Statistica Sinica **13**(2003), 739-762

A KERNEL-BASED METHOD FOR ESTIMATING ADDITIVE PARTIALLY LINEAR MODELS

Yanqin Fan and Qi Li

Vanderbilt University and Texas A & M University

Abstract: In this paper we propose a kernel-based method for estimating additive partially linear models. Our method makes use of the partially linear model structure at the initial stage when estimating the individual nonparametric components. Monte Carlo simulations show that our proposed estimator performs quite well for moderate sample sizes. In addition, we provide a consistent estimator for the asymptotic variance of the estimator of the parameter in the linear part of the model, where the linear component variable can be discrete or continuous. This facilitates inferential procedures based on our proposed estimator for the finite dimensional parameter. Our result also leads to a simple identification condition for the finite dimensional parameter.

Key words and phrases: Additive partially linear model, finite sample efficiency, identification, kernel smoother, simulations.

1. Introduction

Recently, the marginal integration approach to estimating individual components in additive regression models has attracted much attention among statisticians and econometricians, see Linton and Nielsen (1995), Newey (1994) and Tjostheim and Auestad (1994), among others. To elaborate on the idea of this approach, consider the following additive regression model with two regressors:

$$Y_i = \beta_0 + g_1(Z_{1i}) + g_2(Z_{2i}) + U_i,\ \ i = 1, \ldots, n, \tag{1}$$

where $\{Y_i, Z_{1i}, Z_{2i}\}_{i=1}^n$ are independently and identically distributed (i.i.d.) as $\{Y, Z_1, Z_2\}$, $E(U_i|Z_{1i}, Z_{2i}) = 0$, β_0 is an unknown parameter, $g_1(\cdot)$ and $g_2(\cdot)$ are unknown univariate functions. The individual components $g_1(\cdot)$ and $g_2(\cdot)$ are identified under the condition that $E[g_1(Z_1)] = 0$ and $E[g_2(Z_2)] = 0$.

Stone (1985, 1986) has shown that the additive components $g_\alpha(\cdot)$, $\alpha = 1, 2$, in (1) can be consistently estimated at the same rate as a fully nonparametric regression model with only one regressor. Furthermore, it is shown in Fan, Härdle and Mammen (1998) and Mammen, Linton and Nielsen (1999) that an additive component can be estimated as well as if the other components were known in advance. Hence additive regression models circumvent the 'curse of dimensionality' that afflicts the estimation of fully nonparametric regression models. Early algorithms for calculating the additive components are based on iterative procedures using backfitting, which makes its asymptotic properties difficult to analyze. Only recently has important progress been made in the development of the asymptotic theory of backfitting, see Opsomer and Ruppert (1997), and Mammen, Linton and Nielsen (1999). In contrast, the asymptotic theory of the marginal integration method is relatively easy to establish.

In the marginal integration approach, one estimates $g_\alpha(z_\alpha)$, $\alpha = 1, 2$, by marginally integrating (we use the term 'marginally integrating' interchangeably with 'marginally averaging') a linear smoother of $a(z_1, z_2)$ such as a local linear estimator, where $a(z_1, z_2) = E(Y|Z_1 = z_1, Z_2 = z_2)$. Specifically, let $\hat{a}(z_1, z_2)$ be a nonparametric local linear estimator of $a(z_1, z_2)$. Then an estimator of $[g_1(z_1) + \beta_0]$ can be obtained by integrating $\hat{a}(z_1, z_2)$ over z_2, i.e., $\tilde{m}_1(z_1) = n^{-1} \sum_{j=1}^n \hat{a}(z_1, Z_{2j})$. Using the identification condition, one obtains the estimator of $g_1(z_1)$ by subtracting the sample mean of $\tilde{m}_1(\cdot)$ from $\tilde{m}_1(z_1)$, i.e., $\bar{g}_1(z_1) = \tilde{m}_1(z_1) - n^{-1} \sum_{i=1}^n \tilde{m}_1(Z_{1i})$. An estimator for $g_2(z_2)$ can be similarly obtained. Other nonparametric estimators of $a(z_1, z_2)$ such as the Nadaraya-Watson kernel estimator may also be used. However as shown in Fan (1992, 1993) the local linear approach has a number of advantages over the local linear kernel approach, including design adaptivity, automatic boundary carpentry, and minimax efficiency.

Linton and Nielsen (1995), Newey (1994), Tjostheim and Auestad (1994) and Chen, Härdle, Linton and Severance-Lossin (1995) have shown that the marginal integration estimator $\bar{g}_\alpha(z_\alpha)$, $\alpha = 1, 2$, achieves the one-dimensional optimal convergence rate. Based on the same idea and the local polynomial approach, Severance-Lossin and Sperlich (1996) have constructed estimators of derivatives of individual components in additive regression models, Sperlich, Tjostheim and Yang (2002) have dealt with the additive models with interaction terms, and Yang (2002) has considered additive models with proportional components.

Fan, Härdle and Mammen (1998) extended the idea of marginal integration to weighted marginal integration and showed that, with an appropriate choice of the weight function, the additive components can be efficiently estimated: an additive component can be estimated with the same asymptotic bias and variance as if the other components were known. They also applied the weighted marginal integration approach to more general models including additive partially linear models. In the context of additive partially linear models, when the linear component variable is continuous, the asymptotic variance of their estimator of the finite dimensional parameter in the linear part has a complicated structure for which no explicit expression is given. In this paper we propose an alternative approach to estimating additive partially linear models. We provide an explicit expression for the asymptotic variance of our estimator of the finite dimensional parameter, and also provide a consistent estimator for it. This facilitates inferential procedures based on the proposed estimator for the parameter in the linear part of the model. As will be seen in Section 2, this also leads to a simple identification condition for identifying the finite dimensional parameter of the model and allows one to estimate an additive partially linear model with interaction terms entering the model parametrically.

The paper is organized as follows. In Section 2 we first review the estimation methods of Fan, Härdle and Mammen (1998), then introduce our approach and establish the asymptotic distribution of our proposed estimator. Section 3 reports some Monte Carlo simulation results on the finite sample performance of the proposed estimator.

2. Estimating Additive Partially Linear Models

In this section we consider the additive partially linear model

$$Y_i = \beta_0 + X_i'\beta + g_1(Z_{1i}) + g_2(Z_{2i}) + \ldots + g_p(Z_{pi}) + U_i, \quad (2)$$

where X_i is a $q \times 1$ vector of random variables, $\beta = (\beta_1, \ldots, \beta_q)'$ is a $q \times 1$ vector of unknown parameters, β_0 is a scalar parameter, the $Z_{\alpha i}$'s are univariate continuous variables, and $g_\alpha(\cdot)$, $\alpha = 1, \ldots, p$, are unknown smooth functions. The observations $\{Y_i, X_i', Z_{1i}, \ldots, Z_{pi}\}_{i=1}^n$ are i.i.d. We impose the condition that $E[g_\alpha(Z_\alpha)] = 0$ for $\alpha = 1, \ldots, p$, so that the individual components $g_1(\cdot), \ldots, g_p(\cdot)$ are identified. Model (2) is essentially the same as the additive partially linear model studied in Fan, Härdle and Mammen (1998).

Let $Z_{\underline{\alpha}i} = (Z_{1i}, \ldots, Z_{\alpha-1,i}, Z_{\alpha+1,i}, \ldots, Z_{pi})$, where in $Z_{\alpha i}$ is removed from $(Z_{1i}, Z_{2i}, \ldots, Z_{pi})$. Define $G_{\underline{\alpha}}(z_{\underline{\alpha}}) = g_1(z_1) + \cdots + g_{\alpha-1}(z_{\alpha-1}) + g_{\alpha+1}(z_{\alpha+1}) + \cdots + g_p(z_p)$. Then (2) can be written as

$$Y_i = \beta_0 + X_i'\beta + g_\alpha(Z_{\alpha i}) + G_{\underline{\alpha}}(Z_{\underline{\alpha}i}) + U_i. \quad (3)$$

Fan, Härdle and Mammen (1998) considered the case where X_i is a $q \times 1$ vector of discrete variables and $(Z_{1i}, \ldots, Z_{pi}) = (Z_{\alpha i}, Z_{\underline{\alpha}i})$ is a $1 \times p$ vector of continuous variables. They suggested two ways of estimating model (3). These are briefly summarized below (for notational simplicity, we only provide their estimators without weighting).

(I) The Indicator Method

We use subscript I to denote this estimation method. Let $a(z_\alpha, z_{\underline{\alpha}}, x) = E(Y_i | Z_{\alpha i} = z_\alpha, Z_{\underline{\alpha}i} = z_{\underline{\alpha}}, X_i = x)$. As in Linton and Nielsen (1995), Fan, Härdle and Mammen (1998) proposed to estimate the regression function $a(z_\alpha, z_{\underline{\alpha}}, x)$ first by $\tilde{a}_{I\alpha} = \tilde{a}_{I\alpha}(z_\alpha, z_{\underline{\alpha}}, x)$: the solution in a to the minimization problem

$$\min_{\{a,b\}} \sum_{l=1}^{n} [Y_l - a - (Z_{\alpha l} - z_\alpha)b]^2 K_{h_\alpha}(Z_{\alpha l} - z_\alpha) L_{h_{\underline{\alpha}}}(Z_{\underline{\alpha} l} - z_{\underline{\alpha}}) I(X_l - x), \tag{4}$$

where $K_{h_\alpha}(u) = h_\alpha^{-1} K(u/h_\alpha)$, $L_{h_{\underline{\alpha}}}(u) = h_{\underline{\alpha}}^{-(p-1)} L(u/h_{\underline{\alpha}})$, K: $R \to R$ and L: $R^{p-1} \to R$ are kernel functions, I is the indicator function, h_α is the smoothing parameter for Z_α, and $h_{\underline{\alpha}}$ is the smoothing parameter for $Z_{\underline{\alpha}}$ which excludes Z_α from $(Z_1, Z_2, \ldots, Z_p)$. The solution $\tilde{a}_{I\alpha}(z_\alpha, z_{\underline{\alpha}}, x)$ is a special case of the local linear estimator of $E(Y_i | Z_{\alpha i} = z_\alpha, Z_{\underline{\alpha} i} = z_{\underline{\alpha}}, X_i = x)$. Then $g_\alpha(z_\alpha)$ is estimated by

$$\hat{g}_{I\alpha}(z_\alpha) = \hat{m}_{I\alpha}(z_\alpha) - n^{-1} \sum_{i=1}^{n} \hat{m}_{I\alpha}(Z_{\alpha i}), \tag{5}$$

where $\hat{m}_{I\alpha}(z_\alpha) = n^{-1} \sum_{j=1}^{n} \tilde{a}_{I\alpha}(z_\alpha, Z_{\underline{\alpha} j}, X_j)$.

After obtaining $\hat{g}_{I\alpha}(Z_{\alpha i})$, $\alpha = 1, \ldots, p$, $\sqrt{n}$-consistent estimators of β_0 and β can be obtained by regressing $[Y_i - \sum_{\alpha=1}^{p} \hat{g}_{I\alpha}(Z_{\alpha i})]$ on $(1, X_i')$, i.e., $\begin{pmatrix} \hat{\beta}_{0I} \\ \hat{\beta}_I \end{pmatrix} = (\tilde{X}'\tilde{X})^{-1} \tilde{X}'(\tilde{Y} - \sum_{\alpha=1}^{p} \hat{g}_{I\alpha})$, where $\tilde{X}$ is of dimension $n \times (q+1)$ with the ith row given by $(1, X_i')$, $\tilde{Y} = (Y_1, \ldots, Y_n)'$, and $\hat{g}_{I\alpha}$ is an $n \times 1$ vector with the ith element given by $\hat{g}_{I\alpha}(Z_{\alpha i})$, $\alpha = 1, \ldots, p$. The $\sqrt{n}$ asymptotic normality result of $\hat{\beta}_I$ as well as the asymptotic variance of $\hat{\beta}_I$ are given in Theorem 4 in Fan, Härdle and Mammen (1998).

Fan, Härdle and Mammen (1998) observed that when X_i takes many different values, the 'indicator method' can be difficult to use because, for each fixed value of x, few data points are available for computing $\tilde{a}_{I\alpha}(z_\alpha, z_{\underline{\alpha}}, x)$ using (4).

(II) The Linear Method

We use subscript L to denote this estimation method. The 'linear method' first estimates the nonparametric component of the regression function, i.e, $[\beta_0 + g_\alpha(z_\alpha) + G_{\underline{\alpha}}(z_{\underline{\alpha}})]$, by choosing a, b and β to minimize

$$\sum_{l=1}^{n} [Y_l - a - (Z_{\alpha l} - z_\alpha)b - X_l'\beta]^2 K_{h_\alpha}(Z_{\alpha l} - z_\alpha) L_{h_{\underline{\alpha}}}(Z_{\underline{\alpha} l} - z_{\underline{\alpha}}). \tag{6}$$

Let $\hat{a}_{L\alpha}$, $\hat{b}_{L\alpha}$ and $\hat{\beta}_{L\alpha}$ be the solution to (6). Then $\hat{a}_{L\alpha}(z_\alpha, z_{\underline{\alpha}}) = \hat{a}_{L\alpha}$ is a nonparametric estimator of $[\beta_0 + g_\alpha(z_\alpha) + G_{\underline{\alpha}}(z_{\underline{\alpha}})]$. An estimator of $g_\alpha(z_\alpha)$ can be obtained by first averaging $\hat{a}_{L\alpha}$ over $Z_{\underline{\alpha} j}$ to obtain $\hat{m}_{L\alpha}(z_\alpha) = n^{-1} \sum_{j=1}^{n} \hat{a}_{L\alpha}(z_\alpha, Z_{\underline{\alpha}} j)$, and then subtracting the sample mean of $\hat{m}_{L\alpha}(\cdot)$ from it. Hence, we get $\hat{g}_{L\alpha}(z_\alpha) = \hat{m}_{L\alpha}(z_\alpha) - n^{-1} \sum_{i=1}^{n} \hat{m}_{L\alpha}(Z_{\alpha i})$.

Similar to the 'indicator method', after obtaining $\hat{g}_{L\alpha}(Z_{\alpha i})$ $(\alpha = 1, \ldots, p)$, $\sqrt{n}$-consistent estimators of β_0 and β can be obtained by regressing $[Y_i - \sum_{\alpha=1}^{p} \hat{g}_{I\alpha}(Z_{\alpha i})]$ on $(1, X_i')$, $\begin{pmatrix} \hat{\beta}_{0L} \\ \hat{\beta}_L \end{pmatrix} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'(\tilde{Y} - \sum_{\alpha=1}^{p} \hat{g}_{L\alpha})$, where $\hat{g}_{L\alpha}$ is an $n \times 1$ vector with the ith element given by $\hat{g}_{L\alpha}(Z_{\alpha i})$, $\alpha = 1, \ldots, p$. The asymptotic distribution of $(\hat{\beta}_{0L}, \hat{\beta}_L')'$ is stated in Theorem 6 in Fan, Härdle and Mammen (1998). However the explicit expression of the asymptotic variance is quite complicated and is not given.

Note that, unlike the indicator method, the linear method does not depend on how many different values that X_i can take. However, like the indicator method, the linear method does not make full use of the information that X_i enters the regression function linearly because, in the initial estimation stage, it treats the linear part $(X_i'\beta)$ nonparametrically (locally) in the sense that the solution for β is $\hat{\beta}_{L\alpha} = \hat{\beta}_{L\alpha}(z_\alpha, z_{\underline{\alpha}})$ (see (6)) which depends on z_α and $z_{\underline{\alpha}}$. This motivates us to consider an alternative estimation approach.

(III) An Alternative Approach

Our approach is a two-step estimation procedure which applies to the case where X_i contains both discrete and continuous elements. We call it the two-step method. When estimating the individual components $g_\alpha(z_\alpha)$, our approach takes into account the fact that X_i enters into the model linearly.

In the first step of the method, we estimate β by using an idea similar to that of Robinson (1988) who considered the problem of estimating a semiparametric partially linear model without additive structure. However, in order to obtain weak identification of β and thus allow one to estimate additive partially linear models with interaction terms entering the model parametrically (see Remark 1 below), we exploit the additive structure of the nonparametric component in (2).

Define $\xi(z_\alpha, z_{\underline{\alpha}}) = E(Y_i | Z_{\alpha i} = z_\alpha, Z_{\underline{\alpha} i} = z_{\underline{\alpha}})$, $\xi_\alpha(z_\alpha) = E[\xi(z_\alpha, Z_{\underline{\alpha} i})]$, $\eta(z_\alpha, z_{\underline{\alpha}}) = E(X_i | Z_{\alpha i} = z_\alpha, Z_{\underline{\alpha} i} = z_{\underline{\alpha}})$, $\eta_\alpha(z_\alpha) = E[\eta(z_\alpha, Z_{\underline{\alpha} i})]$, $\alpha = 1, \ldots, p$. Denote $\xi_{\alpha i} = \xi_\alpha(Z_{\alpha i})$ and $\eta_{\alpha i} = \eta_\alpha(Z_{\alpha i})$, $\alpha = 1, \ldots, p$. Then, applying the linear operator $E[\cdot | Z_{\alpha i} = z_\alpha, Z_{\underline{\alpha} i} = z_{\underline{\alpha}}]$ to both sides of (3), one gets

$$\xi(z_\alpha, z_{\underline{\alpha}}) = \beta_0 + [\eta(z_\alpha, z_{\underline{\alpha}})]'\beta + g_\alpha(z_\alpha) + G_{z_{\underline{\alpha}}}(z_{\underline{\alpha}}). \tag{7}$$

Integrating both sides of (7) over $z_{\underline{\alpha}}$ leads to

$$\xi_\alpha(z_\alpha) = \beta_0 + [\eta_\alpha(z_\alpha)]'\beta + g_\alpha(z_\alpha), \tag{8}$$

where we used the identification condition $E[G_{z_{\underline{\alpha}}}(Z_{\underline{\alpha}})] = 0$. Replacing z_α in (8) by $Z_{\alpha i}$ and then summing both sides of (8) gives

$$\sum_{s=1}^{p} \xi_{\alpha i} = p\beta_0 + \sum_{\alpha=1}^{p} \eta_{\alpha i}'\beta + \sum_{\alpha=1}^{p} g_\alpha(Z_{\alpha i}). \tag{9}$$

We now subtract (9) from (2) to eliminate $\sum_{\alpha=1}^{p} g_\alpha(Z_{\alpha i})$ which yields

$$Y_i - \sum_{\alpha=1}^{p} \xi_{\alpha i} = (1-p)\beta_0 + (X_i - \sum_{\alpha=1}^{p} \eta_{\alpha i})'\beta + U_i. \tag{10}$$

Let $\mathcal{Y}_i = Y_i - \sum_{\alpha=1}^{p} \xi_{\alpha i}$ and $\mathcal{X}_i' = (1, (X_i - \sum_{\alpha=1}^{p} \eta_{\alpha i})')$. Then in vector notation, (10) can be written as

$$\mathcal{Y} = \mathcal{X}\delta + U, \tag{11}$$

where $\mathcal{Y}$ and $\mathcal{X}$ are $n \times 1$ and $n \times (q+1)$ matrices with ith rows given by $\mathcal{Y}_i$ and $\mathcal{X}_i'$, respectively, $U = (U_1, \ldots, U_n)'$ and $\delta = (\alpha_0, \beta')'$ with $\alpha_0 = (1-p)\beta_0$. Applying ordinary least squares (OLS) to (11) leads to

$$\bar{\delta} \equiv \begin{pmatrix} \bar{\alpha}_0 \\ \bar{\beta} \end{pmatrix} = (\mathcal{X}'\mathcal{X})^{-1}\mathcal{X}'\mathcal{Y} = \delta + (\mathcal{X}'\mathcal{X})^{-1}\mathcal{X}'U. \tag{12}$$

Using standard arguments (The Law of Large Numbers and the Central Limit Theorem), one can easily show that $\bar{\delta}$ is a $\sqrt{n}$-consistent estimator of δ.

The above estimator $\bar{\delta}$ is infeasible because $\sum_{\alpha=1}^{p} \xi_{\alpha i}$ and $\sum_{\alpha=1}^{p} \eta_{\alpha i}$ are unknown. However, they can be consistently estimated. A consistent estimator of $\xi_{\alpha i} = \xi_\alpha(Z_{\alpha i})$ is given by

$$\begin{aligned} \hat{Y}_{\alpha i} &= \frac{1}{n}\sum_{j=1}^{n}\{\frac{\sum_{l\neq j}^{n} Y_l K_{h_\alpha}(Z_{\alpha l} - Z_{\alpha i}) L_{h_{\underline{\alpha}}}(Z_{\underline{\alpha} l} - Z_{\underline{\alpha} j})}{\sum_{s\neq j}^{n} K_{h_\alpha}(Z_{\alpha s} - Z_{\alpha i}) L_{h_{\underline{\alpha}}}(Z_{\underline{\alpha} s} - Z_{\underline{\alpha} j})}\} \\ &= \sum_{l=1}^{n} W_{\alpha i,l} Y_l, \end{aligned} \tag{13}$$

$$W_{\alpha i,l} = \frac{1}{n}\sum_{j\neq l}^{n} \frac{K_{h_\alpha}(Z_{\alpha l} - Z_{\alpha i}) L_{h_{\underline{\alpha}}}(Z_{\underline{\alpha} l} - Z_{\underline{\alpha} j})}{\sum_{s\neq j}^{n} K_{h_\alpha}(Z_{\alpha s} - Z_{\alpha i}) L_{h_{\underline{\alpha}}}(Z_{\underline{\alpha} s} - Z_{\underline{\alpha} j})}, \tag{14}$$

with $K_{h_\alpha}(v) = h_\alpha^{-1} K(v/h_\alpha)$ and $L_{h_{\underline{\alpha}}}(v) = h_{\underline{\alpha}}^{-(p-1)} L(v/h_{\underline{\alpha}})$, K and L are kernel functions, h_α and $h_{\underline{\alpha}}$ are smoothing parameters. We see from (13) that $\hat{Y}_{\alpha i}$ is a weighted average of Y_l's (note that $\sum_{l=1}^{n} W_{\alpha i,l} = 1$). The reason for using $\sum_{l\neq j}^{n}$ and $\sum_{s\neq j}^{n}$ (leave-one-out) instead of $\sum_{l=1}^{n}$ and $\sum_{s=1}^{n}$ in (13) is purely for simplicity of proofs. The results of the paper do not change if one uses $\sum_{l=1}^{n}$ and $\sum_{s=1}^{n}$ in (13) rather than $\sum_{l\neq j}^{n}$ and $\sum_{s\neq j}^{n}$.

Similarly a consistent estimator of $\eta_{\alpha i}$ is given by

$$\hat{X}_{\alpha i} = \sum_{l=1}^{n} W_{\alpha i,l} X_l. \tag{15}$$

A feasible estimator of δ, $\hat{\delta}$ (say) is obtained from (12) by replacing $\mathcal{Y}_i$ and $\mathcal{X}_i$ by $\hat{\mathcal{Y}}_i = Y_i - \sum_{\alpha=1}^{p} \hat{Y}_{\alpha i}$ and $\hat{\mathcal{X}}_i = (1, (X_i - \sum_{\alpha=1}^{p} \hat{X}_{\alpha i})')'$, respectively. However, care must be taken to handle the observations near the boundary of the support of z. At the boundary $\hat{f}(z)$ is not a consistent estimator of $f(z)$ because the bias term will not go to zero even as $n \to \infty$. Therefore we need to trim observations near the boundary. We assume that $z_\alpha \in [c_\alpha, d_\alpha]$, where $c_\alpha < d_\alpha$ are both finite constants, $\alpha = 1, \ldots, p$. Let $\mathcal{D}_n = \prod_{\alpha=1}^{p} [c_\alpha + a_n, d_\alpha - a_n]$ be the trimming set, where $a_n = ch^\epsilon$ for some $c > 0$, $0 < \epsilon < 1$, $h = max\{h_\alpha, h_{\underline{\alpha}}\}$. Let $I_i = I(Z_i \in \mathcal{D}_n)$. We estimate δ by

$$\hat{\delta} = \left[\sum_i \hat{\mathcal{X}}'\hat{\mathcal{X}} I_i\right]^{-1} \sum_i \hat{\mathcal{X}}'\hat{\mathcal{Y}} I_i \equiv \begin{pmatrix} \hat{\alpha}_0 \\ \hat{\beta} \end{pmatrix} = (\hat{\mathcal{X}}'\hat{\mathcal{X}})^{-1}\hat{\mathcal{X}}'\hat{\mathcal{Y}}, \tag{16}$$

where $\hat{\mathcal{Y}}$ and $\hat{\mathcal{X}}$ are of dimensions $n \times 1$ and $n \times (q+1)$ with their ith rows given by $I_i\hat{\mathcal{Y}}_i$ and $I_i\hat{\mathcal{X}}_i'$, respectively. Note that the use of a trimming set is more of theoretical importance. In practice, trimming may or may not be important.

The following theorem shows that $\hat{\beta}$ is a $\sqrt{n}$-consistent estimator of β and presents its asymptotic distribution.

Theorem 2.1. *Under condition A given in the appendix A, we have* $\sqrt{n}(\hat{\beta} - \beta) \to N(0, \Sigma)$ *in distribution, provided* $\Phi \stackrel{def}{=} \mathrm{Var}\,[X_1 - \sum_{\alpha=1}^{p} \eta_\alpha(Z_{\alpha 1})]$ *is positive definite, where* $\Sigma = \Phi^{-1}\Omega\Phi^{-1}$, $\Omega = E[U_1^2 D_1 D_1']$ *with* $D_1 = V_1 + (\epsilon_1 - E(\epsilon_1))(1 - \sum_{\alpha=1}^{p} \psi_{\alpha i})$, $V_1 = X_1 - E(X_1|Z_1)$, $\epsilon_1 = \eta(Z_1) - \sum_{\alpha=1}^{p} \eta_\alpha(Z_{\alpha 1})$, *and* $\psi_{\alpha 1} = f_\alpha(Z_{\alpha 1}) f_{\underline{\alpha}}(Z_{\underline{\alpha} 1}) / f(Z_{\alpha 1}, Z_{\underline{\alpha} 1})$. *Here* $f_\alpha(\cdot)$, $f_{\underline{\alpha}}(\cdot)$ *and* $f(\cdot, \cdot)$ *are the density functions of* Z_α, $Z_{\underline{\alpha}}$ *and* $(Z_\alpha, Z_{\underline{\alpha}})$, *respectively.*

A consistent estimator of Σ is given in the Appendix A, as is the proof of Theorem 2.1.

Remark 1. That Φ is positive definite is an identification condition for β. It allows X_i to be a deterministic function of $(Z_{1i}, \ldots, Z_{pi})$ provided it is not an additive function. More specifically, consider a simple case of $p = 2$ with $X_i = Z_{1i}Z_{2i}$. Then (2) becomes

$$Y_i = \beta_0 + (Z_{1i}Z_{2i})\beta + g_1(Z_{1i}) + g_2(Z_{2i}) + U_i. \tag{17}$$

Model (17) does not suffer from the "curse of dimensionality" problem since it only involves one-dimensional nonparametric functions $g_\alpha(.)$, $\alpha = 1, 2$. Also, it is more general than an additive model that does not have any interaction terms. Thus, in general, (2) allows interaction terms to enter the model parametrically.

Remark 2. If $\eta(Z_i)$ ($\equiv E(X_i|Z_i)$) is also an additive function in its arguments, then $\eta(Z_i) = \sum_{\alpha=1}^{p} \eta_\alpha(Z_{\alpha i})$. Consequently $\epsilon_i \equiv 0$, and Ω simplifies to $\Omega =$

$E\{U_i^2 V_i V_i'\}$. This includes the case that X_i and Z_i are independent of each other.

The second step of the two-step method is to estimate the individual nonparametric components $g_\alpha(z_\alpha)$. Given the $\sqrt{n}$-consistent estimator $\hat{\beta}$, one can rewrite (2) as

$$Y_i - X_i'\hat{\beta} = \beta_0 + \sum_{\alpha=1}^{p} g_\alpha(Z_{\alpha i}) + U_i + X_i'(\beta - \hat{\beta}) = \beta_0 + \sum_{\alpha=1}^{p} g_\alpha(Z_{\alpha i}) + \text{ error }. \quad (18)$$

The intercept term β_0 can be $\sqrt{n}$-consistently estimated by $\hat{\beta}_0 = \bar{Y} - \bar{X}'\hat{\beta}$, where $\bar{Y} = n^{-1}\sum_{i=1}^n Y_i$ and $\bar{X} = n^{-1}\sum_{i=1}^n X_i$.

Note that (18) is essentially an additive regression model with $(Y_i - X_i'\hat{\beta})$ as the new dependent variable and $[U_i + X_i'(\beta - \hat{\beta})]$ as the new (composite) error. Note further that $\hat{\beta}$ is a $\sqrt{n}$-consistent estimator of β, a faster rate of convergence than that of any nonparametric estimator. Therefore the asymptotic distribution of *any* nonparametric estimator of $g_\alpha(z_\alpha)$ based on (18) will remain the same if one replaces $\hat{\beta}$ by β.

One simple way to obtain a consistent estimator of $g_\alpha(z_\alpha)$ is to replace β in (8) by $\hat{\beta}$, i.e., one can estimate $g_\alpha(z_\alpha)$ up to a constant by $\tilde{g}_{LC,\alpha}(z_\alpha) \equiv \hat{Y}_{\alpha,z_\alpha} - \hat{X}'_{\alpha,z_\alpha}\hat{\beta}$, where $\hat{Y}_{\alpha z_\alpha}$ and $\hat{X}_{\alpha,z_\alpha}$ are the estimators of $\xi_\alpha(z_\alpha)$ and $\eta_\alpha(z_\alpha)$ defined in (13) and (15) with z_α in place of $Z_{\alpha i}$. By subtracting the sample mean from it, we obtain a local constant estimator of $g_\alpha(z_\alpha)$ that satisfies the identification condition: $\hat{g}_{LC,\alpha}(z_\alpha) = \tilde{g}_{LC,\alpha}(z_\alpha) - n^{-1}\sum_i \tilde{g}_{LC,\alpha}(Z_{\alpha i})$. Since this is a marginal-integration local constant fitting, it can be easily shown that the resulting estimator has an asymptotic normal distribution (see Chen et al. (1995)).

One can also estimate $g_\alpha(z_\alpha)$ by weighted marginal integration coupled with the local linear method based on (18). Let $\hat{g}_{LL,\alpha}(z_\alpha)$ denote the resulting estimator of $g_\alpha(z_\alpha)$. Then Theorem 3 of Fan, Härdle and Mammen (1998) gives the asymptotic distribution of $\hat{g}_{LL,\alpha}(z_\alpha)$. As shown in Fan, Härdle and Mammen (1998), with an appropriate choice of the weight function, $\hat{g}_{LL,\alpha}(z_\alpha)$ efficiently estimates $g_\alpha(z_\alpha)$.

Finally, one can estimate $g_\alpha(z_\alpha)$ by the backfitting projection algorithm based on (18) and the asymptotic distribution of the resulting estimator will be the same as in Mammen, Linton and Nielsen (1999). For details of the estimation, consult that source. Linton (1997) and Mammen, Linton and Nielsen (1999) have shown that the backfitting method can lead to efficient estimation of $g_\alpha(z_\alpha)$ in the sense that the asymptotic variance of the resulting estimator is the same as when other components were known. Thus, if one uses the backfitting method to estimate $g_\alpha(z_\alpha)$ based on (18), the corresponding estimator will be efficient.

As emphasized before, the estimator $\hat{\beta}$ makes use of the additive structure of the nonparametric component of the model. Alternatively, as suggested by an anonymous referee, one can ignore this additive structure and estimate β by the method of Robinson (1988). Let $\hat{\beta}_R$ denote this estimator. Then one can use $Y_i - X_i'\hat{\beta}_R$ as the new dependent variable to estimate $g_\alpha(\cdot)$ by following the same steps as discussed earlier except that $\hat{\beta}$ is replaced by $\hat{\beta}_R$. We use $\hat{g}_{R\alpha}(\cdot)$ to denote the corresponding estimator of $g_\alpha(\cdot)$.

One potential problem with $\hat{\beta}_R$ is that it does not allow X_i to be a deterministic function of Z_i in which case $E(X_i|Z_i) = X_i$. To see this, consider (17). Clearly β is not identified if one ignores the additive structure of $g(Z_{1i}, Z_{2i}) = g_1(Z_{1i}) + g_2(Z_{2i})$. Then Robinson's method is not applicable.

Semiparametric efficiency analysis

Chamberlain (1992) has derived the semiparametric efficiency bound of β for model (2). Let $E_A(X_i) \equiv E_A(X_i|Z_i)$ denote the projection of X_i on the additive functional space (conditional on Z_i). Here the projection is defined in the mean square error sense, i.e., $E\{[X_i - E_A(X_i)][X_i - E_A(X_i)]'\} = \inf_{\{g(z)=\sum_{\alpha=1}^{p}\xi_\alpha(z_\alpha)\}} E\{[X_i - g(Z_i)][X_i - g(Z_i)]'\}$. Chamberlain (1992, p.579) has shown that the semiparametric efficiency bound for the inverse of the asymptotic variance of an estimator of β is

$$J = E\left\{[X_i - E_A(X_i)]\left[\mathrm{Var}\left(U_i|X_i, Z_i\right)\right]^{-1}[X_i - E_A(X_i)]'\right\}. \tag{19}$$

Comparing J with the inverse of our asymptotic variance Σ, we see that Σ^{-1} differs from the semiparametric efficiency bound stated in (19). However, when the error is conditionally homoskedastic *and* $E(X_i|Z_i)$ is an additive function in $Z_{\alpha i}$, Σ^{-1} coincides with J, i.e., when $E(U_i^2|X_i, Z_i) = \sigma_u^2$ and $E(X_i|Z_i) = E_A(X_i)$. It is easy to see that in this case $\epsilon_i = 0$, and consequently we have $\Sigma^{-1} = E\{[X_i - E(X_i|Z_i)][X_i - E(X_i|Z_i)]'\}/\sigma_u^2 = J$.

When the error is conditionally homoskedastic, we conjecture that if one replaces the marginal integration method by the backfitting projection algorithm of Mammen, Linton and Nielsen (1999) in the first step, the resulting estimator of β will be semiparametrically efficient. The intuition for this conjecture is that jointly estimating the finite-dimensional parameter and the infinite-dimensional unknown (additive) functions simultaneously will usually lead to efficient estimation of the finite-dimenstional parameter when the error is conditionally homoskedastic (e.g., Carroll, Fan, Gijbels and Wang (1997), Severini and Wong (1993) and Speckman (1988)).

Using spline and series estimation method, Schick (1996) and Li (2000) have shown that one can obtain efficient estimators of β when the error is conditionally homoskedastic. However, it is known that kernel-based method often out-performs spline and series methods in out-of-sample forecasting with time series data (like forecasting inflation rate). Therefore, a kernel-based method of estimating a partially linear additive model should be viewed as a complement to the series estimation method, and should be useful to applied researchers.

3. Monte Carlo Results

This section reports some simulation results on the finite sample performances of the four estimation methods: the 'indicator' method, the 'linear' method, Robinson's method and our proposed two-step method. We first use the following data generating processes (DGP). A similar DGP was used by Fan, Härdle and Mammen (1998).

$$(DGP1): \ Y_i = \beta_0 + X_{1i}\beta_1 + g_1(Z_{1i}) + g_2(Z_{2i}) + U_i,$$
$$(DGP2): \ Y_i = \beta_0 + X_{1i}\beta_1 + X_{2i}\beta_2 + g_1(Z_{1i}) + g_2(Z_{2i}) + U_i,$$

where X_1 is a 0-1 dummy variable with $P(X_1 = 1) = 0.5$, X_2 is uniform on $\{1, 2, 3, 4\}$, Z_1 and Z_2 are bivariate normal variables each having zero mean and unit variance and $cov(Z_1, Z_2) = 0.4$, the two additive functions are $g_1(z_1) = 1 + z_1 - z_1^2$ and $g_2(z_2) = 0.5z_2 + sin(-z_2)$, $U_i = \sigma(Z_{1i}, Z_{2i})\epsilon_i$ with $\sigma(z_1, z_2) = \{(1 + \sqrt{z_1^2 + z_2^2})/4\}^{1/2}$, and ϵ_i a standard normal random variable independent of $X_{\alpha j}$ and $Z_{\alpha j}$ for all i, j and α. The parameters are $(\beta_0, \beta_1, \beta_2) = (1.5, 0.3, 0.5)$. We use the product normal kernel and the smoothing parameters $h_\alpha = cz_{\alpha,sd}n^{-1/3}$, where $z_{\alpha,sd}$ is the sample standard deviation of $\{Z_{\alpha i}\}_{i=1}^n$, α, $s = 1, 2$. For $c = 0.8$, 1 and 1.2 the results are quantitatively similar, so we only report the case of $c = 1$. We take $n = 100$ and 200 and the number of replications is 1000.

The computational time of estimating an additive model is roughly n times the computational time of estimating a (non-additive) nonparametric regression model. This is because in estimating an additive model, say for $p = 2$, one needs to estimate $E(Y_i|Z_{1i}, Z_{2j})$ for each i and j, $1 \leq i, j \leq n$. This usually involves two loops, while in estimating a standard nonparametric regression model one only needs to estimate $E(Y_i|Z_{1i}, Z_{2i})$ for $i = 1, \ldots, n$, which is done in one loop.

Note that the difference between DGP1 and DGP2 is that DGP2 has an extra discrete regressor X_{2i} that takes four different values. Hence, one would expect that the performance of the 'indicator method' is affected more for DGP2 than for DGP1. On the other hand, our approach is expected to work equally well for both DGP1 and DGP2.

Tables 1 and 2 report the estimated mean average square errors (MASE) of $g_{\alpha,n}(\cdot)$ and estimated mean square errors (MSE) of $\beta_{s,n}$, $\alpha = 1, 2$, where $g_{\alpha,n}(\cdot)$ is an estimator of $g_\alpha(\cdot)$ (either $\hat{g}_{I\alpha}(\cdot)$ of the indicator method, $\hat{g}_{L\alpha}(\cdot)$ of the linear method, $\hat{g}_{R\alpha}(\cdot)$ of Robinson's method, or $\hat{g}_\alpha(\cdot)$ of the two-step method based on a (un-weighted) local linear fitting), and $\beta_{s,n}$ is either the sth component of $\hat{\beta}_I$, $\hat{\beta}_L$, $\hat{\beta}_R$, or $\hat{\beta}$, $s = 1, 2$.

Table 1. MASE of $g_{\alpha,n}$ and MSE of $\beta_{s,n}$ ($n = 100$).

Est. Method	MASE($g_{1,n}$)	MASE($g_{2,n}$)	MSE($\beta_{1,n}$)	MSE($\beta_{2,n}$)
DGP 1				
Indicator	0.17423	0.10208	0.03626	N|A
Linear	0.21767	0.10921	0.03349	
Robinson	0.08798	0.08412	0.02906	
Two-Step	0.08591	0.08240	0.02894	
DGP 2				
Indicator	0.38996	0.24830	0.00982	0.04940
Linear	0.40474	0.24075	0.00983	0.04401
Robinson	0.09806	0.08793	0.00696	0.03345
Two-Step	0.08871	0.08207	0.00685	0.03214

Table 2. MASE of $g_{\alpha,n}$ and MSE of $\beta_{s,n}$ ($n = 200$).

Est. Method	MASE($g_{1,n}$)	MASE($g_{2,n}$)	MSE($\beta_{1,n}$)	MSE($\beta_{2,n}$)
DGP 1				
Indicator	0.14686	0.06643	0.01537	N|A
Linear	0.19840	0.07180	0.01552	
Robinson	0.08403	0.05607	0.01422	
Two-Step	0.08322	0.05588	0.01387	
DGP 2				
Indicator	0.33060	0.15634	0.00396	0.02051
Linear	0.32984	0.14463	0.00391	0.02114
Robinson	0.08019	0.05543	0.00316	0.01467
Two-Step	0.07912	0.05491	0.00299	0.01455

From Tables 1 and 2, we observe that the two-step and Robinson's estimators of $g_\alpha(\cdot)$ have significantly smaller MASE than those obtained from the indicator and linear methods. Although the MSE of $\hat{\beta}$ and $\hat{\beta}_R$ are also smaller than those obtained from the other two methods, the relative difference is less significant than the relative difference between $\hat{g}_\alpha(\cdot)$ and $\hat{g}_{I\alpha}(\cdot)$ (or $\hat{g}_{L\alpha}(\cdot)$).

The above results can be explained as follows. For DGP1, the indicator method only uses about half of the sample when estimating $E(Y_i|Z_{1i}, Z_{2i}, X_i)$ in the initial stage because $X_i = X_{1i}$ takes two different values. Hence, at this stage, the indicator method gives a less efficient estimator of $E(Y_i|Z_{1i}, Z_{2i}, X_i)$ compared with the two-step method. Using the method of marginal integration (averaging) to estimate $g_\alpha(\cdot)$ may pick up some, but not all, of the initial-stage efficiency loss of the indicator method in finite sample applications. When we move from DGP1 to DGP2, the performance of the indicator method deteriorates because in DGP2, X_i has two components and takes $(2)(4) = 8$ different values. The estimator $\hat{g}_{L\alpha}(z_\alpha)$ is obtained by treating $X_i'\beta$ nonparametrically in the initial estimation stage, and hence does not perform as well as $\hat{g}_\alpha(\cdot)$ or $\hat{g}_{R\alpha}(\cdot)$ in finite samples. In contrast the two-step and Robinson's estimators of $g_\alpha(z_\alpha)$ are obtained by treating $X_i'\beta$ parametrically. Their performance in finite samples is basically not affected by how many different values X_i can take, and is in general better than the estimators of $g_\alpha(z_\alpha)$ based on the other two methods.

The MSE of $\beta_{s,n}$, $s = 1, 2$, is less affected by the accuracy of the point-wise estimate $g_{\alpha,n}(\cdot)$, $\alpha = 1, 2$, because $\beta_{s,n}$ depends on the average of $g_{\alpha,n}(Z_{\alpha i})$, $i = 1, \ldots, n$. The average values of $g_{\alpha,n}(Z_{\alpha i})$'s do not differ as much as the point-wise estimates among the four estimation methods.

A referee points out that, when estimating $g_\alpha(z_\alpha)$, the better performance of the two-step method over the linear methods is because the former fixed β at $\hat{\beta}$, given that $\hat{\beta}_L$ is close to $\hat{\beta}$, will imply that a further step based on Fan, Härdle and Mammen (1998) can also achieve a similar performance as the two-step method, i.e., if one replaces β in (6) by $\hat{\beta}_L$ (rather then by $\hat{\beta}$), the resulting estimates of $g_\alpha(z_\alpha)$ will be similar to the two-step estimation results. This observation is indeed correct. For example, for DGP2 with $n = 100$, an extra iteration on the linear method gives $MASE(g_{1,n}) = 0.09876$ and $MASE(g_{2,n}) = 0.08581$, which is similar to estimation result of Robinson's (non-iterative) method. Thus, in practice one can also use a further iteration of the linear method to obtain improved estimates of $g_\alpha(z_\alpha)$. One remaining problem is that the asymptotic variance of $\hat{\beta}_L$ is not yet available.

The Monte Carlo results reported above show that the two-step and Robinson's estimators perform relatively well. However, it provides no indication that $\hat{\beta}_R$ of Robinson's method is dominated by $\hat{\beta}$ of the two-step method. The reason for this is that, for DGP1 and DGP2, we generated X and Z independently, which implies that $E(X|Z) = E_A(X) = E(X) =$ a constant (vector). Hence $\hat{\beta}$ and $\hat{\beta}_R$ have the same asymptotic variance (see Remark 2 of Section 2). Next we consider a DGP for which $\hat{\beta}$ is more efficient that $\hat{\beta}_R$. Consider

$$(DGP3): \qquad Y_i = \beta_0 + X_{1i}\beta_1 + g_1(Z_{1i}) + g_2(Z_{2i}) + U_i,$$

where $X_i = a_0 Z_{1i} Z_{2i} + V_i$, Z_{1i}, Z_{2i} and V_i are i.i.d. $N(0,1)$, U_i is i.i.d. $N(0,\sigma^2)$. We choose $\sigma = 0.8$, $a_0 = 1$ or 2, β_0, β_1, $g_1(\cdot)$ and $g_2(\cdot)$ are the same as in DGP1. For DGP3, it is easy to see that $\psi_{\alpha i} = 1$, $\eta_{\alpha i} = 0$ for $\alpha = 1, 2$. Hence, $avar(\sqrt{n}\hat{\beta}) = \sigma^2[E(X_i^2)]^{-1} \equiv \sigma^2\{E[(a_0 Z_{1i} Z_{2i} + V_i)^2]\}^{-1}$, while $avar(\sqrt{n}\hat{\beta}_R) = \sigma^2[E(V_i^2)]^{-1}$, since $X_i - E(X_i|Z_i) = V_i$. Hence $\hat{\beta}$ is asymptotically more efficient than $\hat{\beta}_R$, and we expect to observe some finite sample efficiency gains by using $\hat{\beta}$ over $\hat{\beta}_R$. Note that for DGP3, the indicator method is not appropriate since

X_i is a continuous variable. Hence, we only use the remaining three estimation methods. The simulation results are given in Table 3.

The results in Table 3 are consistent with our theoretical analysis: both the $\hat{\beta}$ of the two-step method and $\hat{\beta}_L$ of the linear method have much smaller MSEs than Robinson's $\hat{\beta}_R$.

Table 3. MASE of $g_{\alpha,n}$ and MSE of $\beta_{s,n}$ (DGP3, $n = 100$).

Est. Method	MASE($g_{1,n}$)	MASE($g_{2,n}$)	MSE($\beta_{1,n}$)
$a_0 = 1$			
Linear	0.13540	0.12058	0.00850
Robinson	0.05328	0.05739	0.01194
Two-Step	0.05300	0.05747	0.00560
$a_0 = 2$			
Linear	0.14006	0.21406	0.00470
Robinson	0.05458	0.05839	0.01131
Two-Step	0.05362	0.05803	0.00263

In summary, our Monte Carlo simulation results are consistent with our theoretical analysis. The proposed two-step method exploits the information that X_i enters the model linearly, and Z_{1i} and Z_{2i} enter the model additively, and thus its overall performance is better than that of other methods. Robinson's method performs well when X_i and Z_i are independent with each other. However, when X_i and Z_i are correlated, Robinson's method can lead to inaccurate estimation of β. Moreover, when X_i is a deterministic non-additive function of Z_i (like $X_i = Z_{1i}Z_{2i}$, see (17)), Robinson's estimator $\hat{\beta}_R$ is not applicable while all other methods are still well defined and lead to consistent estimates.

Acknowledgement

We thank two referees, an associate editor and the editor for their insightful comments that greatly improved the paper. Li's research is supported by the Bush Program in Economics of Public Policy, and the Private Enterprise Research Center, Texas A&M University.

Appendix A

In appendices A and B we omit the indicator function $I_i = I(Z_i \in \mathcal{D}_n)$ for notational simplicity. Thus, $\sum_i A_i$ should be interpreted as $\sum_i A_i I_i$. The use of I_i avoids boundary observations so that the bias in kernel estimation will be of the order of $O(h_\alpha^\nu)$ or $O(h_{\underline{\alpha}}^\nu)$ for all $Z_i \in \mathcal{D}_n$.

We use the class of functions $\mathcal{G}_\nu^\alpha$, $\nu > 0$ and $\alpha > 0$, introduced in Robinson (1988), restate here for the readers' convenience.

Definition. Let $z \in R^d$ be a continuous variable, $\mathcal{G}_\nu^\alpha$, $\nu \geq 1$ a positive integer, $\alpha > 0$, is the class of functions g: $R^d \to R$ satisfying: g is $(\nu-1)$-times differentiable; for some $\rho > 0$, $sup_{y \in \phi_{z\rho}} |g(y) - g(z) - Q_g(y,z)|/||y-z|| \leq G_g(z)$ for all z, where $\phi_{z\rho} = \{y : |y-z| < \rho\}$; Q_g is $(\nu-1)$th degree homogeneous polynomial in $y-z$ with coefficients the partial derivatives of g at z of orders 1 through $\nu-1$; $g(z)$ and its partial derivatives of orders 1 through $\nu-1$ and less, and $G_g(Z)$ all have finite αth moments.

Condition A.

(i) $\{(Y_i, X_i', Z_{1i}, \ldots, Z_{pi})\}_{i=1}^n$ are i.i.d.; (X_i', Z_i') has finite support with the support of Z_i being a product set $\prod_{\alpha=1}^p [c_\alpha, d_\alpha]$; the density function of Z_i is bounded from below by a positive constant on its support; $E(U_i^4)$ is finite.

(ii) Let $f(z_1, \ldots, z_p)$ denote the density function of $(Z_{1i}, \ldots, Z_{pi})$. Then $\xi_\alpha(z_\alpha)$, $\eta_\alpha(z_\alpha)$, $g_\alpha(z_\alpha)$ and $f(z_1, \ldots, z_p)$ all belong to $\mathcal{G}_\nu^4$, where $\nu \geq 2$ is an integer. Let $x = (x^c, x^d)$, where x^c and x^d denote the continuous and discrete components of x, respectively, then for all values of x^d, $\sigma_u^2(x, z_1, \ldots, z_p) = E(U_i^2 | X_i = x, Z_{1i} = z_1, \ldots, Z_{pi} = z_p) \in \mathcal{G}_1^4$.

(iii) The kernel functions K and L are bounded, symmetric, and both are of order ν.

(iv) As $n \to \infty$, $nh^{2\alpha} \to \infty$, $n^{3/2} h_\alpha h_{\underline{\alpha}}^{p-1} \to \infty$ and $n(h_\alpha^{2\nu} + h_{\underline{\alpha}}^{2\nu}) \to 0$.

Note that when $h_\alpha = h_{\underline{\alpha}} \equiv h$, condition (iv) becomes $n^{3/2} h^p \to \infty$ and $nh^{2\nu} \to 0$, which allows second order kernels to be used ($\nu = 2$) if $p \leq 5$. Condition A (iv) also implies that the data needs to be undersmoothed.

Proof of Theorem 2.1. Equation (2) can be written in two forms, both used below:

$$Y_i = \beta_0 + X_i'\beta + \sum_{\alpha=1}^p g_\alpha(Z_{\alpha i}) + U_i, \tag{A.1}$$

$$Y_i = \beta_0 + X_i'\beta + g_\alpha(Z_{\alpha i}) + G_{\underline{\alpha}}(Z_{\underline{\alpha} i}) + U_i. \tag{A.2}$$

Replacing Y_l by $(\beta_0 + X_l'\beta + g_{\alpha l} + G_{\underline{\alpha} l} + U_l)$ on the right-hand-side of (13), we get

$$\hat{Y}_{\alpha i} = \beta_0 + \hat{X}_{\alpha i}'\beta + \hat{g}_{\alpha i} + \hat{G}_{\underline{\alpha} i} + \hat{U}_{\alpha i}, \tag{A.3}$$

where $\hat{X}_{\alpha i}$ is given in (15), $\hat{g}_{\alpha i} = \sum_{l=1}^n W_{\alpha i,l} g_{\alpha l}$, $\hat{G}_{\underline{\alpha} i} = \sum_{l=1}^n W_{\alpha i,l} G_{\underline{\alpha} l}$, and $\hat{U}_{\alpha i} = \sum_{l=1}^n W_{\alpha i,l} U_l$.

In vector-matrix notation, (A.1) and (A.3) can be written as

$$Y = \iota\beta_0 + X\beta + \sum_{\alpha=1}^p g_\alpha + U, \tag{A.4}$$

$$\hat{Y}_\alpha = \iota\beta_0 + \hat{X}_\alpha\beta + \hat{g}_\alpha + \hat{G}_{\underline{\alpha}} + \hat{U}_\alpha, \tag{A.5}$$

where ι is a $n \times 1$ vector of ones.

Define a $n \times 1$ constant vector $\bar{G}_{\underline{\alpha}}$ by

$$\bar{G}_{\underline{\alpha}} = c\,\iota \equiv [\frac{1}{n}\sum_{j=1}^{n} G_{\underline{\alpha}j}]\,\iota. \tag{A.6}$$

Note that $c = n^{-1}\sum_{j=1}^{n} G_{\underline{\alpha}\,j} = O_p(n^{-1/2})$ because $E(G_{\underline{\alpha}\,j}) = 0$.

Summing over α in (A.5), also adding and subtracting $\sum_{\alpha=1}^{p} \bar{G}_{\underline{\alpha}}$ gives

$$\begin{aligned}\sum_{\alpha=1}^{p} \hat{Y}_\alpha &= \iota p\beta_0 + \sum_{\alpha=1}^{p} \hat{X}_\alpha\beta + \sum_{\alpha=1}^{p} \hat{g}_\alpha + \sum_{\alpha=1}^{p} \hat{G}_{\underline{\alpha}} + \sum_{\alpha=1}^{p} \hat{U}_\alpha \\ &= \iota p\beta_0 + \sum_{\alpha=1}^{p} \hat{X}_\alpha\beta + \sum_{\alpha=1}^{p} \hat{g}_\alpha + \sum_{\alpha=1}^{p} \bar{G}_{\underline{\alpha}} + \sum_{\alpha=1}^{p} [\,\hat{G}_{\underline{\alpha}} - \bar{G}_{\underline{\alpha}}\,] + \sum_{\alpha=1}^{p} \hat{U}_\alpha. \end{aligned} \tag{A.7}$$

Subtracting (A.7) from (A.4) leads to

$$\begin{aligned}\hat{\mathcal{Y}} \equiv Y - \sum_{\alpha=1}^{p} \hat{Y}_\alpha &= \iota(1-p)\beta_0 + (X - \sum_{\alpha=1}^{p} \hat{X}_\alpha)'\beta + \sum_{\alpha=1}^{p} (g_\alpha - \hat{g}_\alpha) \\ &\qquad - \sum_{\alpha=1}^{p} \bar{G}_{\underline{\alpha}} - \sum_{\alpha=1}^{p} [\hat{G}_{\underline{\alpha}} - \bar{G}_{\underline{\alpha}}] + U - \sum_{\alpha=1}^{p} \hat{U}_\alpha \\ &= \hat{\mathcal{X}}\delta + \sum_{\alpha=1}^{p} (g_\alpha - \hat{g}_\alpha) - \sum_{\alpha=1}^{p} \bar{G}_{\underline{\alpha}} - \sum_{\alpha=1}^{p} [\hat{G}_{\underline{\alpha}} - \bar{G}_{\underline{\alpha}}] + U - \sum_{\alpha=1}^{p} \hat{U}_\alpha. \end{aligned} \tag{A.8}$$

Using (16) and (A.8) we have

$$\begin{aligned}\sqrt{n}(\hat{\delta} - \delta) = [\hat{\mathcal{X}}'\hat{\mathcal{X}}/n]^{-1}\sqrt{n}\hat{\mathcal{X}}'\{[\sum_{\alpha=1}^{p} (g_\alpha - \hat{g}_\alpha) - \sum_{\alpha=1}^{p} \bar{G}_{\underline{\alpha}} \\ - \sum_{\alpha=1}^{p} [\hat{G}_{\underline{\alpha}} - \bar{G}_{\underline{\alpha}}] + U - \sum_{\alpha=1}^{p} \hat{U}_\alpha]/n\}. \end{aligned} \tag{A.9}$$

Recall that $z = (z_\alpha, z_{\underline{\alpha}})$, $\eta(z) = E(X_1|Z_1 = z)$, $\eta_\alpha(z_\alpha) = E[\eta(z_\alpha, Z_{\underline{\alpha}})]$. Define $V_i = X_i - E(X_i|Z_i) \equiv X_i - \eta(Z_i)$ and

$$\hat{V}_{\alpha i} = \sum_{l=1}^{n} W_{\alpha i,l} V_l. \tag{A.10}$$

Use the short-hand notation $\eta_i = \eta(Z_i) \equiv E(X_i|Z_i)$ and $\eta_{\alpha i} = \eta_\alpha(Z_{\alpha i})$. Also define $\epsilon_i = \eta_i - \sum_{\alpha=1}^{p} \eta_{\alpha i}$. Then we have $X_i = \eta_i + V_i = \epsilon_i + V_i + \sum_{\alpha=1}^{p} \eta_{\alpha i}$ or, in matrix notation,

$$X = \eta + V = \epsilon + V + \sum_{\alpha=1}^{p} \eta_\alpha, \tag{A.11}$$

where X denotes the $n \times q$ matrix with ith row given by X_i'. Hence

$$\hat{\mathcal{X}} = [\iota, X - \sum_{\alpha=1}^{p} \hat{X}_\alpha] = [\iota, \epsilon + V + \sum_{\alpha=1}^{p} (\eta_\alpha - \hat{X}_\alpha)]. \qquad \text{(A.12)}$$

First, we show that the following two results hold:

(I) $n^{-1/2}\hat{\mathcal{X}}'\{\sum_{\alpha=1}^{p}(g_\alpha - \hat{g}_\alpha) - \sum_{\alpha=1}^{p}[\hat{G}_{\underline{\alpha}} - \bar{G}_{\underline{\alpha}}]\} = o_p(1)$,

(II) $$[\hat{\mathcal{X}}'\hat{\mathcal{X}}/n] = \begin{bmatrix} 1, & E(X_1 - \sum_{\alpha=1}^{p} \eta_{\alpha 1})' \\ E(X_1 - \sum_{\alpha=1}^{p} \eta_{\alpha 1}), & E[(X_1 - \sum_{\alpha=1}^{p} \eta_{\alpha 1})(X_1 - \sum_{\alpha=1}^{p} \eta_{\alpha 1})'] \end{bmatrix} + o_p(1).$$

Proof of (I). Using (A.12), we have

$$\begin{aligned} & n^{-1/2}\hat{\mathcal{X}}'\Big\{\sum_{\alpha=1}^{p}(g_\alpha - \hat{g}_\alpha) - \sum_{\alpha=1}^{p}[\hat{G}_{\underline{\alpha}} - \bar{G}_{\underline{\alpha}}]\Big\} \\ &= n^{-1/2}\Big\{\iota'[\sum_{\alpha=1}^{p}(g_\alpha - \hat{g}_\alpha) - \sum_{\alpha=1}^{p}(\hat{G}_{\underline{\alpha}} - \bar{G}_{\underline{\alpha}})], \\ & [\epsilon + V + \sum_{\alpha=1}^{p}(\eta_\alpha - \hat{X}_\alpha)]'[\sum_{\alpha=1}^{p}(g_\alpha - \hat{g}_\alpha) - \sum_{\alpha=1}^{p}(\hat{G}_{\underline{\alpha}} - \bar{G}_{\underline{\alpha}})]\Big\}. \qquad \text{(A.13)} \end{aligned}$$

Noting that $n^{-1/2}\epsilon'(g_\alpha - \hat{g}_\alpha) = o_p(1)$ by Lemma B.1 (i), $n^{-1}\epsilon'(\hat{G}_{\underline{\alpha}} - \bar{G}_{\underline{\alpha}}) = o_p(1)$ by Lemma B.2. Since the above results hold with ϵ replaced by ι we obtain (I), because all the remaining terms on the right side of (A. 13) are of even smaller order by Lemma B.4 and Lemma B.5, together with Lemmas B.1 and B.2.

Proof of (II). Using $\hat{\mathcal{X}} = [\iota, X - \sum_{\alpha=1}^{p} \hat{X}_\alpha]$, we get

$$\begin{aligned} \hat{\mathcal{X}}'\hat{\mathcal{X}}/n &= n^{-1}\begin{pmatrix} n, & \iota'[X - \hat{X}_\alpha] \\ [X - \hat{X}_\alpha]'\iota, & [X - \hat{X}_\alpha]'[X - \hat{X}_\alpha] \end{pmatrix} \\ &= \begin{pmatrix} 1, & E\epsilon_1' \\ E\epsilon_1, & E(\epsilon_1 + V_1)(\epsilon_1 + V_1)' \end{pmatrix} + o_p(1), \qquad \text{(A.14)} \end{aligned}$$

where we have used the facts that $n^{-1}\iota'[X - \hat{X}_\alpha] = n^{-1}\iota'[\epsilon + V + \sum_{\alpha=1}^{p}(\eta_\alpha - \hat{X}_\alpha)] = E(\epsilon_1) + o_p(1)$ by Lemma B.5, and $n^{-1}[X - \hat{X}_\alpha]'[X - \hat{X}_\alpha] = n^{-1}[X - \sum_{\alpha=1}^{p}\eta_\alpha][X - \sum_{\alpha=1}^{p}\eta_\alpha]' + o_p(1) \equiv E[(\epsilon_1 + V_1)(\epsilon_1 + V_1)'] + o_p(1)$ by Lemma B.5. In the above we also used (see (A.11)) $\epsilon_i + V_i = X_i - \sum_{\alpha=1}^{p}\eta_{\alpha i}$ and $E\epsilon_1 = E(X_1 - \sum_{\alpha=1}^{p}\eta_{\alpha 1})$. Thus, (A.14) proves (II).

Using (A.14) and a partitioned inverse yields

$$[\frac{\hat{\mathcal{X}}'\hat{\mathcal{X}}}{n}]^{-1} = \begin{bmatrix} [1 + (E\epsilon_1)'\Phi^{-1}(E\epsilon_1)], & -(E\epsilon_1)'\Phi^{-1} \\ -\Phi^{-1}(E\epsilon_1), & \Phi^{-1} \end{bmatrix} + o_p(1), \qquad \text{(A.15)}$$

where $\Phi = E[(\epsilon_1 + V_1)(\epsilon_1 + V_1)'] - (E\epsilon_1)(E\epsilon_1)' = \text{Var}\,(V_1 + \epsilon_1)$ (as defined in Theorem 2.1).

Applying (I) to (A.9) leads to

$$\sqrt{n}(\hat{\delta} - \delta) = [\frac{\hat{\mathcal{X}}'\hat{\mathcal{X}}}{n}]^{-1}\{n^{-1/2}\hat{\mathcal{X}}'[U - \sum_{\alpha=1}^{p}\hat{U}_\alpha - \sum_{\alpha=1}^{p} G_{\underline{\alpha}}]\} + o_p(1). \tag{A.16}$$

Next, using Lemma B.5 we get

$$\begin{aligned}
&n^{-1/2}\hat{\mathcal{X}}'[U - \sum_{\alpha=1}^{p}\hat{U}_\alpha - \sum_{\alpha=1}^{p}\bar{G}_{\underline{\alpha}}] \\
&= n^{-1/2}\Big\{\iota'[U - \sum_{\alpha=1}^{p}\hat{U}_\alpha - \sum_{\alpha=1}^{p}\bar{G}_{\underline{\alpha}}], [\epsilon + V + \sum_{\alpha=1}^{p}(\eta_\alpha - \hat{X}_\alpha)]'[U - \sum_{\alpha=1}^{p}\hat{U}_\alpha - \sum_{\alpha=1}^{p}\bar{G}_{\underline{\alpha}}]\Big\} \\
&= n^{-1/2}\Big\{\iota'[U - \sum_{\alpha=1}^{p}\hat{U}_\alpha - \sum_{\alpha=1}^{p}\bar{G}_{\underline{\alpha}}], [V + \epsilon]'[U - \sum_{\alpha=1}^{p}\hat{U}_\alpha - \sum_{\alpha=1}^{p}\bar{G}_{\underline{\alpha}}]\Big\} + o_p(1). \qquad (A.17)
\end{aligned}$$

Hence (A.15)–(A.17) lead to

$$\begin{aligned}
\sqrt{n}(\hat{\beta} - \beta) &= -\Phi^{-1}n^{-1/2}(E\epsilon_1)\iota'(U - \sum_{\alpha=1}^{p}\hat{U}_\alpha - \sum_{\alpha=1}^{p}\bar{G}_{\underline{\alpha}}) \\
&\quad +\Phi^{-1}n^{-1/2}[V + \epsilon]'[U - \sum_{\alpha=1}^{p}\hat{U}_\alpha - \sum_{\alpha=1}^{p}\bar{G}_{\underline{\alpha}}] + o_p(1) \\
&= \Phi^{-1}n^{-1/2}\Big\{[V + \epsilon - (E\epsilon_1)\iota]'(U - \sum_{\alpha=1}^{p}\hat{U}_\alpha - \sum_{\alpha=1}^{p}\bar{G}_{\underline{\alpha}})\Big\} + o_p(1) \\
&= \Phi^{-1}n^{-1/2}\{V'U + [\epsilon - (E\epsilon_1)\iota]'(U - \sum_{\alpha=1}^{p}\hat{U}_\alpha)\} + o_p(1), \qquad (A.18)
\end{aligned}$$

where in the last equality we have used the facts that $n^{-1/2}(V + \epsilon - E\epsilon)'\bar{G}_{\underline{\alpha}} = [n^{-1/2}\sum_i(V_i + \epsilon_i - E\epsilon_i)][n^{-1}\sum_j G_{\underline{\alpha}j}] = O_p(n^{-1/2}) = o_p(1)$ since $E(G_{\underline{\alpha}j}) = 0$, and $n^{-1/2}V'\hat{U}_\alpha = o_p(1)$ because $E[||n^{-1/2}V'\hat{U}_\alpha||^2] \le CE[(\hat{U}_{\alpha i})^2] = o_p(1)$ by Lemma B.3 (ii).

Now using $1/\hat{f}(Z_{\alpha i}, Z_{\underline{\alpha}j}) = 1/f(Z_{\alpha i}, Z_{\underline{\alpha}j}) + [1/\hat{f}(Z_{\alpha i}, Z_{\underline{\alpha}j}) - 1/f(Z_{\alpha i}, Z_{\underline{\alpha}j})]$, we get

$$\begin{aligned}
&n^{-1}[\epsilon - E(\epsilon_1)\iota]'\hat{U}_\alpha = n^{-1}\sum_i(\epsilon_i - E\epsilon_i)\hat{U}_{\alpha i} \\
&= \frac{1}{n^3}\sum_i\sum_j\sum_{l\ne j}(\epsilon_i - E\epsilon_i)U_l K_a(Z_{\alpha i} - Z_{\alpha l})L_{h_{\underline{\alpha}}}(Z_{\underline{\alpha}j} - Z_{\underline{\alpha}l})[f(Z_{\alpha i}, Z_{\underline{\alpha}j})] + (s.o.), \qquad (A.19)
\end{aligned}$$

where $(s.o.)$ means it has an order smaller than the first term. It is easy to see that the leading term of $n^{-1}(\epsilon - E\epsilon_1\iota)'\hat{U}_\alpha$ corresponds to $i \neq j \neq l$, which can be written as a third order U-statistic

$$n^{-1}(\epsilon - E\epsilon_1\iota)'\hat{U}_\alpha = \frac{1}{6}\binom{n}{3}^{-1}\sum\sum\sum_{1\leq i<j<l\leq n} H_n(\zeta_i,\zeta_j,\zeta_l) + (s.o.), \quad \text{(A.20)}$$

where $H_n(.,.,.)$ is a symmetrized version of $(\epsilon_i - E\epsilon_i)U_l K_h(Z_{\alpha i} - Z_{\alpha l})L(Z_{\underline{\alpha} j} - Z_{\underline{\alpha} l})/[f(Z_{\alpha i}, Z_{\underline{\alpha} j})]$ and $\zeta_i = (Z_i, U_i)$.

By (A.19) and H-decomposition we have

$$n^{-1}(\epsilon - E\epsilon_1\iota)'\hat{U}_\alpha = \frac{1}{n}\sum_{i=1}^{n} H_{1n}(\zeta_i) + (s.o.), \quad \text{(A.21)}$$

where $H_{1n}(\zeta_i)$ is the leading term of $E[H_n(\zeta_i,\zeta_j,\zeta_l)|\zeta_i]$ given by

$$H_{1n}(\zeta_i) = \frac{U_i f_\alpha(Z_{\alpha i}) f_{\underline{\alpha}}(Z_{\underline{\alpha} i})(\epsilon_i - E\epsilon_i)}{f(Z_{\alpha i}, Z_{\underline{\alpha} i})} \overset{def}{=} U_i(\epsilon_i - E\epsilon_i)\psi_{\alpha i}. \quad \text{(A.22)}$$

with $\psi_{\alpha i} = f_\alpha(Z_{\alpha i}) f_{\underline{\alpha}}(Z_{\underline{\alpha} i)})/[f(Z_{\alpha i}, Z_{\underline{\alpha} i})]$.

Using (A.18), (A.21) and (A.22), we get

$$\begin{aligned}\sqrt{n}(\hat{\beta} - \beta) &= \Phi^{-1} n^{-1/2}\sum_{i=1}^{n} U_i[V_i + (\epsilon_i - E\epsilon_i)(1 - \sum_{\alpha=1}^{p}\psi_{\alpha i})] \\ &\to N(0, \Phi^{-1}\Omega\Phi^{-1}),\end{aligned} \quad \text{(A.23)}$$

by the Lindeberg Central Limit Theorem, where $\Omega = E[U_i^2 D_i D_i']$, $D_i = V_i + (\epsilon_i - E\epsilon_i)(1 - \sum_{\alpha=1}^{p}\psi_{\alpha i})$ and $\epsilon_i = \eta_i - \sum_{\alpha=1}^{p}\eta_{\alpha i}$.

A consistent estimator for Σ

Let $\hat{E}(X_i|Z_i)$ denote a kernel estimator of $E(X_i|Z_i)$, $\hat{f}_{\alpha i}$, $\hat{f}_{\underline{\alpha} i}$ and $\hat{f}_i$ denote the kernel estimators of $f_\alpha(Z_{\alpha i})$, $f_{\underline{\alpha}}(Z_{\underline{\alpha} i})$ and $f(Z_{\alpha i}, Z_{\underline{\alpha} i})$, respectively.

Specifically, let $K_{\alpha,ij} \equiv K_{h_\alpha}(Z_{\alpha j} - Z_{\alpha i})$ and $L_{\underline{\alpha},ij} \equiv L_{h_{\underline{\alpha}}}(Z_{\underline{\alpha} j} - Z_{\underline{\alpha} i})$, then $\hat{E}(X_i|Z_i) = n^{-1}\sum_j X_j K_{\alpha,ij} L_{\underline{\alpha},ij}$, $\hat{f}_{\alpha i} = n^{-1}\sum_j K_{\alpha,ij}$, $\hat{f}_{\underline{\alpha} i} = n^{-1}\sum_j L_{\underline{\alpha},ij}$ and $\hat{f}_i = n^{-1}\sum_j K_{\alpha,ij} L_{\underline{\alpha},ij}$.

Next, define $\hat{\psi}_{\alpha i} = \hat{f}_{\alpha i}\hat{f}_{\underline{\alpha} i}/\hat{f}_i$, $\hat{\epsilon}_i = \hat{E}(X_i|Z_i) - \sum_{\alpha=1}^{p}\hat{X}_{\alpha i}$, $\bar{\epsilon} = n^{-1}\sum_i\hat{\epsilon}_i$, $\hat{V}_i = X_i - \hat{E}(X_i|Z_i)$, $\hat{D}_i = \hat{V}_i + (\hat{\epsilon}_i - \bar{\epsilon})(1 - \sum_\alpha^2\hat{\psi}_{\alpha i})$, $\hat{U}_i = Y_i - \hat{\beta}_0 - X_i'\hat{\beta} - \sum_{\alpha=1}^{p}\hat{g}_\alpha(Z_{\alpha i})$. Then a consistent estimator of Σ is given by $\hat{\Sigma} = \hat{\Phi}^{-1}\hat{\Omega}\hat{\Phi}^{-1}$, where $\hat{\Phi} = n^{-1}\sum_i(X_i - \sum_{\alpha=1}^{p}\hat{X}_{\alpha i})(X_i - \sum_{\alpha=1}^{p}\hat{X}_{\alpha i})'$ and $\hat{\Omega} = n^{-1}\sum_i\hat{U}_i^2\hat{D}_i\hat{D}_i'$.

Obviously $\hat{E}(X_i|Z_i)$ is a consistent estimator for $E(X_i|Z_i)$, and $\hat{\psi}_{\alpha i}$ is a consistent estimator of $\psi_{\alpha i}$. Also we know that $\hat{U}_i$ is a consistent estimator of U_i

and $\hat{X}_{\alpha i}$ is a consistent estimator of $\eta_{\alpha i}$ by Lemma B.5. The above implies that $\hat{\epsilon}_i$ is a consistent estimator for ϵ_i.

Using these results, one can easily show that $\hat{\Sigma}$ is a consistent estimator of Σ.

Appendix B

As in Appendix A, $A = B + (s.o.)$ means A and B have the same order, $(s.o.)$ means it has an order smaller than B. Also we use the following short-hand notations: $K_{\alpha,ij} = K_{h_\alpha}(Z_{\alpha i} - Z_{\alpha j})$, $L_{\underline{\alpha},ij} = \bar{L}_{h_{\underline{\alpha}}}(Z_{\underline{\alpha}\, i} - Z_{\underline{\alpha}\, j})$, $f_{\alpha i,\underline{\alpha}\, j} = f(Z_{\alpha i}, Z_{\underline{\alpha}\, j})$, and $\hat{f}_{\alpha i,\underline{\alpha}\, j} = \hat{f}(Z_{\alpha i}, Z_{\underline{\alpha}\, j})$, where

$$\hat{f}(Z_{\alpha i}, Z_{\underline{\alpha}\, j}) = \frac{1}{n} \sum_{l \neq j} K_{h_\alpha}(Z_{\alpha l} - Z_{\alpha i}) L_{h_{\underline{\alpha}}}(Z_{\underline{\alpha}\, l} - Z_{\underline{\alpha}\, j}). \tag{B.1}$$

Because the estimator of β only uses data in the trimmed set $\mathcal{D}_n$ (boundary removed), we have $sup_{z \in \mathcal{D}_n} |\hat{f}(z) - f(z)| = O_p(h^\nu + ln(n)(nh^p)^{-1/2})$ and $inf_{z \in \mathcal{D}_n} |1/\hat{f}(z)| = O_p(1)$. Then for any positive integer m, we have

$$\frac{1}{\hat{f}(z)} = \frac{1}{f(z)} + \sum_{s=1}^{m} \frac{(f(z) - \hat{f}(z))^m}{f^{m+1}(z)} + \frac{(f(z) - \hat{f}(z))^{m+1}}{f^{m+1}(z)\hat{f}(z)}. \tag{B.2}$$

Using (B.2) one can show that whenever there is a term involves $1/\hat{f}(Z_i)$, one can replace $1/\hat{f}(Z_i)$ by $1/f(Z_i)$ to obtain its leading term, the remaining terms will be of smaller orders.

Lemma B.1. (i) $n^{-1} \sum_i (\hat{g}_{\alpha i} - g_{\alpha i})\epsilon_i = o_p(n^{-1/2})$, (ii) $n^{-1} \sum_i (\hat{g}_{\alpha i} - g_{\alpha i})^2 = o_p(1)$.

Proof of (i). Start with

$$\begin{aligned}
& n^{-1} \sum_i \epsilon_i (\hat{g}_{\alpha i} - g_{\alpha i}) \\
&= n^{-3} \sum_i \sum_j \sum_{l \neq j} \epsilon_i (g_{\alpha l} - g_{\alpha i}) K_{\alpha,il} L_{\underline{\alpha},jl} / \hat{f}(Z_{\alpha i}, Z_{\underline{\alpha}\, j}) \\
&= n^{-3} \sum_i \sum_j \sum_{l \neq j} \epsilon_i (g_{\alpha l} - g_{\alpha i}) K_{\alpha,il} L_{\underline{\alpha},\, jl} / f(Z_{\alpha i}, Z_{\underline{\alpha}\, j}) + (s.o.) \\
&\equiv A_{1n} + (s.o.) \quad \text{(by using eq. (B.2))}.
\end{aligned}$$

We need only consider A_{1n}. Since $f(.,.)$ is bounded away from zero, A_{1n} has the same order as $B_{1n} = n^{-3} \sum_i \sum_j \sum_{l \neq i, l \neq j} \epsilon_i (g_{\alpha l} - g_{\alpha i}) K_{\alpha,il} L_{\underline{\alpha},\, jl}$. B_{1n} contains three summations, we consider two different cases for B_{1n}: (1) all three summation indices are different from each other, i.e., $i \neq j \neq l$; and (2) $j = i \neq l$. We use $B_{1n(s)}$, $s = 1, 2$, to denote these two cases.

It is easy to see that $E(B_{1n(1)}) = O(h_\alpha^\nu)$, its second moment is

$$\begin{aligned}E[(B_{1n(1)})^2] &= n^{-6}\sum\sum_{i_1\neq i_2\neq i_3}\sum\ \sum\sum_{i_4\neq i_5\neq i_6}\sum E\{[\epsilon_{i_1}(g_{\alpha i_3}-g_{\alpha i_1})K_{\alpha,i_1i_3}L_{\underline{\alpha},\,i_2i_3}]\\ &\quad\times[\epsilon_{i_4}(g_{\alpha i_6}-g_{\alpha i_4})K_{\alpha,i_4i_6}L_{\underline{\alpha},\,i_5i_6}]\}\\ &= (n^6h_\alpha^2h_{\underline{\alpha}}^{2(p-1)})^{-1}\{n^6O(h_\alpha^{2+2\nu}h_{\underline{\alpha}}^{2(p-1)})+n^5O(h_\alpha^{2+2\nu}h_{\underline{\alpha}}^{2(p-1)})\\ &\quad+n^4O(h_\alpha^3h_{\underline{\alpha}}^{2(p-1)}+h_\alpha^{2+2\nu}h_{\underline{\alpha}}^{(p-1)})+(s.o.)\}\\ &= O(h_\alpha^{2\nu})+O(h_\alpha n^{-2})+O(h_\alpha^{2\nu}(n^2h_{\underline{\alpha}}^{p-1})^{-1})+(s.o.)=o(n^{-1}).\end{aligned}$$

We briefly explain the above results. When the six summation indices $i_1,\ldots,i_6$ take at least five different values, using the fact that for $i\neq j$, $E[(g_{\alpha i}-g_{\alpha j})K_{\alpha,ij}] = h_\alpha O(h_\alpha^\nu) = O(h_\alpha^{1+\nu})$, it is easy to see that they are of the order of $(n^6+n^5)O(h_\alpha^{2+2\nu}h_{\underline{\alpha}}^{2(p-1)})$. Next if the six indices take four different values, it is easy to check that if $i_1=i_4$ and $i_3=i_6$, we get $n^4O(h_\alpha^3h_{\underline{\alpha}}^{2(p-1)})$; if $i_2=i_5$ and $i_3=i_6$, we get $n^4O(h_\alpha^{2+2\nu}h_{\underline{\alpha}}^{p-1})$; if $i_1=i_4$ and $i_2=i_5$, we get $n^4O(h_\alpha^{2+2\nu}h_{\underline{\alpha}}^{2(p-1)})$. Finally it is easy to see that when the six indices take three different values, it has a smaller order smaller than those above. Hence, $B_{1n(1)} = o_p(n^{-1/2})$.

Next, $B_{1n(2)} = 2n^{-3}\sum_i\sum_{l>i}(g_{\alpha l}-g_{\alpha i})K_{\alpha,il}\bar{L}_{\underline{\alpha},\,il}$. Its second moment is

$$\begin{aligned}&E[B_{1n(2)}^2]\\ &= 4n^{-6}\sum\sum_{l>i}\sum\sum_{l'>i'}E[(g_{\alpha l}-g_{\alpha i})K_{\alpha,il}L_{\underline{\alpha},\,il}(g_{\alpha l'}-g_{\alpha i'})K_{\alpha,i'l'}L_{\underline{\alpha},\,i'l'}]\\ &= 4(n^6h_\alpha^2h_{\underline{\alpha}}^{2(p-1)})^{-1}\{n^4h_\alpha^2h_{\underline{\alpha}}^{2(p-1)}O(h_\alpha^{2\nu})+n^3h_\alpha^2h_{\underline{\alpha}}^{p-1}O(h_\alpha^{2\nu})+n^2h_\alpha h_{\underline{\alpha}}^{p-1}O(h_\alpha^2)\}\\ &= n^{-1}\{O(n^{-1}h_\alpha^{2\nu})+O(h_\alpha(n^3h_{\underline{\alpha}}^{p-1})^{-1})\}=o(n^{-1}).\end{aligned}$$

Hence, $B_{1n(2)} = o_p(n^{-1/2})$.

Thus we have shown that $B_{1n} = o_p(n^{-1/2})$ which implies $n^{-1}\sum_i(\hat{g}_{\alpha i}-g_{\alpha i})\epsilon_i = o_p(n^{-1/2})$. Obviously the above result holds true if one replaces ϵ_i by 1.

Proof of (ii). By using (B.2) one can easily show that $n^{-1}\sum_i(\hat{g}_{\alpha i}-g_{\alpha i})^2 = A_{2n}+(s.o.)$, where $A_{2n} = n^{-1}\sum_i(\hat{g}_{\alpha i}-g_{\alpha i})^2\hat{f}_{\underline{\alpha}i}^2/\hat{f}_{\alpha i}^2$ is the same as $n^{-1}\sum_i(\hat{g}_{\alpha i}-g_{\alpha i})^2$ except that the random denominator $1/\hat{f}(Z_j)$ is replaced by $1/f(Z_j)$, which can be further replaced by 1, i.e., $\hat{g}_{\alpha i}-g_{\alpha i} = n^{-2}\sum_j\sum_{l\neq j}(g_{\alpha l}-g_{\alpha i})K_{\alpha,il}L_{\underline{\alpha},\,jl}/\hat{f}_{\alpha i,\underline{\alpha}\,j}$ can be replaced by $I_{1n,i} \stackrel{def}{=} n^{-2}\sum_j\sum_{l\neq j}(g_{\alpha l}-g_{\alpha i})K_{\alpha,il}L_{\underline{\alpha},\,jl}$

We use $E[n^{-1}\sum_i I_{1n,i}^2] = E[I_{1n,1}^2]$ to bound the leading term of $n^{-1}\sum_i(\hat{g}_{\alpha i}-g_{\alpha i})^2$.

$$E[I_{1n,1}^2] \le Cn^{-4}\sum_i\sum_{j\neq i}\sum_{i'}\sum_{j'\neq i'}E[(g_{\alpha j}-g_{\alpha 1})K_{\alpha,1j}L_{\underline{\alpha},\,ij}(g_{\alpha j'}-g_{\alpha 1})K_{\alpha,1j'}L_{\underline{\alpha},\,i'j'}]$$

$$= (n^4 h_\alpha^2 h_{\underline{\alpha}}^{2(p-1)})^{-1}\{n^4 h_\alpha^2 h_{\underline{\alpha}}^{2(p-1)} O(h_\alpha^{2\nu}) + n^3 h_\alpha^2 h_{\underline{\alpha}}^{2(p-1)} O(h_\alpha^{2\nu}) + n^2 h_\alpha h_{\underline{\alpha}}^{p-1} O(h_\alpha^2)\}$$
$$= O(h_\alpha^{2\nu}) + O(h_\alpha (n^2 h_{\underline{\alpha}}^{p-1})^{-1}) = o(1).$$

This implies the leading term of $n^{-1}\sum_i (\hat{g}_{\alpha i} - g_{\alpha i})^2$ is $o_p(1)$.

Lemma B.2. $n^{-1}(\hat{G}_{\underline{\alpha}} - \bar{G}_{\underline{\alpha}})'\epsilon = o_p(n^{-1/2})$.

Proof. Note that $\bar{G}_{\underline{\alpha}} = c\,\iota$ with $c = n^{-1}\sum_j G_{\underline{\alpha}\,j}$ (see (A.6)). Then

$$\begin{aligned}
n^{-1}(\hat{G}_{\underline{\alpha}} - \bar{G}_{\underline{\alpha}})'\epsilon &= n^{-1}\sum_i \epsilon_i[\hat{G}_{\underline{\alpha}\,i} - c] \\
&= n^{-3}\sum_i\sum_j\sum_{l\neq j} \epsilon_i (G_{\underline{\alpha}\,l} - G_{\underline{\alpha}\,j}) K_{\alpha,il} L_{\underline{\alpha},\,jl}/\hat{f}(Z_{\alpha i}, Z_{\underline{\alpha}\,j}) \\
&= n^{-3}\sum_i\sum_j\sum_{l\neq j} \epsilon_i (G_{\underline{\alpha}\,l} - G_{\underline{\alpha}\,j}) K_{\alpha,il} L_{\underline{\alpha},\,jl}/f(Z_{\alpha i}, Z_{\underline{\alpha}\,j}) + (s.o.) \\
&\equiv A_{2n} + (s.o.).
\end{aligned}$$

Obviously A_{2n} has the same order as $B_{2n} = n^{-31}\sum_i\sum_j\sum_{l\neq j}\epsilon_i(G_{\underline{\alpha}\,l} - G_{\underline{\alpha}\,j})K_{\alpha,il}$ $L_{\underline{\alpha},\,jl}$.

B_{2n} contains three summations, we consider three different cases for B_{3n}: (1) all three summation indices are different from each other, (2) $i = l \neq j$ and (3) $i = j \neq l$. We use $B_{2n(s)}$, $s = 1, 2, 3$, to denote these cases.

Similar to the proof of $E[(B_{1n(1)})^2] = o_p(n^{-1})$ in the proof of Lemma B.1, one can easily show that

$$\begin{aligned}
&E[(B_{2n(1)})^2] \\
&= n^{-6}\sum\sum\sum_{i_1\neq i_2\neq i_3}\sum\sum\sum_{i_4\neq i_5\neq i_6} E\{[\epsilon_{i_1}(G_{\underline{\alpha}\,i_3} - G_{\underline{\alpha}\,i_2})K_{\alpha,i_1 i_3}L_{\underline{\alpha},\,i_2 i_3}] \\
&\quad \times[\epsilon_{i_4}(G_{\underline{\alpha}\,i_6} - G_{\underline{\alpha}\,i_5})K_{\alpha,i_4 i_6}L_{\underline{\alpha},\,i_5 i_6}]\} \\
&= (n^6 h_\alpha^2 h_{\underline{\alpha}}^{2(p-1)})^{-1}\{n^6 O(h_\alpha^2 h_{\underline{\alpha}}^{2\nu}) + n^5 O(h_\alpha^2 h_{\underline{\alpha}}^{2\nu}) + n^4 O(h_\alpha^2 h_{\underline{\alpha}}^3 + h_\alpha h_{\underline{\alpha}}^{2\nu}) + (s.o.)\} \\
&= O(h_{\underline{\alpha}}^{2\nu}) + O((n^2 h_{\underline{\alpha}}^{(p-4)}) + (s.o.) = o(n^{-1}).
\end{aligned}$$

Hence, $B_{2n(1)} = o_p(n^{-1/2})$.

Next, $B_{2n(2)} = 2(n^3 h_\alpha)^{-1}K(0)\sum_i\sum_{j>i}\epsilon_i(G_{\underline{\alpha}\,i} - G_{\underline{\alpha}\,j})L_{\underline{\alpha},\,ij}$. Its second moment is

$$\begin{aligned}
E[B_{2n(2)}^2] &\leq C(n^6 h_\alpha^2)^{-1}\sum\sum_{j>i}\sum\sum_{j'>i'} E[(G_{\underline{\alpha}\,i} - G_{\underline{\alpha}\,j})L_{\underline{\alpha},\,ij}(G_{\underline{\alpha}\,i'} - G_{\underline{\alpha}\,j'})L_{\underline{\alpha},\,i'j'}] \\
&= 4(n^6 h_\alpha^2 h_{\underline{\alpha}}^{2(p-1)})^{-1}\{n^4 h_{\underline{\alpha}}^{2(p-1)} O(h_{\underline{\alpha}}^{2\nu}) + n^3 h_{\underline{\alpha}}^{2(p-1)} O(h_{\underline{\alpha}}^{2\nu}) + n^2 h_{\underline{\alpha}}^{p-1} O(h_{\underline{\alpha}}^2)\} \\
&= n^{-1}\{O(h_{\underline{\alpha}}^{2\nu}(nh_\alpha)^{-2}) + O((n^3 h_\alpha^2 h_{\underline{\alpha}}^{(p-3)})^{-1})\} = o(n^{-1}).
\end{aligned}$$

Hence, $B_{2n(2)} = o_p(n^{-1/2})$.

Finally, $B_{2n(3)} = 2n^{-3}\sum_i \sum_{l>i} \epsilon_i (G_{\underline{\alpha}\, l} - G_{\underline{\alpha}\, i}) K_{\underline{\alpha},\, il} L_{\underline{\alpha},\, il}$. Its second moment is bounded by

$$\begin{aligned}
&E[B_{2n(3)}^2]\\
&\le Cn^{-6}\sum_{l>i}\sum \sum_{l'>i'}\sum E[(G_{\underline{\alpha}\, l} - G_{\underline{\alpha}\, i})K_{\underline{\alpha},\, il}L_{\underline{\alpha},\, il}(G_{\underline{\alpha}\, l'} - G_{\underline{\alpha}\, i'})K_{\underline{\alpha},\, i'l'}L_{\underline{\alpha},\, i'l'}]\\
&= 4(n^6 h_\alpha^2 h_{\underline{\alpha}}^{2(p-1)})^{-1}\{n^4 h_\alpha^2 h_{\underline{\alpha}}^{2(p-1)} O(h_{\underline{\alpha}}^{2\nu}) + n^3 h_\alpha^2 h_{\underline{\alpha}}^{p-1} O(h_{\underline{\alpha}}^{2\nu})\\
&\quad + n^2 h_\alpha h_{\underline{\alpha}}^{p-1} O(h_{\underline{\alpha}}^2)\} = o(n^{-1})
\end{aligned}$$

(this term has an smaller order than $E[B_{2n(2)}^2]$). Hence, $B_{2n(3)} = o_p(n^{-1/2})$.

Thus we have shown that $n^{-1}(\hat{G}_{\underline{\alpha}} - \bar{G}_{\underline{\alpha}})'\epsilon = o_p(n^{-1/2})$. Obviously the above proof holds true if one replaces ϵ_i by 1.

Lemma B.3. (i) $n^{-1}||\hat{V}_\alpha||^2 = o_p(n^{-1/2})$, (ii) $n^{-1}||\hat{U}_\alpha||^2 = o_p(n^{-1/2})$.

Proof of (i). $n^{-1}||\hat{V}_\alpha||^2 = n^{-1}\sum_i \hat{V}'_{\alpha i}\hat{V}_{\alpha i}$. Using (B.2) we can replace $\hat{V}_{\alpha i} = n^{-2}\sum_j \sum_{l\ne j} K_{\alpha, il} L_{\underline{\alpha},\, jl} V_l / \hat{f}(Z_{\alpha i}, Z_{\underline{\alpha}\, j})$ by $D_{\alpha,i} = n^{-2}\sum_j \sum_{l\ne j} K_{\alpha,il} L_{\underline{\alpha},\, jl} V_l / f(Z_{\alpha i}, Z_{\underline{\alpha}\, j})$ to obtain the leading term of $\hat{V}_{\alpha i}$, where $D_{\alpha,i}$ is the same as $\hat{V}_{\alpha,i}$ except that $1/\hat{f}(Z_{\alpha i}, Z_{\underline{\alpha}\, j})$ is replaced by $1/f(Z_{\alpha i}, Z_{\underline{\alpha}\, j})$. Hence, $n^{-1}\sum_i ||\hat{V}_{\alpha 1}||^2$ has the same order as $n^{-1}\sum_i ||D_{\alpha,i}||^2$. We can bound $n^{-1}\sum_i ||D_{\alpha,i}||^2$ by $E[n^{-1}\sum_i ||D_{\alpha,i}||^2] = E[||D_{\alpha,1}||^2]$, moreover

$$\begin{aligned}
E[||D_{\alpha,1}||^2] &\le Cn^{-4}\sum_{j\ne i}\sum \sum_{j'\ne i'}\sum E[K_{\alpha,1j} L_{\underline{\alpha},\, ij} K_{\alpha,1j'} L_{\underline{\alpha},\, i'j'} V'_j V_{j'}]\\
&= Cn^{-4}\sum_{j\ne i}\sum \sum_{i'\ne j} E[(K_{\alpha,1j})^2 L_{\underline{\alpha},\, ij} L_{\underline{\alpha},\, i'j} V'_j V_j]\\
&= C(n^2 h_\alpha h_{\underline{\alpha}}^{p-1})^{-2}\{n^3 h_\alpha h_{\underline{\alpha}}^{2(p-1)} O(1) + n^2 h_\alpha h_{\underline{\alpha}}^{p-1} O(1)\}\\
&= n^{-1/2}\{O((n^{1/2}h_\alpha)^{-1}) + O((n^{3/2} h_\alpha h_{\underline{\alpha}}^{p-1})^{-1})\} = o(n^{-1/2}).
\end{aligned}$$

Summarizing, we have shown that $E[||n^{-1}\hat{V}_{\alpha 1}||^2] = o(n^{-1/2})$, which implies that $n^{-1}||\hat{V}_\alpha||^2 = o_p(n^{-1/2})$.

The proof of (ii). follows exactly the same steps as in (i) above.

Lemma B.4. $n^{-1}(\hat{g}_\alpha - g_\alpha)'V = o_p(n^{-1/2})$.

Proof. Let $\mathcal{Z}_n = \{Z_i\}_{i=1}^n$. We have $E[||n^{-1}(\hat{g}_\alpha - g_\alpha)'V||^2|\mathcal{Z}_n] = n^{-2}\sum_i E[(\hat{g}_{\alpha i} - g_{\alpha i})^2 V'_i V_i|\mathcal{Z}_n] \le Cn^{-2}\sum_i(\hat{g}_{\alpha i} - g_{\alpha i})^2 = o_p(n^{-1})$ by Lemma B.1 (ii). Hence, $n^{-1}(\hat{g}_\alpha - g_\alpha)'V = o_p(n^{-1/2})$.

Lemma B.5. $n^{-1}||\eta_\alpha - \hat{X}_\alpha||^2 = o_p(n^{-1/2})$.

Proof. Define $\tilde{\eta}_{\alpha i} = n^{-1}\sum_j \eta(Z_{\alpha i}, Z_{\underline{\alpha}\, j})$, and $\hat{\eta}_{\alpha i} = \sum_l W_{\alpha i,l}\, \eta(Z_l)$. We first prove two intermediate results:

(i) $n^{-1}||\tilde{\eta}_\alpha - \eta_\alpha||^2 = O_p(n^{-1}) = o_p(n^{-1/2})$, and (ii) $n^{-1}||\hat{\eta}_\alpha - \tilde{\eta}_\alpha||^2 = o_p(n^{-1/2})$.

Proof of (i).

$$\begin{aligned}
&E[n^{-1}||\tilde{\eta}_\alpha - \eta_\alpha||^2]\\
&= E[(\tilde{\eta}_{\alpha 1} - \eta_{\alpha 1})'(\tilde{\eta}_{\alpha 1} - \eta_{\alpha 1})]\\
&= n^{-2}\sum_{i\neq 1}\sum_{j} E[(\eta(Z_{\alpha 1}, Z_{\underline{\alpha}\, i}) - \eta_{\alpha 1})'(\eta(Z_{\alpha 1}, Z_{\underline{\alpha}\, j}) - \eta_{\alpha 1})] + O(n^{-1})\\
&= n^{-2}\sum_{i\neq 1} E[(\eta(Z_{\alpha 1}, Z_{\underline{\alpha}\, i}) - \eta_{\alpha 1})'(\eta(Z_{\alpha 1}, Z_{\underline{\alpha}\, i}) - \eta_{\alpha 1})]\\
&\quad + n^{-2}\sum_{i\neq 1}\sum_{j\neq i} E[\eta(Z_{\alpha 1}, Z_{\underline{\alpha}\, i}) - \eta_{\alpha 1}]' E[\eta(Z_{\alpha 1}, Z_{\underline{\alpha}\, j}) - \eta_{\alpha 1}] + O(n^{-1})\\
&= n^{-2}\{O(n) + 0\} + O(n^{-1}) = O(n^{-1}),
\end{aligned}$$

because $E[\eta(Z_{\alpha 1}, Z_{\underline{\alpha}\, i}) - \eta_\alpha(Z_{\alpha 1})] = 0$ for $i \neq 1$.

Proof of (ii) Note that $n^{-1}||\hat{\eta} - \tilde{\eta}||^2 = n^{-1}\sum_i (\hat{\eta}_{\alpha i} - \tilde{\eta}_{\alpha i})'(\hat{\eta}_{\alpha i} - \tilde{\eta}_{\alpha i})$; $\hat{\eta}_{\alpha 1} - \tilde{\eta}_{\alpha 1} = n^{-2}\sum_j\sum_{l\neq j}(\eta_{\alpha 1,\underline{\alpha}\, l} - \eta_{\alpha 1,\underline{\alpha}\, j})K_{\alpha,1l}L_{\underline{\alpha},\, jl}/\hat{f}_{\alpha 1,\underline{\alpha}\, j}$ can be replaced by $F_{1n,i} = n^{-2}\sum_j\sum_{l\neq j}(\eta_{\alpha 1,\underline{\alpha}\, l} - \eta_{\alpha 1,\underline{\alpha}\, j})K_{\alpha,1l}L_{\underline{\alpha},\, jl}/f_{\alpha 1,\underline{\alpha}\, j}$ to obtain the leading term of $\hat{\eta}_{\alpha 1} - \tilde{\eta}_{\alpha 1}$; $n^{-1}||\hat{\eta} - \tilde{\eta}||^2$ has the same order as $n^{-1}\sum_i ||F_{1n,i}||^2$. We bound $n^{-1}\sum_i ||F_{1n,i}||^2$ by $E[n^{-1}\sum_i ||F_{1n,i}||^2] = E[||F_{1n,1}||^2]$, then note that

$$\begin{aligned}
&E[||F_{1n,1}||^2]\\
&\leq Cn^{-4}\sum_j\sum_{l\neq j}\sum_j\sum_{l'\neq j'} E[(\eta_{\alpha 1,\underline{\alpha}\, l} - \eta_{\alpha 1,\underline{\alpha}\, j})'K_{\alpha,1l}L_{\underline{\alpha},\, jl}\\
&\qquad\qquad \times(\eta_{\alpha 1,\underline{\alpha}\, l'} - \eta_{\alpha 1,\underline{\alpha}\, j'})K_{\alpha,1l'}\bar{L}_{\underline{\alpha},\, j'l'}\\
&= (n^2 h_\alpha h_{\underline{\alpha}}^{p-1})^{-2}\{n^4 h_\alpha^2 h_{\underline{\alpha}}^{2(p-1)} O(h_{\underline{\alpha}}^{2\nu}) + n^4 h_\alpha^2 h_{\underline{\alpha}}^{2(p-1)} O(h_{\underline{\alpha}}^{2\nu}) + n^2 h_\alpha h_{\underline{\alpha}}^{p-1} O(h_{\underline{\alpha}}^2)\}\\
&= \{O(h_{\underline{\alpha}}^{2\nu}) + O((n^2 h_\alpha h_{\underline{\alpha}}^{p-3})^{-1})\} = o(n^{-1/2}).
\end{aligned}$$

We now prove Lemma B.5. From $X_i = \eta(Z_i) + V_i$, we have $\hat{X}_\alpha = \hat{\eta}_\alpha + \hat{V}_\alpha$. Hence,

$n^{-1}||\eta_\alpha - \hat{X}_\alpha||^2 = n^{-1}||\eta_\alpha - \hat{\eta}_\alpha - \hat{V}_\alpha||^2 \leq Cn^{-1}\{||\eta_\alpha - \tilde{\eta}_\alpha||^2 + ||\tilde{\eta}_\alpha - \hat{\eta}_\alpha||^2 + ||\hat{V}_\alpha||^2\} = o_p(n^{-1/2})$ by (i) and (ii) above, and by Lemma B.3 (i).

References

Carroll, R. J., Fan, J., Gijbels and Wang, M. P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92**, 477-489.

Chamberlain, G. (1992). Efficiency bounds for semiparametric regression. *Econometrica* **60**, 567-596.

Chen, R., Härdle, W., Linton, O. and Severance-Lossin, E. (1995). Estimation and variable selection in additive nonparametric regression models. Discussion paper, SFB 373, Humboldt-Universität zu Berlin.

Fan, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* **87**, 998-1004.

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

Fan, J., Härdle, W. and Mammen, E. (1998). Direct estimation of low dimensional components in additive models. *Ann. Statist.* **26**, 943-971.

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

Li, Q. (2000). Efficient estimation of additive partially linear models. *Internat. Econom. Rev.* **41**, 1073-1092.

Linton, O. (1997). Efficient estimation of additive nonparametric regression models. *Biometrika* **84**, 469-474.

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

Mammen, E., Linton, O. B. and Nielsen, J. P. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Ann. Statist.* **27**, 1443-1490.

Newey, W. K. (1994). Kernel estimation of partial means in a general variance estimator. *Econom. Theory* **10**, 233-253.

Nielsen, J. P. and Linton, O. (1998). An optimization interpretation of integration and backfitting estimators for separable nonparametric models. *J. Roy. Statist. Soc. Ser. B.* **60**, 217-222.

Opsomer, J. D. and Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *Ann. Statist.* **25**, 186-211.

Robinson, P. (1988). Root-N consistent semiparametric regression. *Econometrica* **56**, 931-954.

Schick, A. (1996). Root-n-consistent and efficient estimation in semiparametric regression models. *Statist. Probab. Lett.* **30**, 45-51.

Severini, T.A. and Wang, W.H. (1992). Generalized profile likelihood and conditional parametric models. *Ann. Statist.* **20**, 1768-1802.

Severance-Lossin, E. and Sperlich, S. (1996). Estimation of derivatives for additive separable models. Technical report.

Speckman, P. (1988). Kernel smoothing in partially linear models. *J. Roy. Statist. Assoc. Ser. B* **50**, 413-436.

Sperlich, Tjosteim and Yang (2002), Nonparametric estimation and testing of interaction in additive models. *Econom. Theory* **18**, 197-251.

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

Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.* **14**, 592-606.

Tjostheim, D. and Auestad, B. H. (1994). Nonparametric identification of nonlinear time series: projections. *J. Amer. Statist. Assoc.* **89**, 1398-1409.

Yang, L. (2002). Direct estimation in an additive model when the components are proportional. *Statist. Sinica* **12**, 801-821.

Department of Economics, Vanderbilt University, Nashville TN 37235-1819, U.S.A.

E-mail: yanqin.fan@vanderbilt.edu

Department of Economics, Texas A&M University, College Station, TX 77843-4228, U.S.A.

E-mail: qi@econmail.tamu.edu

(Received February 2001; accepted February 2003)