Submitted to *Econometric Theory*, November, 1997

Average Regression Surface for Dependent Data

ZONGWU CAI
Department of Mathematics
Southwest Missouri State University
Springfield, MO 65804

JIANQING FAN
Department of Statistics
University of North Carolina
Chapel Hill, NC 27599-3260

November 10, 1997

ABSTRACT

We study the estimation of the additive components in additive regression models, based on the weighted sample average of regression surface, for stationary α-mixing processes. Explicit expression of this method makes possible a fast computation and allows an asymptotic analysis. The estimation procedure is especially useful for additive modeling. In this paper, it is shown that the average surface estimator shares the same optimality as the ideal estimator and has the same ability of estimating the additive component as the ideal case where other components are known. Formulas for the asymptotic bias and normality of the estimator are established.

Keywords: Additive models; α-mixing; Asymptotic bias; Asymptotic normality; Kernel estimates; Partially local linear estimate.

1. Introduction

Let $\{Y_i, \underline{X}_i\}_{i=-\infty}^{\infty}$ be jointly stationary processes with $\underline{X}_i$ taking values in $\Re^d$, where the dimension $d \geq 1$. Assume $E|Y_1| < \infty$ and define the multivariate regression function:

$$m(x_1, \ldots, x_d) = E\left(Y \mid X_1 = x_1, \ldots, X_d = x_d\right), \tag{1.1}$$

where $(Y, \underline{X})$ has the same distribution as $(Y_i, \underline{X}_i)$. The regression function $m(\cdot)$ plays an important role in data analysis, for example, the l-step prediction $(Y_i = X_{i+l})$ of time series. Additive regression model is a useful statistical tool for high-dimensional data analysis. In this paper, we focus on the following additive model:

$$m(x_1, \ldots, x_d) = E\left(Y \,\middle|\, \underline{X}^{(1)} = \underline{x}_1, \underline{X}^{(2)} = \underline{x}_2\right) = \mu + f_1(\underline{x}_1) + f_2(\underline{x}_2), \tag{1.2}$$

where $\underline{X}^{(1)} = (X_1, \ldots, X_r)^T$, $\underline{X}^{(2)} = (X_{r+1}, \ldots, X_{r+q})^T$ with $r+q=d$, and μ is a constant. Note that $\underline{X}^T$ denotes the transpose of $\underline{X}$. In the above additive model, we assume that the variables $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$ are continuous and take values in $\Re^r$ and $\Re^q$, respectively. For identifiability, we assume without loss of generality that $E\left\{f_1\left(\underline{X}^{(1)}\right)\right\} = E\left\{f_2\left(\underline{X}^{(2)}\right)\right\} = 0$. This general setup was considered by Fan, Härdle and Mammen (1997; henceforth FHM) in an independent and identically distributed (i.i.d.) setting. The need for nonlinear time series modeling and forecasting (see Tong (1990)) motivates us to consider the above model for dependent data.

The above setup is wide enough to include many useful statistical models. Some of these are as follows:

a) First, consider

$$X_t = f(X_{t-1}, \ldots, X_{t-i_1}) + g(X_{t-1}, \ldots, X_{t-i_2})\,\nu_t, \qquad t \geq 1, \tag{1.3}$$

where f and g are Lebesgue measurable functions, and $\{\nu_i\}$ are a sequence of i.i.d. random variables (r.v.s) with mean zero and a finite second moment. It is easy to show that the conditional mean and variance are respectively given by

$$E(X_t \mid X_{t-1}, \ldots, X_{t-i_0}) = f(X_{t-1}, \ldots, X_{t-i_1}), \tag{1.4}$$

and

$$\mathrm{Var}(X_t \mid X_{t-1}, \ldots, X_{t-i_0}) = g^2(X_{t-1}, \ldots, X_{t-i_2})\,\mathrm{Var}(\nu_1), \tag{1.5}$$

where $i_0 = \max(i_1, i_2)$. The classes defined by (1.3) include many of more familiar nonlinear parametric models commonly encountered in econometrics (see, e.g., Tjøstheim and Auestad (1994), hereafter referred to as TA): the threshold model and its various modifications; the autoregressive conditional heteroscedastic (ARCH) model as defined in Tong (1990); the exponential autoregressive model introduced by Engle (1982); and the multivariate adaptive splines (MARS) models (see, e.g., Lewis and Stevens (1991)). The nonparametric kernel-type estimation of the conditional mean (see (1.4)) and the conditional variance (see (1.5)) was studied in detail by TA (1994) using the projection method. The appeal of imposing the additive structure on f or g is to avoid so-called "curse of dimensionality". In particular, the following additive model is included in our setting via taking $Y_t = X_t$ and $\underline{X}_t = (X_{t-1}, \ldots, X_{t-i_1})$:

$$X_t = \mu + f_1(X_{t-1}) + \cdots + f_{i_1}(X_{t-i_1}) + g(X_{t-1}, \ldots, X_{t-i_2})\, \nu_t.$$

The model is a useful extension of the classical autoregressive model. Our approach will enable one to construct an explicit estimator of $f_j(\cdot)$ which possesses certain optimality criterion, and hence to predict the future value of the series. The general setup also includes partial autoregressive models such as

$$X_t = \mu + f_1(X_{t-1}) + \cdots + f_{i_1}(X_{t-i_1}) + \beta_1 X_{t-i_1-1} + \cdots + \beta_{i_2} X_{t-i_1-i_2} + g(X_{t-1}, \ldots, X_{t-i_3})\, \nu_t.$$

The models have flexibility of modeling some of components nonparametrically (reducing possible modeling bias) and other components linearly (reducing effective number of parameters).

b) Secondly, model (1.2) includes the additive model in the nonparametric regression with independent data:

$$m(x_1, \ldots, x_q) = \mu + \sum_{j=1}^{q} f_j(x_j). \tag{1.6}$$

A thorough discussion of this model can be found in Baju, Hastie and Tibshirani (1989) and Hastie and Tibshirani (1990) for the i.i.d. setting. The additive components f_j can be estimated with the one-dimensional nonparametric rate, see, e.g., Stone (1985, 1994) for details. In most papers, for estimation of the additive components, algorithms have been proposed, based on the iterative backfitting procedures. Their asymptotic properties are not well understood due to the implicit definition of the estimator. Also, computation can be slow in the high-dimensional case. To avoid the drawbacks of the iterative procedures, in Auestad and Tjøstheim (1990), TA (1994), Linton and Nielsen (1995), and FHM (1997), a direct method has been proposed which is based on "average regression surface". The procedure was referred as "projection method" in

Auestad and Tjøstheim (1990) and TA (1994) and as "marginal integration method" by Linton and Nielsen (1995) and Linton (1997). As pointed out by FHM (1997), the direct method has some advantages: This method does not use iterations. Fast computation can be implemented. Furthermore, the explicit definition allows detailed asymptotic analysis. Efficient estimation of additive components was studied in Linton (1997) and FHM (1997) by using two independent approaches. Masry and Tjøstheim (1997) extended the applicability of average surface idea to additive nonlinear ARX time series. A useful modification of the average surface idea is given in Hargartener (1996).

c) Finally, model (1.2) covers additive partially linear models:

$$m(x_1, \ldots, x_{q+r}) = \mu + \sum_{j=1}^{q} f_j(x_j) + \underline{x}_3^T \underline{\beta}^* \tag{1.7}$$

and partial interaction model:

$$m(x_1, \ldots, x_d) = \mu + f_{12}(x_1, x_2) + \sum_{j=3}^{d} f_j(x_j). \tag{1.8}$$

Further discussions on models (1.7) and (1.8) can be found in FHM (1997) for the i.i.d. setup.

The aim of this paper is to estimate the low dimensional additive component f_1 in (1.2). Analogously, f_2 can be estimated in the same fashion. The basic idea for estimating f_1 is to first estimate directly the high-dimensional regression surface $m(x_1, \ldots, x_d)$ and then average the regression surface over variables $\underline{X}^{(2)}$ to stabilize the variance. The regression surface is estimated by using local polynomial fitting, which has been studied extensively for example Tsybakov (1986), Fan (1993), Ruppert and Wand (1994), Fan and Gijbels (1996), Masry (1996) and Masry and Fan (1997). It has advantages over the Nadaraya-Watson regression estimator. In particular, it reduces bias of the Nadaraya-Watson estimator and copes well with the edge effect. For more details, see, e.g., Fan (1993) and Fan and Gijbels (1996) for the i.i.d. case, and Masry and Fan (1997) for the dependent situations.

The paper is organized as follows. In the next section, we introduce our estimation procedure. In Section 3 the main results of the paper, asymptotic bias and normality, are also formulated in the same section; their proofs are deferred to Section 5, based on some lemmas, which are proved in the Appendix. An application to additive model is discussed in Section 4. Finally, in Section 5 the assumptions used throughout the paper are gathered together for easy reference, followed by some brief comments.

2. Average of Regression Surface

First of all, let us introduce some notation. Denote by

$$\underline{X}_j^{(1)} = \begin{pmatrix} X_{1,j} \\ \vdots \\ X_{r,j} \end{pmatrix}, \qquad \underline{X}_j^{(2)} = \begin{pmatrix} X_{r+1,j} \\ \vdots \\ X_{d,j} \end{pmatrix}, \qquad \underline{X}_j = \begin{pmatrix} X_{1,j} \\ \vdots \\ X_{d,j} \end{pmatrix} = \begin{pmatrix} \underline{X}_j^{(1)} \\ \underline{X}_j^{(2)} \end{pmatrix},$$

and $m(\underline{x}) = E(Y \mid \underline{X} = \underline{x})$. Let $W : \Re^q \to \Re$ be a known weight function with $E\left\{W\left(\underline{X}^{(2)}\right)\right\} = 1$. Observe that, under model (1.2),

$$\begin{aligned} E\left\{m\left(\underline{x}_1, \underline{X}^{(2)}\right) W\left(\underline{X}^{(2)}\right)\right\} &= \mu + f_1(\underline{x}_1) + E\left\{f_2\left(\underline{X}^{(2)}\right) W\left(\underline{X}^{(2)}\right)\right\} \\ &= \mu_0 + f_1(\underline{x}_1) \equiv f_1^*(\underline{x}_1), \end{aligned} \tag{2.1}$$

where

$$\mu_0 = \mu + E\left\{f_2\left(\underline{X}^{(2)}\right) W\left(\underline{X}^{(2)}\right)\right\}. \tag{2.2}$$

Thus, f_1 can be constructed, within a constant shift, via averaging the regression surface with respect to variables $\underline{X}^{(2)}$. This in turn suggests a direct estimation procedure: Estimate the regression function m first and then average out the estimated regression surface with respect to the variables $\underline{X}^{(2)}$. The constant factor is not related to the final estimator, since $f_1(\cdot)$, in practice, is centered to have mean zero for identifiability purpose. This kind of averaging idea was studied by TA (1994) under time series models, and by Linton and Nielsen (1995) for the i.i.d. setting, and was further extended by FHM (1997). The purpose of introducing a weight function here is to optimize the estimation procedure.

Consider the local linear approximation of f_1 at a fixed point $\underline{x}_1$,

$$f_1(\underline{u}_1) \approx a(\underline{x}_1) + \underline{b}^T(\underline{x}_1)(\underline{u}_1 - \underline{x}_1),$$

where $\underline{u}_1$ lies in a neighborhood of $\underline{x}_1$. Also, the local constant approximation of f_2 at a fixed point $\underline{x}_2$ is applied:

$$f_2(\underline{u}_2) \approx c(\underline{x}_2) \qquad \text{for} \qquad \underline{u}_2 \approx \underline{x}_2.$$

Thus, we can approximate m locally by a linear term in a neighborhood of $(\underline{x}_1, \underline{x}_2)$,

$$m(\underline{u}_1, \underline{u}_2) \approx \gamma + \underline{\beta}^T(\underline{u}_1 - \underline{x}_1) \tag{2.3}$$

for some γ and $\underline{\beta}$, depending on $\underline{x}_1$ and $\underline{x}_2$. The reason for introducing the local constant approximation for the "nuisance function" f_2 is to reduce the number of local parameters so that it is easier to implement for practical purpose. Higher order approximation can also be employed for the function f_2 at the expenses of introducing more local parameters and the theoretical results continue to hold.

Let K and L be the kernel functions and let h_1 and h_2 be the bandwidths. Given the observations $\{Y_i,\ \underline{X}_i\}_{i=1}^n$, consider the multivariate weighted least squares

$$\sum_{i=1}^{n}\left[Y_i-\gamma-\underline{\beta}^T\left(\underline{X}_i^{(1)}-\underline{x}_1\right)\right]^2 K_{h_1}\left(\underline{X}_i^{(1)}-\underline{x}_1\right)L_{h_2}\left(\underline{X}_i^{(2)}-\underline{x}_2\right), \tag{2.4}$$

where

$$K_{h_1}(\underline{u})=\frac{1}{h_1^r}K\left(\frac{\underline{u}}{h_1}\right) \qquad \text{and} \qquad L_{h_2}(\underline{u})=\frac{1}{h_2^q}L\left(\frac{\underline{u}}{h_2}\right). \tag{2.5}$$

Minimizing (2.4) with respect to γ and $\underline{\beta}$ gives the estimates of γ and $\underline{\beta}$, respectively. Let $\widehat{\gamma}(\underline{x})$ and $\widehat{\underline{\beta}}(\underline{x})$ be the solution to (2.4). Thus, our partially local linear estimator of m is $\widehat{m}(\underline{x})=\widehat{\gamma}(\underline{x})$. By computing the weighted sample average of $\widehat{m}$, the following average regression surface estimator is proposed in FHM (1997):

$$\widehat{f}_1^*(\underline{x}_1)=\frac{1}{n}\sum_{i=1}^{n}\widehat{m}\left(\underline{x}_1,\ \underline{X}_i^{(2)}\right)W\left(\underline{X}_i^{(2)}\right), \tag{2.6}$$

$$\widehat{f}_1(\underline{x}_1)=\widehat{f}_1^*(\underline{x}_1)-\overline{f}_1^*, \qquad \text{and} \qquad \overline{f}_1^*=\frac{1}{n}\sum_{i=1}^{n}\widehat{f}_1^*\left(\underline{X}_i^{(1)}\right). \tag{2.7}$$

This is a functional of $\widehat{m}(\underline{x})$, and it turns out to possess good properties. Note that when the local constant fit is employed (i.e. $\underline{\beta}=\underline{0}$) in (2.3), the resulting estimate $\widehat{\gamma}$ is the multivariate kernel regression estimator, which was discussed in TA (1994). For details, see the relations (5) and (6) in TA (1994).

Let $X_{n\times(r+1)}=X(\underline{x}_1)$ be the matrix with the i^{th} row $\left(1,\ (\underline{X}_i^{(1)}-\underline{x}_1)^T\right)$ and $W^*_{n\times n}=W^*(\underline{x})$ be the diagonal weight matrix with the i^{th} diagonal element $W_i^*(\underline{x})=K_h^*(\underline{X}_i-\underline{x})$, where

$$K_h^*(\underline{x})=K_{h_1}(\underline{x}_1)L_{h_2}(\underline{x}_2),$$

to the least-squares problem (2.4). Then,

$$\begin{pmatrix}\widehat{\gamma}\\ \widehat{\underline{\beta}}\end{pmatrix}=\left(X^TW^*X\right)^{-1}X^TW^*\underline{Y}, \tag{2.8}$$

where $\underline{Y}=(Y_1,\ldots,Y_n)^T$, and the simple algebra shows that $\widehat{m}(\underline{x})$ can be expressed as

$$\widehat{m}(\underline{x})=\sum_{i=1}^{n}K_n(\underline{X}_i-\underline{x})Y_i, \tag{2.9}$$

where with $S_n(\underline{x}) = X^T(\underline{x}_1) W^*(\underline{x}) X(\underline{x}_1)$ and $\underline{e}_1^T = (1,\, 0,\, \ldots,\, 0)$,

$$K_n(\underline{t} - \underline{x}) = K_n(\underline{t}_1 - \underline{x}_1,\, \underline{t}_2 - \underline{x}_2) = \underline{e}_1^T\, S_n^{-1}(\underline{x}) \begin{pmatrix} 1 \\ \underline{t}_1 - \underline{x}_1 \end{pmatrix} K_h^*(\underline{t} - \underline{x}). \tag{2.10}$$

It follows from the least-squares theory that, for all $\underline{x}$,

$$\sum_{i=1}^{n} K_n(\underline{X}_i - \underline{x}) = 1, \qquad \text{and} \qquad \sum_{i=1}^{n} K_n(\underline{X}_i - \underline{x}) \left(\underline{X}_i^{(1)} - \underline{x}_1\right) = 0. \tag{2.11}$$

3. Main Results

Before we state our main result, we introduce the mixing coefficient. Let $\mathcal{F}_a^b$ be the σ-algebra of events generated by $\{Y_i,\, \underline{X}_j;\; a \le j \le b\}$. The stationary processes $\{Y_j,\, \underline{X}_j\}_{i=-\infty}^{\infty}$ are called strongly mixing (α-mixing), if

$$\sup_{\substack{A \in \mathcal{F}_{-\infty}^0 \\ B \in \mathcal{F}_k^\infty}} |P(AB) - P(A)P(B)| = \alpha(k) \quad \downarrow \quad 0,$$

as $k \to \infty$, and $\alpha(k)$ is called the strong mixing coefficient.

Among various mixing conditions used in literature, α-mixing is reasonably weak, and has many practical applications. Many stochastic processes and time series are known to be α-mixing. Gorodetskii (1977) and Withers (1981) had obtained various conditions for linear process to be α-mixing. Under certain weak assumptions autoregressive and more generally bilinear time series models are strongly mixing with mixing coefficients decaying exponentially fast. Auestad and Tjøstheim (1990) provided illuminating discussions on the role of α-mixing (including geometric ergodicity) for model identification in nonlinear time series analysis. Under some mild conditions, Masry and Tjøstheim (1995) showed that the ARCH process is stationary α-mixing.

For easy reference, we introduce the following notation. Let $p(\underline{x}_1,\, \underline{x}_2)$ be the joint density of $(\underline{X}^{(1)},\, \underline{X}^{(2)})$, and $p_1(\underline{x}_1)$ and $p_2(\underline{x}_2)$ be the marginal densities of $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$, respectively. Let

$$||K||^l = \int |K(\underline{u})|^l d\underline{u} \qquad \text{and} \qquad \mu_2(K) = \int \underline{u}\,\underline{u}^T K(\underline{u}) d\underline{u}.$$

All limits will be taken as $n \to \infty$; this will not be mentioned explicitly in the body of the paper. Then, under Assumptions **(1)**-**(9)** stated in Section 5, we have the following theorem which generalizes one of the main results in FHM (1997) to the dependent case.

Theorem 1 Under Assumptions **(1)**-**(9)**, if the bandwidths are chosen such that $h_1 \to 0$, $h_2 \to 0$ in such a way that

$$n\,h_1^{r+4} = O(1), \qquad h_2^{l_1}/h_1^2 \to 0, \qquad \text{and} \qquad n\,h_1^r\,h_2^q/\log n \to \infty, \tag{3.1}$$

then,

$$(nh_1^r)^{1/2}\left[\widehat{f}_1^*(\underline{x}_1) - f_1^*(\underline{x}_1) - \frac{1}{2}h_1^2\,\mathrm{tr}\,\{f_1''(\underline{x}_1)\mu_2(K)\}\right] \stackrel{d}{\longrightarrow} N(0,\ v(\underline{x}_1)), \tag{3.2}$$

where

$$v(\underline{x}_1) = ||K||^2\,p_1(\underline{x}_1)\,E\left[\Gamma^2(\underline{X})\,\sigma^2(\underline{X})\,|\,\underline{X}^{(1)} = \underline{x}_1\right] \tag{3.3}$$

with

$$\Gamma(\underline{x}_1,\,\underline{x}_2) = \frac{p_2(\underline{x}_2)\,W(\underline{x}_2)}{p(\underline{x}_1,\,\underline{x}_2)}, \tag{3.4}$$

and

$$\sigma^2(\underline{x}) = \mathrm{Var}(Y\,\big|\,\underline{X} = \underline{x}). \tag{3.5}$$

Remark 1. If we consider weight function $W(\cdot)$ that minimizes $v(\underline{x}_1)$, the optimal weight function is

$$W\left(\underline{X}^{(2)}\right) = c^{-1}\frac{p\left(\underline{x}_1,\,\underline{X}^{(2)}\right)p_1(\underline{x}_1)}{\sigma^2\left(\underline{x}_1,\,\underline{X}^{(2)}\right)p_2\left(\underline{X}^{(2)}\right)}, \tag{3.6}$$

where

$$c = p_1(\underline{x}_1)^2 E\{\sigma^{-2}(\underline{X})\,|\,\underline{X}^{(1)} = \underline{x}_1\}.$$

The optimal minimal variance is

$$\min_W v(\underline{x}_1) = \frac{||K||^2}{p_1(\underline{x}_1)}\left[E\{\sigma^{-2}(\underline{X})\,|\,\underline{X}^{(1)} = \underline{x}_1\}\right]^{-1}, \tag{3.7}$$

and $\min_W v(\underline{x}_1) = ||K||^2\sigma^2/p_1(\underline{x}_1)$ if $\sigma^2(\underline{x}_1) = \sigma^2$. For details, see, e.g., (3.4)-(3.6) in FHM (1997). The optimal weight function (3.6) depends on unknown functions and a method on how to choose the optimal weight function based on the data was discussed in FHM (1997) in greater detail. In the ideal situation where $f_2(\underline{x}_2)$ is known, one can estimate $f_1(\underline{x}_1)$, by directly regressing $Y - f_2\left(\underline{X}^{(2)}\right)$ on $\underline{X}^{(1)}$ and such an ideal estimator is optimal in an asymptotic minimax sense (see Fan (1993)). Surprisingly, the average surface estimator (2.6) has the same asymptotic bias and variance as the ideal estimator when $\sigma^2(\underline{x})$ is a constant, even though the former does not rely on the knowledge of f_2. For details, see Remarks 5-8 in FHM (1997).

4. An Application to Additive Model

We now consider the following additive model as (1.6):

$$m(\underline{u}) = E\left(Y \,|\, \underline{U} = \underline{u}\right) = \mu + g_1(u_1) + \ldots + g_q(u_q), \tag{4.1}$$

where $g_1(\cdot), \ldots, g_q(\cdot)$ are univariate functions satisfying the identifiability condition

$$E\{g_i(U_i)\} = 0, \quad i = 1, \ldots, q,$$

μ is an unknown parameter, and $\underline{U} = (U_1, \ldots, U_q)^T$ is a continuous random vector having a joint density p. Our goal is to estimate each additive component g_k using the average surface method. As in (2.1), let g_k^*, $k = 1, \ldots, q$, be the average of regression function:

$$g_k^*(u_k) \equiv E\left\{m\left(\underline{U}^{\,k}\right) W_k\left(\underline{U}^{\,-k}\right)\right\} = g_k(u_k) + \mu_k, \tag{4.2}$$

where $\mu_k = \mu + E\left[\sum_{j \neq k} g_j(U_j) W_k\left(U^{-k}\right)\right]$, $\underline{U}^{\,k} = (U_1, \ldots, U_{k-1}, u_k, U_{k+1}, \ldots, U_q)^T$, $\underline{U}^{\,-k} = (U_1, \ldots, U_{k-1}, U_{k+1}, \ldots, U_q)^T$ having the density p^{-k} and $W_k : \Re^{q-1} \to \Re$ is the weight function such that $E\left\{W_k\left(\underline{U}^{\,-k}\right)\right\} = 1$. For the given observations $\{Y_i, \underline{U}_{\,i}\}_{i=1}^n$, the average surface estimator $\widehat{g}_k^*$ is defined as in (2.6) but now using the bandwidths $h_1 = h_{1k}$ and $h_2 = h_{2k}$,

$$\widehat{g}_k^*(u_k) = \frac{1}{n} \sum_{j=1}^{n} \widehat{m}_k\left(\underline{U}_{\,j}^{\,k}\right) W_k\left(\underline{U}_{\,j}^{\,-k}\right), \tag{4.3}$$

where $\underline{U}_{\,j}^{\,k} = (U_{1,j} \ldots, U_{k-1,j}, u_k, U_{k+1,j}, \ldots, U_{q,j})^T$, $\widehat{m}_k(\cdot)$ is the partially local linear estimator, defined as in (2.9), and $\underline{U}_{\,j}^{\,-k} = (U_{1,j}, \ldots, U_{k-1,j}, U_{k+1,j}, \ldots, U_{q,j})^T$.

Theorem 2 Suppose that the conditions of Theorem 1 hold for each component k. Then we have the following the joint asymptotic normality:

$$\begin{pmatrix} \sqrt{nh_{11}}\left[\widehat{g}_1^*(u_1) - g_1^*(u_1) - \frac{1}{2}\, h_{11}^2 g_1''(u_1)\mu_2(K)\right] \\ \vdots \\ \sqrt{nh_{1q}}\left[\widehat{g}_q^*(u_q) - g_q^*(u_q) - \frac{1}{2}\, h_{1q}^2 g_q''(u_q)\mu_2(K)\right] \end{pmatrix} \xrightarrow{d} N(0, \,\Sigma), \tag{4.4}$$

where $\Sigma = ||K||^2 \mathrm{diag}\left\{\sigma_1^2(u_1), \ldots, \sigma_q^2(u_q)\right\}$ with

$$\sigma_k^2(u_k) = p_k(u_k) E\left[\frac{\sigma^2(\underline{U})p^{-k}(\underline{U}^{\,-k})^2 W_k^2(\underline{U}^{\,-k})}{p^2(\underline{U})} \,\Big|\, U_k = u_k\right]. \tag{4.5}$$

Remark 2. If the ideal weight function in (3.6) applies to each additive component, the weight function W_k should become

$$W_k(\underline{U}^{\,-k}) = \frac{p(\underline{U}^{\,k})p_k(u_k)}{\sigma^2(\underline{U}^{\,k})p^{-k}(\underline{U}^{\,-k})}\left[\int \frac{p(\underline{U}^{\,k})p_k(u_k)d\underline{U}^{\,-k}}{\sigma^2(\underline{U}^{\,k})}\right]^{-1}, \tag{4.6}$$

and the ideal variance is $\|K\|^2/\left[E\left\{\sigma^{-2}(\underline{U}) \,\middle|\, U_k = u_k\right\} p_k(u_k)\right]$, which is $||K||^2\sigma^2/p_k(u_k)$ if $\sigma^2(\underline{u}) = \sigma^2$.

5. Conditions and Proofs

Before we embark on the proofs of theorems, let us list the assumptions used throughout the paper.

Assumptions:

(1) Suppose that the functions W and f_2 are bounded on the support D of W. The weight function $W(\underline{x}_2)$ is uniformly continuous.

(2) The kernel functions K and L are symmetric and have bounded supports. Furthermore, L is an order l_1 kernel.

(3) f_1 has a bounded second derivative in a neighborhood of $\underline{x}_1$ and $f_2(\underline{x}_2)$ has a bounded l_1^{th} order derivative. Furthermore, for $\underline{u}_1$ in a neighborhood of $\underline{x}_1$ and $\underline{u}_2 \in D$, the density $p(\underline{u}_1, \underline{u}_2)$ has bounded partial derivatives up to order two with respect to $\underline{u}_1$ and up to order l_1 with respect to $\underline{u}_2$, and it also satisfies

$$\inf_{\substack{\underline{u}_1 \in N(\underline{x}_1) \\ \underline{x}_2 \in D}} p(\underline{u}_1, \underline{x}_2) > 0,$$

where $N(\underline{x}_1)$ is a neighborhood of $\underline{x}_1$.

(4) Suppose that $\sigma^2(\underline{u})$ and $b(\underline{u}) = E\left(|Y - m(\underline{u})|^{2+\delta} \,|\, \underline{X} = \underline{u}\right)$ are continuous at the point $\underline{u}_1 = \underline{x}_1$, and

$$E\left(b(\underline{X})|\Gamma(\underline{X})|^{2+\delta} \,\middle|\, \underline{X}^{(1)} = \underline{x}_1\right) < \infty$$

is bounded for all $\underline{x}_1$ for some $\delta > 0$.

(5) Suppose that the joint conditional density $f_{(\underline{X}_1, \underline{X}_j)\,|\,(Y_1, Y_j)}$ of $(\underline{X}_1, \underline{X}_j)$ given (Y_1, Y_j) satisfies, for all $j > 1$ and all values of arguments involved,

$$\left|f_{(\underline{X}_1, \underline{X}_j)\,|\,(Y_1, Y_j)}(\underline{u}, \underline{v} \,|\, y_1, y_2)\right| \le M < \infty$$

for some positive constant M.

(6) The processes $\{Y_i, \underline{X}_i\}$ are strongly mixing with $\sum_{i=1}^{\infty} i^a[\alpha(i)]^{\frac{\delta}{2+\delta}} < \infty$ for some $a > \delta/(2+\delta)$, where δ is given in Assumption **(4)**.

(7) Assume that there is a sequence of positive integers satisfying $v_n \to \infty$ and $v_n = o\left(\sqrt{nh_1^r}\right)$ such that $(n/h_1^r)^{1/2}\alpha(v_n) \to 0$.

(8) The conditional distribution of $G(y \mid \underline{u})$ of Y given $\underline{X} = \underline{u}$ is continuous at the point $\underline{u}_1 = \underline{x}_1$.

(9) $n\, h_1^r\, h_2^{2q} / \log^2 n \to \infty$ and $h_1^4 \log n / h_2^q \to 0$.

Remark 3. Consider the popular choice for the bandwidth $h_1^r = d n^{-\theta^*}$ ($d > 0$, $0 < \theta^* < 1$), one can show that a sufficient condition for Assumption **(7)** is $\alpha(n) = O\left(n^{-\rho}\right)$ with $\rho > (1 + \theta^*)/(1 - \theta^*)$. In particular, if $\theta^* = 1/5$, then $\rho > 3/2$. Also, a sufficient condition for Assumption **(6)** is $\alpha(n) = O\left(n^{-\rho'}\right)$ with $\rho' > 2 + 2/\delta$. Therefore, if $\theta^* = 1/5$, and $\delta = 2$, then a sufficient condition for Assumptions **(6)** and **(7)** is $\alpha(n) = O\left(n^{-\rho''}\right)$ with $\rho'' > 3$. For details, see Masry and Fan (1997).

Note that by the dominated convergence theorem, it can be easily shown from Assumption **(8)** that for any $J > 0$, the functions

$$m_J(\underline{u}) = E\left(Y I(|Y| \le J) \,\middle|\, \underline{X} = \underline{u}\right) \qquad \text{and} \qquad \sigma_J^2(\underline{u}) = \mathrm{Var}\left(Y I(|Y| \le J) \,\middle|\, \underline{X} = \underline{u}\right)$$

are continuous at the point $\underline{u}_1 = \underline{x}_1$. Also, for each $L > 0$, $\widetilde{\sigma}_J^2(\underline{u}) = \mathrm{Var}\left(Y I(|Y| > J) \,\middle|\, \underline{X} = \underline{u}\right)$ is continuous at the point $\underline{u}_1 = \underline{x}_1$.

Proof of Theorem 1:

Let $W_i = W\left(\underline{X}_i^{(2)}\right)$ and $\underline{X}^i = \begin{pmatrix} \underline{x}_1 \\ \underline{X}_i^{(2)} \end{pmatrix}$. Then, by (2.1) and Assumptions **(1)** and **(6)**, and applying the central limit theorem for stationary α-mixing sequences (see, for example, Theorem 18.5.3 in Ibragimov and Linnik (1971, p.346)), we have

$$\frac{1}{n} \sum_{i=1}^{n} m\left(\underline{X}^i\right) W_i = f_1^*(\underline{x}_1) + O_p\left(n^{-1/2}\right). \tag{5.1}$$

Thus,

$$\widehat{f}_1^*(\underline{x}_1) - f_1^*(\underline{x}_1) = \frac{1}{n} \sum_{i=1}^{n} \left[\widehat{m}\left(\underline{X}^i\right) - m\left(\underline{X}^i\right)\right] W_i + O_p\left(n^{-1/2}\right). \tag{5.2}$$

Let $\varepsilon_i = Y_i - E(Y_i \mid \underline{X}_i) = Y_i - m(\underline{X}_i)$ and $s_j(\underline{x}) = m(\underline{X}_j) - m(\underline{x}) - f_1'(\underline{x}_1)^T \left(\underline{X}_j^{(1)} - \underline{x}_1\right)$. Then, it follows from (2.9)-(2.11) that

$$\begin{aligned}
\widehat{m}(\underline{x}) - m(\underline{x}) &= \sum_{j=1}^{n} K_n\left(\underline{X}_j - \underline{x}\right)\left[\varepsilon_j + s_j(\underline{x})\right] \\
&= \underline{e}_1^T S_n^{-1}(\underline{x}) \begin{pmatrix} 1 & \cdots & 1 \\ \underline{X}_1^{(1)} - \underline{x}_1 & \cdots & \underline{X}_n^{(1)} - \underline{x}_1 \end{pmatrix} W^*(\underline{x})\, \underline{\varepsilon} \\
&\quad + \underline{e}_1^T S_n^{-1}(\underline{x}) \begin{pmatrix} 1 & \cdots & 1 \\ \underline{X}_1^{(1)} - \underline{x}_1 & \cdots & \underline{X}_n^{(1)} - \underline{x}_1 \end{pmatrix} W^*(\underline{x})\, \underline{s}(\underline{x}),
\end{aligned} \tag{5.3}$$

where $\underline{\varepsilon} = (\varepsilon_j)_{n\times 1}$ and $\underline{s}(\underline{x}) = (s_j(\underline{x}))_{n\times 1}$. Let $H = \mathrm{diag}\left\{1,\ h_1^{-1},\ \ldots,\ h_1^{-1}\right\}$ be a $(r+1)\times(r+1)$ diagonal matrix and $a_n = (n\,h_1^r\,h_2^q/\log n)^{-1/2}$. Then, owing to the uniform weak convergence of the kernel density estimator (c.f., Theorem 1 in Masry (1996)) and by Assumptions **(2)**-**(6)**, we have

$$S_n^*(\underline{x}) \equiv \frac{1}{n} H\, S_n(\underline{x})\, H = \frac{1}{n}\sum_{j=1}^{n} W_j^*(\underline{x}) \begin{pmatrix} 1 \\ \frac{\underline{X}_j^{(1)} - \underline{x}_1}{h_1} \end{pmatrix} \begin{pmatrix} 1 \\ \frac{\underline{X}_j^{(1)} - \underline{x}_1}{h_1} \end{pmatrix}^T$$

converges to $S(\underline{x}) = p(\underline{x}) \begin{pmatrix} 1 & 0 \\ 0 & \mu_2(K) \end{pmatrix}$ in probability uniformly in $\underline{x}$, and

$$\begin{aligned}
\frac{1}{n} H\, S_n(\underline{x})\, H &= \frac{1}{n}\sum_{j=1}^{n} W_j^*(\underline{x}) \begin{pmatrix} 1 \\ \frac{\underline{X}_j^{(1)} - \underline{x}_1}{h_1} \end{pmatrix} \begin{pmatrix} 1 \\ \frac{\underline{X}_j^{(1)} - \underline{x}_1}{h_1} \end{pmatrix}^T \\
&= E\left\{ W_1^*(\underline{x}) \begin{pmatrix} 1 \\ \frac{\underline{X}_1^{(1)} - \underline{x}_1}{h_1} \end{pmatrix} \begin{pmatrix} 1 \\ \frac{\underline{X}_1^{(1)} - \underline{x}_1}{h_1} \end{pmatrix}^T \right\} + O_p(a_n) \\
&= \begin{pmatrix} p(\underline{x}) & h_1\, p^{(1,0)}(\underline{x})^T\, \mu_2(K) \\ h_1\, \mu_2(K)\, p^{(1,0)}(\underline{x}) & p(\underline{x})\, \mu_2(K) \end{pmatrix} + O_p(c_n),
\end{aligned}$$

uniformly in $\underline{x}$, where $c_n = h_1^2 + h_2^{l_1} + a_n$ and $p^{(1,0)}$ denotes the vector of partial derivatives of p with respect to $\underline{x}_1$. Now note that

$$\begin{pmatrix} p(\underline{x}) & h_1\, p^{(1,0)}(\underline{x})^T\, \mu_2(K) \\ h_1\, \mu_2(K)\, p^{(1,0)}(\underline{x}) & p(\underline{x})\, \mu_2(K) \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} p(\underline{x}) & 0 \\ 0 & p(\underline{x})\, \mu_2(K) \end{pmatrix}^{-1} + \frac{h_1}{p(\underline{x})} \begin{pmatrix} 0 & p^{(1,0)}(\underline{x})^T\, \mu_2(K) \\ \mu_2(K)\, p^{(1,0)}(\underline{x}) & 0 \end{pmatrix} + O_p(h_1^2).$$

Therefore,

$$n\, \underline{e}_1^T\, S_n^{-1}(\underline{x})\, H^{-1} = p^{-1}(\underline{x}) \left(1,\ h_1\, p^{(1,0)}(\underline{x})^T\, \mu_2(K)\right) + O_p(c_n). \tag{5.4}$$

Likewise, using the same argument as above, and by Assumptions **(2)**-**(6)**, one has

$$\begin{aligned}
&\frac{1}{n} H \begin{pmatrix} 1 & \cdots & 1 \\ \underline{X}_1^{(1)} - \underline{x}_1 & \cdots & \underline{X}_n^{(1)} - \underline{x}_1 \end{pmatrix} W^*(\underline{x})\ \underline{s}(\underline{x}) \\
&= \frac{1}{n}\sum_{j=1}^{n} W_j^*(\underline{x})\ s_j(\underline{x}) \begin{pmatrix} 1 \\ \frac{\underline{X}_j^{(1)} - \underline{x}_1}{h_1} \end{pmatrix} = O_p(c_n)
\end{aligned} \tag{5.5}$$

uniformly in $\underline{x}$. Substitute (5.4) and (5.5) into (5.3) to obtain

$$\underline{e}_1^T\, S_n^{-1}(\underline{x}) \begin{pmatrix} 1 & \cdots & 1 \\ \underline{X}_1^{(1)} - \underline{x}_1 & \cdots & \underline{X}_n^{(1)} - \underline{x}_1 \end{pmatrix} W^*(\underline{x})\ \underline{s}(\underline{x})$$

$$= p^{-1}(\underline{x}) \left\{ \frac{1}{n} \sum_{j=1}^{n} W_j^*(\underline{x}) \; s_j(\underline{x}) \right\}$$

$$+ p^{-1}(\underline{x}) \, p^{(1,0)}(\underline{x})^T \mu_2(K) \left\{ \frac{1}{n} \sum_{j=1}^{n} W_j^*(\underline{x}) \; s_j(\underline{x}) \left(\underline{X}_j^{(1)} - \underline{x}_1 \right) \right\} + O_p(c_n^2)$$

$$= \frac{1}{2} h_1^2 \operatorname{tr}\{f_1''(\underline{x}_1)\mu_2(K)\} + p^{-1}(\underline{x}) \, B_n(\underline{x}) + o_p(h_1^2) + O_p(c_n^2), \tag{5.6}$$

where

$$B_n(\underline{x}) = \frac{1}{n} \sum_{j=1}^{n} W_j^*(\underline{x}) \left[f_2\left(\underline{X}_j^{(2)}\right) - f_2(\underline{x}_2) \right] \left\{ 1 + p^{(1,0)}(\underline{x})^T \mu_2(K) \left(\underline{X}_j^{(1)} - \underline{x}_1 \right) \right\}.$$

Also, note that

$$S_n^{*-1}(\underline{x}) = S^{-1}(\underline{x}) \left[I + (S(\underline{x}) - S_n^*(\underline{x})) \, S_n^{*-1}(\underline{x}) \right].$$

Then, $S_n^{*-1}(\underline{x})$ converges to $S^{-1}(\underline{x})$ in probability uniformly in $\underline{x}$. It then follows that $S_n^{*-1}(\underline{x}) = S^{-1}(\underline{x})(1+o_p(1))$, where $o_p(1)$ is uniform in $\underline{x}$, and $n \, S_n^{-1}(\underline{x}) \, H^{-1} = H \, S^{-1}(\underline{x})(1+o_p(1))$. Therefore,

$$n \, \underline{e}_1^T \, S_n^{-1}(\underline{x})(\underline{x}) \, H^{-1} = p^{-1}(\underline{x}) \, \underline{e}_1^T \, (1 + o_p(1)). \tag{5.7}$$

Substituting (5.7) into (5.3), one obtains

$$\underline{e}_1^T \, S_n^{-1}(\underline{x}) \begin{pmatrix} 1 & \cdots & 1 \\ \underline{X}_1^{(1)} - \underline{x}_1 & \cdots & \underline{X}_n^{(1)} - \underline{x}_1 \end{pmatrix} W^*(\underline{x}) \; \underline{\varepsilon} = (1 + o_p(1)) \, p^{-1}(\underline{x}) \, T_n(\underline{x}), \tag{5.8}$$

where

$$T_n(\underline{x}) = \frac{1}{n} \sum_{j=1}^{n} W_j^*(\underline{x}) \, \varepsilon_j.$$

Substituting (5.6) and (5.8) into (5.3), after some algebra, we obtain

$$\widehat{m}(\underline{x}) - m(\underline{x}) = \frac{1}{2} h_1^2 \operatorname{tr}\{f_1''(\underline{x}_1)\mu_2(K)\} + p^{-1}(\underline{x}) \, \{T_n(\underline{x}) + B_n(\underline{x})\} + o_p(h_1^2) + O_p(c_n^2). \tag{5.9}$$

Thus, by (5.2), (5.9) and the strong law of large numbers (see, for example, Cai and Roussas (1992)), we have

$$\widehat{f}_1^*(\underline{x}_1) - f_1^*(\underline{x}_1) - \frac{1}{2} h_1^2 \operatorname{tr}\{f_1''(\underline{x}_1)\mu_2(K)\} = T_n^*(\underline{x}_1) + B_n^*(\underline{x}_1) + o_p(h_1^2) + O_p\left(c_n^2 + n^{-1/2}\right), \tag{5.10}$$

where with $A(\underline{x}) = W(\underline{x}_2)/p(\underline{x})$,

$$T_n^*(\underline{x}_1) = \frac{1}{n} \sum_{i=1}^{n} T_n(\underline{x}^i) \, A(\underline{x}^i), \qquad \text{and} \qquad B_n^*(\underline{x}_1) = \frac{1}{n} \sum_{i=1}^{n} B_n(\underline{x}^i) \, A(\underline{x}^i). \tag{5.11}$$

A simple algebra leads to

$$T_n^*(\underline{x}_1) = \frac{1}{n}\sum_{j=1}^{n} K_{h_1}\left(\underline{X}_j^{(1)} - \underline{x}_1\right)\,\Gamma\left(\underline{x}^j\right)\,\varepsilon_j$$
$$+ \quad \frac{1}{n}\sum_{j=1}^{n} K_{h_1}\left(\underline{X}_j^{(1)} - \underline{x}_1\right)\,\left\{\Gamma_n\left(\underline{x}^j\right) - \Gamma\left(\underline{x}^j\right)\right\}\,\varepsilon_j \equiv G_n(\underline{x}_1) + G_n^*(\underline{x}_1), \quad (5.12)$$

where Γ is defined in (3.4) and

$$\Gamma_n(\underline{x}_1,\, \underline{x}_2) = \frac{1}{n}\sum_{i=1}^{n} L_{h_2}\left(\underline{X}_i^{(2)} - \underline{x}_2\right)\,A(\underline{x}^i). \quad (5.13)$$

Substituting (5.12) into (5.10), one has

$$\widehat{f}_1^*(\underline{x}_1) - f_1^*(\underline{x}_1) - \frac{1}{2}\,h_1^2\,\mathrm{tr}\,\{f_1''(\underline{x}_1)\mu_2(K)\}$$
$$= G_n(\underline{x}_1) + G_n^*(\underline{x}_1) + B_n^*(\underline{x}_1) + o_p(h_1^2) + O_p\left(c_n^2 + n^{-1/2}\right). \quad (5.14)$$

In order to complete the proof, we need the following two lemmas but their proofs are relegated to the Appendix since they are quite involved. To this end, let

$$\varepsilon_j^* = \Gamma(\underline{x}^j)\,\varepsilon_j = \frac{\{Y_j - m(\underline{X}_j)\}\;p_2\left(\underline{X}_j^{(2)}\right)W\left(\underline{X}_j^{(2)}\right)}{p\left(\underline{x}_1,\,\underline{X}_j^{(2)}\right)}, \quad (5.15)$$

and

$$\zeta_j = \zeta_j(\underline{x}_1) = K_{h_1}\left(\underline{X}_j^{(1)} - \underline{x}_1\right)\varepsilon_j^*, \quad (5.16)$$

Then, by (5.12) and stationarity,

$$G_n(\underline{x}_1) = \frac{1}{n}\sum_{j=1}^{n}\zeta_j, \qquad \text{and} \qquad n\,\mathrm{Var}(G_n(\underline{x}_1)) = \mathrm{Var}(\zeta_1) + 2\sum_{j=2}^{n}\left(1 - \frac{j}{n}\right)\mathrm{Cov}(\zeta_1,\,\zeta_j).$$

Lemma 1 Under the assumptions of Theorem 1, we have,

$$n\,h_1^r\,\mathrm{Var}(G_n(\underline{x}_1)) \;\to\; v(\underline{x}_1), \qquad \text{and} \qquad h_1^r\sum_{j=2}^{n}\left|\mathrm{Cov}(\zeta_1,\,\zeta_j)\right| \;\to\; 0,$$

where $v(\underline{x}_1)$ is defined in (3.3).

Lemma 2 Under the assumptions of Theorem 1, we have,

$$G_n^*(\underline{x}_1) = o_p\left((n\,h_1^r)^{-1/2}\right), \qquad \text{and} \qquad B_n^*(\underline{x}_1) = o_p\left((n\,h_1^r)^{-1/2}\right).$$

It follows from (5.14), Lemma 2 and the conditions on the bandwidths (see (3.1) and Assumption **(9)**) that

$$\widehat{f}_1^*(\underline{x}_1) - f_1^*(\underline{x}_1) - \frac{1}{2} h_1^2 \operatorname{tr}\{f_1''(\underline{x}_1)\mu_2(K)\} = G_n(\underline{x}_1) + o_p\left((n\,h_1^r)^{-1/2}\right). \tag{5.17}$$

We remark that the third term in the left hand side of (5.17) can be viewed as the "asymptotic bias" of $\widehat{f}_1^*(\underline{x}_1)$, and the "asymptotic variance" of $\widehat{f}_1^*(\underline{x}_1)$ is $v(\underline{x}_1)$ defined in (3.3).

We now turn to show (3.2). This is equivalent to demonstrating the asymptotic normality of $G_n(\underline{x}_1)$ in (5.17). In discussing the convergence in (3.2), we use the familiar technique of "big block – small block" procedure. More precisely, partition the set $\{1, \ldots, n\}$ into $2\,k_n+1$ subsets with large block of size u_n and small block of size v_n, where $k = k_n = \lfloor n/(u_n+v_n)\rfloor$. Now we first consider the choices of the block sizes. Assumption **(7)** implies that there is a sequence of positive constants $\gamma_n \to \infty$ such that

$$\gamma_n\, v_n = o\left(\sqrt{nh_1^r}\right), \qquad \text{and} \qquad \gamma_n (n/h_1^r)^{1/2}\, \alpha(v_n) \to 0. \tag{5.18}$$

Define the large block size u_n by $u_n = \lfloor (nh_1^r)^{1/2}/\gamma_n \rfloor$ and the small block size v_n. Then, it can easily be shown from (5.18) that, as $n \to \infty$,

$$v_n/u_n \to 0, \qquad u_n/n \to 0, \qquad u_n\,(nh_1^r)^{-1/2} \to 0, \qquad \text{and} \qquad (n/u_n)\,\alpha(v_n) \to 0. \tag{5.19}$$

Ignore the dependence on $\underline{x}_1$, and for $j = 1, \ldots, k_n$, set $r_j^* = (j-1)(u_n+v_n)$, and

$$\xi_j = \sum_{i=r_j^*+1}^{r_j^*+u_n} \zeta_i, \qquad \eta_j = \sum_{i=r_j^*+u_n+1}^{r_{j+1}^*} \zeta_i, \qquad \text{and} \qquad \xi_{k+1} = \sum_{i=r_{k+1}^*+1}^{n} \zeta_i.$$

Write

$$n\,G_n(\underline{x}_1) = \sum_{j=1}^{k} \xi_j + \sum_{j=1}^{k} \eta_j + \xi_{k+1} = G_{n,1} + G_{n,2} + G_{n,3}. \tag{5.20}$$

It will be shown that, as $n \to \infty$,

$$\frac{h_1^r}{n}\left\{E[G_{n,2}]^2 + E[G_{n,3}]^2\right\} \to 0, \tag{5.21a}$$

$$\frac{h_1^r}{n} \sum_{j=1}^{k} E(\xi_j^2) \to v(\underline{x}_1), \tag{5.21b}$$

$$\left| E[\exp(itG_{n,1})] - \prod_{j=1}^{k} E[\exp(it\xi_j)] \right| \to 0, \tag{5.21c}$$

and

$$\frac{h_1^r}{n}\sum_{j=1}^{k}E\left[\xi_j^2 I\left(|\xi_j|>\varepsilon\sqrt{\frac{v(\underline{x}_1)\,n}{h_1^r}}\right)\right]\to 0 \qquad (5.21d)$$

for every $\varepsilon > 0$. (5.21a) implies that $G_{n,2}$ and $G_{n,3}$ are asymptotically negligible in probability; (5.21c) shows that the summands $\{\xi_j\}$ in $G_{n,1}$ are asymptotically independent; and (5.21b) and (5.21d) are the standard Lindeberg-Feller conditions for asymptotic normality of $G_{n,1}$ for the independent setup.

Let us first establish (5.21a). Observe that

$$E(G_{n,2})^2=\sum_{j=1}^{k}\mathrm{Var}(\eta_j)+2\sum_{1\le i<j\le k}\mathrm{Cov}(\eta_i,\,\eta_j)\equiv I_1+I_2. \qquad (5.22)$$

It follows from stationarity and Lemma 1 that

$$I_1=k_n\,\mathrm{Var}(\eta_1)=k_n\,\mathrm{Var}\left(\sum_{j=1}^{v_n}\zeta_j\right)=k_n\,v_n\,h_1^{-r}\,[v(\underline{x}_1)+o(1)]. \qquad (5.23)$$

Next consider the second term I_2 in the right hand side of (5.22). Since $r_j^*-r_i^*\ge u_n$ for all $j>i$, we therefore have

$$|I_2|\le 2\sum_{1\le i<j\le k}\sum_{j_1=1}^{v_n}\sum_{j_2=1}^{v_n}\left|\mathrm{Cov}(\zeta_{r_i+u_n+j_1},\,\zeta_{r_j+u_n+j_2})\right|\le 2\sum_{j_1=1}^{n-u_n}\sum_{j_2=j_1+u_n}^{n}\left|\mathrm{Cov}(\zeta_{j_1},\,\zeta_{j_2})\right|.$$

By stationarity and Lemma 1, one obtains

$$|I_2|\le 2n\sum_{j=u_n+1}^{n}\left|\mathrm{Cov}(\zeta_1,\,\zeta_j)\right|=o\left(n\,h_1^{-r}\right). \qquad (5.24)$$

Hence by (5.22)-(5.24), we have

$$\frac{h_1^r}{n}E[G_{n,2}]^2=O\left(k_n\;v_n\;n^{-1}\right)+o(1)=o(1). \qquad (5.25)$$

It follows from stationarity, (5.19) and Lemma 1 that

$$\mathrm{Var}(G_{n,3})=\mathrm{Var}\left(\sum_{j=1}^{n-k_n(u_n+v_n)}\zeta_j\right)=O\left((n-k_n(u_n+v_n))h_1^{-r}\right)=o\left(nh_1^{-r}\right). \qquad (5.26)$$

Combining (5.22), (5.25) and (5.26), we establish (5.21a). As for (5.21b), by stationarity, (5.19) and Lemma 1, it is easily seen that

$$\frac{h_1^r}{n}\sum_{j=1}^{k_n}E(\xi_j^2)=\frac{h_1^r\,k_n}{n}E(\xi_1^2)=\frac{k_n\,u_n}{n}\cdot\frac{h_1^r}{u_n}\mathrm{Var}\left(\sum_{j=1}^{u_n}\zeta_j\right)\to v(\underline{x}_1).$$

In order to establish (5.21c), we make use of Lemma 1.1 in Volkonskii and Rozanov (1959) (see also Ibragimov and Linnik (1971, p.338)) to obtain

$$\left| E[\exp(itG_{n,1})] - \prod_{j=1}^{k_n} E[\exp(it\xi_j)] \right| \leq 16\,(n/u_n)\,\alpha(v_n)$$

tending to zero by (5.19).

Finally, we will establish (5.21d). To this end, we now employ a truncation technique as follows. Let $b_J(y) = yI(|y| \leq J)$, where J is a fixed positive number, and $\zeta_{n,j}^J = \zeta_{n,j}^J(\underline{x}_1) = [b_J(Y_j) - m_J(\underline{X}_j)]\,K_{h_1}\left(\underline{X}_j^{(1)} - \underline{x}_1\right)\,\Gamma(\underline{x}^j)$, then, $G_n(\underline{x}_1) = G_n^J(\underline{x}_1) + \widetilde{G}_n^J(\underline{x}_1)$, where

$$G_n^J(\underline{x}_1) = n^{-1}\sum_{j=1}^{n} \zeta_{n,j}^J, \qquad \text{and} \qquad \widetilde{G}_n^J(\underline{x}_1) = n^{-1}\sum_{j=1}^{n}\left[\zeta_j - \zeta_{n,j}^J\right].$$

By using the same arguments as those employed in the proof of Lemma 1, one has, as $n \to \infty$,

$$v_{n,J}(\underline{x}_1) = n\,h_1^r \mathrm{Var}\left(G_n^J\right) \to v_J(\underline{x}_1) = ||K||^2 p_1(\underline{x}_1) E\left\{\sigma_J^2(\underline{X}_1)\Gamma^2(\underline{x}^1)\,\middle|\,\underline{X}_1^{(1)} = \underline{x}_1\right\}.$$

The boundedness of K and Γ implies that $|\zeta_{n,j}^J| \leq B/h_1^r$ for some $B > 0$. This in turn implies that $\sqrt{n\,h_1^r}\ \max_{1\leq j\leq k_n}|\xi_j^J| \leq B\ q_1/\sqrt{n\,h_1^r} \to 0$, by (5.19). Therefore, the set $\left\{|\xi_j^J| > \varepsilon\sqrt{v_J(\underline{x}_1)\ n/h_1^r}\right\}$ is an empty set when n is large sufficiently. Hence, it follows that (5.21d) holds true for ξ_j^J. Consequently, we have established the following asymptotic normality:

$$\sqrt{nh_1^r}\ G_n^J(\underline{x}_1) \quad \xrightarrow{d} \quad N(0, v_J(\underline{x}_1)) \tag{5.27}$$

as $n \to \infty$. Observe that, for any $t \in \Re^1$,

$$\begin{aligned}
&\left| E\exp\left(it\sqrt{nh_1^r}\ G_n\right) - \exp\left(-v(\underline{x}_1)t^2/2\right)\right| \\
&\leq \left| E\exp\left(it\sqrt{nh_1^r}\ (G_n^J + \widetilde{G}_n^J)\right) - \exp\left(-v_J(\underline{x}_1)t^2/2\right)\right| \\
&\quad + \left|\exp\left(-v_J(\underline{x}_1)t^2/2\right) - \exp\left(-v(\underline{x}_1)t^2/2\right)\right| \\
&\leq \left| E\exp\left(it\sqrt{nh_1^r}\ G_n^J\right) - \exp\left(-v_J(\underline{x}_1)t^2/2\right)\right| + E\left|\exp\left(it\sqrt{nh_1^r}\ \widetilde{G}_n^J\right) - 1\right| \\
&\quad + \left|\exp\left(-v_J(\underline{x}_1)t^2/2\right) - \exp\left(-v(\underline{x}_1)t^2/2\right)\right|.
\end{aligned}$$

As $n \to \infty$, the first term goes to 0 by (5.27) for each $J > 0$ and the third term also goes to 0 as $J \to \infty$ by the dominated convergence theorem. In order to complete the proof, it suffices to show that the second term goes to 0 as $n \to \infty$ and then $J \to \infty$. To this end, using the fact that $|e^{ix} - 1| \leq 2|x|$ for all $x \in \Re$ and the Cauchy-Schwartz's inequality, we have

$$E\left|\exp\left(it\sqrt{nh_1^r}\,\widetilde{G}_n^J\right) - 1\right| \leq 2\sqrt{n\,h_1^r\,\mathrm{Var}(\widetilde{G}_n^J)}.$$

Note that $\widetilde{G}_n^J$ has the same structure as G_n except that Y_j is replaced by $Y_j I(|Y_j| > J)$. Then, using the same arguments as those used in the proof of Lemma 1, we obtain, as $n \to \infty$,

$$nh_1^r\,\mathrm{Var}(\widetilde{G}_n^J) \to ||K||^2 p_1(\underline{x}_1) E\left\{\widetilde{\sigma}_J^2(\underline{X}_1)\Gamma^2(\underline{x}^1)\,\middle|\,\underline{X}_1^{(1)} = \underline{x}_1\right\}.$$

It can be easily shown that the right hand side goes to 0 as $J \to \infty$. This completes the proof of the theorem. ■

Proof of Theorem 2:

By (5.17), the direct estimator of each component $\widehat{g}_i^*(u_i)$ has the following stochastic representation

$$\widehat{g}_i^*(u_i) - g_i^*(u_i) = T_{i,n} + \frac{1}{2}\,h_{1i}^2\,g_i''(u_i)\,\mu_2(K) + o_p\left((n\,h_{1i})^{-1/2}\right), \tag{5.28}$$

where

$$T_{i,n} = \frac{1}{n}\sum_{j=1}^{n} K_{h_{1i}}(U_{i,j} - u_i)\,\Gamma_i(\underline{U}_j^i)\,\varepsilon_j, \qquad \text{and} \qquad \Gamma_i(\underline{U}^i) = \frac{W_i(\underline{U}^{-i})\,p^{-i}(\underline{U}^{-i})}{p(\underline{U}^i)}.$$

In order to show (4.4), it suffices to show from (5.28) that

$$\begin{pmatrix} \sqrt{n\,h_{11}}\;T_{1,n} \\ \vdots \\ \sqrt{nh_{1q}}\;T_{q,n} \end{pmatrix} \to N(0,\,\Sigma), \tag{5.29}$$

It suffices to show from Theorem 1 that the asymptotic covariance between $T_{i,n}$ and $T_{l,n}$ should be zero for $i \neq l$. In other words, we will show that

$$\sqrt{n^2\,h_{1i}\,h_{1l}}\;\mathrm{Cov}(T_{i,n},\,T_{l,n}) \to 0. \tag{5.30}$$

To this effect, by stationarity, we have,

$$\begin{aligned}\mathrm{Cov}(T_{i,n},\,T_{l,n}) &= \frac{1}{n}\,E\,\left[K_{h_{1i}}(U_i - u_i)\,K_{h_{1l}}(U_l - u_l)\,\Gamma_i(\underline{U}^i)\Gamma_l(\underline{U}^l)\,\varepsilon^2\right] \\ &\quad + \frac{1}{n^2}\sum_{j_1 \neq j_2}^{n} E\,\left[K_{h_{1i}}(U_{i,j_1} - u_i)\,K_{h_{1l}}(U_{l,j_2} - u_l)\,\Gamma_i(\underline{U}_{j_1}^i)\Gamma_l(\underline{U}_{j_2}^l)\,\varepsilon_{j_1}\,\varepsilon_{j_2}\right] \\ &\equiv F_1 + F_2. \end{aligned} \tag{5.31}$$

It is easily seen by Theorem 1 in Sun (1984) that

$$F_1 = O\left(n^{-1}\right). \tag{5.33}$$

Employing the same arguments as those used in the proof of (A.5), and by stationarity, we have

$$|F_2| \le \frac{1}{n} \sum_{j=2}^{n} \left\{ \left| E\left[K_{h_{1i}}(U_{i,1} - u_i)\, K_{h_{1l}}(U_{l,j} - u_l)\, \Gamma_i(\underline{U}_1^i) \Gamma_l(\underline{U}_j^l)\, \varepsilon_1\, \varepsilon_j \right] \right.\right.$$
$$\left.\left. + E\left[K_{h_{1i}}(U_{i,j} - u_i)\, K_{h_{1l}}(U_{l,1} - u_l)\, \Gamma_i(\underline{U}_j^i) \Gamma_l(\underline{U}_1^l)\, \varepsilon_1\, \varepsilon_j \right] \right| \right\} = O\left(n^{-1}\right).$$

This, in conjunction with (5.31) and (5.32), concludes that (5.30) holds true. Therefore, this completes the proof of the theorem. ∎

APPENDIX

Proof of Lemma 1 It is easily seen by stationarity that

$$n\,\mathrm{Var}\left(G_n(\underline{x}_1)\right) = \mathrm{Var}(\zeta_1) + 2 \sum_{j=2}^{n} \left(1 - \frac{j}{n}\right) \mathrm{Cov}(\zeta_1, \zeta_j) \equiv J_1 + J_2. \tag{A.1}$$

For J_1, since $E(\zeta_j) = 0$, we have

$$h_1^r\, J_1 = h_1^r\, E(\zeta_1^2) = \int K^2(\underline{u}_1) \sigma^2(\underline{x}_1 + h_1 \underline{u}_1, \underline{u}_2) \Gamma^2(\underline{x}_1, \underline{u}_2)\, p(\underline{x}_1 + h_1 \underline{u}_1, \underline{u}_2) d\underline{u}_1 d\underline{u}_2.$$

By Theorem 1 in Sun (1984) and Assumptions **(2)**-**(4)**, we then have

$$h_1^r\, J_1 \to ||K||^2 \int \sigma^2(\underline{x}_1, \underline{u}_2) \Gamma^2(\underline{x}_1, \underline{u}_2)\, p(\underline{x}_1, \underline{u}_2)\, d\underline{u}_2$$
$$= ||K||^2 p_1(\underline{x}_1) E\left[\sigma^2(\underline{X}) \Gamma^2(\underline{X}) \mid \underline{X}^{(1)} = \underline{x}_1 \right] = v(\underline{x}_1). \tag{A.2}$$

It remains to show that $h_1^r\, J_2 \to 0$. To this end, choose a sequence of positive integers satisfying $\pi_n = O\left(h_1^{-\delta\, p/a(2+\delta)}\right)$, where a is given in Assumption **(6)**. Then

$$\pi_n = o\left(h_1^{-r}\right) \qquad \text{and} \qquad h_1^{-\frac{\delta r}{2+\delta}} \sum_{j \ge \pi_n} \alpha^{\frac{\delta}{2+\delta}}(j) \to 0. \tag{A.3}$$

We decompose the sum into three terms due to the possible overlap between $\underline{X}_1$ and $\underline{X}_j$,

$$\sum_{j=2}^{n} |\mathrm{Cov}(\zeta_1, \zeta_j)| = \sum_{j=2}^{d-1} + \sum_{j=d}^{\pi_n} + \sum_{j=\pi_n+1}^{n} \equiv J_{21} + J_{22} + J_{23}, \tag{A.4}$$

For J_{21}, there is an overlap between the components of $\underline{X}_1$ and $\underline{X}_j$ but not in J_{22} or J_{23}. Let us consider J_{22} first. By Assumptions **(1)**-**(5)**, and the Cauchy-Schwartz's inequality, we have

$$
\begin{aligned}
&|\mathrm{Cov}(\zeta_1,\,\zeta_j)| = |E(\zeta_1\,\zeta_j)| \\
&\le M \int \frac{1}{h_1^r} K\left(\frac{\underline{u}_1 - \underline{x}_1}{h_1}\right) \left|\Gamma(\underline{x}_1,\,\underline{u}_2)\right| \frac{1}{h_1^r} K\left(\frac{\underline{v}_1 - \underline{x}_1}{h_1}\right) \left|\Gamma(\underline{x}_1,\,\underline{v}_2)\right| \\
&\qquad \times \left[E(Y_1^2) + |m(\underline{u}) + m(\underline{v})|\, E|Y_1| + |m(\underline{u})m(\underline{v})|\right] d\underline{u}_1 d\underline{u}_2 d\underline{v}_1 d\underline{v}_2 \\
&\le \text{const.} \quad \int \left|W(\underline{u}_2)W(\underline{v}_2)\right| \Big[E(Y_1^2) + |f_2(\underline{u}_2) + f_2(\underline{v}_2)|[E|Y_1| + |f_1(\underline{x}_1)|] \\
&\quad + 2|f_1(\underline{x}_1)|\, E|Y_1| + f_1^2(\underline{x}_1) + |f_1(\underline{x}_1)||f_2(\underline{u}_2) + f_2(\underline{v}_2)| + |f_2(\underline{u}_2) f_2(\underline{v}_2)|\Big] \\
&\quad \times p_2(\underline{u}_2) p_2(\underline{v}_2)\, d\underline{u}_2\, d\underline{v}_2 < \infty. \qquad (A.5)
\end{aligned}
$$

Hence,

$$
h_1^r\,|J_{22}| \le \text{const.} \quad \sum_{j=d}^{\pi_n} h_1^r = O\left(\pi_n h_1^r\right) = o(1) \qquad (A.6)
$$

by the choice of π_n. Next, work with J_{21}. To this end, let $\tilde{r}$ be the number of the common elements in $\left(\underline{X}_1^{(1)},\, \underline{X}_j^{(1)}\right)$. Employing the exactly same arguments as those used in the proof of (6.10) in Masry and Tjøstheim (1997) and (A.5), one can show that

$$
h_1^r\,|J_{21}| \le \text{const.} \quad \sum_{j=2}^{d-1} h_1^{r-\tilde{r}} = O(h_1) = o(1). \qquad (A.7)
$$

For J_{23}, we apply Davydov's inequality (see, e.g., Hall and Heyde (1980), Corollary A.2) to obtain,

$$
|\mathrm{Cov}(\zeta_1,\,\zeta_j)| \le 8\, \alpha^{\frac{\delta}{2+\delta}}(j-1) \left(E|\zeta_1|^{2+\delta}\right)^{\frac{2}{2+\delta}}.
$$

It is easily seen that

$$
\begin{aligned}
h_1^{(1+\delta)r} E|\zeta_1|^{2+\delta} &= \sum_{\underline{u}_3} \int |K(\underline{u}_1)|^{2+\delta} b(\underline{x}_1 + h_1 \underline{u}_1,\, \underline{u}_2) |\Gamma(\underline{x}_1,\, \underline{u}_2)|^{2+\delta}\, p(\underline{x}_1 + h_1\underline{u}_1,\, \underline{u}_2) d\underline{u}_1 d\underline{u}_2 \\
&\to\ ||K||^{2+\delta}\, p_1(\underline{x}_1)\, E\left(b(\underline{X})|\Gamma(\underline{X})|^{2+\delta}\,|\, \underline{X}^{(1)} = \underline{x}_1\right) < \infty
\end{aligned}
$$

as $n \to \infty$. Therefore,

$$
E|\zeta_1|^{2+\delta} \le \text{const.} \quad h_1^{-(1+\delta)r}.
$$

Thus,

$$
|\mathrm{Cov}(\zeta_1,\,\zeta_j)| \le \text{const.} \quad \alpha^{\frac{\delta}{2+\delta}}(j-1) h_1^{-\frac{2(1+\delta)r}{2+\delta}}. \qquad (A.8)
$$

Thus,

$$
h_1^r\,|J_{23}| \le \text{const.} \quad h_1^{-\frac{\delta r}{2+\delta}} \sum_{j \ge \pi_n} \alpha^{\frac{\delta}{2+\delta}}(j) \to 0, \qquad (A.9)
$$

as $n \to \infty$ by (A.3). Thus, by (A.4), (A.6), (A.7) and (A.9), we have

$$h_1^r \sum_{j=2}^{n} |\mathrm{Cov}(\zeta_1,\ \zeta_j)| \to 0.$$

Consequently,

$$h_1^r\,|J_2| \to 0. \tag{A.10}$$

Combining the above expression with (A.1) and (A.2) completes the proof of the lemma. ∎

Proof of Lemma 2 By (5.12) and (5.13),

$$\begin{aligned} G_n^*(\underline{x}_1) &= \frac{1}{n}\sum_{j=1}^{n} K_{h_1}\left(\underline{X}_j^{(1)} - \underline{x}_1\right)\ \left\{\Gamma_n^*\left(\underline{x}^j\right) - \Gamma\left(\underline{x}^j\right)\right\}\ \varepsilon_j \\ &+ \quad \frac{1}{n}\sum_{j=1}^{n} K_{h_1}\left(\underline{X}_j^{(1)} - \underline{x}_1\right)\ \left\{\Gamma_n\left(\underline{x}^j\right) - \Gamma_n^*\left(\underline{x}^j\right)\right\}\ \varepsilon_j \\ &\equiv G_{n,1}^*(\underline{x}_1) + G_{n,2}^*(\underline{x}_1), \end{aligned} \tag{A.11}$$

where Γ_n is defined in (5.13) and

$$\Gamma_n^*(\underline{x}) = E\{\Gamma_n(\underline{x})\} = \int_{\Re^q} L(\underline{u})\ A(\underline{x}_1,\ \underline{x}_2 + h_2\,\underline{u})p_2(\underline{u})\,d\,\underline{u}\,. \tag{A.12}$$

Clearly, as $n \to \infty$,

$$\Gamma_n^*(\underline{x}) \ \to\ \Gamma(\underline{x}). \tag{A.13}$$

Let $\tau_j = \tau_{n,j}(\underline{x}_1) = K_{h_1}\left(\underline{X}_j^{(1)} - \underline{x}_1\right)\ \left\{\Gamma_n^*\left(\underline{x}^j\right) - \Gamma\left(\underline{x}^j\right)\right\}\ \varepsilon_j$. Then,

$$n\,h_1^r\,\mathrm{Var}(G_{n,1}^*(\underline{x}_1)) = h_1^r\,\mathrm{Var}(\tau_1) + 2\,h_1^r\sum_{j=2}^{n}\left(1 - \frac{j}{n}\right)\mathrm{Cov}(\tau_1,\ \tau_j) \equiv F_3 + F_4. \tag{A.14}$$

A simple algebra gives

$$\begin{aligned} F_3 &= h_1^r\,E\left[K_{h_1}^2\left(\underline{X}_1^{(1)} - \underline{x}_1\right)\ \left\{\Gamma_n^*\left(\underline{x}^1\right) - \Gamma\left(\underline{x}^1\right)\right\}^2\ \sigma^2(\underline{X}_1)\right] \\ &= \int K^2(\underline{u}_1)\ \{\Gamma_n^*(\underline{x}_1,\ \underline{u}_2) - \Gamma(\underline{x}_1,\ \underline{u}_2)\}^2 \\ &\quad \times\ p(\underline{x}_1 + h_1\underline{u}_1,\ \underline{u}_2)\,\sigma^2(\underline{x}_1 + h_1\underline{u}_1,\ \underline{u}_2)\,d\underline{u}_1\,d\underline{u}_2 \\ &= o(1) \end{aligned} \tag{A.15}$$

by (A.13). Similar to (A.4), we decompose the sum into three terms due to the possible overlap between $\underline{X}_1$ and $\underline{X}_j$,

$$\sum_{j=2}^{n} |\mathrm{Cov}(\tau_1,\ \tau_j)| = \sum_{j=2}^{d-1} + \sum_{j=d}^{\pi_n} + \sum_{j=\pi_n+1}^{n} \equiv F_{41} + F_{42} + F_{43}, \tag{A.16}$$

For F_{41}, there is an overlap between the components of $\underline{X}_1$ and $\underline{X}_j$ but not in F_{42} or F_{43}. For F_{41}, by the Cauchy-Schwartz's inequality and (A.15), we have

$$h_1^r F_{41} = o(1). \tag{A.17}$$

Following the same lines as those employed in the proof of (A.6) and (A.10), we have

$$h_1^r F_{42} = o(1), \qquad \text{and} \qquad h_1^r F_{43} = o(1).$$

This, in conjunction with (A.14)-(A.17), implies that

$$n\, h_1^r \operatorname{Var}(G_{n,1}^*(\underline{x}_1)) = o(1). \tag{A.18}$$

Next we show that $G_{n,2}^*(\underline{x}_1)$ is negligible. To this end, let F_n denote the empirical distribution of $\left\{\underline{X}_j^{(2)}\right\}_{j=1}^n$, and let F be the distribution of $\underline{X}^{(2)}$. By (5.13) and (A.12), we obtain

$$\Gamma_n(\underline{x}) - \Gamma_n^*(\underline{x}) = \int_{\Re^q} L_{h_2}(\underline{u} - \underline{x}_2)\, A(\underline{x}_1, \underline{u})\, d\{F_n(\underline{u}) - F(\underline{u})\}. \tag{A.19}$$

Let $\widetilde{L}$ be the Fourier transform of L. Substitute

$$L(\underline{u}) = (2\pi)^{-q} \int_{\Re^q} e^{-i\,\underline{\lambda}\cdot\underline{u}}\, \widetilde{L}(\underline{\lambda})\, d\underline{\lambda}$$

into (A.19) to obtain

$$\begin{aligned}
\Gamma_n(\underline{x}) - \Gamma_n^*(\underline{x}) &= (2\pi)^{-q} \int_{\Re^q} \widetilde{L}(\underline{\lambda})\, e^{i\,\underline{\lambda}\cdot\underline{x}_2/h_2}\, d\,\underline{\lambda} \int_{\Re^q} \frac{1}{h_2^q}\, e^{-i\,\underline{\lambda}\cdot\underline{u}/h_2}\, A(\underline{x}_1, \underline{u})\, d\{F_n(\underline{u}) - F(\underline{u})\} \\
&= \int_{\Re^q} \widetilde{L}(\underline{\lambda})\, e^{i\,\underline{\lambda}\cdot\underline{x}_2/h_2}\, I_{12}(\underline{x}_1, \underline{\lambda})\, d\,\underline{\lambda}\,, \qquad (A.20)
\end{aligned}$$

where

$$I_{12}(\underline{x}_1, \underline{\lambda}) = (2\pi)^{-q} \int_{\Re^q} \frac{1}{h_2^q}\, e^{i\,\underline{\lambda}\cdot\underline{u}/h_2}\, A(\underline{x}_1, \underline{u})\, d\{F_n(\underline{u}) - F(\underline{u})\}.$$

Substituting (A.20) into $G_{n,2}^*(\underline{x}_1)$ of (A.11), we have

$$G_{n,2}^*(\underline{x}_1) = \int_{\Re^q} I_{11}(\underline{x}_1, \underline{\lambda})\, I_{12}(\underline{x}_1, \underline{\lambda})\, \widetilde{L}(\underline{\lambda})\, d\,\underline{\lambda}\,, \tag{A.21}$$

where

$$I_{11}(\underline{x}_1, \underline{\lambda}) = \frac{1}{n} \sum_{j=1}^n e^{i\,\underline{\lambda}\cdot\underline{X}_j^{(2)}/h_2}\, K_{h_1}(\underline{X}_j^{(1)} - \underline{x}_1)\, \varepsilon_j.$$

$I_{11}(\underline{x}_1, \underline{\lambda})$ can be analyzed by following the same lines as those employed in the proof of Lemma 1 to obtain

$$\sup_{\underline{\lambda}\in\Re^q} E\left\{\left|I_{11}^2(\underline{x}_1, \underline{\lambda})\right|\right\} = O\left((n\, h_1^r)^{-1}\right). \qquad (A.22)$$

By (6.61) in Masry and Tjøstheim (1997),

$$\sup_{\underline{\lambda}\in\Re^q} E\left\{\left|I_{12}^2(\underline{x}_1, \underline{\lambda})\right|\right\} = O\left((n\, h_2^{2q})^{-1}\right). \qquad (A.23)$$

By the Cauchy-Schwartz inequality, (A.21)-(A.23) and Assumptions **(2)** and **(9)**,

$$\begin{aligned} E\left|G_{n,2}^*(\underline{x}_1)\right| &\le \sup_{\underline{\lambda}\in\Re^q} \left[E\left\{\left|I_{11}^2(\underline{x}_1, \underline{\lambda})\right|\right\}\, E\left\{\left|I_{12}^2(\underline{x}_1, \underline{\lambda})\right|\right\}\right]^{1/2} \int_{\Re^q} \left|\widetilde{L}(\underline{\lambda})\right| d\underline{\lambda} \\ &= O\left((n^2\, h_1^r\, h_2^{2q})^{-1/2}\right) = o\left((n\, h_1^r)^{-1/2}\right). \qquad (A.24) \end{aligned}$$

This completes the proof of the first part of lemma. Finally, as in FHM (1997), by calculation of the first two moments in the manner of the proofs of Lemma 1 and the first part of this lemma, one can show that

$$B_n^*(\underline{x}_1) = o_p\left((n\, h_1^r)^{-1/2}\right).$$

This concludes the lemma. ■

REFERENCES

Auestad, B. and Tjøstheim, D.(1990). Identification of nonlinear time series: First order characterization and order determination. *Biometrika* **77**, 669–687.

Baju, A., Hastie, T.J. and Tibshirani, R.J.(1989). Linear smoothers and additive models (with discussion). *Ann. Statist.* **17**, 453–510.

Cai, Z. and Roussas, G.G.(1992). Uniform strong estimation under α-mixing, with rates. *Statist. Probab. Lett.* **15**, 47–55.

Engle, R.F.(1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50**, 987–1008.

Fan, J.(1993). Local linear regression smoothers and their minimax. *Ann. Statist.* **21**, 196–216.

Fan, J. and Gijbels, I.(1996). *Local Polynomial Modeling and Its Applications.* Chapman and Hall, London.

Fan, J., Härdle, W. and Mammen, E.(1997). Direct estimation of low dimensional components in additive models. *Ann. Statist.*, to appear.

Gorodetskii, V.V.(1977). On the strong mixing property for linear sequences. *Th. Probab. Appl.* **22**, 411–413.

Hall, P. and Heyde, C.C.(1980). *Martingale Limit Theory and its Applications.* Academic Press, New York.

Hargartner, N.W.(1996). Rate optimal estimation of additive regression via the integration method in the presence of many covariate. *Manuscript.*

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

Ibragimov, I.A. and Linnik, Yu.V.(1971). *Independent and Stationary Sequences of Random Variables.* Walters-Noordhoff, Groningen, the Netherlands.

Lewis, P.A.W. and Stevens, J.G.(1991). Nonlinear modeling of time series using multivariate adaptive splines (MARS). *J. Amer. Statist. Assoc.* **86**, 864–877.

Linton, O.B.(1997). Efficient estimation of additive nonparametric regression models. *Biometrika* **84** 469–473.

Linton, O.B. and Nielsen, J.P.(1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82**, 93–100.

Masry, E.(1996). Multivariate local polynomial regression estimation for time series: Uniform strong consistency and rates. *J. Time Series Analy.* **17**, 571–599.

Masry, E. and Fan, J.(1997). Local polynomial estimation of regression functions for mixing processes. *Scand. J. Statist.* **24**, 165–179.

Masry, E. and Tjøstheim, D.(1995). Nonparametric estimation and identification of nonlinear ARCH time series: Strong convergence and asymptotic normality. *Econometric Theory* **11**, 258–289.

Masry, E. and Tjøstheim, D.(1997). Additive nonlinear ARX time series and projection estimates. *Econometric Theory* **13**, 214–252.

Ruppert, D. and Wand, M.P.(1994). Multivariate weighted least squares estimation. *Ann. Statist.* **22**, 1346–1370.

Stone, C.J.(1985). Additive regression and other nonparametric models. *Ann. Statist.* **13**, 685–705.

Stone, C.J.(1994). The use of polynomial splines and their tensor products in multivariate function (with discussion). *Ann. Statist.* **22**, 118-184.

Sun, Z.(1984). Asymptotic unbiased and strong consistency for density function estimator. *Acta Math Sinica* **27**, 769–782.

Tjøstheim, D. and Auestad, B.(1994). Nonparametric identification of nonlinear time series: Projections. *J. Amer. Statist. Assoc.* **89**, 1398–1409.

Tong, H.(1990). *Nonlinear Time Series: A Dynamical System Approach.* Oxford: Oxford University Press.

Tsybakov, A.B.(1986). Robust reconstruction of functions by the local approximation method. *Prob. Inform. Trans.* **22**, 133-146.

Volkonskii, V.A. and Rozanov, Yu.A.(1959). Some limit theorems for random functions. I. *Theory Probab. Appl.* **4**, 178–197.

Withers, C.S.(1981). Conditions for linear processes to be strong mixing. *Z. Wahrsch. verw. Gebiete* **57**, 477–480.