Econometric Theory, **18**, 2002, 197–251. Printed in the United States of America.
DOI: 10.1017.S0266466602182016

NONPARAMETRIC ESTIMATION AND TESTING OF INTERACTION IN ADDITIVE MODELS

STEFAN SPERLICH
Universidad Carlos III de Madrid

DAG TJØSTHEIM
University of Bergen

LIJIAN YANG
Michigan State University

We consider an additive model with second-order interaction terms. Both marginal integration estimators and a combined backfitting-integration estimator are proposed for all components of the model and their derivatives. The corresponding asymptotic distributions are derived. Moreover, two test statistics for testing the presence of interactions are proposed. Asymptotics for the test functions and local power results are obtained. Because 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 developed. 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.

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

$$Y = m(X) + \sigma(X)\varepsilon, \tag{1}$$

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 $\mathrm{Var}(\varepsilon) = 1$. Although in principle this

The authors thank Stefan Profit and Oliver Linton for inspiration and helpful discussions. The comments from two anonymous referees and co-editor Donald Andrews have resulted in substantial extension of the work and improved presentation. 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 BEC2001-1270. Yang's research was supported in part by grant DMS 9971186 from the United States National Science Foundation. Address correspondence to: Stefan Sperlich, Universidad Carlos III de Madrid, Departamento de Estadística y Econometría, c/Madrid 126, 28903 Getafe-Madrid, Spain.

 0266-4666/02 $9.50

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.,

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

where the functions f_α are unknown. Additive models were discussed already by Leontief (1947). He analyzed so-called separable functions, i.e., functions that 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, e.g., 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).

In statistics, the usefulness of additive modeling has been emphasized by among others Stone (1985) and Hastie and Tibshirani (1990). Additive models constitute 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, purely 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) presents 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 of interest for economists (Severance-Lossin and Sperlich, 1999).

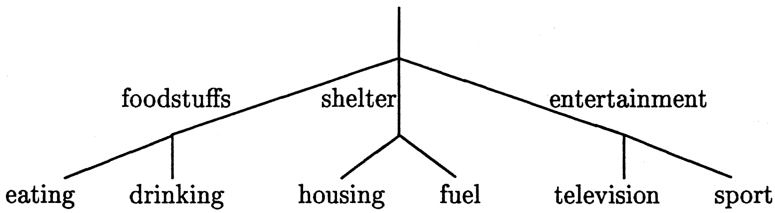
A weakness of the purely additive model is that interactions between the explanatory variables are completely ignored, and in certain econometric contexts—production function modeling being one of them—the absence of interaction terms has been criticized. In this paper we allow for second-order interactions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

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

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

and the main objective of the paper is to consider estimation and testing in such models, mainly using marginal integration techniques. Notice that introducing higher order interaction would gradually bring back problems of interpretation and the curse of dimensionality.

In the sequel the model (3) will be referred to as an additive model with interactions as opposed to the purely additive model (2). Actually, models with interactions are not uncommon in economics, but parametric functions have typically been used to describe them, 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 that we want to model utility for a household and consider the utility tree



Example: utility tree for households.

In a nonparametric approach this would lead to the model

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

where the x_α's stand for the inputs of the bottom line in the tree (counted from left to right). The interaction function f_{12} stands for interaction in foodstuffs, f_{34} in shelter, and f_{56} for entertainment; other interactions are assumed to be nonexistent.

In the context of production function estimation various (parametric) functional forms including interaction have been proposed as alternatives to the classic Cobb–Douglas model, resulting in the

Generalized Cobb–Douglas $$\ln Y = c + \sum_{\alpha=1}^{d} \sum_{\beta=1}^{d} c_{\alpha\beta} \ln\left(\frac{X_\alpha + X_\beta}{2}\right),$$

Translog $$\ln Y = c + \sum_{\alpha=1}^{d} c_\alpha \ln X_\alpha + \sum_{\alpha=1}^{d} \sum_{\beta=1}^{d} c_{\alpha\beta} (\ln X_\alpha)(\ln X_\beta),$$

Generalized Leontief $$Y = c + \sum_{\alpha=1}^{d} c_\alpha \sqrt{X_\alpha} + \sum_{\alpha=1}^{d} \sum_{\beta=1}^{d} c_{\alpha\beta} \sqrt{X_\alpha X_\beta},$$

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

$$\text{Quadratic} \qquad Y = c + \sum_{\alpha=1}^{d} c_\alpha X_\alpha + \sum_{\alpha=1}^{d} \sum_{\beta=1}^{d} c_{\alpha\beta} X_\alpha X_\beta,$$

$$\text{Generalized concave} \qquad Y = \sum_{\alpha=1}^{d} \sum_{\beta=1}^{d} c_{\alpha\beta} X_\beta f_{\alpha\beta}\left(\frac{X_\alpha}{X_\beta}\right),$$

$f_{\alpha\beta}$ known and concave.

Although parametric, they all have a functional form encompassed by (3). For further discussion and references see Section 7.3, where we present a detailed nonparametric example.

Turning to the estimation of model (3), we can construct an asymptotic theory for marginal integration and also for a one-step efficient estimator analogous to that of Linton (1997). However, extending the remarkable work of Mammen et al. (1999) on the asymptotic theory of backfitting seems difficult as a result of its strong dependence on projector theory, which would be hard to carry through for the interaction terms.

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

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

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

can be applied relatively straightforwardly to our framework with interaction terms included. Such mixed models are interesting from a practical and also from a theoretical point of view, and they permit estimating θ with the parametric $\sqrt{n}$ rate. Also, an extension to generalized additive models should be possible. We refer to Linton and Härdle (1996) and Härdle, Huet, Mammen, and Sperlich (1998) for a more detailed description of these models.

Coming finally to the issue of testing, it should be noted that sometimes economic reasoning is used as a justification for omitting interaction terms, as in the utility example. However, from a general statistical point of view one would like to *test* for the potential presence of interactions. Additivity tests developed so far have mostly been focused on testing whether a function $m(x_1,\ldots,x_d)$ is purely additive or not in the sense of (2). However, if pure additivity is rejected, the empirical researcher would like to know exactly which interaction terms are relevant. A main point of the present paper is to test directly for such interactions (cf. the example of Section 7.3), and we propose two basic functionals for doing this for a pair of variables (x_α, x_β). The most obvious one is to estimate $f_{\alpha\beta}$ of (3) and then use a test functional

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

where π is an appropriate nonnegative 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 previous 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.

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 modeling m by a purely additive model as in equation (2) as 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 by Derbort, Dette, and Munk (2002), who both only consider fixed designs.

Our paper is divided into two main parts concerned with 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, Sections 5 and 7 provide several simulation experiments and an application to real data. 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

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

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

Here in general, $X = (X_1, X_2, \ldots, X_d)$ represents a sequence of independent and 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 $\mathrm{Var}(\varepsilon) = 1$. We permit heteroskedasticity, and the variance function is denoted by $\sigma^2(X)$. In the previous expression c is a constant, $\{f_\alpha(\cdot)\}_{\alpha=1}^d$ and $\{f_{\alpha\beta}(\cdot)\}_{1\le\alpha<\beta\le d}$ are real-valued unknown functions. Clearly, the representation (5) is not unique, but it can be made so by introducing for $\alpha = 1,2,\ldots,d$, the identifiability conditions

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

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

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

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

It is important to observe that equations (6) and (7) do not represent restrictions on our model. Indeed, if a representation as given in (3) or (5) does not satisfy (6) and (7), one can easily change it so that it conforms to these identifiability conditions by taking the following steps.

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

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

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

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

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

In the sequel, unless otherwise stated, each f_α and $f_{\alpha\beta}$ will be assumed to satisfy (6) and (7).

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 $X = (X_1, \ldots, X_d)$ 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 density of X_α, $X_{\underline{\alpha}}$, $X_{\underline{\alpha\beta}}$, and X by $\varphi_\alpha(x_\alpha)$, $\varphi_{\underline{\alpha}}(x_{\underline{\alpha}})$, $\varphi_{\underline{\alpha\beta}}(x_{\underline{\alpha\beta}})$, and $\varphi(x)$, respectively.

We now define by marginal integration

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

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

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

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

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

$$D_{\alpha\alpha} = \{(\gamma,\delta)|1 \le \gamma < \delta \le d, \gamma \in D_\alpha, \delta \in D_\alpha\}, \qquad \text{and}$$
$$D_{\alpha\beta} = \{(\gamma,\delta)|1 \le \gamma < \delta \le d, \gamma \in D_\alpha \cap D_\beta, \delta \in D_\alpha \cap D_\beta\}.$$

The quantities F_α and $F_{\alpha\beta}$ do not satisfy the identifiability conditions. Actually, (6) and (7) entail the following lemma.

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

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

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

3. $$c_{\alpha\beta} = \int \{F_{\alpha\beta}(u, x_\beta) - F_\alpha(u)\}\varphi_\alpha(u)du - F_\beta(x_\beta) + \int m(x)\varphi(x)dx$$
$$f_{\alpha\beta}(x_\alpha, x_\beta) = F_{\alpha\beta}(x_\alpha, x_\beta) - F_\alpha(x_\alpha) - \int \{F_{\alpha\beta}(u, x_\beta) - F_\alpha(u)\}\varphi_\alpha(u)du$$

We define another auxiliary function:

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

In Section 3 we will estimate F_α and $f^*_{\alpha\beta}$. These quantities are more convenient to work with than f_α and $f_{\alpha\beta}$, and as shown by Lemma 1 and the definition of $f^*_{\alpha\beta}$ they can be identified with f_α and $f_{\alpha\beta}$ up to a constant. That $f^*_{\alpha\beta}$ is a convenient substitute for $f_{\alpha\beta}$ when it comes to testing is shown by the following corollary.

COROLLARY 1. *Let $f^*_{\alpha\beta}(x_\alpha, x_\beta)$ and $f_{\alpha\beta}(x_\alpha, x_\beta)$ be as defined previously. Then*

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

The corollary suggests the use of the following functional for testing of additivity of the αth and βth directions.

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

where

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

with estimates $\hat{F}_{\alpha\beta}$, $\hat{F}_\alpha$, and $\hat{F}_\beta$ 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

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

In practice, because $\varphi_{\alpha\beta}$ is unknown, the integral in (9) is replaced by an empirical average (cf. Section 6).

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

$$\int \{F^{(1,1)}_{\alpha\beta}(x_\alpha, x_\beta)\}^2 \pi(x_\alpha, x_\beta)\,dx_\alpha\,dx_\beta$$

is zero, because, under the identifiability condition (7), $F^{(1,1)}_{\alpha\beta} = 0$ 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 (9), estimators of the interaction terms must be prescribed. Imagine the X-variables to be scaled so that we can choose the same bandwidth h for the directions represented by α, β, and g for $\underline{\alpha\beta}$. Further, let K and L be kernel functions and define $K_h(\cdot) = h^{-1}K(\cdot/h)$ and $L_g(\cdot) = g^{-1}L(\cdot/g)$. 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.

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

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

$$\hat{F}_{\alpha\beta}(x_\alpha, x_\beta) = \frac{1}{n}\sum_{l=1}^{n} \hat{m}(x_\alpha, x_\beta, X_{l\underline{\alpha\beta}}), \quad \hat{F}_\alpha(x_\alpha) = \frac{1}{n}\sum_{l=1}^{n} \hat{m}(x_\alpha, X_{l\underline{\alpha}}), \tag{11}$$

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

The estimator $\hat{m}(x_\alpha, x_\beta, X_{l\underline{\alpha\beta}})$ will be called the preestimator 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

$$\sum_{i=1}^{n} \{Y_i - a_0 - a_1(X_{i\alpha} - x_\alpha) - a_2(X_{i\beta} - x_\beta)\}^2 K_h(X_{i\alpha} - x_\alpha, X_{i\beta} - x_\beta) \times L_g(X_{i\underline{\alpha\beta}} - X_{l\underline{\alpha\beta}}) \tag{12}$$

for each fixed l. Accordingly we define

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

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

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

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

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 preestimator $\hat{m}(x_\alpha, X_{l\underline{\alpha}})$, with $e_1 = (1,0)$, we define

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

in which

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

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

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.

To derive the asymptotics of these estimators we make use of the concept of equivalent kernels; see Ruppert and Wand (1994) and Fan, Gasser, Gijbels, Brockmann, and Engel (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

$$K_\nu^\star(u) := \sum_{t=0}^{p} s_{\nu t} u^t K(u) \tag{13}$$

in the one-dimensional case, where $S = (\int u^{t+s} K(u)\,du)_{0 \le t, s \le p}$ and $S^{-1} = (s_{\nu t})_{0 \le \nu, t \le p}$ 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, e.g., $p = 1$, we have

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

with $\mu_j = \int u^j K(u)\,du$.

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

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 with the nonnegative $K(\cdot)$ satisfying $\int K(u)\,du = 1$. The $(d-1)$-dimensional kernel $L(\cdot)$ is a product of univariate kernels $L(u)$ of order $q \ge 2$, i.e.,

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

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

(A3) The functions $f_\alpha, f_{\alpha\beta}$ have bounded Lipschitz continuous derivatives of order q.

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

Remark 2. For ease of presentation we will use local linear smoothers in Theorems 1–3. But from Severance-Lossin and Sperlich (1999) it is clear that this can easily be extended to arbitrary degrees $p \geq 1$. Then, assumption (A02) would change to $nhg^{(d-1)}/\ln(n) \to \infty$, $g^q/h^{p+1} \to 0$, and $h = h_0 n^{-(1/2p+3)}$, whereas in (A3) one would have to require that the functions $f_\alpha, f_{\alpha\beta}$ have bounded Lipschitz continuous derivatives of order $\max\{p+1, q\}$.

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

$$\sqrt{nh}\{\hat{F}_\alpha(x_\alpha) - F_\alpha(x_\alpha) - h^2 b_1(x_\alpha)\} \xrightarrow{\mathcal{L}} N\{0, v_1(x_\alpha)\}, \tag{14}$$

where F_α is given by (8) and Lemma 1, $\hat{F}_\alpha$ by (11). The variance is

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

and the bias

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

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

$$K_\nu^*(u,v) := K(u,v)s_\nu(1,u,v)^T, \tag{15}$$

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

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

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 (10). Since we have a two-dimensional problem, the assumptions have to be adjusted accordingly.

(A1) The kernels $K(\cdot)$ and $L(\cdot)$ are bounded, symmetric, compactly supported, and Lipschitz continuous. The bivariate kernel K is a product kernel

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

such that (with some abuse of notation) $K(u,v) = K(u)K(v)$, where $K(u)$ and $K(v)$ are identical functions with the nonnegative $K(\cdot)$ satisfying $\int K(u)du = 1$. 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 $nh^2g^{(d-2)}/\ln^2(n) \to \infty$ and $nhg^{(d-1)}/\ln^2(n) \to \infty$, $g^q/h^2 \to 0$ and $h = h_0 n^{-1/6}$.

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

$$\sqrt{nh^2}\{\widehat{f^*_{\alpha\beta}}(x_\alpha,x_\beta) - f^*_{\alpha\beta}(x_\alpha,x_\beta) - h^2B_1(x_\alpha,x_\beta)\} \xrightarrow{\mathcal{L}} N\{0, V_1(x_\alpha,x_\beta)\}, \quad \textbf{(16)}$$

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

$$V_1(x_\alpha,x_\beta) = \|K_0^*\|_2^2 \int \sigma^2(x) \frac{\varphi^2_{\underline{\alpha\beta}}(x_{\underline{\alpha\beta}})}{\varphi(x)}\, dx_{\underline{\alpha\beta}}$$

and

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

Theorems 1 and 2 are concerned with the individual components. The last result of this section (whose proof essentially follows from Theorems 1 and 2 and will be omitted) states the asymptotics of the combined regression estimator $\widetilde{m}(x)$ of $m(x)$ given by

$$\widetilde{m}(x) = \sum_{\alpha=1}^{d} \hat{F}_\alpha(x_\alpha) + \sum_{1\le\alpha<\beta\le d} \widehat{f^*_{\alpha\beta}}(x_\alpha,x_\beta) - (d-1)\,\frac{1}{n}\sum_{i=1}^{n} Y_i. \quad \textbf{(17)}$$

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

$$\sqrt{nh^2}\{\widetilde{m}(x) - m(x) - h^2B_m(x)\} \xrightarrow{\mathcal{L}} N\{0, V_m(x)\}, \quad \textbf{(18)}$$

where h *is as in (A2),*

$$B_m(x) = \sum_{1\le\alpha<\beta\le d} B_1(x_\alpha,x_\beta) \quad \textit{and} \quad V_m(x) = \sum_{1\le\alpha<\beta\le d} V_1(x_\alpha,x_\beta).$$

3.2. Estimation of Derivatives

Because 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 additiv-

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

ity without second-order interaction because $F_{\alpha\beta}^{(1,1)} = 0$ 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 $(x_\alpha, x_\beta, X_{i\underline{\alpha\beta}})$ 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

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

and accordingly define our estimator by

$$\hat{F}_{\alpha\beta}^{(1,1)}(x_\alpha, x_\beta) = \frac{1}{n}\sum_{i=1}^{n} e_4(Z_{\alpha\beta}^T W_{i,\alpha\beta} Z_{\alpha\beta})^{-1} Z_{\alpha\beta}^T W_{i,\alpha\beta} Y, \tag{19}$$

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

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

This estimator is bivariate locally quadratic for the directions α and β and locally constant elsewhere.

Recalling the approach of the preceding 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, $0 \le \nu \le 5$, row of

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

The relationship between $\mathcal{S}^{-1}$ and $(Z_{\alpha\beta}^T W_{l,\alpha\beta} Z_{\alpha\beta})^{-1}$ is given in Lemma A2 in the Appendix.

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

$$\int uvK_3^*(u,v)\,dudv = 1,$$

$$\int u^i K_3^*(u,v)\,dudv = \int v^i K_3^*(u,v)\,dudv = 0 \qquad \text{for } i = 0,1,2,3,\ldots,$$

$$\int u^2 v^i K_3^*(u,v)\,dudv = \int u^i v^2 K_3^*(u,v)\,dudv = 0 \qquad \text{for } i = 0,1,2,3,\ldots.$$

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

Further, using a local quadratic smoother, Assumption (A3) changes to the following assumption.

(A7) The functions $f_\alpha, f_{\alpha\beta}$ have bounded Lipschitz continuous derivatives of order $\max\{3,q\}$.

Then we have the following theorem.

THEOREM 4. *Under conditions (A1), and (A4)–(A7),*

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

where

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

and

$$B_2(x_\alpha,x_\beta) = \mu_4\mu_2^{-1}\Bigg[\frac{1}{2}\{f_{\alpha\beta}^{(2,1)}(x_\alpha,x_\beta)\varrho_\beta + f_{\alpha\beta}^{(1,2)}(x_\alpha,x_\beta)\varrho_\alpha\} + \frac{1}{3!}\{f_{\alpha\beta}^{(3,1)}(x_\alpha,x_\beta) + f_{\alpha\beta}^{(1,3)}(x_\alpha,x_\beta) + f_{\alpha\beta}^{(3,0)}(x_\alpha,x_\beta)\varrho_\beta + f_{\alpha\beta}^{(0,3)}(x_\alpha,x_\beta)\varrho_\alpha + f_\alpha^{(3)}(x_\alpha)\varrho_\beta + f_\beta^{(3)}(x_\beta)\varrho_\alpha\}\Bigg]$$

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

with

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

and ϱ_β *defined analogously.*

4. A ONE-STEP EFFICIENT ESTIMATOR

It is known that for purely additive models of the form

$$E[Y|X = x] = m(x_1, \ldots, x_d) = c + \sum_{\alpha=1}^{d} f_\alpha(x_\alpha) \tag{20}$$

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

$$U_{i1} = Y_i - c - \sum_{\alpha=2}^{d} f_\alpha(X_{i\alpha}). \tag{21}$$

Basically, this is the idea of the (iterative) backfitting procedure. Linton (1997, 2000) suggests an estimator combining the backfitting with the marginal integration idea. He first performs the marginal integration procedure to obtain $\hat{f}_\alpha, \ldots, \hat{f}_d$, and then derives the partial residuals

$$\widetilde{U}_{i1} = Y_i - c - \sum_{\alpha=2}^{d} \hat{f}_\alpha(X_{i\alpha}). \tag{22}$$

Subsequently, he applies a one-dimensional local linear smoother on the $\widetilde{U}_{i\alpha}$. This is equivalent to a one-step backfit.

Assuming that we already know the true underlying model, we consider an extension of his approach to our additive interaction models of the form (5). This ought to be of some interest, because 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 could be constructed. Hastie and Tibshirani discuss this point but only for one interaction term, and they merely supply some heuristic motivation for their methods.

At the outset we do not restrict ourselves to homoskedastic errors but let $\sigma_\alpha^2(x_\alpha) = \mathrm{Var}[Y - m(x)|X_\alpha = x_\alpha]$, with $\sigma_{\alpha\beta}(x_\alpha, x_\beta)$ defined analogously, and assume the existence of finite second moments for these quantities. We present an analogue to the efficient estimator of Linton (1997), but, as in his paper, we can claim a smaller variance than for the classic marginal integration only in case of homoskedasticity.

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

Consider model (5) and the two partial residuals

$$U_{i\alpha} = Y_i - \sum_{\gamma \neq \alpha} f_\gamma(X_{i\gamma}) - \sum_{1 \le \gamma < \beta \le d} f_{\gamma\beta}(X_{i\gamma}, X_{i\beta}) + \sum_{(\gamma,\delta) \in D_{\alpha\alpha}} c_{\gamma\delta}$$

$$= Y_i - m(X_i) + F_\alpha(X_{i\alpha}), \tag{23}$$

$$U_{i\alpha\beta} = Y_i - \sum_{\gamma=1}^{d} f_\gamma(X_{i\gamma}) - \sum_{\substack{1 \le \gamma < \delta \le d \\ (\gamma,\beta) \neq (\alpha,\beta)}} f_{\gamma\delta}(X_{i\gamma}, X_{i\delta}) + c_{\alpha\beta}$$

$$= Y_i - m(X_i) + f^*_{\alpha\beta}(X_{i\alpha}, X_{i\beta}). \tag{24}$$

For the estimation of the functional form it does not matter whether we correct for the constants ($c_{\alpha\beta}$) 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 of this issue can also be found in Linton (1997, 2000).

Let now $\check{F}_\alpha^{opt}$ be the local linear regressor of $U_{i\alpha}$ in (23) with respect to X_α, and $\check{f}_{\alpha\beta}^{*opt}$ the one of $U_{i\alpha\beta}$ in (24) versus (X_α, X_β). From Fan (1993) and Ruppert and Wand (1994) we know that under standard regularity conditions the asymptotic properties are

$$\sqrt{nh_e}\{\check{F}_\alpha^{opt}(x_\alpha) - F_\alpha(x_\alpha) - h_e^2 b_e(x_\alpha)\} \to N\{0, v_e(x_\alpha)\}, \tag{25}$$

$$\sqrt{nh_e^2}\{\check{f}_{\alpha\beta}^{*opt}(x_\alpha, x_\beta) - f^*_{\alpha\beta}(x_\alpha, x_\beta) - h_e^2 B_e(x_\alpha, x_\beta)\} \to N\{0, V_e(x_\alpha, x_\beta)\} \tag{26}$$

with

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

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

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

where J is the one- or two-dimensional kernel (corresponding to (23) and (24), respectively) with $\mu_2(J) = \int u^2 J(u)du$ in the one-dimensional case and h_e is the associated bandwidth.

Suppose now that the kernel J is the same as the kernel K used to define $\hat{F}_{\alpha\beta}(x_\alpha, x_\beta)$ and $\hat{F}_\alpha(x_\alpha)$ in (11). Then the bias expressions are the same for the efficient estimators $\check{F}_\alpha^{opt}(x_\alpha)$ and $\check{f}_{\alpha\beta}^{*opt}(x_\alpha, x_\beta)$ and the original marginal integration estimators. Using the simple trick of Linton (1997), it is also straightforward to verify that if $\sigma(x) \equiv \sigma_0 > 0$ is a constant, then $\check{F}_\alpha^{opt}(x_\alpha)$ and $\check{f}_{\alpha\beta}^{*opt}(x_\alpha, x_\beta)$ have smaller variances than $\hat{F}_\alpha(x_\alpha)$ and $\hat{F}_{\alpha\beta}(x_\alpha, x_\beta)$, respectively. The appellation "efficient estimator" is justified in the sense that this estimator has the same asymptotic variance as if the other components were known.

Now we replace $U_{i\alpha}$, $U_{i\alpha\beta}$ by $\widetilde{U}_{i\alpha}$, $\widetilde{U}_{i\alpha\beta}$ by replacing the real functions $F_\gamma, f^*_{\gamma\delta}$ 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

$$\min_{c_0,c_1} \sum_{i=1}^{n} \{\widetilde{U}_{i\alpha} - c_0 - c_1(X_{i\alpha} - x_\alpha)\}^2 J\left(\frac{X_{i\alpha} - x_\alpha}{h_e}\right). \tag{27}$$

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

$$\min_{c_0,c_1,c_2} \sum_{i=1}^{n} \{\widetilde{U}_{i\alpha\beta} - c_0 - c_1(X_{i\alpha} - x_\alpha) - c_2(X_{i\beta} - x_\beta)\}^2 J\left(\frac{X_{i\alpha} - x_\alpha}{h_e}, \frac{X_{i\beta} - x_\beta}{h_e}\right). \tag{28}$$

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

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

$$\sqrt{nh_e}\{\check{F}_\alpha(x_\alpha) - \check{F}^{opt}_\alpha(x_\alpha)\} \xrightarrow{p} 0, \tag{29}$$

$$\sqrt{nh_e^2}\{\check{f}^*_{\alpha\beta}(x_\alpha, x_\beta) - \check{f}^{*opt}_{\alpha\beta}(x_\alpha, x_\beta)\} \xrightarrow{p} 0 \tag{30}$$

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

Apart from the theoretical differences made apparent by (25), (26), (29), and (30), there is also another, substantial, difference between backfitting, marginal integration, and this efficient estimator. The backfitting consists in estimating the additive components after a projection of the regression function into the space of purely 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, and Härdle 1999). The efficient estimator is a mixture and thus suffers from the lack of interpretability if the assumed model structure is not completely fulfilled. This could be a disadvantage in empirical research. Moreover, in the context of testing model structure this may lead to problems, especially if we use bootstrap realizations 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 $n = 169$ observations. Certainly, an intensive computational investigation of not only ours but also

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

alternative estimation procedures for additive models would be of interest, but this 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 et al. (1999) but without interaction terms.

Here, we present a small illustration to indicate how these procedures behave in small samples. The data have been generated from the model

$$m(x) = E(Y|X=x) = c + \sum_{j=1}^{3} f_j(x_j) + f_{12}(x_1,x_2), \tag{31}$$

where

$$f_1(u) = 2u, \qquad f_2(u) = 1.5\sin(-1.5u), \tag{32}$$

$$f_3(u) = -u^2 + E(X_3^2) \quad \text{and} \quad f_{12}(u,v) = auv$$

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 $(15/16)(1-u^2)^2 \mathbb{1}\{|u| \le 1\}$ for $K(u)$ and $L(u)$, and we used 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 et al. (1999), but it must be admitted that a complete and practically useful solution to this problem remains to be found. This is particularly true (also for the bandwidth h_e) if applying the one-step efficient estimator.

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

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_{12} is given in the lower left window. For 100 repetitions we estimated the functions on a grid with the previously 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. 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

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

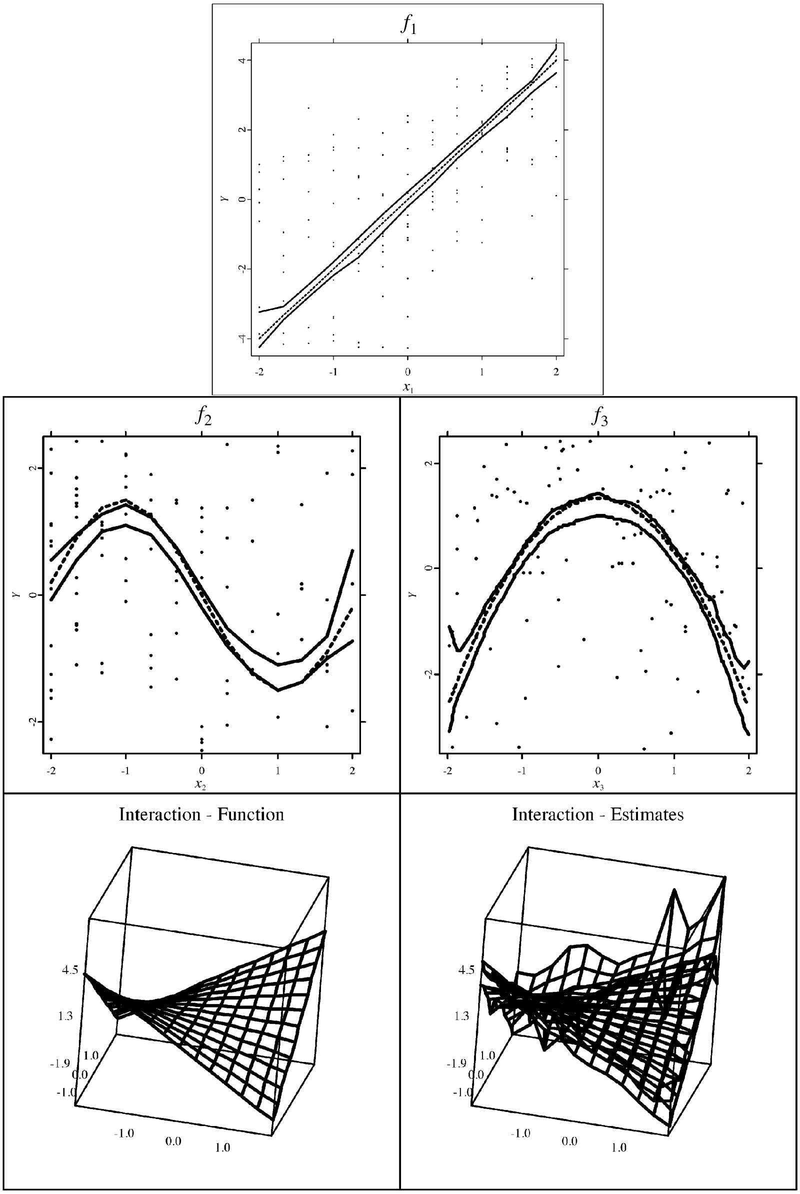


FIGURE 1. Performance of the "simple" marginal integration estimator. Real functions (dashed) and extreme points for 100 of their estimates (solid). For the first run also the response variable Y (points) is given. Position: f_1 (top), f_2 (upper left), f_3 (upper right), f_{12} (lower left), and extreme points of its estimates after 100 simulation runs (lower right).

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

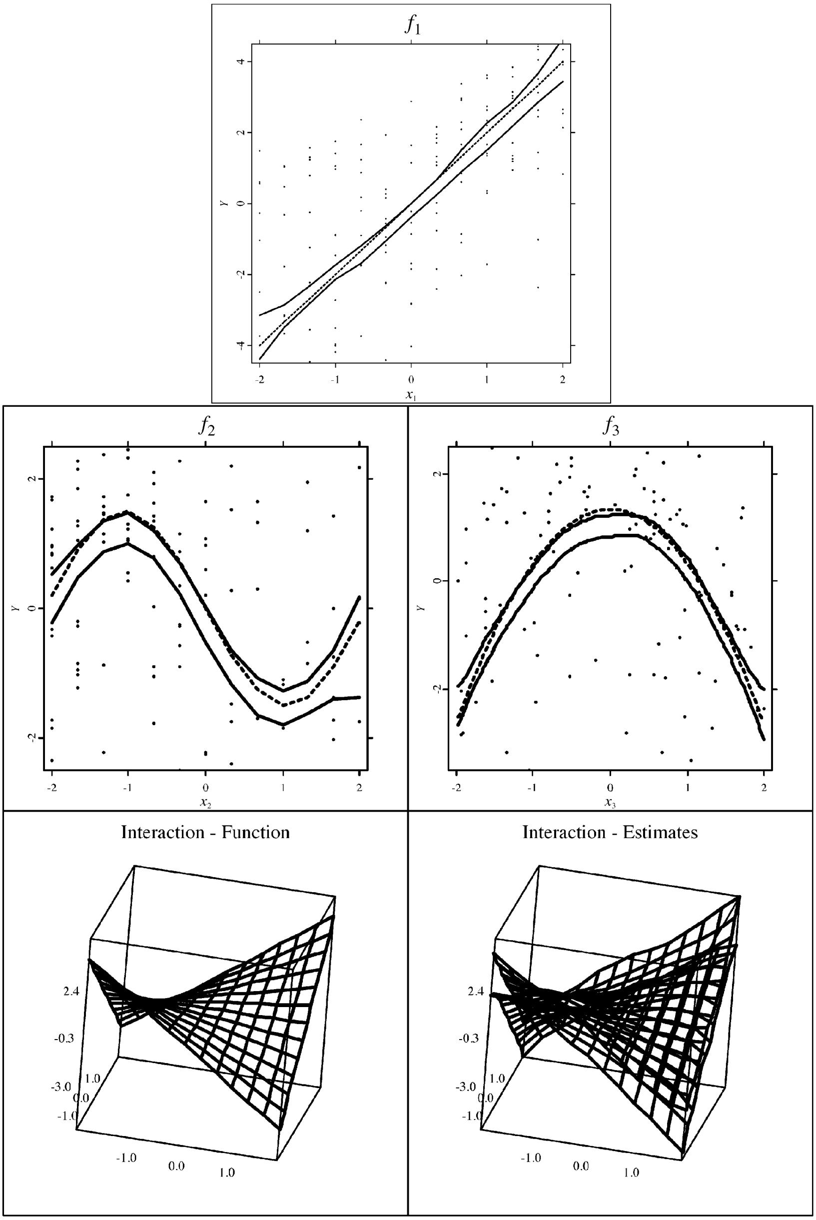


FIGURE 2. Performance of the "efficient" estimator. Real functions (dashed) and extreme points for 100 of their estimates (solid). For the first run also the response variable Y (points) is given. Position: f_1 (top), f_2 (upper left), f_3 (upper right), f_{12} (lower left), and extreme points of its estimates after 100 simulation runs (lower right).

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

both and are similar. All in all, for a sample of this size the two estimators give roughly the same results.

6. TESTING FOR INTERACTION

The second major objective of this paper is to provide tests of second-order interaction. For the model (3) we consider the null hypothesis $H_{0,\alpha\beta}: f_{\alpha\beta} = 0$; 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_γ, X_δ, $1 \le \gamma < \delta \le d$ this can be regarded as a test for separability in the regression model.

6.1. Considering the Interaction Function

We will briefly sketch how the test statistic can be analyzed and then state the theorem giving the asymptotics. The detailed proof is given in 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, because $\varphi_{\alpha\beta}$ is unknown, this functional will subsequently be replaced by an empirical average in (37). Note first that by Theorem 2, equation (16), and some tedious calculations we get the following decomposition:

$$
\begin{aligned}
&\int \widehat{f^*_{\alpha\beta}}^{\,2}(x_\alpha, x_\beta)\varphi_{\alpha\beta}(x_\alpha, x_\beta)\,dx_\alpha\,dx_\beta \\
&\quad = 2\sum_{1\le i<j\le n} H(X_i,\varepsilon_i,X_j,\varepsilon_j) + \sum_{i=1}^{n} H(X_i,\varepsilon_i,X_i,\varepsilon_i) \\
&\qquad + \int f^{*2}_{\alpha\beta}(x_\alpha, x_\beta)\varphi_{\alpha\beta}(x_\alpha, x_\beta)\,dx_\alpha\,dx_\beta \\
&\qquad + 2h^2 \int f^*_{\alpha\beta}(x_\alpha, x_\beta) B_1(x_\alpha, x_\beta)\varphi_{\alpha\beta}(x_\alpha, x_\beta)\,dx_\alpha\,dx_\beta + o_p(h^2),
\end{aligned}
$$

where

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

with weights $w_{i\alpha}$, $w_{i\beta}$, and $w_{i\alpha\beta}$ defined in the Appendix, equations (A.2) and (A.5).

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 (see the proof of Theorem 6 in the Appendix) the following theorem.

THEOREM 6. *Under Assumptions (A1)–(A5),*

$$nh\int \widehat{f^{*}_{\alpha\beta}}^{2}(x_\alpha,x_\beta)\varphi_{\alpha\beta}(x_\alpha,x_\beta)dx_\alpha dx_\beta - \frac{2\{K^{(2)}(0)\}^2}{h}$$

$$\times \int \frac{\varphi_{\alpha\beta}(z_\alpha,z_\beta)\varphi^2_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}})}{\varphi(z)}\sigma^2(z)dz$$

$$- nh\int f^{*2}_{\alpha\beta}(x_\alpha,x_\beta)\varphi_{\alpha\beta}(x_\alpha,x_\beta)dx_\alpha dx_\beta$$

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

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

in which

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

$$\times \sigma^2(z_1)\sigma^2(z_{1\alpha},z_{1\beta},z_{2\underline{\alpha\beta}})dz_1 dz_{2\underline{\alpha\beta}},$$

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

It should be noted that under the null hypothesis of no pairwise interactions, the terms involving $f^{*}_{\alpha\beta}$ vanish identically. Thus it is not necessary to estimate the bias term B_1.

To derive the local power, denote by $S_{\alpha\beta}$ the support of the density $\varphi_{\alpha\beta}$ and let $\mathcal{B}_{\alpha\beta}(M)$ be the function class consisting of functions $f_{\alpha\beta}$ satisfying

$$\|f_{\alpha\beta}\|_{H^2(S_{\alpha\beta})} \le M,$$

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

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

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 as follows.

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

$$T_n \geq C(\eta; h, K, \varphi, \sigma), \tag{33}$$

where the test functional is

$$T_n = nh \int \widehat{f_{\alpha\beta}^*}^2(x_\alpha, x_\beta) \varphi_{\alpha\beta}(x_\alpha, x_\beta) dx_\alpha dx_\beta \tag{34}$$

and the critical value

$$C(\eta; h, K, \varphi, \sigma) = \Phi^{-1}(1-\eta) V^{1/2}(K, \varphi, \sigma) + \frac{2\{K^{(2)}(0)\}^2}{h} \int \frac{\varphi_{\alpha\beta}(z_\alpha, z_\beta) \varphi_{\underline{\alpha\beta}}^2(z_{\underline{\alpha\beta}})}{\varphi(z)} \sigma^2(z) dz, \tag{35}$$

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

THEOREM 7. *Under assumptions (A1)–(A5), let for* $1 \leq i \leq n$

$$Y_{i,n} = c + \sum_{\gamma=1}^{d} f_\gamma(X_{i\gamma,n}) + f_{\alpha\beta,n}(X_{i\alpha,n}, X_{i\beta,n}) + \sum_{\substack{1 \leq \gamma < \delta \leq d \\ (\gamma,\delta) \neq (\alpha,\beta)}} f_{\gamma\delta}(X_{i\gamma,n}, X_{i\delta,n}) + \sigma(X_{i,n}) \varepsilon_{i,n} \tag{36}$$

be the data array generated from the i.i.d. array $(X_{i,n}, \varepsilon_{i,n}), 1 \leq i \leq n$, *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* $(f_{\alpha\beta,n})_{n=1}^\infty$ *a sequence of functions such that* $f_{\alpha\beta,n} \in \mathcal{F}_{\alpha\beta}(a_n)$ *where* $\{a_n\}$ *is a sequence satisfying* $a_n^{-1} = o(nh + h^{-2}) = o(n^{5/6})$ *as* $n \to \infty$. *Denote by* p_n *the probability of rejecting* $H_{0,\alpha\beta}: f_{\alpha\beta,n}(x_\alpha, x_\beta) \equiv 0$ *in favor of the local alternative* $H_{1,\alpha\beta}(a_n): f_{\alpha\beta,n} \in \mathcal{F}_{\alpha\beta}(a_n)$ *based on the data* $(X_{i,n}, Y_{i,n}), 1 \leq i \leq n$ *as defined in (36). Then* $\lim_{n\to\infty} p_n = 1$.

Theorem 7 guarantees that asymptotically, the proposed test procedure (33) is able to detect an interaction term of the magnitude $n^{-5/6}$ with probability 1.

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

To implement the test procedure (33), the critical value $C(\eta; h, K, \varphi, \sigma)$ can be obtained as the wild bootstrap quantiles of the test statistic $T_n = nh\int \widehat{f^*_{\alpha\beta}}^2(x_\alpha, x_\beta)\varphi_{\alpha\beta}(x_\alpha, x_\beta)dx_\alpha dx_\beta$. Because the density $\varphi_{\alpha\beta}$ is unknown, T_n is approximated, using a law of large numbers argument, by

$$\widetilde{T}_n = nh\sum_{l=1}^{n} \widehat{f^*_{\alpha\beta}}^2(X_{l\alpha}, X_{l\beta})/n = h\sum_{l=1}^{n} \widehat{f^*_{\alpha\beta}}^2(X_{l\alpha}, X_{l\beta}). \quad \textbf{(37)}$$

The following theorem ensures that this substitution is asymptotically feasible.

THEOREM 8. *Under Assumptions (A1)–(A5),*

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

Hence, Theorem 6 and test rule (33) are not affected by replacing T_n with $\widetilde{T}_n$. Further, this holds for Theorem 7 also, 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 functional is $\int \hat{F}^{(1,1)2}_{\alpha\beta}\ \varphi_{\alpha\beta}(x_\alpha, x_\beta)\ dx_\alpha \times dx_\beta = \int \hat{f}^{(1,1)2}_{\alpha\beta}\varphi_{\alpha\beta}(x_\alpha, x_\beta)\ dx_\alpha dx_\beta$.

As can be seen from the proofs of Theorems 1, 2, and 4–6 in the Appendix, 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 $K_1^\star$ 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 looked at. In fact, as will be seen in the simulations in Section 7, large samples are needed to approximate the asymptotic properties, where *large* can mean thousands of observations. So, even if the test procedure proposed in Section 6.1 should at some point beat the one we consider here, this is not clear for a small or moderate sample, which is 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 (A4)–(A7),*

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

where B_2 is defined in the formulation of Theorem 4.

Again, we note that under the hypothesis of no interactions the terms containing $f_{\alpha\beta}^{(1,1)}$ drop out and consequently the bias term B_2 need not be estimated.

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

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

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 $a > 0$,

$$
\begin{aligned}
\mathcal{F}_{\alpha\beta}(a) = \Bigg\{ f_{\alpha\beta} \in \mathcal{B}_{\alpha\beta}(M) : \|f_{\alpha\beta}^{(1,1)}\|_{L^2(S_{\alpha\beta}, \varphi_{\alpha\beta})} \\
= \sqrt{\int_{S_{\alpha\beta}} f_{\alpha\beta}^{(1,1)2}(x_\alpha, x_\beta)\varphi_{\alpha\beta}(x_\alpha, x_\beta)dx_\alpha dx_\beta} \ge a \Bigg\}.
\end{aligned}
$$

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

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

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

$$nh^5 \int \hat{F}_{\alpha\beta}^{(1,1)2}(x_\alpha, x_\beta)\varphi_{\alpha\beta}(x_\alpha, x_\beta)dx_\alpha dx_\beta$$

$$- \frac{2\{K_1^{\star(2)}(0)\}^2}{h} \int \frac{\varphi_{\alpha\beta}(z_\alpha, z_\beta)\varphi_{\underline{\alpha\beta}}^2(z_{\underline{\alpha\beta}})}{\varphi(z)} \sigma^2(z)dz$$

$$\geq \frac{\Phi}{(1-\eta)} \sqrt{\begin{array}{l} 2\|K_1^{\star(2)}\|_2^4 \int \dfrac{\varphi_{\alpha\beta}^2(z_{1\alpha}, z_{1\beta})\varphi_{\underline{\alpha\beta}}^2(z_{1\underline{\alpha\beta}})\varphi_{\underline{\alpha\beta}}^2(z_{2\underline{\alpha\beta}})}{\varphi(z_1)\varphi(z_{1\alpha}, z_{1\beta}, z_{2\underline{\alpha\beta}})} \\ \quad \times \sigma^2(z_1)\sigma^2(z_{1\alpha}, z_{1\beta}, z_{2\underline{\alpha\beta}})dz_1 dz_{2\underline{\alpha\beta}} \end{array}}. \quad \mathbf{(38)}$$

The following result concerns the local power of the preceding test.

THEOREM 10. *Under Assumptions (A1) and (A4)–(A7), let* $Y_{i,n}, 1 \leq i \leq n$ *be the same data array as in Theorem 7 but with the* $\alpha\beta$*th interaction* $f_{\alpha\beta,n} \in \mathcal{F}_{\alpha\beta}(a_n)$ *where* $\{a_n\}$ *is a sequence satisfying* $a_n^{-1} = o(nh^5 + h^{-2}) = o(n^{1/2})$ *as* $n \to \infty$. *Denote by* p_n *the probability of rejecting* $H_{0,\alpha\beta}: f_{\alpha\beta,n}(x_\alpha, x_\beta) \equiv 0$ *in favor of the local alternative* $H_{1,\alpha\beta}(a_n): f_{\alpha\beta,n} \in \mathcal{F}_{\alpha\beta}(a_n)$ *based on the data* $(X_{i,n}, Y_{i,n}), 1 \leq i \leq n$ *as defined in (36). Then* $\lim_{n\to\infty} p_n = 1$.

Thus Theorem 10 guarantees that asymptotically with probability 1, the proposed test procedure (38) is able to detect an interaction term whose mixed derivative is of the magnitude $n^{-1/2}$. The proof of Theorem 10 is similar to that of Theorem 7 and is therefore omitted. Also, Theorem 8 can be extended to the test rule (38), but we have omitted its statement because of 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 multicolinearity among the explanatory variables, as in the classical linear regression context, it may happen 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

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

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,

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

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, Secs. 3.5, 3.9, 5.4.5, 6.8.3) looking at the trace of the smoother matrices. However, theory is lacking, and Sperlich, Linton, and 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 much simpler testing problems than considered here. Further it was briefly discussed in Härdle, Mammen, and Müller (1998), also in a different context of testing.

7. AN EMPIRICAL INVESTIGATION OF THE TEST PROCEDURES

In nonparametric statistics one has to be cautious when using the asymptotic distribution for small and moderate sample sizes. We have the additional problem of having complicated expressions for the bias and variance of the test statistics, which means that asymptotic critical values are hard to obtain. Moreover, we are dealing with a type of nonparametric test functional that has been known (Hjellvik et al., 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 et al. (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 variance 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 $G_{(a,b),i}$ that has mean zero, variance equal to the square of the residual, and third moment equal to the cube of the residual for all $i = 1,2,\ldots,n$. 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), $1 \le \alpha < \beta \le d$, and construct the residuals $\tilde{u}_i = \tilde{u}_{i,\alpha\beta} = Y_i - \widehat{m_0}(X_i)$, for $i = 1,2,\ldots,n$.

(2) For each X_i, draw a bootstrap residual u_i^* from the distribution $G_{(a,b),i}$ such that for $U \sim G_{(a,b),i}$,

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

(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, e.g., 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 (2000). Dalelane showed via strong approximation that it holds in supremum norm, whereas Achmus proved that the wild bootstrap works 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 (31), (32) 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 the error terms are normally distributed 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 before. 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 preestimation to do the wild bootstrap we used $h = 1.0$ and $g = 1.2$. To calculate the test statistic based on the joint derivative $f_{12}^{(1,1)}$ *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 $H_{0,12}: f_{12}(u) \equiv 0$ 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.

This conclusion is consistent with the results of Hjellvik et al. (1998) for a similar type of functional designed for testing of linearity. For that functional roughly 100,000 observations were needed to obtain a good approximation. The reason seems to be that for a functional of type

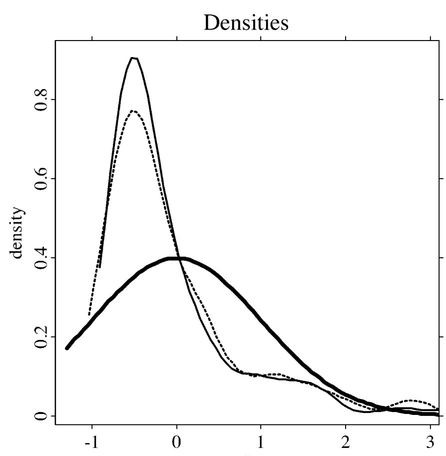


FIGURE 3. Densities of the test statistics; direct method (solid), testing derivatives (dashed), and normal density (thick, solid)

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

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

$\int \widehat{f^*_{\alpha\beta}}^2(x_\alpha, x_\beta)\pi(x_\alpha, x_\beta)dx_\alpha dx_\beta$ 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 et al. (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, 8, and 9, considered the test statistics

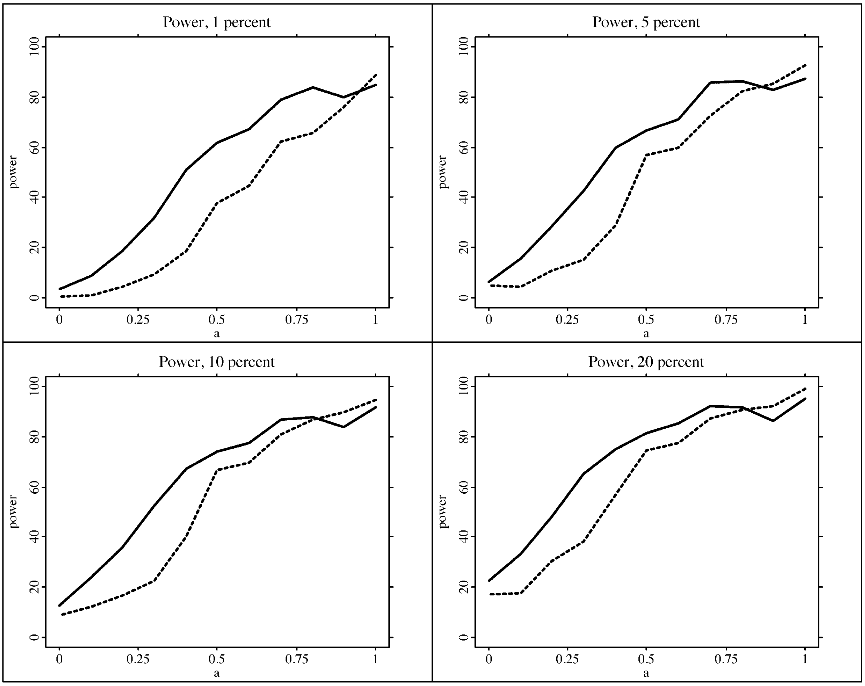


FIGURE 4. Power functions at the 1, 5, 10, and 20% significance levels for both procedures: direct method (solid) (see equation (39)) and testing derivatives (dashed) (see equation (40)). We used 249 bootstrap replications and performed 300 iterations at each point (0.0, 0.1, 0.2, ...) to determine the probability to reject.

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{f_{12}^{*}}^{2}(X_1,X_2)\ \mathbb{1}\{|X_k|\le 1.6 \quad \text{for } k=1,2\} \tag{39}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{F_{12}^{(1,1)}}^{2}(X_1,X_2)\ \mathbb{1}\{|X_k|\le 1.6 \quad \text{for } k=1,2\}, \tag{40}$$

respectively; i.e. we have integrated numerically over the empirical distribution function and used 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 et al., 1998).

Table 1 presents the error of the first kind for both methods and at different significance levels; the rejection probability was determined by performing 500 iterations.

Obtaining an accurate error of the first kind with the aid of the wild bootstrap depends on a proper choice of bandwidth, although the results are fairly robust for a reasonably 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 (32) is displayed for both methods and for different levels in Figure 4. Both procedures appear to work well. For this particular model the power function of the *direct method* is steeper (also when the tests are adjusted to have the same level). 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. For instance, the derivative test is more designed for detecting a less smooth interaction term having high-frequency components.

Obviously a much more detailed simulation study would be of interest, in particular concerning the interplay between model complexity and (optimal) choice of bandwidth. However, it is beyond the scope of the present paper. 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. There has been much discussion in the past whether production functions can be taken to be additive (strongly separable)[1] for a particular data set. It goes back at least to Denny and Fuss (1977), Fuss et al. (1978), and 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, 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 1,000 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 that 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 1 to 10 years) X_5.

The underlying purely additive model is of the form

$$\ln(y) = c + \sum_{\alpha=1}^{d} f_\alpha\{\ln(x_\alpha)\}. \tag{41}$$

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

$$\ln(y) = c + \sum_{\alpha=1}^{d} f_\alpha\{\ln(x_\alpha)\} + \sum_{1\le\alpha<\beta\le d} f_{\alpha\beta}\{\ln(x_\alpha), \ln(x_\beta)\}, \tag{42}$$

and the assumed strong separability (additivity) can be checked by testing the null hypothesis $H_{0,\alpha\beta}: f_{\alpha\beta} = 0$ for all α, β using the approach of Section 3.

First we estimated all functions of type f_α and $f_{\alpha\beta}$. The estimation results are given in Figures 5–7. Because we want to plot the functional forms jointly with the real data point cloud, we depict $\hat{F}_\alpha$ for the one-dimensional impact functions; for the interactions we plotted $\widehat{f^*_{\alpha\beta}}$, the functions we will use for the test procedure. 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 reasonably smooth estimates. However, we know by experience that the integration estimator is quite robust against a relatively wide range of

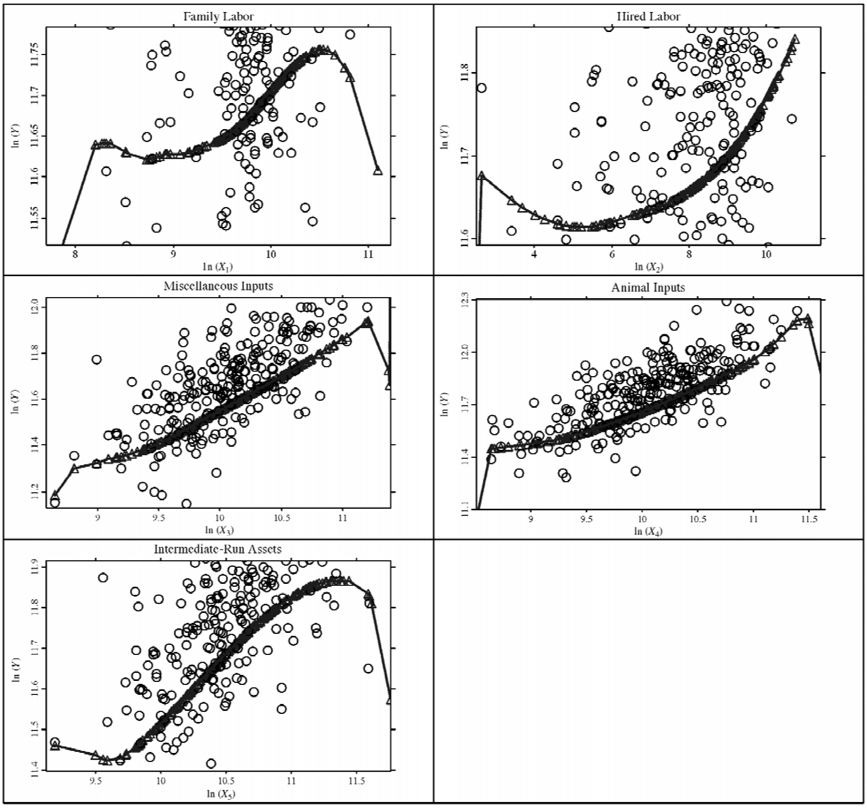


FIGURE 5. Function estimates for the univariate additive components and partial residuals

choices of bandwidths. For a detailed discussion of the bandwidth choice and robustness we refer to Sperlich et al. (1997).

In Figure 5 the univariate function estimates (not centered to zero as in (6)) are displayed together with a kind of partial residuals $\hat{r}_{i\alpha} := y_i - \sum_{\gamma \neq \alpha} \hat{F}_\gamma(X_{i\gamma})$. 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 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 repli-

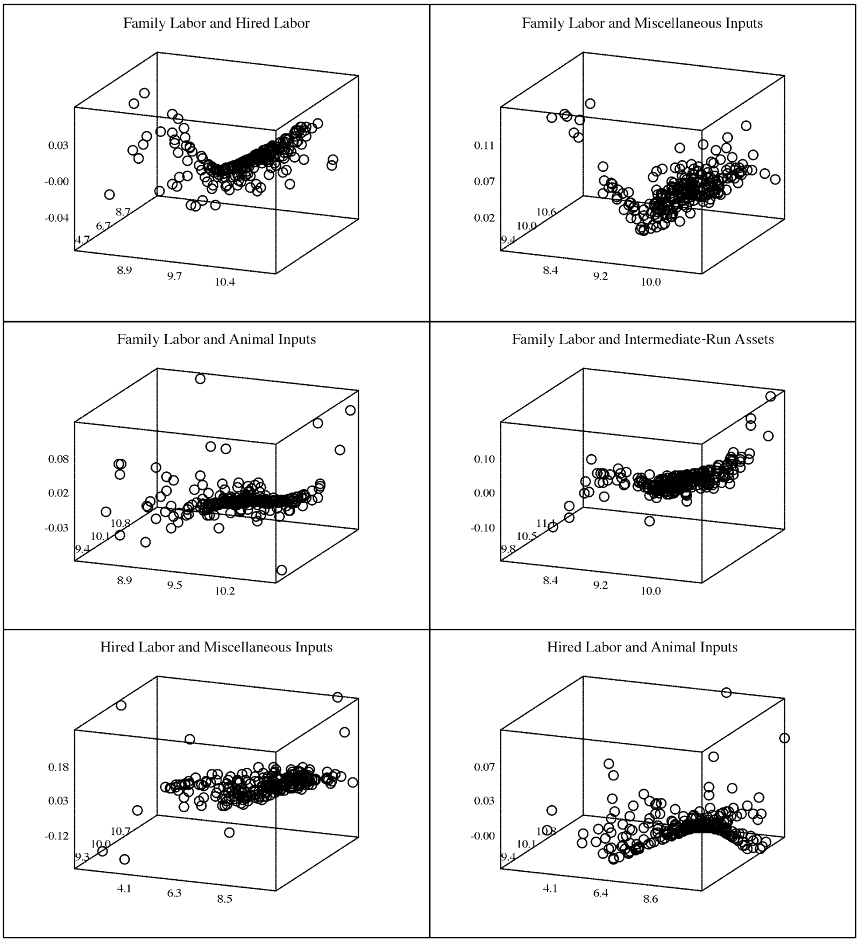


FIGURE 6. Estimates of the first six interaction terms

cations. 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 interaction term directly, we used $h = 1.3$–2.1, $g = 2.9$–3.7 for the preestimation to get estimates for the bootstrap and $h = 1.6$–2.4, $g = 3.1$–3.9 to calculate the test statistics. For the second method, which involves the mixed derivatives of the interaction term, we used $h = 1.6$–2.4, $g = 3.1$–3.9 for the preestimation to get estimates for the bootstrap and $h = 2.1$–2.9, $g = 3.1$–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

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

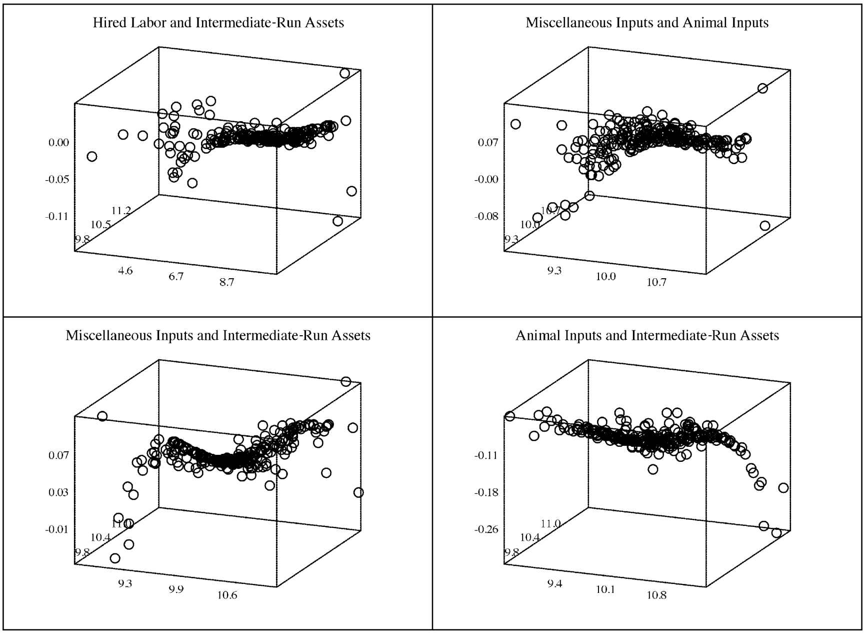


FIGURE 7. Estimates of the last four interaction terms

term $f_{\alpha\beta}$ including all the other functions f_γ, $1 \le \gamma \le d$, and $f_{\gamma\delta}$, $1 \le \gamma < \delta \le d$, with $(\gamma, d) \neq (\alpha, \beta)$ in the model (42). 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 previously. 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_{13} (family labor and miscellaneous inputs) was significant at a 5% level with a p-value of about 2%. Of the other terms f_{35} and f_{15} came closest to being significant.

For the second method, considering the derivatives, f_{15} (family labor and intermediate-run assets) and f_{35} (miscellaneous inputs and intermediate-run assets) had the lowest p-values, f_{15} having a p-value of less than 1%. Among the others, f_{13} came closest.

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_{13} and f_{15} are not multiplicative in their input factors. This would make it difficult for many parametric tests to detect the interaction as they usually are based on multiplicative combinations of the regressors.

NOTE

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

REFERENCES

Achmus, S. (2000) *Nichtparametrische additive Modelle.* Thesis at Technische Universität, Braunschweig, Germany.

Andrews, D.W.K. & Y.-J. Whang (1990) Additive interactive regression models: Circumvention of the curse of dimensionality. *Econometric Theory* 6, 466–479.

Auestad, B.H. & D. Tjøstheim (1991) Functional identification in nonlinear time series. In G.G. Roussas (ed.), *Nonparametric Functional Estimation and Related Topics*, 483–507. Amsterdam: Kluwer.

Berndt, E.R. & L.R. Christensen (1973) The internal structure of functional relationships: Separability, substitution, and aggregation. *Review of Economic Studies* 40, 403–410.

Chen, R., J.S. Liu, & R.S. Tsay (1995) Additivity tests for nonlinear autoregression. *Biometrika* 82, 369–383.

Dalelane, C. (1999) *Bootstrap Confidence Bands for the Integration Estimator in Additive Models.* Diploma thesis, Humboldt-Universität zu Berlin, Faculty of Mathematics, Germany.

Deaton, A. & J. Muellbauer (1980) *Economics and Consumer Behavior.* New York: Cambridge University Press.

Denny, M. & M. Fuss (1977) The use of approximation analysis to test for separability and the existence of consistent aggregates. *American Economic Review* 67, 404–418.

Derbort, S., H. Dette, & A. Munk (2002) A test for additivity in nonparametric regression. *Annals of the Institute of Mathematical Statistics*, in press.

Eubank, R.L., J.D. Hart, D.G. Simpson, & L.A. Stefanski (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., T. Gasser, I. Gijbels, M. Brockmann, & J. Engel (1993) Local Polynomial Fitting: A Standard for Nonparametric Regression. CORE Discussion Paper 9315.

Fan, J., W. Härdle, & E. Mammen (1998) Direct estimation of low dimensional components in additive models. *Annals of Statistics* 26, 943–971.

Fuss, M., D. McFadden, & Y. Mundlak (1978) A survey of functional forms in the economic analysis of production. In Fuss, M. & D. McFadden (eds.), *Production Economics: A Dual Approach to Theory and Applications*, pp. 219–268.

Gozalo, P.L. & O.B. Linton (1997) Testing Additivity in Generalized Nonparametric Regression Models. Preprint, Yale University.

Hall, P. (1984) Central limit theorem for integrated square error of multivariate nonparametric density estimators, *Journal of Multivariate Analysis* 14, 1–16.

Härdle, W., S. Huet, E. Mammen, & S. Sperlich (1998) Semiparametric Additive Indices for Binary Response and Generalized Additive Models. Discussion Paper 95, Sonderforschungsbereich 373.

Härdle, W. & E. Mammen (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926–1947.

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

Härdle, W., E. Mammen, & M. Müller (1998) Testing parametric versus semiparametric modelling in generalized linear models. *Journal of the American Statistical Association* 93, 1461–1474.

Hastie, T.J. & R.J. Tibshirani (1990) *Generalized Additive Models.* London: Chapman and Hall.

Hjellvik, V., Q. Yao, & D. Tjøstheim (1998) Linearity testing using local polynomial approximation. *Journal of Statistical Planning and Inference* 68, 295–321.

Johnson, R.A. & D.W. Wichern (1988) *Applied Multivariate Statistical Analysis*, 2nd ed. Englewood Cliffs, New Jersey: Prentice-Hall.

Kreiss, J.-P., M.H. Neumann, & Q. Yao (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. (2000) Efficient estimation of generalized additive nonparametric regression models. *Econometric Theory* 16, 502–523.

Linton, O.B. & W. Härdle (1996) Estimation of additive regression models with known links. *Biometrika* 83, 529–540.

Linton, O.B. & J.P. Nielsen (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., O.B. Linton, & J.P. Nielsen (1999) The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Annals of Statistics* 27, 1443–1490.

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, pp. 221–230.

Newey, W.K. (1994) Kernel estimation of partial means. *Econometric Theory* 10, 233–253.

Newey, W.K. (1995) Convergence rates for series estimators. In G.S. Maddala, P.C.B. Phillips, & T.N. Srinavsan (eds.), *Statistical Methods of Economics and Quantitative Economics: Essays in Honor of C.R. Rao*, pp. 254–275. Cambridge: Blackwell.

Opsomer, J.D. & D. Ruppert (1997) Fitting a bivariate additive model by local polynomial regression. *Annals of Statistics* 25, 186–211.

Ruppert, D. & M.P. Wand (1994) Multivariate locally weighted least squares regression. *Annals of Statistics* 22, 1346–1370.

Severance-Lossin, E. & S. Sperlich (1999) Estimation of derivatives for additive separable models. *Statistics* 33, 241–265.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Sperlich, S., O.B. Linton, & W. Härdle (1997) A Simulation Comparison between Integration and Backfitting Methods of Estimating Separable Nonparametric Regression Models. Discussion Paper 66, Sonderforschungsbereich 373.

Sperlich, S., O.B. Linton, & W. Härdle (1999) Integration and backfitting methods in additive models: Finite sample properties and comparison. *Test* 8, 419–458.

Stone, C.J. (1985) Additive regression and other nonparametric models. *Annals of Statistics* 13, 689–705.

Stone, C.J., M.H. Hansen, C. Kooperberg, & Y. Truong (1997) Polynomial splines and their tensor products in extended linear modeling (with discussion). *Annals of Statistics* 25, 1371–1470.

Tjøstheim, D. & B.H. Auestad (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*, 2nd printing. Philadelphia: Society for Industrial and Applied Mathematics.

Wu, C.F.J. (1986) Jackknife, bootstrap, and other resampling methods in regression analysis (with discussion). *Annals of Statistics* 14, 1261–1350.

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

APPENDIX

Proof of Lemma 1.

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

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

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

$$F_{\alpha\beta}(x_\alpha,x_\beta) - F_\alpha(x_\alpha) - F_\beta(x_\beta) + \int m(x)\varphi(x)dx$$

$$= f_{\alpha\beta}(x_\alpha,x_\beta) + \sum_{1\le\gamma<\delta\le d} c_{\delta\gamma} + \sum_{(\gamma,\delta)\in D_{\alpha\beta}} c_{\delta\gamma} - \sum_{(\gamma,\delta)\in D_{\alpha\alpha}} c_{\delta\gamma} - \sum_{(\gamma,\delta)\in D_{\beta\beta}} c_{\delta\gamma}$$

$$= f_{\alpha\beta}(x_\alpha,x_\beta) + c_{\alpha\beta}.$$

3. We only need to integrate both sides of the equation in item 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 item 2. ■

Proof of Corollary 1. First assume that $f^*_{\alpha\beta}(x_\alpha,x_\beta) \equiv 0$. By Lemma 1, $F_{\alpha\beta}(x_\alpha,x_\beta) - F_\alpha(x_\alpha) - F_\beta(x_\beta) + \int m(x)\varphi(x)dx \equiv 0$ 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 $f_{\alpha\beta}(x_\alpha,x_\beta) + c_{\alpha\beta} \equiv 0$. ■

Proof of Theorems 1 and 2. The proof of Theorems 1 and 2 makes use of the following two lemmas, whose proofs are not difficult. We refer to Silverman (1986) or Fan et al. (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 $B_n = (b_{ij})_{1\le i,j\le d}$ where $b_{ij} = O_p(\delta_n)$ with d fixed, independent of n. Then $D_n^{-1} = A^{-1}(I + C_n)$ where $C_n = (c_{ij})_{1\le i,j\le d}$ 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 $H = \mathrm{diag}(h^{i-1})_{i=1,\dots,p+1}$. Then*

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

and

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

Define $E_i[\cdot] = E[\cdot|X_{i1},\ldots,X_{id}]$ and $E_*[\cdot] = E[\cdot|\mathbf{X}]$, where $\mathbf{X}$ is the design matrix $\{X_{i\alpha}\}_{i,\alpha=1,1}^{n,d}$. The proofs of Theorems 1 and 2 can now be divided into two parts corresponding to the estimators $\hat{F}_\alpha$ and $\hat{f}_{\alpha\beta}^*$, respectively.

(i) We start by considering the univariate estimator $\hat{F}_\alpha$. This is also a component of the estimator $\widehat{f_{\alpha\beta}^*}$ 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

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

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 $(x_\alpha, X_{l\underline{\alpha}})$, we obtain

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

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

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

where

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

and the expression in the brackets $[\dots]$ is as in formula (A.1). It remains to work with the first-order approximations.

Let

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

For the bias part we prove that

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

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

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

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

$$= \frac{1}{\varphi(x_\alpha, X_{i\underline{\alpha}})} \int K(u)L(v)\varphi(x_\alpha + uh, X_{i\underline{\alpha}} + vg)$$

$$\times \Bigg[\frac{(uh)^2}{2} \Bigg\{ f_\alpha^{(2)}(x_\alpha) + \sum_{\gamma \in D_\alpha} f_{\alpha\gamma}^{(2,0)}(x_\alpha, X_{l\gamma}) \Bigg\} + O_p\{(uh)^3\}$$

$$+ \sum_{\gamma \in D_\alpha} \{ f_\gamma(X_{i\gamma} + gv_\gamma) - f_\gamma(X_{i\gamma}) \}$$

$$+ \sum_{(\gamma,\delta) \in D_{\alpha\alpha}} \{ f_{\gamma\delta}(X_{i\gamma} + gv_\gamma, X_{i\delta} + v_\delta g) - f_{\gamma\delta}(X_{i\gamma}, X_{i\delta}) \} \Bigg] dvdu + o_p(1)$$

$$= h^2 \mu_2(K) \frac{1}{2} \Bigg\{ f_\alpha^{(2)}(x_\alpha) + \sum_{\gamma \in D_\alpha} f_{\alpha\gamma}^{(2,0)}(x_\alpha, X_{l\gamma}) \Bigg\} + o_p(h^2) + O_p(g^q)$$

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

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

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

and note that

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

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

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

$$w_{i\alpha} = \frac{1}{n} K_h(x_\alpha - X_{i\alpha}) \frac{\varphi_{\underline{\alpha}}(X_{i\underline{\alpha}})}{\varphi(x_\alpha, X_{i\underline{\alpha}})}, \tag{A.2}$$

we have

$$T_{2n} = \sum_{i=1}^{n} w_{i\alpha} \sigma(X_i) \varepsilon_i + o_p\{(nh)^{-1/2}\} \tag{A.3}$$

and hence

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

It is easily seen that the preceding expressions imply Theorem 1 (cf. Severance-Lossin and Sperlich, 1999).

(ii) Analogous to the univariate case of $\hat{F}_\alpha$, we proceed for the bivariate case considering $\hat{F}_{\alpha\beta}$. We need the following definition.

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

where $\{\cdots\}$ 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

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

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

$$\begin{aligned} &\hat{F}_{\alpha\beta}(x_\alpha, x_\beta) - F_{\alpha\beta}(x_\alpha, x_\beta) \\ &\quad = \frac{1}{n} \sum_{i=1}^n \frac{1}{\varphi(x_\alpha, x_\beta, X_{i\underline{\alpha\beta}})} \frac{1}{n} \sum_{l=1}^n K_h(X_{l\alpha} - x_\alpha) K_h(X_{l\beta} - x_\beta) L_g(X_{l\underline{\alpha\beta}} - X_{i\underline{\alpha\beta}}) \\ &\qquad \times \left\{ I + O_p\left(h^2 + \sqrt{\frac{\ln n}{nh^2 g^{d-2}}} \right) \right\} \\ &\qquad \times \Bigg[\frac{(X_{l\alpha} - x_\alpha)^2}{2} \left\{ f_\alpha^{(2)}(x_\alpha) + \sum_{\gamma \in D_{\alpha,\beta}} f_{\alpha\gamma}^{(2,0)}(x_\alpha, X_{l\gamma}) + f_{\alpha\beta}^{(2,0)}(x_\alpha, x_\beta) \right\} \\ &\qquad\quad + \frac{(X_{l\beta} - x_\beta)^2}{2} \left\{ f_\beta^{(2)}(x_\beta) + \sum_{\gamma \in D_{\alpha,\beta}} f_{\beta\gamma}^{(2,0)}(x_\beta, X_{l\gamma}) + f_{\alpha\beta}^{(0,2)}(x_\alpha, x_\beta) \right\} \\ &\qquad\quad + \sum_{\gamma \in D_{\alpha,\beta}} \{ f_\gamma(X_{l\gamma}) - f_\gamma(X_{i\gamma}) \} + \sum_{(\gamma,\delta) \in D_{\alpha\beta}} \{ f_{\gamma\delta}(X_{l\gamma}, X_{l\delta}) - f_{\gamma\delta}(X_{i\gamma}, X_{i\delta}) \} \\ &\qquad\quad + (X_{l\alpha} - x_\alpha)(X_{l\beta} - x_\beta) f_{\alpha\beta}^{(1,1)}(x_\alpha, x_\beta) + O_p\{(X_{l\alpha} - x_\alpha)^3\} \\ &\qquad\quad + o_p\{(X_{l\alpha} - x_\alpha)(X_{l\beta} - x_\beta)\} + O_p\{(X_{l\beta} - x_\beta)^3\} + \sigma(X_l)\varepsilon_l \Bigg] + O_p(n^{-1/2}). \quad \textbf{(A.4)} \end{aligned}$$

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

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

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

where

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

and $[\ldots]$ is the expression in the large brackets of equation (A.4).

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

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

We first prove that

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

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

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

$$= \frac{1}{\varphi(x_\alpha, x_\beta, X_{i\underline{\alpha\beta}})} \int K(u_\alpha)K(u_\beta)L(w)\varphi(x_\alpha + u_\alpha h, x_\beta + u_\beta h, X_{i\underline{\alpha\beta}} + vg)$$

$$\times \Bigg[\frac{(hu_\alpha)^2}{2} \Bigg\{ f_\alpha^{(2)}(x_\alpha) + \sum_{\gamma \in D_{\alpha,\beta}} f_{\alpha\gamma}^{(2,0)}(x_\alpha, X_{l\gamma}) + f_{\alpha\beta}^{(2,0)}(x_\alpha, x_\beta) \Bigg\}$$

$$+ \frac{(hu_\beta)^2}{2} \Bigg\{ f_\beta^{(2)}(x_\beta) + \sum_{\gamma \in D_{\alpha,\beta}} f_{\beta\gamma}^{(2,0)}(x_\beta, X_{l\gamma}) + f_{\alpha\beta}^{(0,2)}(x_\alpha, x_\beta) \Bigg\}$$

$$+ \sum_{\gamma \in D_{\alpha,\beta}} \{ f_\gamma(X_{i\gamma} + gv_\gamma) - f_\gamma(X_{i\gamma}) \}$$

$$+ \sum_{(\gamma,\delta) \in D_{\alpha\beta}} \{ f_{\gamma\delta}(X_{i\gamma} + gv_\gamma, X_{i\delta} + gv_\delta) - f_{\gamma\delta}(X_{i\gamma}, X_{i\delta}) \}$$

$$+ (hu_\alpha)(hu_\beta) f_{\alpha\beta}^{(1,1)}(x_\alpha, x_\beta) + O_p(h^3) \Bigg] dvdu + o_p(1)$$

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

$$+ \sum_{\gamma \in D_{\alpha,\beta}} f_{\beta\gamma}^{(2,0)}(x_\beta, X_{i\gamma}) + f_{\alpha\beta}^{(2,0)}(x_\alpha, x_\beta) + f_{\alpha\beta}^{(0,2)}(x_\alpha, x_\beta) \Bigg\}$$

$$+ o_p(h^2) + O_p(g^q)$$

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

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

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

$$+ \sum_{\gamma \in D_{\alpha,\beta}} \frac{1}{n} \sum_{i=1}^n f_{\beta\gamma}^{(2,0)}(x_\beta, X_{i\gamma}) + f_{\alpha\beta}^{(2,0)}(x_\alpha, x_\beta) + f_{\alpha\beta}^{(0,2)}(x_\alpha, x_\beta) \Bigg\}$$

$$+ o_p(h^2).$$

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:

$$h^2 B_1 = h^2 \mu_2(K) \frac{1}{2} \Bigg\{ f_{\alpha\beta}^{(2,0)}(x_\alpha, x_\beta) - \int f_{\alpha\beta}^{(2,0)}(x_\alpha, u_\beta) \varphi_{\underline{\alpha}}(u)\, du$$

$$+ f_{\alpha\beta}^{(0,2)}(x_\alpha, x_\beta) - \int f_{\alpha\beta}^{(0,2)}(u_\alpha, x_\beta) \varphi_{\underline{\beta}}(u)\, du \Bigg\} + o_p(h^2).$$

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

$$w_{i\alpha\beta} = \frac{1}{n} K_h(x_\alpha - X_{i\alpha}, x_\beta - X_{i\beta}) \frac{\varphi_{\underline{\alpha\beta}}(X_{i\underline{\alpha\beta}})}{\varphi(x_\alpha, x_\beta, X_{i\underline{\alpha\beta}})}, \tag{A.5}$$

$$\mathcal{T}_{2n} = \sum_{i=1}^{n} w_{i\alpha\beta}\sigma(X_i)\varepsilon_i + o_p\{(nh^2)^{-1/2}\}, \tag{A.6}$$

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 (10) is asymptotically normal.

Finally, we want to calculate the variance of the combined estimator $\hat{F}_{\alpha\beta}(x_\alpha, x_\beta) - \hat{F}_\alpha(x_\alpha) - \hat{F}_\beta(x_\beta)$. Because of the faster rate of the stochastic term in (i), 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

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

Theorem 2 follows. ■

Proof of Theorem 4. This proof is analogous to that of Theorems 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 $K^*_\nu(\cdot)$; i.e., $K^*_3(u,v) = K(u,v)uv\mu_2^{-2}$ and the weights are

$$w_{i\alpha\beta} = \frac{1}{nh^3} K^*_{3,h}(x_\alpha - X_{i\alpha}, x_\beta - X_{i\alpha}) \frac{\varphi_{\underline{\alpha\beta}}(X_{i\underline{\alpha\beta}})}{\varphi(x_\alpha, x_\beta, X_{i\underline{\alpha\beta}})}, \tag{A.7}$$

where $K^*_{3,h}(\cdot,\cdot) = (1/h^2)K^*_3(\cdot/h, \cdot/h)$. ■

Proof of Theorem 5. Because the proof follows the arguments of Linton (1997, 2000) we give only a sketch here. Further, we only discuss the statement in (29), because for (30) the reasoning is the same. We have

$$\check{F}^{opt}_\alpha(x_\alpha) - \check{F}_\alpha(x_\alpha)$$
$$= \sum_{i=1}^{n} \frac{w_i}{\sum_{j=1}^{n} w_j} \left[\sum_{\gamma \neq \alpha} \{F_\gamma(X_{i\gamma}) - \hat{F}_\gamma(X_{i\gamma})\} + \sum_{\gamma < \beta} \{f^*_{\gamma\beta}(X_{i\gamma}, X_{i\beta}) - \widehat{f^*_{\gamma\beta}}(X_{i\gamma}, X_{i\beta})\} \right]$$
$$+ O_p(n^{-1/2})$$

with

$$w_i = J\left(\frac{x_\alpha - X_{i\alpha}}{h_e}\right)\{s_{n2} - (x_\alpha - X_{i\alpha})s_{n1}\}$$

and

$$s_{nl} = \sum_{k=1}^{n} J\left(\frac{x_\alpha - X_{k\alpha}}{h_e}\right)(x_\alpha - X_{k\alpha})^l, \qquad l = 1,2.$$

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

From Theorem 1 and its proof in this Appendix we know that the leading terms of $F_\gamma(X_{i\gamma}) - \hat{F}_\gamma(X_{i\gamma})$ are

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

with

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

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 $w_i/\sum_{j=1}^n w_j$ and summing we still get a term of order h^2, which by assumption is $o(n^{-2/5})$.

Also, if we consider the stochastic part of $F_\gamma(X_{i\gamma}) - \hat{F}_\gamma(X_{i\gamma})$, we can see that for $l = 0,1$

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

$$= \frac{1}{n}\sum_{j=1}^{n} \Delta_{nlj\gamma}\,\varphi_{\underline{\gamma}}(X_{j\underline{\gamma}})\,\sigma(X_j)\varepsilon_j \tag{A.8}$$

where we have

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

But $\Delta_{nlj\gamma}$ is bounded, and the whole expression (A.8) is of order $O_p(n^{-1/2})$.

For the interaction terms the reasoning is virtually the same. The leading terms of $f_{\gamma\delta}^*(X_{i\gamma}, X_{i\delta}) - \widehat{f_{\gamma\delta}^*}(X_{i\gamma}, X_{i\delta})$ are

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

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 $w_i/\sum_{j=1}^n w_j$ this term stays of order $h^2 = o(n^{-2/5})$.

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

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

where now γ is also allowed to take the value α (compare (23)). Nevertheless the same arguments of boundedness apply as previously.

The rest of the proof is along the lines of Linton (1997). For the interaction term in (30) all of the arguments are the same (compare (24) and (28)), but the rate is slower (by one bandwidth) because of having one dimension more to estimate. ■

Proof of Theorem 6. Consider the decomposition

$$\begin{aligned}
&\int \widehat{f^*_{\alpha\beta}}^{\,2}(x_\alpha, x_\beta)\varphi_{\alpha\beta}(x_\alpha, x_\beta)dx_\alpha dx_\beta \\
&\quad = \sum_{1\le i\ne j\le n} H(X_i, \varepsilon_i, X_j, \varepsilon_j) + \sum_{i=1}^{n} H(X_i, \varepsilon_i, X_i, \varepsilon_i) \\
&\qquad + \int f^{*2}_{\alpha\beta}(x_\alpha, x_\beta)\varphi_{\alpha\beta}(x_\alpha, x_\beta)dx_\alpha dx_\beta \\
&\qquad + 2h^2 \int f^*_{\alpha\beta}(x_\alpha, x_\beta)B_1(x_\alpha, x_\beta)\varphi_{\alpha\beta}(x_\alpha, x_\beta)dx_\alpha dx_\beta + o_p(h^2)
\end{aligned}$$

in which

$$\begin{aligned}
&H(X_i, \varepsilon_i, X_j, \varepsilon_j) \\
&\quad = \varepsilon_i \varepsilon_j \int \frac{1}{n^2}(w_{i\alpha\beta} - w_{i\alpha} - w_{i\beta})(w_{j\alpha\beta} - w_{j\alpha} - w_{j\beta})\sigma(X_i)\sigma(X_j)\varphi_{\alpha\beta}(x_\alpha, x_\beta)dx_\alpha dx_\beta
\end{aligned} \tag{A.9}$$

with $w_{i\alpha}$, $w_{i\beta}$, and $w_{i\alpha\beta}$ as in equations (A.2) and (A.5).

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

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

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

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

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

where

$$H_1 = \frac{\varepsilon_i\varepsilon_j\sigma(X_i)\sigma(X_j)}{n^2h^2} K^{(2)}\left(\frac{X_{i\alpha}-X_{j\alpha}}{h}\right)K^{(2)}\left(\frac{X_{i\beta}-X_{j\beta}}{h}\right)\varphi_{\underline{\alpha\beta}}(X_{i\underline{\alpha\beta}})\varphi_{\underline{\alpha\beta}}(X_{j\underline{\alpha\beta}}) \times\left\{\frac{\varphi_{\alpha\beta}(X_{i\alpha},X_{i\beta})}{\varphi(X_i)\varphi(X_{i\alpha},X_{i\beta},X_{j\underline{\alpha\beta}})}+\frac{\varphi_{\alpha\beta}(X_{j\alpha},X_{j\beta})}{\varphi(X_j)\varphi(X_{j\alpha},X_{j\beta},X_{i\underline{\alpha\beta}})}\right\},$$

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

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

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

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

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 H_2 to H_5 are of higher order and thus negligible. Now we calculate

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

Introducing the change of variable

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

we obtain

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

or

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

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

$$E\{G_1^2(X_1,\varepsilon_1,X_2,\varepsilon_2)\}+n^{-1}E\{H_1^4(X_1,\varepsilon_1,X_2,\varepsilon_2)\}=o[\{EH_1^2(X_1,\varepsilon_1,X_2,\varepsilon_2)\}^2],$$

where

$$G_1(x,\varepsilon,y,\delta)=E\{H_1(X_1,\varepsilon_1,x,\varepsilon)H_1(X_1,\varepsilon_1,y,\delta)\}\tag{A.10}$$

(see Hall, 1984).

LEMMA A3. *As* $h\to 0$ and $nh^2\to\infty$,

$$n^{-1}E\{H_1^4(X_1,\varepsilon_1,X_2,\varepsilon_2)\}=O(n^{-9}h^{-6})=o[\{EH_1^2(X_1,\varepsilon_1,X_2,\varepsilon_2)\}^2].$$

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

$$E\{H_1^4(X_1,\varepsilon_1,X_2,\varepsilon_2)\}$$

$$= \frac{h^2}{n^8h^8}\int K^{(2)\,4}(u)K^{(2)\,4}(v)\varphi^4_{\underline{\alpha\beta}}(z_{1\underline{\alpha\beta}})\varphi^4_{\underline{\alpha\beta}}(z_{2\underline{\alpha\beta}})\sigma^4(z_1)\sigma^4(z_{1\alpha},z_{1\beta},z_{2\underline{\alpha\beta}})$$

$$\times\left\{\frac{\varphi_{\alpha\beta}(z_{1\alpha},z_{1\beta})}{\varphi(z_1)\varphi(z_{1\alpha},z_{1\beta},z_{2\underline{\alpha\beta}})}+\frac{\varphi_{\alpha\beta}(z_{1\alpha},z_{1\beta})}{\varphi(z_{1\alpha},z_{1\beta},z_{2\underline{\alpha\beta}})\varphi(z_1)}\right\}^4$$

$$\times\,\varphi(z_1)\varphi(z_{1\alpha},z_{1\beta},z_{2\underline{\alpha\beta}})dz_1\,dudvdz_{2\underline{\alpha\beta}}\{1+o_p(1)\},$$

which implies that

$$n^{-1}E\{H_1^4(X_1,\varepsilon_1,X_2,\varepsilon_2)\} = O(n^{-9}h^{-6}) = \{EH_1^2(X_1,\varepsilon_1,X_2,\varepsilon_2)\}^2O(n^{-1}h^{-2}),$$

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

LEMMA A4. *As $h\to 0$ and $nh^2\to\infty$,*

$$G_1(x,\varepsilon,y,\delta)$$

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

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

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

Proof. According to the definition of G_1

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

$$= \frac{\varepsilon\delta}{n^4h^4}E\Bigg[K^{(2)}\left(\frac{X_{1\alpha}-x_\alpha}{h}\right)K^{(2)}\left(\frac{X_{1\beta}-x_\beta}{h}\right)$$

$$\times\left\{\frac{\varphi_{\alpha\beta}(x_\alpha,x_\beta)}{\varphi(x)\varphi(x_\alpha,x_\beta,X_{\underline{\alpha\beta}})}+\frac{\varphi_{\alpha\beta}(X_{1\alpha},X_{1\beta})}{\varphi(X_1)\varphi(X_{1\alpha},X_{1\beta},x_{\underline{\alpha\beta}})}\right\}$$

$$\times\,\varphi_{\underline{\alpha\beta}}(X_{1\underline{\alpha\beta}})\varphi_{\underline{\alpha\beta}}(x_{\underline{\alpha\beta}})\sigma(X_1)\sigma(x)$$

$$\times\,K^{(2)}\left(\frac{X_{1\alpha}-y_\alpha}{h}\right)K^{(2)}\left(\frac{X_{1\beta}-y_\beta}{h}\right)$$

$$\times\left\{\frac{\varphi_{\alpha\beta}(y_\alpha,y_\beta)}{\varphi(y)\varphi(y_\alpha,y_\beta,X_{\underline{\alpha\beta}})}+\frac{\varphi_{\alpha\beta}(X_{1\alpha},X_{1\beta})}{\varphi(X_1)\varphi(X_{1\alpha},X_{1\beta},y_{\underline{\alpha\beta}})}\right\}$$

$$\times\,\varphi_{\underline{\alpha\beta}}(X_{1\underline{\alpha\beta}})\varphi_{\underline{\alpha\beta}}(y_{\underline{\alpha\beta}})\sigma(X_1)\sigma(y)\Bigg]$$

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

or

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

Introducing the change of variable

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

we obtain

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

Using convolution notation, one has

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

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

or

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

$$\times K^{(4)}\left(\frac{x_\alpha - y_\alpha}{h}\right)K^{(4)}\left(\frac{x_\beta - y_\beta}{h}\right)$$

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

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

which is what we set out to prove. ■

By techniques used in the two previous lemmas, we have the following lemma.

LEMMA A5. *As* $h \to 0$ *and* $nh^2 \to \infty$,

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

Lemmas A3 and A5 and the martingale central limit theorem of Hall (1984) imply the following propositions.

PROPOSITION A1. *As* $h \to 0$ *and* $nh^2 \to \infty$,

$$nh\sum_{1\le i<j\le n} H(X_i,\varepsilon_i,X_j,\varepsilon_j)$$

$$\xrightarrow{\mathcal{L}} N\left\{0,2\|K^{(2)}\|_2^4\int \frac{\varphi^2_{\alpha\beta}(z_{1\alpha},z_{1\beta})\varphi^2_{\underline{\alpha\beta}}(z_{1\underline{\alpha\beta}})\varphi^2_{\underline{\alpha\beta}}(z_{2\underline{\alpha\beta}})}{\varphi(z_1)\varphi(z_{1\alpha},z_{1\beta},z_{2\underline{\alpha\beta}})}\sigma^2(z_1)\sigma^2(z_{1\alpha},z_{1\beta},z_{2\underline{\alpha\beta}})dz_1\,dz_{2\underline{\alpha\beta}}\right\}.$$

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

PROPOSITION A2. *As* $h \to 0$ *and* $nh^2 \to \infty$,

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

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. ■

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

$$\|B_1\|_{L^2(S_{\alpha\beta},\varphi_{\alpha\beta})} = \sqrt{\int B_1^2(x_\alpha,x_\beta)\varphi_{\alpha\beta}(x_\alpha,x_\beta)dx_\alpha\,dx_\beta} \le CM \qquad \textbf{(A.11)}$$

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

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

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press

is the bias function of Theorem 2. Meanwhile, because

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

it follows that

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

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

$$\|f^*_{\alpha\beta}\|^2_{L^2(S_{\alpha\beta}, \varphi_{\alpha\beta})} = \int f^{*2}_{\alpha\beta}(x_\alpha, x_\beta)\varphi_{\alpha\beta}(x_\alpha, x_\beta)dx_\alpha dx_\beta \geq a^2. \tag{A.12}$$

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

$$\begin{aligned}T'_n = nh\int \widehat{f^*_{\alpha\beta,n}}^2(x_\alpha, x_\beta)\varphi_{\alpha\beta}(x_\alpha, x_\beta)dx_\alpha dx_\beta \\ - \frac{2\{K^{(2)}(0)\}^2}{h}\int \frac{\varphi_{\alpha\beta}(z_\alpha, z_\beta)\varphi^2_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}})}{\varphi(z)}\sigma^2(z)dz \\ - nh\int f^{*2}_{\alpha\beta,n}(x_\alpha, x_\beta)\varphi_{\alpha\beta}(x_\alpha, x_\beta)dx_\alpha dx_\beta \\ - 2nh^3\int f^*_{\alpha\beta,n}(x_\alpha, x_\beta)B_{1n}(x_\alpha, x_\beta)\varphi_{\alpha\beta}(x_\alpha, x_\beta)dx_\alpha dx_\beta,\end{aligned}$$

where $(f_{\alpha\beta,n})_{n=1}^\infty$ is the sequence in Theorem 7 and B_{1n} the corresponding bias coefficients. Note that although the function $f^*_{\alpha\beta,n}(x_\alpha, x_\beta)$ 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 $f^*_{\alpha\beta,n}(x_\alpha, x_\beta)$ is bounded uniformly for $n = 1, 2, \ldots$, 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

$$T'_n \xrightarrow{\mathcal{L}} N\{0, V(K, \varphi, \sigma)\} \tag{A.13}$$

as $n \to \infty$. Now let

$$t_n = nh \int f^{*2}_{\alpha\beta,n}(x_\alpha, x_\beta)\varphi_{\alpha\beta}(x_\alpha, x_\beta)dx_\alpha dx_\beta$$
$$+ 2nh^3 \int f^{*}_{\alpha\beta,n}(x_\alpha, x_\beta)B_{1n}(x_\alpha, x_\beta)\varphi_{\alpha\beta}(x_\alpha, x_\beta)dx_\alpha dx_\beta.$$

Then

$$t_n \geq nh\|f^*_{\alpha\beta,n}\|^2_{L^2(S_{\alpha\beta},\varphi_{\alpha\beta})} - 2nh^3\|f^*_{\alpha\beta,n}\|_{L^2(S_{\alpha\beta},\varphi_{\alpha\beta})}\|B_{1n}\|_{L^2(S_{\alpha\beta},\varphi_{\alpha\beta})}$$
$$= nh\|f^*_{\alpha\beta,n}\|_{L^2(S_{\alpha\beta},\varphi_{\alpha\beta})}\{\|f^*_{\alpha\beta,n}\|_{L^2(S_{\alpha\beta},\varphi_{\alpha\beta})} - 2h^2\|B_{1n}\|_{L^2(S_{\alpha\beta},\varphi_{\alpha\beta})}\}$$
$$t_n \geq nha_n\{a_n - 2h^2CM\},$$

which, using the condition that $a_n^{-1} = o(nh + h^{-2})$, entails that $t_n \to \infty$ as $n \to \infty$. By the definition of the test (33)

$$p_n = P[T_n' + t_n \geq \Phi^{-1}(1-\eta)V^{1/2}(K,\varphi,\sigma)]. \quad \textbf{(A.14)}$$

Now (A.13), (A.14), and $t_n \to \infty$ yield $\lim_{n\to\infty} p_n = 1$. ■

Proof of Theorem 8. In parallel to the proof of Theorem 6, one can decompose

$$\sum_{l=1}^{n} \widehat{f^*_{\alpha\beta}}^{\,2}(X_{l\alpha}, X_{l\beta})/n = \sum_{1\leq i\neq j\leq n} \tilde{H}(X_i,\varepsilon_i,X_j,\varepsilon_j) + \sum_{i=1}^{n}\tilde{H}(X_i,\varepsilon_i,X_i,\varepsilon_i)$$
$$+ \sum_{l=1}^{n} f^{*2}_{\alpha\beta}(X_{l\alpha},X_{l\beta})/n + 2h^2\sum_{l=1}^{n} f^*_{\alpha\beta}(X_{l\alpha},X_{l\beta})B_1(X_{l\alpha},X_{l\beta})/n$$
$$+ o_p(h^2),$$

in which

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

with

$$\widetilde{w}_{i\alpha,l} = \frac{1}{n}K_h(X_{l\alpha} - X_{i\alpha})\frac{\varphi_{\underline{\alpha}}(X_{i\underline{\alpha}})}{\varphi(X_{l\alpha}, X_{i\underline{\alpha}})}, \quad \textbf{(A.15)}$$

$$w_{i\alpha\beta,l} = \frac{1}{n}K_h(X_{l\alpha} - X_{i\alpha}, X_{l\beta} - X_{i\beta})\frac{\varphi_{\underline{\alpha\beta}}(X_{i\underline{\alpha\beta}})}{\varphi(X_{l\alpha}, X_{l\beta}, X_{i\underline{\alpha\beta}})}. \quad \textbf{(A.16)}$$

It is directly verified that for $w_{i\alpha}$ defined in (A.2) and $w_{i\alpha\beta}$ defined in (A.5)

$$\sum_{l=1}^{n}\frac{1}{n}(\widetilde{w}_{i\alpha\beta,l} - \widetilde{w}_{i\alpha,l} - \widetilde{w}_{i\beta,l})(\widetilde{w}_{j\alpha\beta,l} - \widetilde{w}_{j\alpha,l} - \widetilde{w}_{j\beta,l})$$
$$= \int (w_{i\alpha\beta} - w_{i\alpha} - w_{i\beta})(w_{j\alpha\beta} - w_{j\alpha} - w_{j\beta})\varphi_{\alpha\beta}(x_\alpha, x_\beta)dx_\alpha dx_\beta\{1 + O_p(n^{-1/2})\}$$

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

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

whereas

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

with H as defined in (A.9). 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. ∎

https://doi.org/10.1017/S0266466602182016 Published online by Cambridge University Press