

NONPARAMETRIC ESTIMATION AND TESTING OF INTERACTION
IN ADDITIVE MODELS

Stefan Sperlich, Dag Tjøstheim, Lijian Yang*

Abstract

We consider an additive model with second order interaction terms. Both marginal integration estimators and backfitting-integration efficient estimators are proposed for all components of the model and their derivatives, together with their explicitly derived asymptotic distributions. Moreover, two test statistics for testing the presence of interactions are proposed. Asymptotics for the test functions and local power results are obtained. Since direct implementation of the test procedure based on the asymptotics would produce inaccurate results unless the number of observations is very large, a bootstrap procedure is provided, which is applicable for small or moderate sample sizes. Further, based on these methods a general test for additivity is proposed. Estimation and testing methods are shown to work well in simulation studies. 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.

Key Words

Derivative Estimation; local linear; regression; local power; Sobolev Seminorm; Wild Bootstrap

*Sperlich, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Spain, e-mail: stefan@est-econ.uc3m.es; Tjøstheim, Department of Mathematics, University of Bergen, Norway; Yang, Department of Statistics and Probability, Michigan State University, U.S.A. We thank Stefan Profit and Oliver Linton for inspiration and helpful discussions. This research was financially supported by the SFB 373, Deutsche Forschungsgemeinschaft. Sperlich's work was also funded by the Spanish "Dirección General de Enseñanza Superior"(DGES), reference number PB98-0025. Yang's research was supported in part by NSF grant DMS 9971186.

1 Introduction

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

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

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 modeling complexity is to consider m as being additive, i.e.

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

where the functions f_{α} are unknown. Additive models in this general form was already discussed in Leontief (1947). He analyzed so called separable functions, i.e. functions which are characterized by the independence between the marginal rate of substitution for a pair of inputs and the changes in the level of another input. Subsequently the additivity assumption has been employed in several areas of economic theory, for example in connection with the separability hypothesis of production theory. Today additive models are widely used in both theoretical economics and empirical data analysis. They have a desirable statistical structure allowing econometric analysis for subsets of the regressors, permitting decentralization in optimizing and decision making and aggregation of inputs into indices. For more discussion, motivation and references see e.g. Fuss, McFadden and Mundlak (1978) or Deaton and Muellbauer (1980) which both devote substantial portions of their books to this topic and stress the importance of additive models in economics.

For statistics, especially when starting from a general nonparametric model such as (1), the usefulness of additive modeling has been emphasized among others by Stone, see e.g. Stone (1985). He points out that additive models yield a good compromise between the somewhat conflicting requirements of flexibility, dimensionality and interpretability. In particular, the curse of dimensionality can be treated in a satisfactory manner.

So far, additive models have mostly been estimated using backfitting (Hastie and Tibshirani 1990) combined with splines, 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

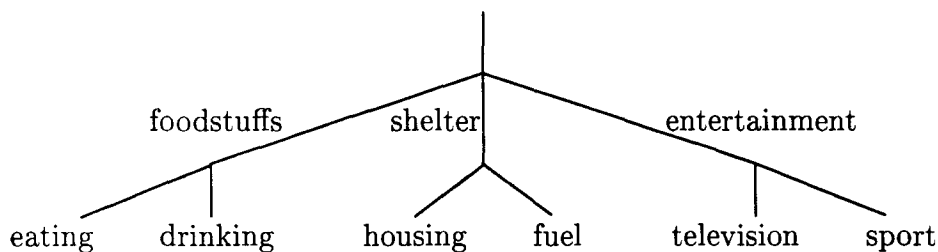
theory can be constructed. Combining marginal integration with a one-step backfit, Linton (1997) presented an efficient estimator. It should be remarked that important progress has also been made recently (Mammen, Linton and Nielsen 1999, Opsomer and Ruppert 1997) in the asymptotic theory of backfitting. Finally, the estimation of derivatives in additive nonparametric models is also of interest for economists, and it has been treated by Severance-Lossin and Sperlich (1999).

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. The lack of interaction terms may partly be due to the absence of appropriate testing procedures for testing simple interactions against purely additive models.

In this paper we allow for second order pairwise interactions resulting in a model of the form

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

Such a model is quite common in economics. However, parametric models have typically been used for the interactions, which may lead to wrong conclusions if the parametric form is incorrect. Examples for demand and utility functions can be found e.g. in Deaton and Muellbauer (1980). Imagine we want to model utility for household and consider the utility tree:



Example for utility tree for households.

In a nonparametric approach this would lead us to a model of the form

$$m(x) = c + \sum_{\alpha=1}^6 f_{\alpha}(x_{\alpha}) + f_{12}(x_1, x_2) + f_{34}(x_3, x_4) + f_{56}(x_5, x_6),$$

where the x_{α} stand for the inputs of the bottom line in the tree (counted from the left to the right). The interaction functions f_{12} stands for interaction in foodstuffs, f_{34} in shelter, and f_{56} for entertainment, whereas other interactions are not included in the estimation as they are considered as nonexistent.

In the context of production function estimation alternative functional forms as well as including interaction into the classic Cobb-Douglas model have been considered, resulting in the

$$\begin{aligned}
\text{Generalized Cobb-Douglas} \quad & \log Y = c + \sum_{\alpha=1}^d \sum_{\beta=1}^d c_{\alpha\beta} \ln \left(\frac{X_{\alpha} + X_{\beta}}{2} \right) \\
\text{Translog} \quad & \log Y = c + \sum_{\alpha=1}^d c_{\alpha} \ln X_{\alpha} + \sum_{\alpha=1}^d \sum_{\beta=1}^d c_{\alpha\beta} (\ln X_{\alpha})(\ln X_{\beta}) \\
\text{Generalized Leontief} \quad & Y = c + \sum_{\alpha=1}^d c_{\alpha} \sqrt{X_{\alpha}} + \sum_{\alpha=1}^d \sum_{\beta=1}^d c_{\alpha\beta} \sqrt{X_{\alpha} X_{\beta}} \\
\text{Quadratic} \quad & Y = c + \sum_{\alpha=1}^d c_{\alpha} X_{\alpha} + \sum_{\alpha=1}^d \sum_{\beta=1}^d c_{\alpha\beta} X_{\alpha} X_{\beta} \\
\text{Generalized Concave} \quad & Y = \sum_{\alpha=1}^d \sum_{\beta=1}^d c_{\alpha\beta} X_{\beta} f_{\alpha\beta} \left(\frac{X_{\alpha}}{X_{\beta}} \right), \quad f_{\alpha\beta} \text{ known and concave,}
\end{aligned}$$

which, although parametric, all have a functional form that is included in our model. For further discussion and references see Section 7.3, where we present a detailed example for production function estimation.

For model (3) we can give a closed asymptotic theory for both marginal integration and also a one-step efficient estimator analogous to that of Linton (1997). However, extending the remarkable work of Mammen, Linton and Nielsen (1999) on the asymptotic theory of backfitting seems difficult because of its strong dependence on projector theory, which would be hard to carry through for the interaction term.

It should be pointed out that estimation in such models has already been mentioned and discussed in the context of series estimator and backfitting with splines. For example Andrews and Whang (1990) give theoretical results using a series estimator. Hastie and Tibshirani (1990) discuss possible algorithms for backfitting with splines. Stone, Hansen, Kooperberg and Troung (1997) developed estimation theory for interaction of any order by polynomial spline methods. For further general references concerning series estimators in additive interaction models and concerning splines, see Newey (1995) and Wahba (1992), respectively. For the marginal integration method interactions have been briefly discussed in Tjøstheim and Auestad (1994).

The main objective of this paper is to consider estimation and testing in additive interaction models using marginal integration techniques. Again, the latter make it possible to give closed form expressions for the estimators, also for the derivatives, and to construct an explicit asymptotic theory. In addition we present simulation studies and an application to a production model.

It should be mentioned that the approach of Fan, Härdle, Mammen (1998) in estimating an additive partially linear model

$$m(x, z) = z^T \beta + c + \sum_{\alpha=1}^d f_{\alpha}(x_{\alpha})$$

can be applied relatively straightforwardly to our framework with interaction terms included. Such mixed models are interesting from a practical, as well as from a theoretical point of view, and they permit estimating β with the parametric \sqrt{n} -rate. Also, an extension to generalized additive and partially additive model should not be difficult to do. We refer to Linton, Härdle (1996) and Härdle, Huet, Mammen, Sperlich (1999) for a closer description of these models.

Although often modelling the regression model additively or at least the neglecting of some interaction terms is already justified by economic theory, from a statistical point of view this usually should be tested. The existing testing methods focus on full additivity, as in the references discussed at the end of this section. However, if full additivity is rejected, the empirical researcher would still like to know exactly which interaction terms are relevant. We propose 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

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

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 test can be carried out without estimating $f_{\alpha\beta}$ itself, 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 does not give 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 refer to the example of Section 7.3.

Other tests of additivity 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) or Derbort, Dette and Munk (1999) who both consider only fixed designs.

Our paper is divided into two main parts devoted to estimation and testing, respectively. In Section 2 we present our model in more detail and state some identifying assumptions. In Section 3 are given the marginal integration estimator for additive components and interactions, for derivatives, and subsequently, in Section 4 the corresponding one-step efficient estimators. The testing problem is introduced in Section 6 with two procedures for testing the significance of single interaction terms; also local power results are given. Finally, Section 7 provides several simulation studies and an application to real data. Most of the technical proofs have been relegated to the Appendix.

2 Some Simple Properties of the Model

In this section some basic assumptions and notations are introduced. We consider the additive interactive regression model

$$(5) \quad Y = c + \sum_{\alpha=1}^d f_{\alpha}(X_{\alpha}) + \sum_{1 \leq \alpha < \beta \leq d} f_{\alpha\beta}(X_{\alpha}, X_{\beta}) + \sigma(X)\varepsilon.$$

Here 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 $Var(\varepsilon) = 1$. We permit heteroskedasticity and the variance function is denoted by $\sigma^2(X)$. In the above expression 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, where for $\alpha = 1, 2, \dots, d$,

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

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

$$(7) \quad \int f_{\alpha\beta}(x_{\alpha}, x_{\beta}) \varphi_{\alpha}(x_{\alpha}) dx_{\alpha} = \int f_{\alpha\beta}(x_{\alpha}, x_{\beta}) \varphi_{\beta}(x_{\beta}) dx_{\beta} = 0,$$

with $\{\varphi_{\alpha}(\cdot)\}_{\alpha=1}^d$ being marginal densities (assumed to exist) of the X_{α} 's.

It is important to understand that equations (6) and (7) are only identifiability conditions and do not represent restrictions on our model. Indeed, if one thinks in a model of the form given in (3) or (5) but not satisfying (6) and (7), what would that mean? Just constant shifts of the additive components which in the end all cancel each other. For this imagine we take the following steps:

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

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

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

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

and adjust the $\{f_\beta(x_\beta)\}_{\beta=1}^d$'s and the constant term c accordingly so that $m()$ remains unchanged;

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 that $m()$ remains unchanged.

So we see that any model of the form (3) or (5) is equivalent to ours.

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

We now define by marginal integration

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

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

$$(9) \quad F_{\alpha\beta}(x_\alpha, x_\beta) = \int m(x_\alpha, x_\beta, x_{\underline{\alpha\beta}}) \varphi_{\underline{\alpha\beta}}(x_{\underline{\alpha\beta}}) dx_{\underline{\alpha\beta}},$$

for every pair $1 \leq \alpha < \beta \leq d$. Denote by D_α the subset of $\{1, 2, \dots, d\}$ with α 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

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

for every pair $1 \leq \alpha < \beta \leq d$. Then (6) and (7) entail the following lemma.

Lemma 1 *For model (5) the following equations for the marginals hold:*

$$\begin{aligned} 1) \quad & F_{\alpha}(x_{\alpha}) = f_{\alpha}(x_{\alpha}) + c + \sum_{(\gamma, \delta) \in D_{\alpha\alpha}} c_{\delta\gamma} \\ & F_{\alpha\beta}(x_{\alpha}, x_{\beta}) = f_{\alpha\beta}(x_{\alpha}, x_{\beta}) + f_{\alpha}(x_{\alpha}) + f_{\beta}(x_{\beta}) + c + \sum_{(\gamma, \delta) \in D_{\alpha\beta}} c_{\delta\gamma} \\ 2) \quad & F_{\alpha\beta}(x_{\alpha}, x_{\beta}) - F_{\alpha}(x_{\alpha}) - F_{\beta}(x_{\beta}) + \int m(x) \varphi(x) dx = f_{\alpha\beta}(x_{\alpha}, x_{\beta}) + c_{\alpha\beta} \\ 3) \quad & c_{\alpha\beta} = \int \{F_{\alpha\beta}(u, x_{\beta}) - F_{\alpha}(u)\} \varphi_{\alpha}(u) du - F_{\beta}(x_{\beta}) + \int m(x) \varphi(x) dx \\ & f_{\alpha\beta}(x_{\alpha}, x_{\beta}) = F_{\alpha\beta}(x_{\alpha}, x_{\beta}) - F_{\alpha}(x_{\alpha}) - \int \{F_{\alpha\beta}(u, x_{\beta}) - F_{\alpha}(u)\} \varphi_{\alpha}(u) du \end{aligned}$$

Proof.

1) Both formulas follow from the definitions of $D_{\alpha\alpha}$, $D_{\alpha\beta}$, $c_{\alpha\beta}$ and equations (8) and (9).

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 (7). The rest follows by the equation in 2). Q.E.D.

We define another auxiliary function

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

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

Corollary 1 *Let $f_{\alpha\beta}^*(x_{\alpha}, x_{\beta})$ and $f_{\alpha\beta}(x_{\alpha}, x_{\beta})$ be as defined above. Then*

$$f_{\alpha\beta}^*(x_{\alpha}, x_{\beta}) \equiv 0 \iff f_{\alpha\beta}(x_{\alpha}, x_{\beta}) \equiv 0.$$

Proof.

First assume that $f_{\alpha\beta}^*(x_\alpha, x_\beta) \equiv 0$. 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 (7) 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 $f_{\alpha\beta}(x_\alpha, x_\beta) \equiv 0$.

On the other hand, 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 $f_{\alpha\beta}(x_\alpha, x_\beta)$; namely the functional

$$\int f_{\alpha\beta}^{*2}(x_\alpha, x_\beta)\pi(x_\alpha, x_\beta)dx_\alpha dx_\beta,$$

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:

$$(10) \quad \int \widehat{f_{\alpha\beta}^{*2}}(x_\alpha, x_\beta)\varphi_{\alpha\beta}(x_\alpha, x_\beta)dx_\alpha dx_\beta$$

where

$$(11) \quad \widehat{f_{\alpha\beta}^*}(x_\alpha, x_\beta) = \widehat{F}_{\alpha\beta}(x_\alpha, x_\beta) - \widehat{F}_\alpha(x_\alpha) - \widehat{F}_\beta(x_\beta) + \frac{1}{n} \sum_{j=1}^n Y_j.$$

with estimates $\widehat{F}_{\alpha\beta}$, \widehat{F}_α and \widehat{F}_β of $F_{\alpha\beta}$, F_α and 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, under the identifiability condition (7), is equivalent to $f_{\alpha\beta} = 0$.

3 Marginal Integration Estimation

3.1 Estimation of the additive components and interactions using marginal integration

To use the marginal integration type statistic (10), 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 $K_h(\cdot) = h^{-1}K(\cdot/h)$ and $L_g(\cdot) = g^{-1}L(\cdot/g)$. We will give more detailed descriptions of the kernels K and L and the bandwidths h and g in subsequent 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. Proofs can be found in the appendix.

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:

$$(12) \quad \widehat{F}_{\alpha\beta}(x_\alpha, x_\beta) = \frac{1}{n} \sum_{l=1}^n \widehat{m}(x_\alpha, x_\beta, X_{l\underline{\alpha\beta}}) \quad , \quad \widehat{F}_\alpha(x_\alpha) = \frac{1}{n} \sum_{l=1}^n \widehat{m}(x_\alpha, X_{l\underline{\alpha}}),$$

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

$$(13) \quad \sum_{i=1}^n \{Y_i - a_0 - a_1(X_{i\alpha} - x_\alpha) - a_2(X_{i\beta} - x_\beta)\}^2 K_h(X_{i\alpha} - x_\alpha, X_{i\beta} - x_\beta) L_g(X_{i\underline{\alpha\beta}} - X_{l\underline{\alpha\beta}})$$

for each fixed l . Accordingly we define

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

where

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

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

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

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

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

in which

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

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

This estimator results from minimizing

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

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 concept 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 a "higher order kernel" given by

$$(14) \quad K_\nu^*(u) := \sum_{t=0}^p s_{\nu t} u^t K(u)$$

in the one-dimensional case, where $S = (\int u^{t+s} K(u) du)_{0 \leq t, s \leq p}$ and $S^{-1} = (s_{\nu t})_{0 \leq \nu, t \leq p}$ and where p is chosen according to need. Estimates of derivatives of m can then be obtained by choosing appropriate rows of S^{-1} . If for instance $p = 1$, we have

$$S^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \mu_2^{-1}(K) \end{pmatrix},$$

with $\mu_2(K) = \int u^2 K(u) du$.

To estimate the functions f_α , (or m) itself ($\nu = 0$) we use a local linear smoother and have simply $K_0^*(u) = K(u)$.

We can now state the first main result for estimation in our additive interactive regression model. For this, we need the following assumptions:

(A01) The kernels $K(\cdot)$ and $L(\cdot)$ are bounded, symmetric, compactly supported and Lipschitz continuous while the nonnegative $K(\cdot)$ satisfies $\int K(u)du = 1$. The $(d-1)$ -dimensional kernel $L(\cdot)$ is a product of univariate kernels $L(u)$ of order $q \geq 2$, i.e.

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

(A02) Bandwidths satisfy $\frac{nhg^{(d-1)}}{\ln(n)} \rightarrow \infty$, $\frac{g^q}{h^2} \rightarrow 0$ and $h = h_0 n^{-\frac{1}{5}}$.

(A3) The functions f_α , $f_{\alpha\beta}$ have bounded Lipschitz continuous derivatives of order $\max(p+1, q)^{\text{th}}$

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

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

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

Theorem 1 Let (x_α) be in the interior of the support of $\varphi_\alpha(\cdot)$. Then under conditions (A01)-(A02), (A3)-(A5),

$$(15) \quad \sqrt{nh} \{ \widehat{F}_\alpha(x_\alpha) - F_\alpha(x_\alpha) - h^2 b_1(x_\alpha) \} \xrightarrow{\mathcal{L}} N\{0, v_1(x_\alpha)\},$$

where F_α is given by (8) and Lemma 1, \widehat{F}_α by (12). The variance is

$$v_1(x_\alpha) = \|K\|_2^2 \int \sigma^2(x) \frac{\varphi_\alpha^2(x_\alpha)}{\varphi(x)} dx_\alpha$$

and the bias

$$b_1(x_\alpha) = \frac{\mu_2(K)}{2} f_\alpha^{(2)}(x_\alpha).$$

We now have almost everything at hand to estimate the interaction terms, again using local linear smoothers. For the two-dimensional local linear ($p = 1$) case the equivalent kernel is

$$(16) \quad K_\nu^*(u, v) := K(u, v) s_\nu(1, u, v)^T,$$

with s_ν , $0 \leq \nu \leq 2$, being the $(\nu + 1)^{\text{th}}$ row of

$$S^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \mu_2^{-1} & 0 \\ 0 & 0 & \mu_2^{-1} \end{pmatrix}.$$

Using a local linear smoother we have $K_0^*(u, v) = K(u, v)$, 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 (11). Since we have a two-dimensional problem, the assumptions have to be adjusted accordingly:

(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 while the nonnegative $K(\cdot)$ satisfies $\int K(u)du = 1$. The $(d - 1)$ -, respectively $(d - 2)$ -dimensional kernel $L(\cdot)$ is also a product of univariate kernels $L(u)$ of order $q \geq 2$.*

(A2) *Bandwidths satisfy $\frac{nh^2g^{(d-2)}}{\ln^2(n)} \rightarrow \infty$, and $\frac{nhg^{(d-1)}}{\ln^2(n)} \rightarrow \infty$, $\frac{g^q}{h^2} \rightarrow 0$ and $h = h_0n^{-\frac{1}{d}}$.*

Theorem 2 *Let (x_α, x_β) be in the interior of the support of $\varphi_{\alpha\beta}(\cdot)$. Then under conditions (A1)-(A5),*

$$(17) \quad \sqrt{nh^2}\{\widehat{f_{\alpha\beta}^*}(x_\alpha, x_\beta) - f_{\alpha\beta}^*(x_\alpha, x_\beta) - h^2B_1(x_\alpha, x_\beta)\} \xrightarrow{\mathcal{L}} N\{0, V_1(x_\alpha, x_\beta)\},$$

where $\widehat{f_{\alpha\beta}^*}$ is given by (11) 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

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

Theorems 1 and 2 are concerned with the individual components. The last result of this sub-section states the asymptotics of the combined regression estimator $\tilde{m}(x)$ of $m(x)$ given by

$$(18) \quad \tilde{m}(x) = \sum_{\alpha=1}^d \widehat{F}_\alpha(x_\alpha) + \sum_{1 \leq \alpha < \beta \leq d} \widehat{f_{\alpha\beta}^*}(x_\alpha, x_\beta) - (d-1) \frac{1}{n} \sum_{i=1}^n Y_i.$$

and state

Theorem 3 *Let x be in the interior of the support of $\varphi(\cdot)$. Then under conditions (A1), (A3)-(A5) and choosing bandwidths as in (A02) for the one-, (A2) for the two dimensional component functions, it holds*

$$(19) \quad \sqrt{nh^2} \left\{ \widehat{m}(x) - m(x) - h^2 B_m(x) \right\} \xrightarrow{\mathcal{L}} N \{0, V_m(x)\},$$

where h is as in (A2), $B_m(x) = \sum_{1 \leq \alpha < \beta \leq d} B_1(x_\alpha, x_\beta)$ and $V_m(x) = \sum_{1 \leq \alpha < \beta \leq d} V_1(x_\alpha, x_\beta)$.

3.2 Estimation of derivatives

Since the estimation of derivatives for additive separable models has already been considered in the paper of Severance-Lossin and Sperlich (1999), in this section we concentrate on estimating the mixed derivatives of the function $F_{\alpha\beta}$. Our interest in this estimator is motivated by testing the hypothesis of additivity without 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 under the identifiability condition (7).

Following the ideas of the previous section at the point $(x_\alpha, x_\beta, X_{i\alpha\beta})$ we implement a special version of the local polynomial estimator. For our purpose it is enough to use a bivariate local quadratic ($p = 2$) estimator. We want to minimize

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

and accordingly define our estimator by

$$(20) \quad \widehat{F}_{\alpha\beta}^{(1,1)}(x_\alpha, x_\beta) = \frac{1}{n} \sum_{i=1}^n e_4 \left(Z_{\alpha\beta}^T W_{i,\alpha\beta} Z_{\alpha\beta} \right)^{-1} Z_{\alpha\beta}^T W_{i,\alpha\beta} Y$$

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

Thus in equation (20), $Z_{\alpha\beta}$ 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 quadratic ones.

Recalling the approach of the preceding sub-section we can now put the equivalent kernel K^* to effective use. Using a local quadratic smoother 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

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

where $\mu_j = \mu_j(K) = \int u^j K(u) du$. The relationship between S^{-1} and $(Z_{\alpha\beta}^T W_{l,\alpha\beta} Z_{\alpha\beta})^{-1}$ 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

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

To state the asymptotics for the joint derivative estimator we need bandwidth conditions that differ slightly from (A2). In fact, more smoothing is required:

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

Then we have:

Theorem 4 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

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

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.

Comparing to Theorems 1 to 3, it is seen that the rate of convergence for the derivative estimator is slower than for the direct estimator.

4 A one-step "Efficient Estimator"

It is known that for additive models of the form

$$(21) \quad E[Y|X = x] = m(x_1, \dots, x_d) = c + \sum_{\alpha=1}^d f_\alpha(x_\alpha)$$

the marginal integration estimator is not efficient if the regressors are correlated. It is inefficient in the sense that if f_2, \dots, f_d are known, then the function f_1 could be estimated with a smaller variance applying a simple one dimensional smoother on the partial residual

$$(22) \quad U_{i1} = Y_i - c - \sum_{\alpha=2}^d f_\alpha(X_{i\alpha}).$$

Basically, this is the idea of the (iterative) backfitting estimation procedure. Linton (1997,1999) suggested an estimator combining the backfitting with the marginal integration idea. He first performed the marginal integration procedure to obtain $\hat{f}_\alpha \forall \alpha$, and then derived the estimated partial residuals

$$(23) \quad \tilde{U}_{i1} = Y_i - c - \sum_{\alpha=2}^d \hat{f}_\alpha(X_{i\alpha}).$$

Finally, he applied a one-dimensional local linear smoother on the $\tilde{U}_{i\alpha}$. This is equivalent to a one-step backfit. Certainly, for this all the theory done in Section 3 is necessary before one can proceed as Linton (1997) suggested.

Assuming that we already know the true underlying model, we consider an extension of his approach to models of the form

$$(24) \quad m(x) = c + \sum_{\alpha=1}^d f_{\alpha}(x_{\alpha}) + \sum_{1 \leq \alpha < \beta \leq d} f_{\alpha\beta}(x_{\alpha}, x_{\beta}) .$$

This ought to be of some interest, since in contradistinction to the case of no interaction, for a pure backfitting procedure, analogous to Hastie and Tibshirani (1990) or Mammen et al (1999), it is not even clear how a consistent estimate should look like. Hastie and Tibshirani discussed this topic but only for one interaction term and they can not give more than some intuitive motivation for their methods.

In contrast to Linton (1997) we do not restrict ourselves to homoskedastic errors but let $\sigma_{\alpha}^2(x_{\alpha}) = \text{Var}[Y - m(x)|X_{\alpha} = x_{\alpha}]$, with $\sigma_{\alpha\beta}(x_{\alpha}, x_{\beta})$ defined analogously and assume the existence of finite second moments for them. Consider model (24) and the two partial residuals

$$(25) \quad \begin{aligned} U_{i\alpha} &= Y_i - \sum_{\gamma \neq \alpha} f_{\gamma}(X_{i\gamma}) - \sum_{1 \leq \gamma < \beta \leq d} f_{\gamma\beta}(X_{i\gamma}, X_{i\beta}) + \sum_{(\gamma, \delta) \in D_{\alpha\alpha}} c_{\gamma\delta} \\ &= Y_i - m(X_i) + F_{\alpha}(X_{i\alpha}) \end{aligned}$$

$$(26) \quad \begin{aligned} U_{i\alpha\beta} &= Y_i - \sum_{\gamma=1}^d f_{\gamma}(X_{i\gamma}) - \sum_{\substack{1 \leq \gamma < \delta \leq d \\ (\gamma, \beta) \neq (\alpha, \beta)}} f_{\gamma\delta}(X_{i\gamma}, X_{i\delta}) + c_{\alpha\beta} \\ &= Y_i - m(X_i) + f_{\alpha\beta}^*(X_{i\alpha}, X_{i\beta}) \end{aligned}$$

For the estimation of the functional form it does not matter whether we correct for the constant before or after calculating the efficient estimator. To be consistent in our presentation with the preceding sections we have chosen the latter option. Further discussion to this topic can also be found in Linton (1997,1999).

Let now \check{F}_{α}^{opt} be the local linear regressor of $U_{i\alpha}$ in (25) with respect to X_{α} , and $\check{f}_{\alpha\beta}^{*opt}$ the one of $U_{i\alpha\beta}$ in (26) versus (X_{α}, X_{β}) . From Fan (1993), Ruppert and Wand (1994) we know under standard regularity conditions the asymptotic properties to be

$$(27) \quad \sqrt{nh_e} \left\{ \check{F}_{\alpha}^{opt}(x_{\alpha}) - F_{\alpha}(x_{\alpha}) - h_e^2 b_e(x_{\alpha}) \right\} \longrightarrow N \{0, v_e(x_{\alpha})\}$$

$$(28) \quad \sqrt{nh_e^2} \left\{ \check{f}_{\alpha\beta}^{*opt}(x_{\alpha}, x_{\beta}) - f_{\alpha\beta}^*(x_{\alpha}, x_{\beta}) - h_e^2 B_e(x_{\alpha}, x_{\beta}) \right\} \longrightarrow N \{0, V_e(x_{\alpha}, x_{\beta})\}$$

$$\text{with } b_e(x_{\alpha}) = \mu_2(J) \frac{1}{2} f^{(2)}(x_{\alpha}), \quad v_e(x_{\alpha}) = \|J\|_2^2 \sigma_{\alpha}^2(x_{\alpha}) \varphi_{\alpha}^{-1}(x_{\alpha})$$

$$B_e(x_{\alpha}, x_{\beta}) = \mu_2(J) \frac{1}{2} \left\{ f^{(2,0)}(x_{\alpha}, x_{\beta}) + f^{(0,2)}(x_{\alpha}, x_{\beta}) \right\},$$

$$V_e(x_{\alpha}, x_{\beta}) = \|J\|_2^2 \sigma_{\alpha\beta}^2(x_{\alpha}, x_{\beta}) \varphi_{\alpha\beta}^{-1}(x_{\alpha}, x_{\beta})$$

where J is the one or two dimensional kernel and h_e the corresponding bandwidth.

Now we replace $U_{i\alpha}$, $U_{i\alpha\beta}$ by $\tilde{U}_{i\alpha}$, $\tilde{U}_{i\alpha\beta}$ by substituting the real functions F_γ , $f_{\gamma\delta}^*$ by their marginal integration estimates defined in the preceding sections.

The efficient estimator $\check{F}_\alpha(x_\alpha)$ for $F_\alpha(x_\alpha)$ is defined as being the solution for c_0 in

$$(29) \quad \min_{c_0, c_1} \sum_{i=1}^n \left\{ \tilde{U}_{i\alpha} - c_0 - c_1(X_{i\alpha} - x_\alpha) \right\}^2 J \left(\frac{X_{i\alpha} - x_\alpha}{h_e} \right).$$

Similarly, for $f_{\alpha\beta}^*(x_\alpha, x_\beta)$ it is defined as being the solution for c_0 in

$$(30) \quad \min_{c_0, c_1, c_2} \sum_{i=1}^n \left\{ \tilde{U}_{i\alpha\beta} - c_0 - c_1(X_{i\alpha} - x_\alpha) - c_2(X_{i\beta} - x_\beta) \right\}^2 J \left(\frac{X_{i\alpha} - x_\alpha}{h_e}, \frac{X_{i\beta} - x_\beta}{h_e} \right).$$

Note that for reasons of notation we use J first as a one and later as a two dimensional kernel as should be obvious from the context.

From the discussion in Linton (1997) it is clear that slightly undersmoothing the marginal integration estimator, i.e. $h, g = o_p(n^{-1/5})$, leads to the desired result that asymptotically the 'efficient estimators' \check{F}_α and $\check{f}_{\alpha\beta}^*$ inherit the properties of \check{F}_α^{opt} , $\check{f}_{\alpha\beta}^{opt}$:

Theorem 5 *Suppose that conditions (A1) to (A5) hold, that the kernel J behaves like the kernel K , and g, h are at least $o_p(n^{-1/5})$, $h_e = Cn^{-1/5}$, $C > 0$. Then we have in probability*

$$(31) \quad \sqrt{nh_e} \left\{ \check{F}_\alpha(x_\alpha) - \check{F}_\alpha^{opt}(x_\alpha) \right\} \rightarrow 0$$

$$(32) \quad \sqrt{nh_e^2} \left\{ \check{f}_{\alpha\beta}^*(x_\alpha, x_\beta) - \check{f}_{\alpha\beta}^{opt}(x_\alpha, x_\beta) \right\} \rightarrow 0$$

for all $\alpha, \beta = 1, \dots, d$.

Notice also that the bias expression is the same for the 'efficient estimator' and the original marginal integration. Since the proof follows the arguments of Linton (1997,1999) we give only a sketch here. In the context of parametric estimation it can also be found in Cox, Hinkley (1974).

Proof.

We only discuss the statement in (27), since for (28) the reasoning is analogous. We have

$$\begin{aligned} & \check{F}_\alpha^{opt}(x_\alpha) - \check{F}_\alpha(x_\alpha) = \\ & = \sum_{i=1}^n \frac{w_i}{\sum_{i=1}^n w_i} \left[\sum_{\gamma \neq \alpha} \left\{ F_\gamma(X_{i\gamma}) - \hat{F}_\gamma(X_{i\gamma}) \right\} + \sum_{\gamma < \beta} \left\{ f_{\gamma\beta}^*(X_{i\gamma}, X_{i\beta}) - \hat{f}_{\gamma\beta}^*(X_{i\gamma}, X_{i\beta}) \right\} \right] \end{aligned}$$

$$\begin{aligned}
& +O_p(n^{-1/2}) \\
\text{with } & w_i = J\left(\frac{x_\alpha - X_{i\alpha}}{h_e}\right) \{s_{n2} - (x_\alpha - X_{i\alpha})s_{n1}\} \\
\text{and } & s_{nl} = \sum_{k=1}^n J\left(\frac{x_\alpha - X_{k\alpha}}{h_e}\right) (x_\alpha - X_{k\alpha})^l, \quad l = 1, 2.
\end{aligned}$$

We consider the differences for the one and two dimensional functions separately.

From the previous sections and the proof in the Appendix A.1 we know that the leading terms of $f_\gamma(X_{i\gamma}) - \hat{f}_\gamma(X_{i\gamma})$ are

$$h^2 b_1(X_{i\gamma}) + \frac{1}{n} \sum_{j=1}^n K_h(X_{j\gamma} - X_{i\gamma}) \frac{\varphi_\gamma(X_{j\gamma})}{\varphi(X_{i\gamma}, X_{j\gamma})} \sigma(X_j) \varepsilon_j$$

with

$$b_1(X_{i\gamma}) = \frac{\mu_2(K)}{2} \{f_\gamma^{(2)}(X_{i\gamma}) + \sum_{\delta \in D_\gamma} \frac{1}{n} \sum_{j=1}^n f_{\gamma\delta}^{(2,0)}(X_{i\gamma}, X_{j\delta})\}.$$

All these terms are additive over γ and the correlation between the marginal integration estimators is of smaller order. Multiplying $h^2 b_1(\cdot)$ with $\frac{w_i}{\sum_{j=1}^n w_j}$ and summing we still get a term of order h^2 , which by assumption is $o_p(n^{-2/5})$.

Also, if we consider the stochastic part, we can see that for $l = 0, 1$

$$\begin{aligned}
(33) \quad & \frac{1}{n} \sum_{i=1}^n J_{h_e}(x_\alpha - X_{i\alpha})(x_\alpha - X_{i\alpha})^l \times \frac{1}{n} \sum_{j=1}^n K_h(X_{j\gamma} - X_{i\gamma}) \frac{\varphi_\gamma(X_{j\gamma})}{\varphi(X_{i\gamma}, X_{j\gamma})} \sigma(X_j) \varepsilon_j = \\
& = \frac{1}{n} \sum_{j=1}^n \Delta_{nlj\gamma} \varphi_\gamma(X_{j\gamma}) \sigma(X_j) \varepsilon_j
\end{aligned}$$

where we have

$$\Delta_{nlj\gamma} = \frac{1}{n} \sum_{i=1}^n J_{h_e}(x_\alpha - X_{i\alpha})(x_\alpha - X_{i\alpha})^l K_h(X_{i\alpha} - X_{j\gamma}) \frac{1}{\varphi(X_{i\gamma}, X_{j\gamma})}$$

But $\Delta_{nlj\gamma}$ is bounded and the whole expression (33) is of order $O_p(n^{-1/2})$.

For the interaction terms it is almost the same. It is already shown that the leading terms of $f_{\gamma\delta}(X_{i\gamma}, X_{i\delta}) - \widehat{f}_{\gamma\delta}^*(X_{i\gamma}, X_{i\delta})$ are

$$h^2 B_1(X_{i\gamma}, X_{i\delta}) + \frac{1}{n} \sum_{j=1}^n K_h(X_{j\gamma} - X_{i\gamma}, X_{j\delta} - X_{i\delta}) \frac{\varphi_{\gamma\delta}(X_{j\gamma\delta})}{\varphi(X_{i\gamma}, X_{i\delta}, X_{j\gamma\delta})} \sigma(X_j) \varepsilon_j$$

with $B_1(\cdot)$ defined in Theorem 2.

Again, all of these terms are additive and asymptotically independent. Further, multiplying $h^2 B_1(\cdot)$ with $\frac{w_i}{\sum_{j=1}^n w_j}$ these terms stay of order $h^2 = o_p(n^{-2/5})$.

For the remaining term we have ($l = 0, 1$)

$$\frac{1}{n} \sum_{i=1}^n J_{h_\epsilon}(x_\alpha - X_{i\alpha})(x_\alpha - X_{i\alpha})^l \times \frac{1}{n} \sum_{j=1}^n K_h(X_{j\gamma} - X_{i\gamma}, X_{j\delta} - X_{i\delta}) \frac{\varphi_{\gamma\delta}(X_{j\gamma\delta})}{\varphi(X_{i\gamma}, X_{i\delta}, X_{j\gamma\delta})} \sigma(X_j) \epsilon_j,$$

where now γ is also allowed to take the value α , compare (25). Nevertheless the same arguments of boundedness apply as above.

For the rest, the proof is along the lines of Linton (1997). For the interaction term all arguments are the same, compare (26) and (30), but the rate is slower (by one bandwidth) due to having one dimension more to estimate.

Apart from these theoretical differences which will be emphasized under the performance point of view in Section 5, there is also another, substantial difference between backfitting, marginal integration and this efficient estimator. The backfitting is estimating the additive components after a projection of the regression problem into the space of additive models, the marginal integration estimator, in contrast, always estimates the marginal influence of the particular regressor, whatever the true model is, see e.g. Sperlich, Linton, Härdle (1999). The efficient estimator now is a mixture of this and thus suffers from the lack of interpretability if the predetermined model structure is not completely fulfilled. This can be a disadvantage for empirical research. Also in the context of testing model structure this leads to problems, especially if we use bootstrap generated with an estimated hypothetical model.

5 Computational Performance of the Estimators

To examine the small sample behavior of the estimators of the previous sections we did a simulation study for a sample size of $n = 150$, respectively $n = 169$ (for 3D-presentation reasons) observations. Certainly, an intensive computational comparison between not only ours but also alternative estimation procedures for additive models would be of interest, but would really require a separate paper. A first detailed investigation and comparison between the backfitting and the marginal integration estimator can be found in Sperlich, Linton, Härdle (1999) but without interaction terms and not examining the robustness when additivity is violated.

Here, we concentrate on a small illustration to see how reasonable these procedures behave in small samples. A more detailed simulation study is carried out for the testing procedures in Section 7. Further, an application to real data is there. 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

$$(34) \quad f_1(u) = 2u \quad , \quad f_2(u) = 1.5 \sin(-1.5u) \quad ,$$

$$(35) \quad f_3(u) = -u^2 + E(u^2) \quad \text{and} \quad f_{1,2}(u, v) = auv$$

with $a = 1$ for the simulations in this section. 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 all calculations we used the quartic kernel $\frac{15}{16}(1 - u^2)^2 \mathbb{1}\{|u| \leq 1\}$ for $K(u)$ as well as for $L(u)$, and product kernels for higher dimensions. We chose different bandwidths depending on the actual situation and on whether the direction was of interest or not (in the previous sections we distinguished them by denoting them h and g). For a discussion of optimal choice of bandwidth, we refer to Sperlich, Linton, Härdle (1999), but it must be admitted that a complete and practically useful solution to this problem remains to be found. This is in particular true for the bandwidth h_e of the one-step efficient estimator.

When we considered the the estimation of the functions f_α , $f_{\alpha\beta}^*$ we used $h = 0.9$, $g = 1.1$. For the pre-estimation with subsequent application of to apply afterwards the one-step backfit (efficient estimator) we used $h = 0.7$ and $g = 0.9$, as we have to undersmooth. For the one step backfit, we selected $h_e = 0.9$.

In Figure 1 we depict the performance of the 'simple' marginal integration estimator, using the local linear smoother. The data generating functions f_1 , f_2 , and f_3 are given as dashed lines in a point cloud that represents the observed responses Y after the first simulation run. The interaction function $f_{1,2}$ is given in the lower left window. For one hundred repetitions we estimated the functions on a grid with the above mentioned bandwidths and kernels and plotted for each grid point the extreme upper and extreme lower value of these estimates. For the one-step efficient estimator we did the same. The results are given in Figure 2.

The results are quite good having in mind that we have used only $n = 169$ observations. Apart from this we can observe several interesting, partly expected behaviors. E.g. as intended, the estimates, at least for the interaction term are smoother for the one-step efficient estimator. The biases can be seen clearly for both and behave the same. All in all, for a sample of this size the two estimators give roughly the same results.

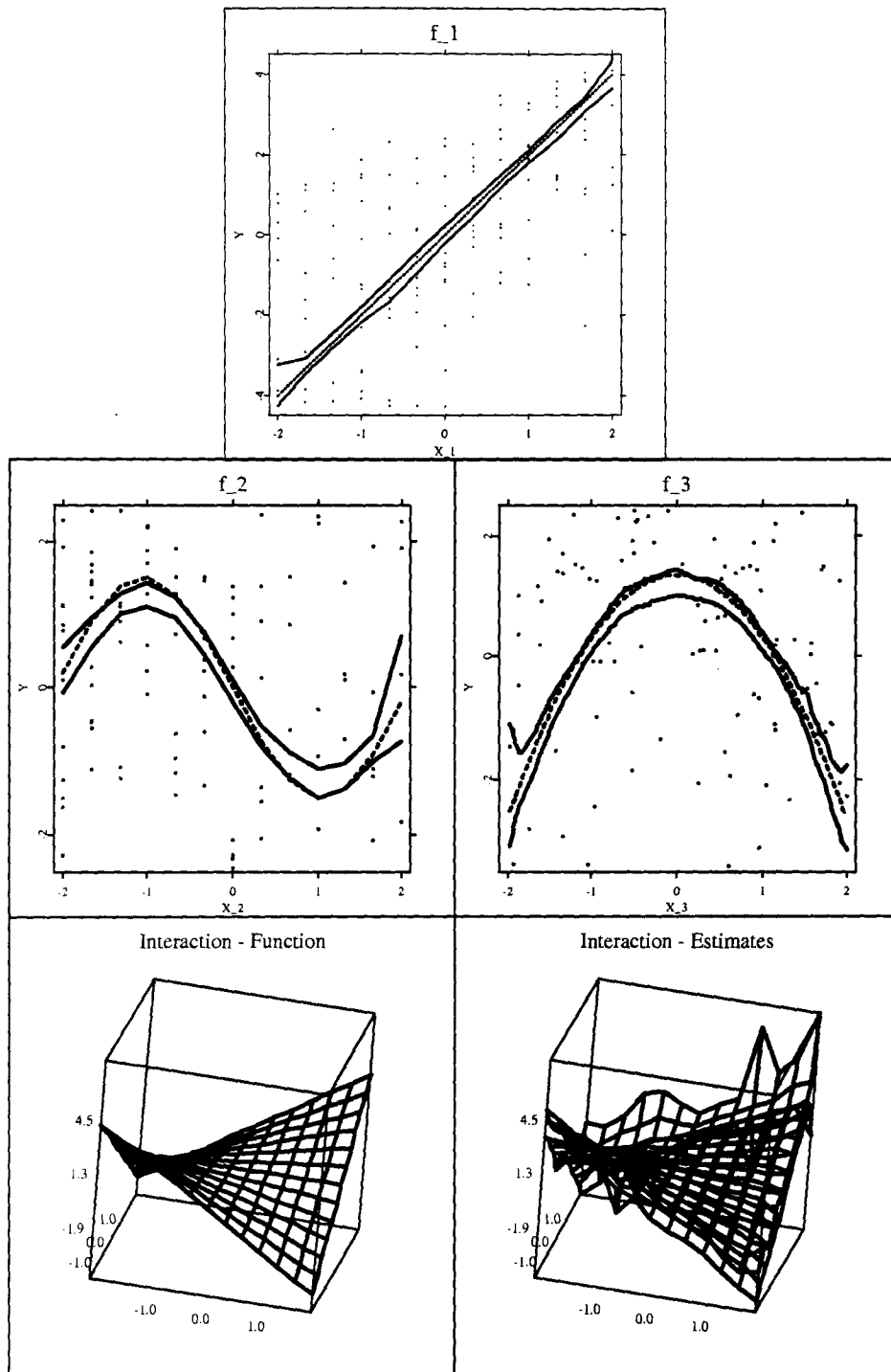


Figure 1: Performance of the 'simple' marginal integration estimator. Real functions (dashed) and extreme points for 100 of their estimates (solid). For the first run also the response variable Y (points) is given. Position: f_1 (top), f_2 (upper left), f_3 (upper right), $f_{1,2}$ (lower left) and the extreme points of the estimates after 100 simulation runs (lower right).

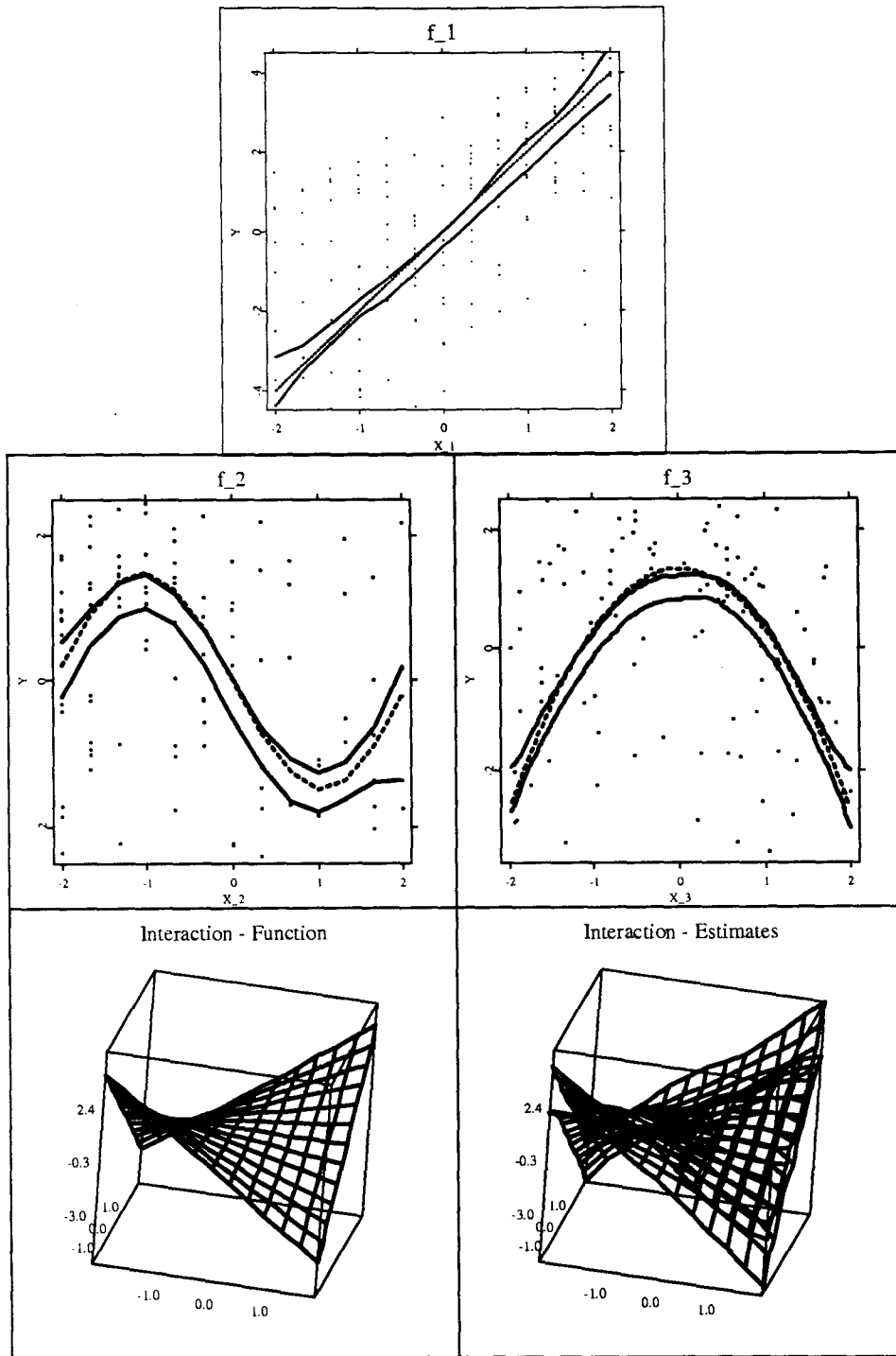


Figure 2: Performance of the 'efficient' estimator. Real functions (dashed) and extreme points for 100 of their estimates (solid). For the first run also the response variable Y (points) is given. Position: f_1 (top), f_2 (upper left), f_3 (upper right), $f_{1,2}$ (lower left) and the extreme points of the estimates after 100 simulation runs (lower right).

6 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 $H_{0,\alpha\beta} : f_{\alpha\beta} \equiv 0$, 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 separability in the regression model.

In Section 2 we pointed out that for this purpose it is equivalent to consider $f_{\alpha\beta}^*$ instead of $f_{\alpha\beta}$. We propose two procedures; the first one is focused on $f_{\alpha\beta}^*$ directly, the second one on the mixed derivative $\tilde{f}_{\alpha\beta}^{(1,1)}$. For reasons discussed in Section 4, we concentrate here on the pure marginal integration estimator.

6.1 Considering the interaction function

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

We consider $\int \widehat{f_{\alpha\beta}^*}^2(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta$. In practice, as will be seen in (40), this functional is replaced by an empirical average. To study the test functional, note first that by Theorem 2, equation (17) and some tedious calculations we get the following decomposition

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

$$\int f_{\alpha\beta}^{*2}(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta + 2h^2 \int f_{\alpha\beta}^*(x_\alpha, x_\beta) B_1(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta + o_p(h^2)$$

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 (44) and (47).

We then calculate the asymptotics of the sums of the $H(X_i, \varepsilon_i, X_i, \varepsilon_i)$'s and the $H(X_i, \varepsilon_i, X_j, \varepsilon_j)$'s, put the results together and obtain (cf. A.3 of the appendix):

Theorem 6 *Under assumptions (A1) to (A5), as $h \rightarrow 0$ and $nh^2 \rightarrow \infty$,*

$$nh \int \widehat{f_{\alpha\beta}^*}^2(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta - \frac{2\{K^{(2)}(0)\}^2}{h} \int \frac{\varphi_{\alpha\beta}(z_\alpha, z_\beta) \varphi_{\alpha\beta}^2(z_{\alpha\beta})}{\varphi(z)} \sigma^2(z) dz$$

$$-nh \int f_{\alpha\beta}^{*2}(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta - 2nh^3 \int f_{\alpha\beta}^*(x_\alpha, x_\beta) B_1(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta$$

$$\xrightarrow{L} N\{0, V(K, \varphi, \sigma)\},$$

in which

$$V(K, \varphi, \sigma) = 2 \|K^{(2)}\|_2^4 \int \frac{\varphi_{\alpha\beta}^2(z_{1\alpha}, z_{1\beta}) \varphi_{\alpha\beta}^2(z_{1\alpha\beta}) \varphi_{\alpha\beta}^2(z_{2\alpha\beta})}{\varphi(z_1) \varphi(z_{1\alpha}, z_{1\beta}, z_{2\alpha\beta})} \sigma^2(z_1) \sigma^2(z_{1\alpha}, z_{1\beta}, z_{2\alpha\beta}) dz_1 dz_{2\alpha\beta}.$$

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

Denote now by $S_{\alpha\beta}$ the support of the density $\varphi_{\alpha\beta}$ and let $\mathcal{B}_{\alpha\beta}(M)$ denote the function class consisting of functions $f_{\alpha\beta}$ satisfying

$$\|f_{\alpha\beta}\|_{H^s(S_{\alpha\beta})} \leq M$$

where one denotes by $\|f_{\alpha\beta}\|_{H^s(S_{\alpha\beta})}$ the Sobolev seminorm

$$\sqrt{\sum_{\gamma=0}^s \int_{S_{\alpha\beta}} \left\{ \frac{\partial^\gamma f_{\alpha\beta}(x_\alpha, x_\beta)}{\partial^\gamma x_\alpha \partial^{s-\gamma} x_\beta} \right\}^2 dx_\alpha dx_\beta}, \quad s = 2, 3, \dots,$$

and $M > 0$ is a constant. Consider the null hypothesis $H_{0,\alpha\beta} : f_{\alpha\beta}(x_\alpha, x_\beta) \equiv 0$ versus the local alternative $H_{1,\alpha\beta}(a) : f_{\alpha\beta} \in \mathcal{F}_{\alpha\beta}(a)$ where, for any $a > 0$

$$\mathcal{F}_{\alpha\beta}(a) = \left\{ f_{\alpha\beta} \in \mathcal{B}_{\alpha\beta}(M) : \|f_{\alpha\beta}\|_{L^2(S_{\alpha\beta}, \varphi_{\alpha\beta})} = \sqrt{\int_{S_{\alpha\beta}} f_{\alpha\beta}^2(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta} \geq a \right\}.$$

Based on Theorem 6, the test rule with asymptotic significance level $1 - \eta$ is:

Reject the null hypothesis $H_{0,\alpha\beta}$ in favor of the alternative $H_{1,\alpha\beta}(a)$ if

$$(36) \quad T_n \geq C(\eta; h, K, \varphi, \sigma)$$

where the test statistic

$$(37) \quad T_n = nh \int \widehat{f_{\alpha\beta}^*}^2(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta$$

and the critical value

$$(38) \quad C(\eta; h, K, \varphi, \sigma) = \Phi^{-1}(1 - \eta) V^{1/2}(K, \varphi, \sigma) + \frac{2\{K^{(2)}(0)\}^2}{h} \int \frac{\varphi_{\alpha\beta}(z_\alpha, z_\beta) \varphi_{\alpha\beta}^2(z_{\alpha\beta})}{\varphi(z)} \sigma^2(z) dz$$

in which Φ is the cumulative distribution function of the standard normal variable. The following result concerns the local power of the above test:

Theorem 7 Under assumptions (A1) to (A5) and as $h \rightarrow 0, nh^2 \rightarrow \infty$, let for $1 \leq i \leq n$

$$(39) Y_{in} = c + \sum_{\gamma=1}^d f_{\gamma}(X_{in\gamma}) + f_{n,\alpha\beta}(X_{in\alpha}, X_{in\beta}) + \sum_{\substack{1 \leq \gamma < \delta \leq d \\ (\gamma, \delta) \neq (\alpha, \beta)}} f_{\gamma\delta}(X_{in\gamma}, X_{in\delta}) + \sigma(X_{in})\varepsilon_{in}$$

be the data array generated from the i.i.d. array $(X_{in}, \varepsilon_{in}), 1 \leq i \leq n$, for each $n = 1, 2, \dots$, with fixed main effects $\{f_{\gamma}\}_{\gamma=1}^d$ and interactions $\{f_{\gamma\delta}\}_{1 \leq \gamma < \delta \leq d, (\gamma, \delta) \neq (\alpha, \beta)}$ and with the $\alpha\beta$ -th interaction $(f_{n,\alpha\beta})_{n=1}^{\infty}$ a sequence of functions such that $f_{n,\alpha\beta} \in \mathcal{F}_{\alpha\beta}(a_n)$ where $\{a_n\}$ is a sequences satisfying $a_n^{-1} = o(nh + h^{-2})$ as $n \rightarrow \infty$. Denote by p_n the probability of rejecting $H_{0,\alpha\beta} : f_{n,\alpha\beta}(x_{\alpha}, x_{\beta}) \equiv 0$ in favor of the local alternative $H_{1,\alpha\beta}(a_n) : f_{n,\alpha\beta} \in \mathcal{F}_{\alpha\beta}(a_n)$ based on the data $(X_{in}, Y_{in}), 1 \leq i \leq n$ as defined in (39). Then $\lim_{n \rightarrow \infty} p_n = 1$.

Theorem 7 guarantees that asymptotically, the proposed test procedure (36) is able to detect an interaction term of the magnitude $n^{-1}h^{-1} + h^2$ with probability 1.

To implement the test procedure (36), the critical value $C(\eta; h, K, \varphi, \sigma)$ can be obtained as the wild bootstrap quantiles of the test statistic $T_n = nh \int \widehat{f_{\alpha\beta}^*}^2(x_{\alpha}, x_{\beta}) \varphi_{\alpha\beta}(x_{\alpha}, x_{\beta}) dx_{\alpha} dx_{\beta}$. Since the density $\varphi_{\alpha\beta}$ is unknown, T_n is approximated by method of moment as

$$(40) \quad \tilde{T}_n = nh \sum_{l=1}^n \widehat{f_{\alpha\beta}^*}^2(X_{l\alpha}, X_{l\beta}) / n = h \sum_{l=1}^n \widehat{f_{\alpha\beta}^*}^2(X_{l\alpha}, X_{l\beta}).$$

The following theorem ensures that this substitution is asymptotically allowable

Theorem 8 Under assumptions (A1) to (A5) and as $h \rightarrow 0, nh^2 \rightarrow \infty$

$$\begin{aligned} \tilde{T}_n - \frac{2\{K^{(2)}(0)\}^2}{h} \int \frac{\varphi_{\alpha\beta}(z_{\alpha}, z_{\beta}) \varphi_{\alpha\beta}^2(z_{\alpha\beta})}{\varphi(z)} \sigma^2(z) dz \\ - nh \int f_{\alpha\beta}^*(x_{\alpha}, x_{\beta}) \varphi_{\alpha\beta}(x_{\alpha}, x_{\beta}) dx_{\alpha} dx_{\beta} - 2nh^3 \int f_{\alpha\beta}^*(x_{\alpha}, x_{\beta}) B_1(x_{\alpha}, x_{\beta}) \varphi_{\alpha\beta}(x_{\alpha}, x_{\beta}) dx_{\alpha} dx_{\beta} \\ \xrightarrow{L} N\{0, V(K, \varphi, \sigma)\}. \end{aligned}$$

Hence, Theorem 6 and test rule (36) are not affected when replacing T_n with \tilde{T}_n . Further, Theorem 7 holds as well, provided the same additional assumptions are true.

6.2 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 6, the derivation of the asymptotics for this test statistic is the same as in the proof of Theorem 6 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. Again, it can be noted that the convergence rate is slower than that obtained in Theorem 6 so it could be asked why this test statistic should be considered. In fact, as will be seen in the simulations, Section 7, the asymptotic properties hold for large samples, where *large* can be many thousands of observations. So, even when the test procedure proposed first should at some point beat the one we consider now, this is not clear for small sample which are typical for many real data sets. Further, it is well known that even though a certain test based on the estimation of a functional form is superior in detecting a general deviation from the hypothetical one, a single peak or bump can often be better detected by tests based on the derivatives.

Theorem 9 *Under assumptions (A1) and (A3)-(A6), as $h \rightarrow 0$ and $nh^6 \rightarrow \infty$,*

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

where B_2 is defined in the formulation of Theorem 4.

Now let $\mathcal{B}_{\alpha\beta}(M)$ denote the function class consisting of functions $f_{\alpha\beta}$ satisfying

$$\|f_{\alpha\beta}\|_{H^4(S_{\alpha\beta})} + \|f_{\alpha\beta}\|_{H^3(S_{\alpha\beta})} \leq M$$

where $M > 0$ is a constant. Consider the null hypothesis $H_{0,\alpha\beta} : f_{\alpha\beta}(x_\alpha, x_\beta) \equiv 0$ versus the local alternative $H_{1,\alpha\beta}(a) : f_{\alpha\beta} \in \mathcal{F}_{\alpha\beta}(a)$ where, for any $a > 0$

$$\mathcal{F}_{\alpha\beta}(a) = \left\{ f_{\alpha\beta} \in \mathcal{B}_{\alpha\beta}(M) : \|f_{\alpha\beta}^{(1,1)}\|_{L^2(S_{\alpha\beta}, \varphi_{\alpha\beta})} = \sqrt{\int_{S_{\alpha\beta}} f_{\alpha\beta}^{(1,1)2}(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta} \geq a \right\}.$$

Based on Theorem 9, the test rule with asymptotic significance level $1 - \eta$ is:

Reject the null hypothesis $H_{0,\alpha\beta}$ in favor of the alternative $H_{1,\alpha\beta}(a)$ if

$$(41) \quad nh^5 \int \hat{F}_{\alpha\beta}^{(1,1)2}(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta - \frac{2\{K_1^{*(2)}(0)\}^2}{h} \int \frac{\varphi_{\alpha\beta}(z_\alpha, z_\beta) \varphi_{\alpha\beta}^2(z_{\alpha\beta})}{\varphi(z)} \sigma^2(z) dz$$

$$\geq \frac{\Phi}{(1-\eta)} \sqrt{2 \|K_1^{*(2)}\|_2^4 \int \frac{\varphi_{\alpha\beta}^2(z_{1\alpha}, z_{1\beta}) \varphi_{\alpha\beta}^2(z_{1\alpha\beta}) \varphi_{\alpha\beta}^2(z_{2\alpha\beta})}{\varphi(z_1) \varphi(z_{1\alpha}, z_{1\beta}, z_{2\alpha\beta})} \sigma^2(z_1) \sigma^2(z_{1\alpha}, z_{1\beta}, z_{2\alpha\beta}) dz_1 dz_{2\alpha\beta}}.$$

The following result concerns the local power of the above test:

Theorem 10 *Under assumptions (A1) to (A5) and as $h \rightarrow 0, nh^6 \rightarrow \infty$, let $Y_{in}, 1 \leq i \leq n$ be the same data array as in Theorem 7 but with the $\alpha\beta$ -th interaction $f_{n,\alpha\beta} \in \mathcal{F}_{\alpha\beta}(a_n)$ where $\{a_n\}$ is a sequences satisfying $a_n^{-1} = o(nh^5 + h^{-2})$ as $n \rightarrow \infty$. Denote by p_n the probability of rejecting $H_{0,\alpha\beta} : f_{n,\alpha\beta}(x_\alpha, x_\beta) \equiv 0$ in favor of the local alternative $H_{1,\alpha\beta}(a_n) : f_{n,\alpha\beta} \in \mathcal{F}_{\alpha\beta}(a_n)$ based on the data $(X_{in}, Y_{in}), 1 \leq i \leq n$ as defined in (39). Then $\lim_{n \rightarrow \infty} p_n = 1$.*

Thus Theorem 10 guarantees that asymptotically with probability 1, the proposed test procedure (41) is able to detect an interaction term whose mixed derivative is of the magnitude $n^{-1}h^{-5} + h^2$. The proof of Theorem 10 is similar to that of Theorem 7, and therefore omitted. Also, Theorem 8 can be extended to test rule (41), but we have omitted its statement due to similarity.

6.3 A possible F-type test

Both Theorems 6 and 9 are used to test pairwise interactions. As remarked by one of the referees, methodologically speaking we propose two individual t -type statistics to check for a given interaction. Because of possible high multicollinearity among the explanatory variables, as in the classical linear regression context, it may be possible that individual test statistics are insignificant, but their joint effect is significant.

To consider such a situation, in general let $G_{\alpha\beta}$ be a functional for testing $f_{\alpha\beta} = 0$. We have shown that

$$g(n, h) \{G_{\alpha\beta} - E(G_{\alpha\beta})\} \xrightarrow{L} N(0, V_{\alpha\beta}),$$

where $g(n, h)$ is a normalizing factor and $V_{\alpha\beta}$ is the asymptotic variance.

Let $G = \{G_{\alpha\beta}, 1 \leq \alpha < \beta \leq d\}$ be the vector obtained by considering all pairwise interactions. It has dimension $p = d(d-1)/2$ corresponding to the number of possible interactions. If it can be proved that G is jointly asymptotically normal,

$$g(n, h) \{G - E(G)\} \xrightarrow{L} N(0, V),$$

where V is a covariance matrix of dimension p , then one would have that

$$g^2(n, h)\{G - E(G)\}^T V^{-1}\{G - E(G)\}$$

is asymptotically χ_p^2 -distributed. But studentizing and by analogy with ordinary multivariate analysis (cf. Johnson and Wichern (1988, p. 171) one might expect that

$$g^2(n, h)\{G - E(G)\}^T \hat{V}^{-1}\{G - E(G)\}$$

should be more accurately described by an F-type statistic. Such a statistic would yield an F -type test for all of the pairwise interactions. It is a natural suggestion, but it is far from trivial to establish, and it is a topic for further research. For example it is not clear how one should choose the number of degrees of freedom. Some discussion of this point is given in a related framework by Hastie and Tibshirani (1990, Sections 3.5, 3.9, 5.4.5 and 6.8.3). However, theory is lacking and Sperlich, Linton, Härdle (1997,1999) found reasons to doubt the generality of these methods, especially for the marginal integration estimator. This was partly confirmed by Müller (1997) in the context of even much simpler testing problems than we consider here. Further it was briefly discussed in Härdle, Mammen, Müller (1998), also in a different context of testing.

7 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. We have the additional problem of having complicated 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, to be described in this Section, 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. For an intensive simulation study of the performance of marginal integration estimation in finite samples see also Sperlich, Linton and Härdle (1999).

7.1 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). The wild bootstrap allows for a heterogeneous vari-

ance of the residuals. Härdle and Mammen (1993) put 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 6 or 9 and let n^* be the number of bootstrap replications. The testing procedure 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 $f_{\alpha\beta} = 0$ 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 , draw a bootstrap residual u_i^* from the distribution $G_{(a,b),i}$ such that for $U \sim G_{(a,b),i}$,

$$E_{G_{(a,b),i}}(U) = 0, \quad E_{G_{(a,b),i}}(U^2) = \tilde{u}_i^2$$

and $E_{G_{(a,b),i}}(U^3) = \tilde{u}_i^3.$

3. Generate a sample $\{(Y_i^*, X_i)\}_{i=1}^n$ with $Y_i^* = \widehat{m}_0 + u_i^*$. 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 values for the rejection region.

For the two-point distribution $G_{(a,b),i}$ we have used the so-called *golden cut* construction, 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 $q = (5 + \sqrt{5})/10$.

For the marginal integration estimator Dalelane (1999) 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 (1999). 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. Important general progress in this area has recently been achieved by Kreiss, Neumann and Yao (1999). There is still some work needed to establish a theory of the wild bootstrap for the test statistic we are using.

7.2 The simulation study

The small sample behavior of the estimators has already been investigated and discussed in Section 5. For testing we again use the model

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

with

$$\begin{aligned} f_1(u) &= 2u & , & & f_2(u) &= 1.5 \sin(-1.5u) \\ f_3(u) &= -u^2 + E(u^2) & , & & f_{1,2}(u, v) &= auv \end{aligned}$$

where $a = 0$ under the null hypothesis and $a = 1$ under the alternative. Again, $X_j \sim U[-2, 2]$ i.i.d. for $j = 1, 2, 3$, and normally distributed error terms with standard deviation 0.5. Sample size is now always $n = 150$.

To calculate the test statistic we used the (product) quartic kernel for $K(u)$ and $L(u)$ as above. When we considered the test statistic based on the estimation of f_{12}^* (*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*), which generally requires more smoothing (cf. (A6)), we selected $h = 1.5, g = 1.6$ and $h = 1.4, g = 1.5$, respectively.

We consider first the null hypothesis $H_{0,12} : f_{1,2}(u) \equiv 0$ and look at the asymptotics. In Figure 3 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.

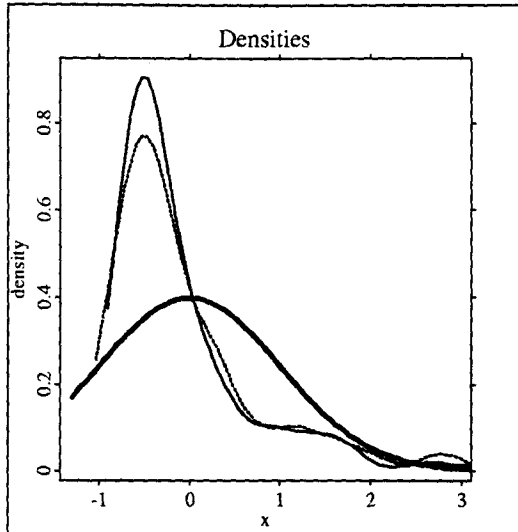


Figure 3: Densities of the test statistics; direct method (solid), testing derivatives (dashed) and normal density (thick, solid).

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 seems to be that for a functional of type $\int \widehat{f_{\alpha\beta}^*}^2(x_\alpha, x_\beta)\pi(x_\alpha, x_\beta)dx_\alpha dx_\beta$ 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.

To get the results of Table 1 and Figure 4, describing the bootstrap version of the tests, we did 249 bootstrap replications and, following Theorems 6, 9, 8, considered the test statistics

$$\frac{1}{n} \sum_{i=1}^n \widehat{f_{12}^*}^2(X_1, X_2) \mathbb{1}\{|X_k| \leq 1.6 \text{ for } k = 1, 2\}$$

and

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

respectively, i.e. we have integrated numerically over the empirical distribution function and using a weight function (the indicator function $\mathbb{1}$) for the test statistic to remove outliers and avoid boundary effects caused by the estimation (cf. Hjellvik, Yao and Tjøstheim 1998).

Table 1 is presenting the error of the first kind for both methods and at different significance levels. As noted above, a detailed simulation study is beyond our paper. So it certainly

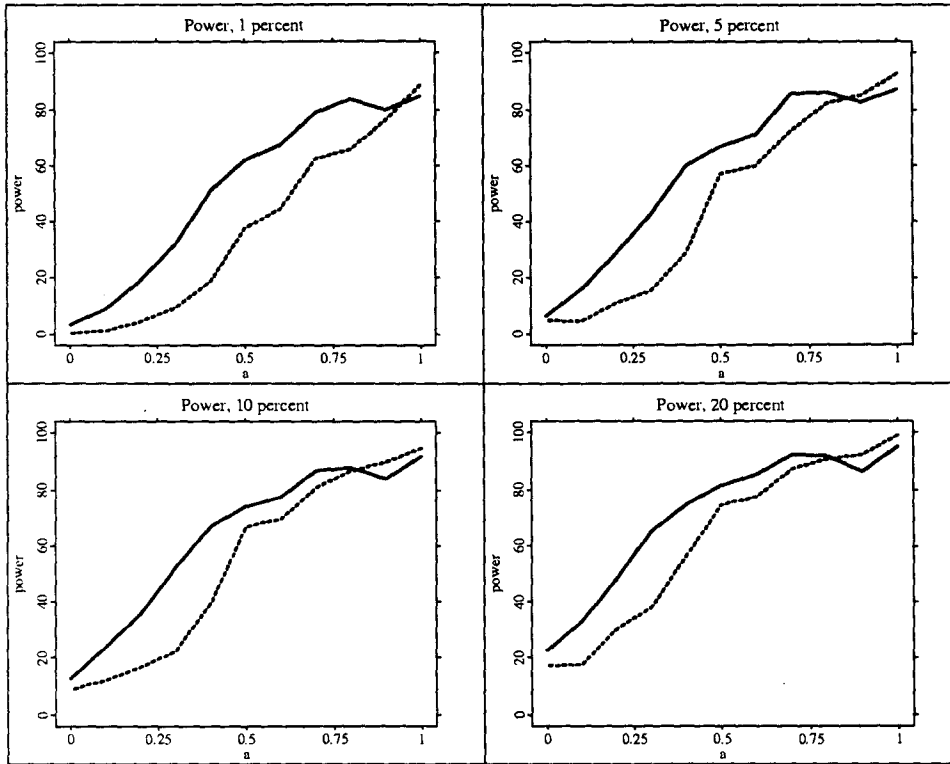


Figure 4: Power functions at the 1, 5, 10 and 20% significance levels for both procedures: direct method (solid) and testing derivatives (dashed).

would be interesting to look on different power results for different bandwidth choices. By no means we state here to have chosen the optimal bandwidth as to find this even for the estimation procedure can be hard, see Sperlich, Linton, Härdle (1999). Thus, all we are interested in for the moment is, to see whether using an in estimation reasonably smoothing bandwidth (see Section 5) also leads to reasonable testing results.

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 although the results are fairly robust for a reasonable wide range of bandwidths. In the absence of an optimal procedure for

choosing the bandwidth, Table 1 must be interpreted with caution as far as a comparison of the two testing procedures is concerned. But it is seen that the wild bootstrap 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 the error of the first kind and the power for a wide range of examples. The power as a function of a in (42) is displayed for both methods and for different levels in Figure 4. Both procedures are working well. For this particular model the power function of the *direct method* is steeper. This is intuitively reasonable as the estimator and the test statistic have smaller asymptotic variance for this method, but for a finite sample it is quite likely that the comparative advantages of the two methods depend on the particular model or design.

Obviously a much more detailed simulation study would be of interest, in particular concerning the interplay between model complexity and (optimal) choice of bandwidth. At the moment bandwidths have been chosen somewhat arbitrarily, but we have been pleased to observe that the same set of bandwidths seems to lead to satisfactory results for both estimation and testing.

7.3 An Application to Production Function Estimation

In this section we apply the estimation and testing procedures to a five dimensional production function.

Separability and additivity of production functions have been discussed since the early paper by Leontief (1947). These assumptions yield many important economic results, for example 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 separable¹) for a particular data set. This discussion goes back at least to Denny and Fuss (1977), Fuss, McFadden 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 (1999) 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 the interest there was focused mainly on the return to scale and hence on derivative estimation,

¹The expression "strong separability" is equivalently used for "additivity" or "generalized additivity"; see Berndt and Christensen (1973).

presently we are more interested in examining the validity of the assumption of absence of interaction terms looking only at second order interactions as these are the only interpretable ones. 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 consist 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 (ignoring interaction) is of the form

$$(42) \quad \ln(y) = c + \sum_{\alpha=1}^d f_{\alpha} \{\ln(x_{\alpha})\} .$$

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

$$(43) \quad \ln(y) = c + \sum_{\alpha=1}^d f_{\alpha} \{\ln(x_{\alpha})\} + \sum_{1 \leq \alpha < \beta \leq d} f_{\alpha\beta} \{\ln(x_{\alpha}), \ln(x_{\beta})\}$$

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

First we estimated all functions f_{α} and $f_{\alpha\beta}$. The estimation results are given in Figures 5 to 7. Again, quartic kernels were employed 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 a relatively wide range of choices of bandwidths. For a detailed discussion of the bandwidth choice and robustness we refer to Sperlich, Linton and Härdle (1997).

In Figure 5 the univariate function estimates (not centered to zero) are displayed together with a kind of partial residuals $\hat{r}_{i\alpha} := y_i - \sum_{j \neq \alpha} \hat{f}_j(X_{ij}) = \hat{c} + \hat{f}_{\alpha}(X_{i\alpha}) + \hat{\varepsilon}_i$. To see clearly the shape of the estimates we display the main part of the point clouds including the function estimates. As suggested already in Severance-Lossin and Sperlich, the graphs in Figure 5 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

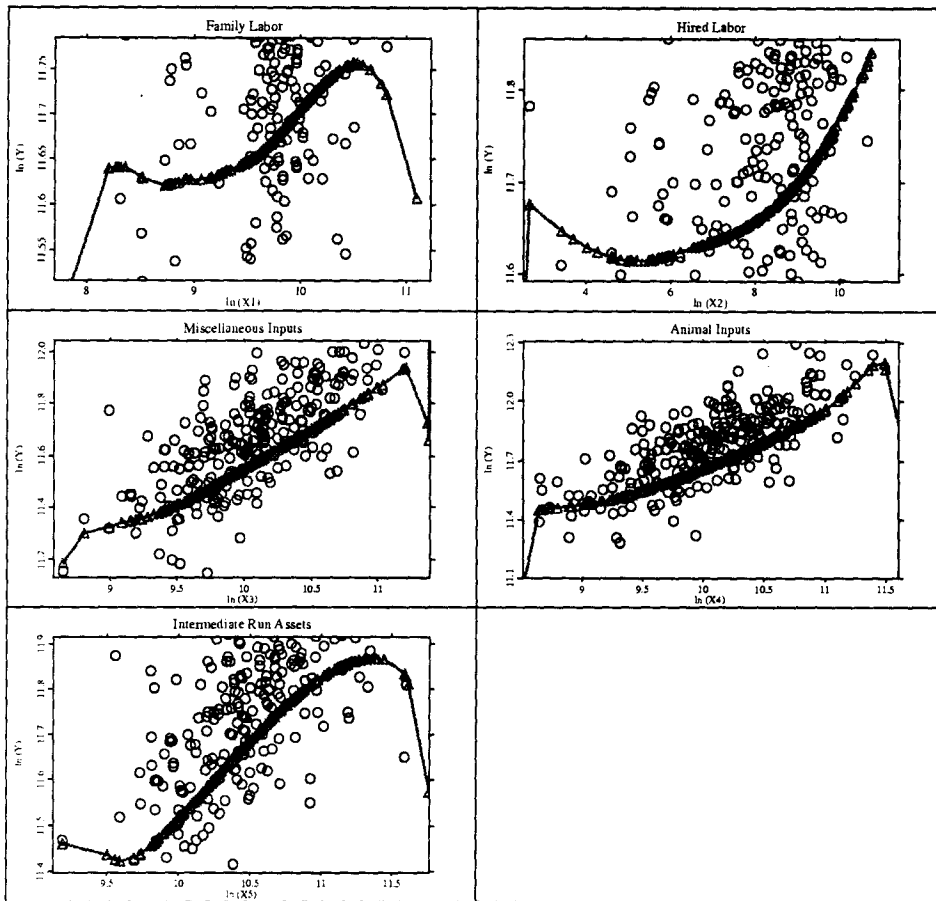


Figure 5: Function estimates for the univariate additive components and partial residuals.

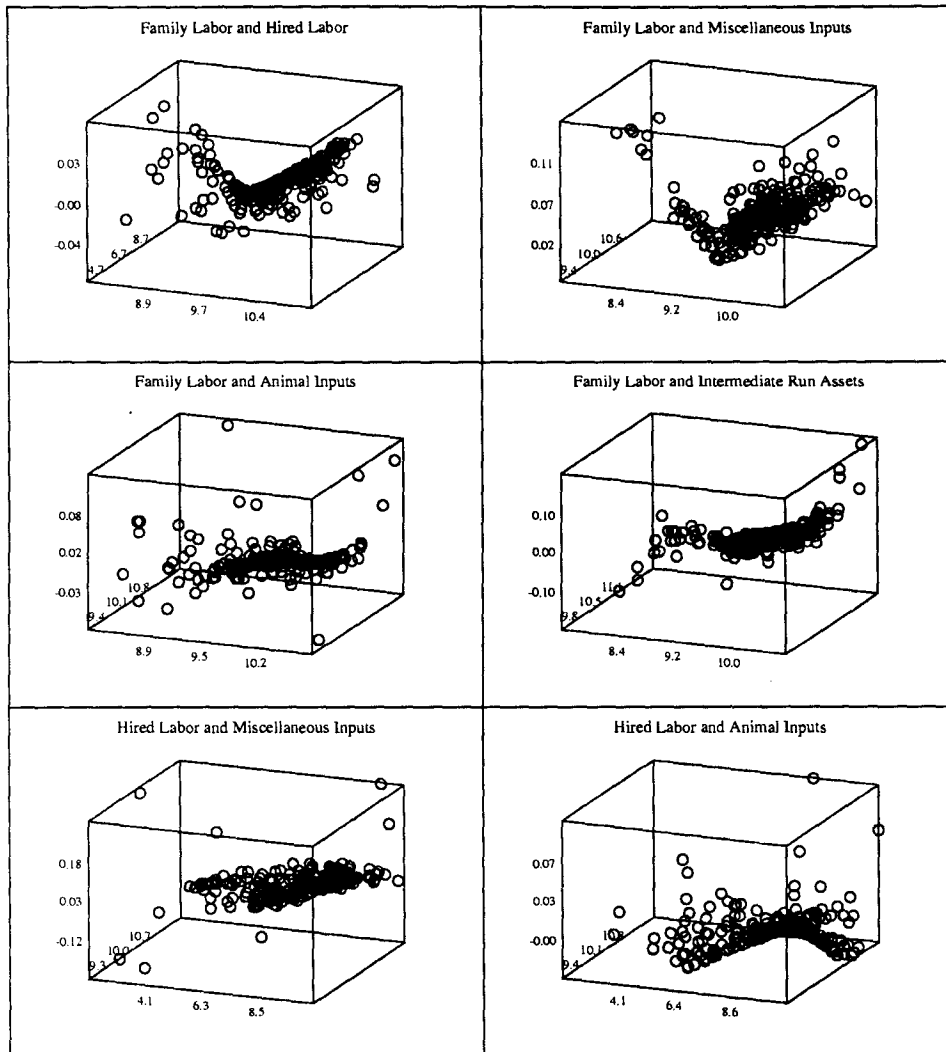


Figure 6: Estimates of the first 6 interaction terms.

lead to increasing returns to scale. An obvious inference from an economic point of view would be that larger farms are more productive.

In Figures 6 and 7 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 observations, and used the quartic kernel.

The same kernel and trimming were used for the testing, and we did 249 bootstrap replications. To examine the sensitivity of the test procedures against choice of bandwidth, we tried a wide range of bandwidths. For the first method, which employs the estimate of the

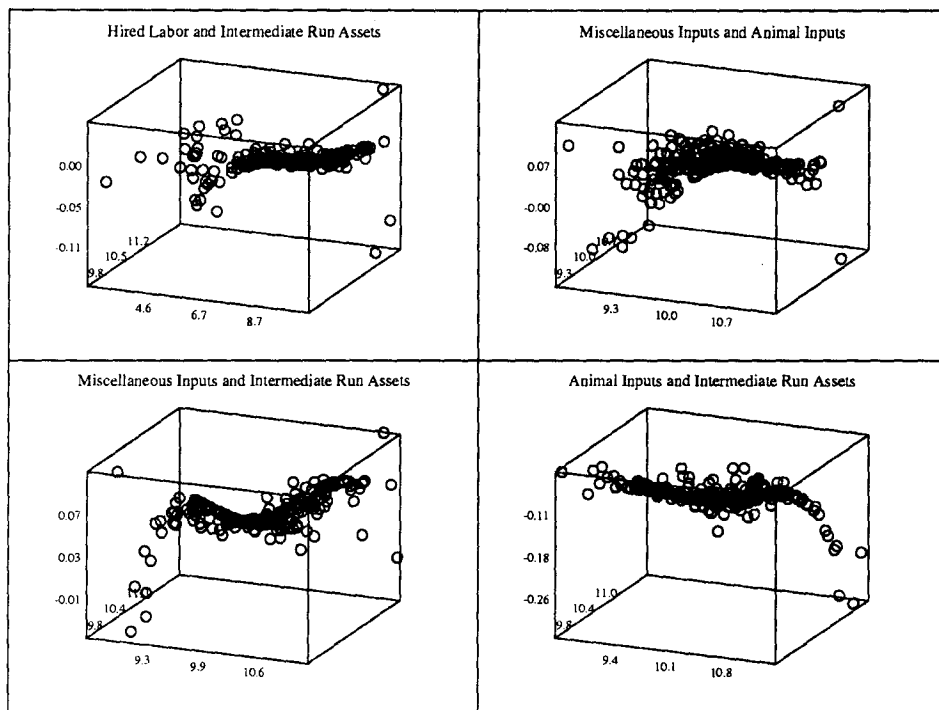


Figure 7: Estimates of the last 4 interaction terms.

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 an iterative model selection procedure: First we calculated the p-values for each interaction term $f_{\alpha\beta}$ including all the other functions f_γ , $1 \leq \gamma \leq d$ and $f_{\gamma\delta}$, $1 \leq \gamma < \delta \leq d$ with $(\gamma, d) \neq (\alpha, \beta)$ in the model (43). 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 eliminating 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 $f_{3,5}$ 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%.

Both procedures suggest that a weak form of interaction is present, and that the variable *family labor* plays a significant role in the interaction. The fact that the two procedures are not entirely consistent in their selection of relevant interaction terms should not be too surprising in view of the moderate sample size and the lack of any strong interactions. There are fairly clear indications from Figures 6, and 7 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 Theorems 1 and 2

The proof of Theorems 1 and 2 make use of the following two lemmas, whose proofs are not difficult. We refer to Silverman (1986) or Fan, Härdle, Mammen (1998).

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

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

and

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

Define $E_i[\cdot] = E[\cdot | X_{i1}, \dots, X_{id}]$ and $E_\star[\cdot] = E[\cdot | \mathbf{X}]$, where \mathbf{X} is the design matrix $\{X_{i\alpha}\}_{i,\alpha=1,1}^{n,d}$. The proofs can now be divided into two parts corresponding to the estimators \hat{F}_α and $\hat{F}_{\alpha\beta}$, respectively.

I) We start by considering the univariate estimator \hat{F}_α . This is also a component of the estimator $\widehat{f_{\alpha\beta}^\star}$ of interest in Theorem 2. 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

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

Computing the matrix product and inserting 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 obtain

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

Separating this expression into a systematic "bias" and a stochastic "variance" we have

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

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 [...] is as in the formula above. 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)

$$\begin{aligned} \frac{E_i(\hat{a}_i)}{\varphi(x_\alpha, X_{i\alpha})} = & \frac{1}{\varphi(x_\alpha, X_{i\alpha})} E_i \left[\frac{1}{n} \sum_{l=1}^n K_h(X_{l\alpha} - x_\alpha) L_g(X_{l\alpha} - X_{i\alpha}) \left[\frac{(X_{l\alpha} - x_\alpha)^2}{2} \right. \right. \\ & \left. \left. \left\{ f_\alpha^{(2)}(x_\alpha) + \sum_{\gamma \in D_\alpha} f_{\alpha\gamma}^{(2,0)}(x_\alpha, X_{l\gamma}) \right\} + \sum_{\gamma \in D_\alpha} \{ f_\gamma(X_{l\gamma}) - f_\gamma(X_{i\gamma}) \} + \right. \right. \\ & \left. \left. \sum_{(\gamma,\delta) \in D_{\alpha\alpha}} \{ f_{\gamma\delta}(X_{l\gamma}, X_{l\delta}) - f_{\gamma\delta}(X_{i\gamma}, X_{i\delta}) \} + O_p\{(X_{l\alpha} - x_\alpha)^3\} \right] \right] \end{aligned}$$

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

since $E_*(\varepsilon_i) = 0$, respectively $E_i(\varepsilon_i) = 0$ 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 by using (A2)

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

and note that

$$\frac{1}{n} \sum_{i=1}^n f_{\alpha\gamma}^{(2,0)}(x_\alpha, X_{i\gamma}) = E f_{\alpha\gamma}^{(2,0)}(x_\alpha, X_\gamma) = \frac{\partial^2}{\partial x_\alpha^2} \int f_{\alpha\gamma}(x_\alpha, u_\gamma) \varphi_\gamma(u_\gamma) du_\gamma = 0$$

by (7), for any $\gamma \in D_\alpha$.

For the stochastic term we use the same technique as in Fan, Härdle, Mammen (1998), Severance-Lossin and Sperlich (1999) to prove that with $w_{i\alpha}$ given by

$$(44) \quad w_{i\alpha} = \frac{1}{n} K_h(x_\alpha - X_{i\alpha}) \frac{\varphi_\alpha(X_{i\alpha})}{\varphi(x_\alpha, X_{i\alpha})},$$

we have

$$(45) \quad T_{2n} = \sum_{i=1}^n w_{i\alpha} \sigma(X_i) \varepsilon_i + o_p\{(nh)^{-1/2}\}$$

and hence

$$\hat{F}_\alpha(x_\alpha) - F_\alpha(x_\alpha) = O_p\{(nh)^{-1/2}\} + O_p(h^2).$$

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

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

As above in I) we do the matrix calculation, replace Y_i by $Y_i = \sigma(X_i)\varepsilon_i + m(X_i)$ and use the Taylor expansion of m around $(x_\alpha, x_\beta, X_{i\alpha\beta})$. Then we obtain

$$\begin{aligned} (46) \quad & \hat{F}_{\alpha\beta}(x_\alpha, x_\beta) - F_{\alpha\beta}(x_\alpha, x_\beta) = \\ & \frac{1}{n} \sum_{i=1}^n \frac{1}{\varphi(x_\alpha, x_\beta, X_{i\alpha\beta})} \frac{1}{n} \sum_{l=1}^n K_h(X_{l\alpha} - x_\alpha) K_h(X_{l\beta} - x_\beta) L_g(X_{l\alpha\beta} - X_{i\alpha\beta}) \times \\ & \left\{ I + O_p \left(h^2 + \sqrt{\frac{\ln n}{nh^2 g^{d-2}}} \right) \right\} \left[\frac{(X_{l\alpha} - x_\alpha)^2}{2} \left\{ f_\alpha^{(2)}(x_\alpha) + \sum_{\gamma \in D_{\alpha,\beta}} f_{\alpha\gamma}^{(2,0)}(x_\alpha, X_{l\gamma}) \right. \right. \\ & \left. \left. + f_{\alpha\beta}^{(2,0)}(x_\alpha, x_\beta) \right\} + \frac{(X_{l\beta} - x_\beta)^2}{2} \left\{ f_\beta^{(2)}(x_\beta) + \sum_{\gamma \in D_{\alpha,\beta}} f_{\beta\gamma}^{(2,0)}(x_\beta, X_{l\gamma}) + f_{\alpha\beta}^{(0,2)}(x_\alpha, x_\beta) \right\} \right. \\ & \left. + \sum_{\gamma \in D_{\alpha,\beta}} \{f_\gamma(X_{l\gamma}) - f_\gamma(X_{i\gamma})\} + \sum_{(\gamma, \delta) \in D_{\alpha\beta}} \{f_{\gamma\delta}(X_{l\gamma}, X_{l\delta}) - f_{\gamma\delta}(X_{i\gamma}, X_{i\delta})\} + \right. \\ & \left. (X_{l\alpha} - x_\alpha)(X_{l\beta} - x_\beta) f_{\alpha\beta}^{(1,1)}(x_\alpha, x_\beta) + O_p\{(X_{l\alpha} - x_\alpha)^3\} + o_p\{(X_{l\alpha} - x_\alpha)(X_{l\beta} - x_\beta)\} \right. \\ & \left. + O_p\{(X_{l\beta} - x_\beta)^3\} + \sigma(X_l)\varepsilon_l \right] + O_p(n^{-\frac{1}{2}}). \end{aligned}$$

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

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

where,

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

and $[\dots]$ is the expression in the same angular brackets of equation (46).

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

Let

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

We first prove that

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

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

$$\begin{aligned} &\frac{E_i(\hat{a}_i)}{\varphi(x_\alpha, x_\beta, X_{i\alpha\beta})} \\ &= \frac{1}{\varphi(x_\alpha, x_\beta, X_{i\alpha\beta})} \int K_h(z_\alpha - x_\alpha) K_h(z_\beta - x_\beta) L_g(w - X_{i\alpha\beta}) \varphi(z, w) \\ &\quad \left[\frac{(z_\alpha - x_\alpha)^2}{2} \left\{ f_\alpha^{(2)}(x_\alpha) + \sum_{\gamma \in D_{\alpha,\beta}} f_{\alpha\gamma}^{(2,0)}(x_\alpha, X_{i\gamma}) + f_{\alpha\beta}^{(2,0)}(x_\alpha, x_\beta) \right\} + \right. \\ &\quad \left. \frac{(z_\beta - x_\beta)^2}{2} \left\{ f_\beta^{(2)}(x_\beta) + \sum_{\gamma \in D_{\alpha,\beta}} f_{\beta\gamma}^{(2,0)}(x_\beta, X_{i\gamma}) + f_{\alpha\beta}^{(0,2)}(x_\alpha, x_\beta) \right\} + \right. \\ &\quad \left. \sum_{\gamma \in D_{\alpha,\beta}} \{f_\gamma(w_\gamma) - f_\gamma(X_{i\gamma})\} + \sum_{\gamma, \delta \in D_{\alpha,\beta}} \{f_{\gamma\delta}(w_\gamma, w_\delta) - f_{\gamma\delta}(X_{i\gamma}, X_{i\delta})\} + \right. \\ &\quad \left. (z_\alpha - x_\alpha)(z_\beta - x_\beta) f_{\alpha\beta}^{(1,1)}(x_\alpha, x_\beta) + O_p\{(z_\alpha - x_\alpha)^3\} + O_p\{(z_\beta - x_\beta)^3\} \right] \end{aligned}$$

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

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 the theorem:

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

We now turn to the variance part \mathcal{T}_{2n} . In Fan, Härdle, Mammen (1998) it is shown that for

$$(47) \quad w_{i\alpha\beta} = \frac{1}{n} K_h(x_\alpha - X_{i\alpha}, x_\beta - X_{i\beta}) \frac{\varphi_{\alpha\beta}(X_{i\alpha\beta})}{\varphi(x_\alpha, x_\beta, X_{i\alpha\beta})},$$

$$(48) \quad \mathcal{T}_{2n} = \sum_{i=1}^n w_{i\alpha\beta} \sigma(X_i) \varepsilon_i + o_p\{(nh^2)^{-1/2}\}$$

where the term

$$\sum_{i=1}^n w_{i\alpha\beta} \sigma(X_i) \varepsilon_i = O_p\{(nh^2)^{-1/2}\}$$

is asymptotically normal and dominates the corresponding stochastic term

$$\sum_{i=1}^n w_{i\alpha} \sigma(X_i) \varepsilon_i = O_p\{(nh)^{-1/2}\}$$

from part I of the proof. This means that $\widehat{f}_{\alpha\beta}^*(x_\alpha, x_\beta)$ as defined by (11) is asymptotically normal.

Finally, we want to calculate the variance of the combined estimator $\widehat{F}_{\alpha\beta}(x_\alpha, x_\beta) - \widehat{F}_\alpha(x_\alpha) - \widehat{F}_\beta(x_\beta)$. Because of the faster rate of the stochastic term in I than the one in II, it is enough to consider II, i.e. $\sum_{i=1}^n w_{i\alpha\beta} \sigma(X_i) \varepsilon_i$. 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 4

This proof is analogous to that of Theorem 1 and 2 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

$$(49) \quad w_{i\alpha\beta} = \frac{1}{nh^3} K_{3,h}^*(x_\alpha - X_{i\alpha}, x_\beta - X_{i\alpha}) \frac{\varphi_{\alpha\beta}(X_{i\alpha\beta})}{\varphi(x_\alpha, x_\beta, X_{i\alpha\beta})},$$

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

QED.

A.3 Proof of Theorem 6

Consider the decomposition

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

$$\int f_{\alpha\beta}^{*2}(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta + 2h^2 \int f_{\alpha\beta}^*(x_\alpha, x_\beta) B_1(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta + o_p(h^2)$$

in which

$$H(X_i, \varepsilon_i, X_j, \varepsilon_j) = \varepsilon_i \varepsilon_j \int \frac{1}{n^2} (w_{i\alpha\beta} - w_{i\alpha} - w_{i\beta}) (w_{j\alpha\beta} - w_{j\alpha} - w_{j\beta}) \sigma(X_i) \sigma(X_j) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta \quad (50)$$

with $w_{i\alpha}$, $w_{i\beta}$ and $w_{i\alpha\beta}$ as in equation (44) and (47).

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

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

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

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

where

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

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

$$\begin{aligned}
H_5 = & \frac{\varepsilon_i \varepsilon_j \sigma(X_i) \sigma(X_j)}{n^2} \left\{ \frac{\varphi_{\alpha}(X_{i\alpha})}{\varphi(X_i)} \frac{\varphi_{\beta}(X_{j\beta})}{\varphi(X_{i\beta}, X_{j\beta})} + \frac{\varphi_{\alpha\beta}(X_{i\beta})}{\varphi(X_i)} \frac{\varphi_{\alpha}(X_{j\alpha})}{\varphi(X_{i\alpha}, X_{j\alpha})} \right\} \varphi_{\alpha\beta}(X_{i\alpha}, X_{i\beta}) \\
& + \frac{\varepsilon_i \varepsilon_j \sigma(X_i) \sigma(X_j)}{n^2} \left\{ \frac{\varphi_{\alpha}(X_{j\alpha})}{\varphi(X_j)} \frac{\varphi_{\beta}(X_{i\beta})}{\varphi(X_{j\beta}, X_{i\beta})} + \frac{\varphi_{\alpha\beta}(X_{j\beta})}{\varphi(X_j)} \frac{\varphi_{\alpha}(X_{i\alpha})}{\varphi(X_{j\alpha}, X_{i\alpha})} \right\} \varphi_{\alpha\beta}(X_{j\alpha}, X_{j\beta})
\end{aligned}$$

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

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

Introducing the change of variable

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

we obtain

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

or

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

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

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

where

$$(51) \quad G_1(x, \varepsilon, y, \delta) = E \{ H_1(X_1, \varepsilon_1, x, \varepsilon) H_1(X_1, \varepsilon_1, y, \delta) \},$$

see Hall (1984).

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

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

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

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

which implies that

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

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

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

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

Proof. According to the definition of G_1

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

or

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

Introducing the change of variable

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

we obtain

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

Using convolution notation, one has

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

or

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

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

which is what we set out to prove. By techniques used in the two previous lemmas, it follows that

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

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

Lemmas A3, A5, and the Martingale Central Limit Theorem of Hall (1984) imply:

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

$$nh \sum_{1 \leq i < j \leq n} H(X_i, \varepsilon_i, X_j, \varepsilon_j) \xrightarrow{L} N \left\{ 0, 2 \left\| K^{(2)} \right\|_2^4 \int \frac{\varphi_{\alpha\beta}^2(z_{1\alpha}, z_{1\beta})\varphi_{\alpha\beta}^2(z_{1\alpha\beta})\varphi_{\alpha\beta}^2(z_{2\alpha\beta})}{\varphi(z_1)\varphi(z_{1\alpha}, z_{1\beta}, z_{2\alpha\beta})} \sigma^2(z_1)\sigma^2(z_{1\alpha}, z_{1\beta}, z_{2\alpha\beta}) dz_1 dz_{2\alpha\beta} \right\}.$$

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 $nh^2 \rightarrow \infty$,

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

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 6 is proved.

QED.

A.4 Proof of Theorem 7

To prove Theorem 7, first note that the support $S_{\alpha\beta}$ of $\varphi_{\alpha\beta}$ is compact, hence there exists a constant $C > 0$ such that

$$(52) \quad \|B_1\|_{L^2(S_{\alpha\beta}, \varphi_{\alpha\beta})} = \sqrt{\int B_1^2(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta} \leq CM$$

for any $f_{\alpha\beta} \in \mathcal{B}_{\alpha\beta}(M)$. Here

$$B_1(x_\alpha, x_\beta) = \mu_2(K) \frac{1}{2} \left\{ f_{\alpha\beta}^{(2,0)}(x_\alpha, x_\beta) + f_{\alpha\beta}^{(0,2)}(x_\alpha, x_\beta) \right\}$$

is the bias function of Theorem 2. Meanwhile, since

$$f_{\alpha\beta}^*(x_\alpha, x_\beta) = f_{\alpha\beta}(x_\alpha, x_\beta) + c_{\alpha\beta}, \quad c_{\alpha\beta} = \int f_{\alpha\beta}(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta$$

it follows that

$$\begin{aligned} \int f_{\alpha\beta}^{*2}(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta &= \int f_{\alpha\beta}^2(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta \\ &\quad + 2c_{\alpha\beta} \int f_{\alpha\beta}(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta + c_{\alpha\beta}^2 \\ &= \int f_{\alpha\beta}^2(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta + 3c_{\alpha\beta}^2 \geq \int f_{\alpha\beta}^2(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta. \end{aligned}$$

Hence for any $f_{\alpha\beta} \in \mathcal{F}_{\alpha\beta}(a)$, one has

$$(53) \quad \|f_{\alpha\beta}^*\|_{L^2(S_{\alpha\beta}, \varphi_{\alpha\beta})}^2 = \int f_{\alpha\beta}^{*2}(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta \geq a^2.$$

Now for $n = 1, 2, \dots$, let

$$\begin{aligned} T'_n &= nh \int \widehat{f_{n,\alpha\beta}^*}^2(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta - \frac{2\{K^{(2)}(0)\}^2}{h} \int \frac{\varphi_{\alpha\beta}(z_\alpha, z_\beta) \varphi_{\alpha\beta}^2(z_{\alpha\beta})}{\varphi(z)} \sigma^2(z) dz \\ &\quad - nh \int f_{n,\alpha\beta}^{*2}(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta - 2nh^3 \int f_{n,\alpha\beta}^*(x_\alpha, x_\beta) B_{n1}(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta \end{aligned}$$

where $(f_{n,\alpha\beta})_{n=1}^\infty$ is the sequence in Theorem 7, and B_{n1} the corresponding bias coefficients. Note that although the function $f_{n,\alpha\beta}^*(x_\alpha, x_\beta)$ is different for each n , a careful review of the proof of Theorem 6 shows that it still holds because the second order Sobolev seminorm of each $f_{n,\alpha\beta}^*(x_\alpha, x_\beta)$ is bounded uniformly for $n = 1, 2, \dots$, and all the main effects $\{f_\gamma\}_{\gamma=1}^d$ and other interactions $\{f_{\gamma\delta}\}_{1 \leq \gamma < \delta \leq d, (\gamma, \delta) \neq (\alpha, \beta)}$ are fixed. Hence

$$(54) \quad T'_n \xrightarrow{L} N\{0, V(K, \varphi, \sigma)\}$$

as $n \rightarrow \infty$. Now let

$$t_n = nh \int f_{n,\alpha\beta}^{*2}(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta + 2nh^3 \int f_{n,\alpha\beta}^*(x_\alpha, x_\beta) B_{n1}(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta$$

then

$$\begin{aligned} t_n &\geq nh \left\| f_{n,\alpha\beta}^* \right\|_{L^2(S_{\alpha\beta, \varphi_{\alpha\beta}})}^2 - 2nh^3 \left\| f_{n,\alpha\beta}^* \right\|_{L^2(S_{\alpha\beta, \varphi_{\alpha\beta}})} \left\| B_{n1} \right\|_{L^2(S_{\alpha\beta, \varphi_{\alpha\beta}})} \\ &= nh \left\| f_{n,\alpha\beta}^* \right\|_{L^2(S_{\alpha\beta, \varphi_{\alpha\beta}})} \left\{ \left\| f_{n,\alpha\beta}^* \right\|_{L^2(S_{\alpha\beta, \varphi_{\alpha\beta}})} - 2h^2 \left\| B_{n1} \right\|_{L^2(S_{\alpha\beta, \varphi_{\alpha\beta}})} \right\} \\ &\quad t_n \geq nha_n \{a_n - 2h^2 CM\} \end{aligned}$$

which, using the condition that $a_n^{-1} = o(nh + h^{-2})$, entails that $t_n \rightarrow \infty$ as $n \rightarrow \infty$. By the definition of the test (36)

$$(55) \quad p_n = P \left[T'_n + t_n \geq \Phi^{-1}(1 - \eta) V(K, \varphi, \sigma) \right].$$

Now (54), (55) and $t_n \rightarrow \infty$ yield $\lim_{n \rightarrow \infty} p_n = 1$.

QED.

A.5 Proof of Theorem 8

In parallel to the proof of Theorem 6, one can decompose

$$\begin{aligned} \sum_{l=1}^n \widehat{f_{\alpha\beta}^*}^2(X_{l\alpha}, X_{l\beta})/n &= \sum_{1 \leq i \neq j \leq n} \widetilde{H}(X_i, \varepsilon_i, X_j, \varepsilon_j) + \sum_{i=1}^n \widetilde{H}(X_i, \varepsilon_i, X_i, \varepsilon_i) + \\ &\quad \sum_{l=1}^n f_{\alpha\beta}^{*2}(X_{l\alpha}, X_{l\beta})/n + 2h^2 \sum_{l=1}^n f_{\alpha\beta}^*(X_{l\alpha}, X_{l\beta}) B_1(X_{l\alpha}, X_{l\beta})/n + o_p(h^2) \end{aligned}$$

in which

$$\widetilde{H}(X_i, \varepsilon_i, X_j, \varepsilon_j) = \varepsilon_i \varepsilon_j \sum_{l=1}^n \frac{1}{n^3} (\tilde{w}_{i\alpha, l} - \tilde{w}_{i\alpha} - \tilde{w}_{i\beta, l}) (\tilde{w}_{j\alpha\beta, l} - \tilde{w}_{j\alpha, l} - \tilde{w}_{j\beta, l}) \sigma(X_i) \sigma(X_j)$$

with

$$(56) \quad \tilde{w}_{i\alpha, l} = \frac{1}{n} K_h(X_{l\alpha} - X_{i\alpha}) \frac{\varphi_{\underline{\alpha}}(X_{i\alpha})}{\varphi(X_{l\alpha}, X_{i\alpha})},$$

$$(57) \quad w_{i\alpha\beta, l} = \frac{1}{n} K_h(X_{l\alpha} - X_{i\alpha}, X_{l\beta} - X_{i\beta}) \frac{\varphi_{\underline{\alpha\beta}}(X_{i\alpha\beta})}{\varphi(X_{l\alpha}, X_{l\beta}, X_{i\alpha\beta})}.$$

It is directly verified that for $w_{i\alpha}$ defined in (44) and $w_{i\alpha\beta}$ defined in (47)

$$\begin{aligned} &\sum_{l=1}^n \frac{1}{n} (\tilde{w}_{i\alpha\beta, l} - \tilde{w}_{i\alpha, l} - \tilde{w}_{i\beta, l}) (\tilde{w}_{j\alpha\beta, l} - \tilde{w}_{j\alpha, l} - \tilde{w}_{j\beta, l}) \\ &= \int (w_{i\alpha\beta} - w_{i\alpha} - w_{i\beta}) (w_{j\alpha\beta} - w_{j\alpha} - w_{j\beta}) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta \left\{ 1 + O_p(n^{-1/2}) \right\} \end{aligned}$$

uniformly for all $1 \leq i, j \leq n$. Note also the fact that $E(\varepsilon_i | X_1, \dots, X_n) = 0$, $E(\varepsilon_i^2 | X_1, \dots, X_n) = 1$, and using the independence of $\varepsilon_1, \dots, \varepsilon_n$, one obtains

$$\sum_{i=1}^n \widetilde{H}(X_i, \varepsilon_i, X_i, \varepsilon_i) = \left\{ 1 + O_p(n^{-1/2}) \right\} \sum_{i=1}^n H(X_i, \varepsilon_i, X_i, \varepsilon_i)$$

while

$$\sum_{1 \leq i \neq j \leq n} \widetilde{H}(X_i, \varepsilon_i, X_j, \varepsilon_j) = \sum_{1 \leq i \neq j \leq n} H(X_i, \varepsilon_i, X_j, \varepsilon_j) + O_p(n^{-1/2})$$

with H as defined in (50). These, plus the trivial facts that

$$\sum_{l=1}^n f_{\alpha\beta}^{*2}(X_{l\alpha}, X_{l\beta})/n = \int f_{\alpha\beta}^{*2}(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta + O_p(n^{-1/2})$$

$$\sum_{l=1}^n f_{\alpha\beta}^*(X_{l\alpha}, X_{l\beta}) B_1(X_{l\alpha}, X_{l\beta})/n = \int f_{\alpha\beta}^*(x_\alpha, x_\beta) B_1(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta + O_p(n^{-1/2})$$

establish Theorem 8.

References

- ACHMUS, S. (1999) Nichtparametrische additive Modelle. *Thesis at Technische Universität Braunschweig, Germany*
- ANDREWS, D.W.K. AND WHANG, Y.-J. (1990) Additive interactive regression models: circumvention of the curse of dimensionality. *Econometric Theory* **6**: 466-479.
- 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. *Annals of Statistics*, **17**: 453-510.
- CHEN, R., LIU, J.S. AND TSAY, R.S. (1995) Additivity tests for nonlinear autoregression. *Biometrika*, **82**: 369-383.
- COX, D.R. AND HINKLEY, D.V. (1974) *Theoretical Statistics*. Chapman and Hall: London
- DALELANE, C. (1999) 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.
- DERBORT, S., DETTE, H. AND MUNK, A. (1999) A test for additivity in nonparametric regression. *Preprint, Ruhr-Universität Bochum, Germany*
- DENNY, M. AND FUSS, M. (1977) The use of approximation analysis to test for separability and the existence of consistent aggregates. *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. *Annals of Statistics*, **23**: 1896-1920.

- FAN, J. (1993) Local Linear Regression Smoothers and their Minimax Efficiencies. *Annals of Statistics*, **21**: 196-216.
- FAN, J., GASSER, T., GIJBELS, I., BROCKMANN, M. AND ENGEL, J. (1993) Local polynomial fitting: A standard for nonparametric regression. *CORE Discussion Paper 9315*.
- FAN, J., HÄRDLE, W. AND MAMMEN, E. (1998) Direct estimation of low dimensional components in additive models. *Annals of Statistics*, **26**: 943-971.
- 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.
- GOLDMAN, S.M. AND UZAWA, H. (1964) A note on separability in demand analysis. *Econometrica*, **32**: 387-398.
- GOZALO, P.L. AND LINTON, O.B. (1997) Testing additivity in generalized nonparametric regression models. *Preprint, Yale University, USA*
- HALL, P. (1984) Central limit theorem for integrated square error of multivariate nonparametric density estimators, *Journal of Multivariate Analysis*, **14**: 1-16.
- HÄRDLE, W., HUET, S., MAMMEN, E. AND SPERLICH, S. (1999) Semiparametric additive indices for binary response and generalized additive models. *Manuscript, Carlos III de Madrid, Spain*.
- HÄRDLE, W. AND MAMMEN, E. (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics*, **21**: 1926-1947.
- HÄRDLE, W., MAMMEN, E. AND MÜLLER, M. (1998) Testing parametric versus semiparametric modelling in generalized linear models. *Journal of the American Statistical Association*, **93**: 1461-1474.
- HASTIE, T.J. AND TIBSHIRANI, R.J. (1990) *Generalized Additive Models*. Chapman and Hall: London
- HJELLVIK, V., YAO, Q. AND TJØSTHEIM, D. (1998) Linearity testing using local polynomial approximation. *Journal of Statistical Planning and Inference*, **68**, 295-321.
- JOHNSON, R.A. AND WICHERN, D.W. (1988) *Applied Multivariate Statistical Analysis*. 2nd ed. Prentice Hall: Englewood Cliffs.

- KREISS, J.-P., NEUMANN, M.H. AND YAO, Q. (1999) Bootstrap tests for simple structures in nonparametric time series regression. *Preprint, Technische Universität Braunschweig, Germany*
- LEONTIEF, W. (1947) Introduction to a theory of the internal structure of functional relationships. *Econometrica*, **15**: 361-373.
- LINTON, O.B. (1997) Efficient Estimation of Additive Nonparametric Regression Models. *Biometrika*, **84**: 469-473.
- LINTON, O.B. (1999) Efficient Estimation of Generalized Additive Nonparametric Regression Models. *Preprint, Yale and LSE*.
- LINTON, O.B. AND HÄRDLE, W. (1996) Estimation of additive regression models with known links. *Biometrika*, **83**: 529-540.
- LINTON, O.B. AND NIELSEN, J.P. (1995) A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, **82**: 93-101.
- LIU, R. (1988) Bootstrap procedures under some non i.i.d. models. *Annals of Statistics*, **16**: 1696-1708.
- MAMMEN, E., LINTON, O.B. AND NIELSEN, J.P. (1999) The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. to appear in *Annals of Statistics*,
- MÜLLER, M. (1997) Computer-assisted generalized partial linear models. *Proceedings of the 29th Symposium of the Interface, Houston, TX, May 14-17*, vol. **91/1**: 221-230.
- NADARAYA, E.A. (1964) On estimating regression. *Theory of Probability and Applications*, **9**: 141-142.
- NEWBY, W.K. (1994) Kernel estimation of partial means. *Econometric Theory*, **10**: 233-253.
- NEWBY, W.K. (1995) Convergence rates for series estimators. In *Maddala, G.S., Phillips, P.C.B., Srinivasan, T.N. (eds.), Statistical Methods of Economics and Quantitative Economics: Essays in Honor of C.R. Rao*. Blackwell, Cambridge, 254-275.
- OPSOMER, J.D. AND RUPPERT, D. (1997) Fitting a bivariate additive model by local polynomial regression. *Annals of Statistics*, **25**: 186-211.
- RUPPERT, D. AND WAND, M.P. (1994) Multivariate locally weighted least squares regression. *Annals of Statistics*, **22**: 1346-1370.

- SEVERANCE-LOSSIN, E. AND SPERLICH, S. (1999) Estimation of derivatives for additive separable models. to appear in *Statistics*, **33**: 241-265.
- SILVERMAN, B.W. (1986) Density Estimation for Statistics and Data Analysis. *Chapman and Hall, London*
- SPERLICH, S., LINTON, O.B. AND HÄRDLE, W. (1999) Integration and Backfitting methods in additive models: Finite sample properties and comparison. to appear in *Test*,
- 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 unabridged version of Integration and Backfitting Methods ...*
- STONE, C.J. (1985) Additive regression and other nonparametric models. *Annals of Statistics*, **13**: 689-705.
- STONE, C.J., HANSEN, M.H., KOOPERBERG, C. AND TRUONG, Y. (1997) Polynomial splines and their tensor products in extended linear modeling (with discussion). *Annals of Statistics*, **25**: 1371-1470.
- TJØSTHEIM, D. AND AUESTAD, B.H. (1994) Nonparametric identification of nonlinear time series: projections. *Journal of the American Statistical Association*, **89**: 1398-1409.
- WAHBA, G. (1992) Spline models for observational data. *Society for Industrial and Applied Mathematics: Philadelphia, Pennsylvania* 2nd printing.
- 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). *Annals of Statistics* , **14**: 1261-1350.