap procedures under some non i.i.d. models. *Ann. Statist.*, **16** : 1696-1708.
- MASRY, E. AND TJØSTHEIM, D. (1995) Nonparametric estimation and identification of nonlinear ARCH time series: strong convergence properties and asymptotic normality. *Econometric Theory*, **11** : 258-289.
- MASRY, E. AND TJØSTHEIM, D. (1996) Additive nonlinear ARX time series and projection estimates. *Econometric Theory*, **13** : 214-252.
- NADARAYA, E.A. (1964) On estimating regression. *Theor. Probab. Appl.*, **9** : 141-142.
- NEWWEY, W. K. (1994) Kernel estimation of partial means. *Econometric Theory*, **10** : 233-253.
- OPSOMER, J.D. AND RUPPERT, D. (1997) Fitting a bivariate additive model by local polynomial regression. *Ann. Statist.*, **25**: 186-211.
- RUPPERT, D. AND WAND, M. P. (1994) Multivariate locally weighted least squares regression. *Ann. Statist.*, **22** : 1346-1370.
- SEVERANCE-LOSSIN, E. AND SPERLICH, S. (1997) Estimation of derivatives for additive separable models. *Discussion Paper 30, sfb373, Humboldt-Universität Berlin, Germany* ,
- SILVERMAN, B.W. (1986) Density Estimation for Statistics and Data Analysis. *Chapman and Hall, London*
- SPERLICH, S., LINTON, O.B. AND HÄRDLE, W. (1997) A simulation comparison between integration and backfitting methods of estimating separable nonparametric regression models. *Discussion Paper 66, sfb373, Humboldt-Universität Berlin, Germany* ,
- STONE, C.J. (1985) Additive regression and other nonparametric models. *Ann. Statist.*, **13** : 689-705.
- STONE, C.J. (1986) The dimensionality reduction principle for generalized additive models. *Ann. Statist.*, **14** : 590-606.

- TJØSTHEIM, D. AND AUESTAD, B.H. (1994) Nonparametric identification of nonlinear time series: projections. *J. American Statistical Association*, **89** : 1398-1409.
- WATSON, G.S. (1964) Smooth regression analysis. *Sankhyā A*, **26** : 359-372.
- WU, C. F. J. (1986) Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist.*, **14** : 1261-1350.