Journal of Multivariate Analysis **71**, 191–240 (1999)
Article ID jmva.1999.1838, available online at http://www.idealibrary.com on IDEAL®

Asymptotic Normality of a Combined Regression Estimator*

Yanqin Fan

University of Windsor, Windsor, Ontario, Canada
E-mail: Fan1@uwindsor.ca

and

Aman Ullah

University of California, Riverside
E-mail: Ullah@mail.ucr.edu

Received August 5, 1997

In this paper, we propose a combined regression estimator by using a parametric estimator and a nonparametric estimator of the regression function. The asymptotic distribution of this estimator is obtained for cases where the parametric regression model is correct, incorrect, and approximately correct. These distributional results imply that the combined estimator is superior to the kernel estimator in the sense that it can never do worse than the kernel estimator in terms of convergence rate and it has the same convergence rate as the parametric estimator in the case where the parametric model is correct. Unlike the parametric estimator, the combined estimator is robust to model misspecification. In addition, we also establish the asymptotic distribution of the estimator of the weight given to the parametric estimator in constructing the combined estimator. This can be used to construct consistent tests for the parametric regression model used to form the combined estimator.

AMS 1991 subject classifications: 62G07, 62G10, 62J12.

Key words and phrases: asymptotic distribution; convex combination; kernel estimation; model specification testing; parametric estimation; semiparametric estimation.

* The first author thanks the Natural Sciences and Engineering Research Council of Canada and the Social Sciences and Humanities Research Council of Canada for their support. The second author thanks the Academic Senate, UCR, for the research support. We are grateful to Oliver Linton for helpful discussions and for providing several important references, and to two anonymous referees for helpful suggestions.

0047-259X/99 $30.00

1. INTRODUCTION

Consider the regression model

$$Y = \theta(X) + \varepsilon, \tag{1.1}$$

where $\theta(x) = E(Y \mid X = x)$, $x \in R^d$, is the true, but unknown regression function. Suppose that n i.i.d. observations $\{Y_i, X_i\}_{i=1}^n$ are available from (1.1).

Let $\{f(\beta, x); \beta \in R^q\}$ be a family of parametric regression functions. If $\theta(x) = f(\beta, x)$ for almost all x and for some β, then we say that the parametric family of regression functions: $\{f(\beta, x): \beta \in R^q\}$ correctly specifies $\theta(\cdot)$ or the parametric regression model given by $Y = f(\beta, X) + \varepsilon$ is correct. It is well known that in this case, one can construct a consistent estimate of β, $\hat{\beta}$ (say), which yields a consistent estimate of $\theta(x)$ given by $f(\hat{\beta}, x)$. In general, if the parametric regression model is incorrect, then $f(\hat{\beta}, x)$ may not be a consistent estimate of $\theta(x)$. However, one can still consistently estimate the unknown regression function $\theta(x)$ by various nonparametric estimation techniques such as kernel, series, and spline, among others. See Härdle (1990) for details. Compared with parametric estimators, nonparametric estimators converge more slowly and suffer from the so-called "Curse of Dimensionality," that is, their rates of convergence depend on d, the dimension of the regressor X, and become slower as d increases.

Intuitively, if the true regression function $\theta(x)$ is not too far from the parametric family $f(\beta, x)$, then the parametric estimate $f(\hat{\beta}, x)$ should contain certain useful information about $\theta(x)$ even though it is not consistent. Recently a number of authors have proposed estimators of the regression function which incorporate the information contained in the parametric model. The most studied approach is the local likelihood approach, where in the context of regression estimation, a parametric regression model such as $f(\beta, x)$ is fitted locally using kernel weights. See Hjort and Jones (1996) for applications of the local likelihood approach to density estimation, Hjort (1997) for hazard rate estimation, and Gozalo and Linton (1995) for regression function estimation. The second approach which multiplies an initial parametric estimate with a nonparametric kernel-type estimate of the necessary correction factor has been used in Hjort and Glad (1995) for density estimation and in Glad (1995) for regression estimation. Instead of combining two estimators multiplicatively as in Hjort and Glad (1995) and Glad (1998), Olkin and Speigelman (1987) (for density estimation), and Ullah and Vinod (1993), and Burman and Chaudhuri (1994) (for regression estimation), combined the parametric and the kernel estimators additively.

All the estimates developed in the above-mentioned papers share similar properties: Under certain conditions, (i) they often have smaller bias than the kernel estimates; (ii) if the parametric model is correct, then the new estimates have the parametric $\sqrt{n}$-convergence rate; (iii) if the parametric model is incorrect, then the new estimates have similar asymptotic behavior to the kernel estimates. Hence, by making use of the information contained in the parametric model, these estimates can never do worse than the corresponding kernel estimate in terms of convergence rate and are more robust to model misspecification than the parametric estimate.

In this paper, we propose an additively combined regression estimate $\hat{\theta}(x)$ (say) which is similar to the estimators in Ullah and Vinod (1993), and Burman and Chaudhuri (1994). It is given by a convex combination of the parametric and the kernel estimators with $\hat{\lambda}$ (say) being the weight given to the parametric estimator. Unlike the local likelihood approach, the computation of $\hat{\theta}(x)$ is straightforward and does not require numerical optimization. The combined estimator $\hat{\theta}(x)$ also adapts to the data (or the parametric model) automatically through $\hat{\lambda}$ in the sense that if the parametric model accurately describes the data, then $\hat{\lambda}$ converges to one, hence $\hat{\theta}(x)$ puts all the weight on the parametric estimate asymptotically and has $\sqrt{n}$-convergence rate; If the parametric model is incorrect, then $\hat{\lambda}$ converges to zero and $\hat{\theta}(x)$ puts all the weight on the kernel estimate and has the nonparametric rate of convergence. In practice, one may use $\hat{\lambda}$ as a measure of the degree of accuracy of the parametric model and hence to test the adequacy of the parametric model. For the estimates based on the first two approaches, however, this role is played by the smoothing parameter and there is not yet available a way of choosing the smoothing parameter such that the resulting estimates adapt to the data in the above sense. Moreover, $\hat{\theta}(x)$ as well as the regression estimators in Ullah and Vinod (1993), and Burman and Chaudhuri (1994) may be advantageous to the one in Glad (1998) in situations where the parametric regression estimator is not strictly bounded away from zero.

Although the convergence rate of the estimator in Burman and Chaudburi (1994) is known, its asymptotic distribution is not yet available, which prevents it from being employed in making inferences about the true regression model. In this paper, we analyze in detail the asymptotic distribution of the proposed estimator $\hat{\theta}(x)$ and of $\hat{\lambda}$, the weight given to the parametric estimator, in all possible cases: the true model is close to the parametric model; far away from the parametric model. As a consequence, we are able to characterize formally situations where the combined estimator performs better than the kernel estimate in terms of convergence rate. The asymptotic distribution of $\hat{\lambda}$ under correct parametric model specification derived in this paper may also be used to construct a consistent test for the parametric regression model. Most of the existing papers have emphasized the bias reduction aspect of their estimates and hence focused on the MSE analysis. Although it is observed in the existing papers that the above-mentioned estimates may perform better than the traditional kernel method in a broad nonparametric vicinity of the parametric model employed, their asymptotic distributions axe not fully studied. We hope that the results of this paper will facilitate the application of the combined estimate $\hat{\theta}(x)$ in practice.

The rest of this paper is organized as follows. In Section 2, we first describe the estimators in Ullah and Vinod (1993), and Burman and Chaudhuri (1994), and then introduce our estimator $\hat{\theta}(x)$. In Section 3, we show the asymptotic normality of $\hat{\theta}(x)$ for the case where the parametric regression model is correct. Section 4 considers the case where the parametric regression model is incorrect and Section 5 deals with the case where the parametric regression model is approximately correct in the sense that the distance between the true regression function $\theta(x)$ and the family of parametric regression functions $f(\beta, x)$, say δ_n, approaches zero as $n \to \infty$. Section 6 provides some Monte Carlo results. Section 7 concludes with some additional comments. The proofs of the results in Sections 3–5 are presented in the last section.

To close this section, we restate two definitions from Robinson (1988).

DEFINITION 1.1. $\mathscr{K}_l$, $l \geqslant 1$, is the class of even functions k: $R \to R$ satisfying

$$\int_R u^i k(u)\, du = \delta_{i0} \qquad (i = 0, 1, ..., l-1)$$

$$k(u) = O((1 + |u|^{l+1+\varepsilon})^{-1}) \qquad \text{some} \quad \varepsilon > 0,$$

where δ_{ij} is the Kronecker's delta.

DEFINITION 1.2. $\mathscr{G}^{\alpha}_{\mu}$, $\alpha > 0$, $\mu > 0$, is the class of functions g: $R^d \to R$ satisfying that g is $(m-1)$-times partially differentiable, for $m - 1 \leqslant \mu \leqslant m$; for some $\rho > 0$, $\sup_{y \in \phi_{z\rho}} |g(y) - g(z) - Q_g(y, z)|/|y - z|^{\mu} \leqslant h_g(z)$ for all z, where $\phi_{z\rho} = \{y\colon |y - z| < \rho\}$; $Q_g = 0$ when $m = 1$; Q_g is a $(m-1)$th degree homogeneous polynomial in $y - z$ with coefficients the partial derivatives of g at z of orders 1 through $m-1$ when $m > 1$; and $g(z)$; its partial derivatives of order $m-1$ and less, and $h_g(z)$, have finite αth moments.

2. THE COMBINED ESTIMATOR

Following the result of Olkin and Speigelman (1987) in the density estimation case, Ullah and Vinod (1993) and Burman and Chaudhuri

(1994) proposed the following combined estimate of $\theta(x)$ by using the parametric estimate $f(\hat{\beta}, x)$ and a nonparametric estimate $\tilde{g}(x)$, say

$$\hat{\theta}^*(x) = \hat{\lambda}^* f(\hat{\beta}, x) + (1 - \hat{\lambda}^*)\, \tilde{g}(x), \tag{2.1}$$

where

$$\hat{\lambda}^* = \frac{n^{-1} \sum_{i=1}^{n} \{f(\hat{\beta}^{(i)}, X_i) - \tilde{g}^{(i)}(X_i)\}\{Y_i - f(\hat{\beta}, X_i)\}}{n^{-1} \sum_{i=1}^{n} \{f(\hat{\beta}, X_i) - \tilde{g}(X_i)\}^2} + 1; \tag{2.2}$$

$\hat{\beta}^{(i)}$ and $\tilde{g}^{(i)}(X_i)$ are leave-one-out versions of $\hat{\beta}$ and $\tilde{g}(X_i)$.

We note that Ullah and Vinod (1993) did not obtain any asymptotic result for $\hat{\lambda}^*$ or $\hat{\theta}^*(x)$, and in their case the nonparametric estimate $\tilde{g}(x)$ is the kernel estimate, $\tilde{g}^{(i)}(X_i)$ is $\tilde{g}(X_i)$ in (2.2), and $f(\hat{\beta}^{(i)}, X_i) = f(\hat{\beta}, X_i) = X_i\hat{\beta}$ is linear. Burman and Chaudhuri (1994) considered the case where the regressor X is fixed taking values in a compact set of R^d and derived the rate of convergence of $\hat{\theta}^*(x)$. Specifically, they showed that the rate of convergence of $\hat{\theta}^*(x)$ depends crucially on how close the parametric model is to the true model: In the case where $f(\beta, x)$ is not too far from $\theta(x)$, the rate of convergence of $\hat{\theta}^*(x)$ is faster than that of the nonparametric estimator and does not depend on d, the dimension of the regressor X (free from "the Curse of Dimensionality"). However, they did not present the asymptotic distributions of $\hat{\theta}^*(x)$ and $\hat{\lambda}^*$. This is the purpose of the present paper.

To facilitate the derivation of the asymptotic distribution of $\hat{\theta}^*(x)$, we will specify the nonparametric estimate $\tilde{g}(x)$ as the kernel estimate $\hat{g}(x)$ defined in (2.3) below and derive the asymptotic distribution of a modified version of $\hat{\theta}^*(x)$ for random regressor X. Throughout, $\hat{\beta}$ denotes the nonlinear least squares estimate under the assumption that the parametric model is correct.

The Nadaraya–Watson kernel estimate is defined as

$$\hat{g}(x) \equiv \frac{\hat{r}(x)}{\hat{h}(x)}, \tag{2.3}$$

with

$$\hat{r}(x) = \frac{1}{na^d} \sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{a}\right), \tag{2.4}$$

$$\hat{h}(x) = \frac{1}{na^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{a}\right), \tag{2.5}$$

where K is a product kernel with univariate kernel function $k(\cdot)$ and $a \equiv a_n \to 0$ is a bandwidth. Note that $\hat{h}(x)$ is the kernel estimate of the density function, $h(x)$, of the regressor X.

Our approach to the formulation of the combined estimator of $\theta(x)$ is more straightforward than that of Burman and Chaudhuri (1994). It is based on the compound model,

$$Y_i = \lambda f(\beta, X_i) + (1-\lambda)\,\theta(X_i) + u_i, \tag{2.6}$$

or equivalently,

$$Y_i - f(\beta, X_i) = (\lambda - 1)[f(\beta, X_i) - \theta(X_i)] + u_i, \tag{2.7}$$

where u_i is the error in the compound model. Note that in (2.6) or (2.7), $\lambda = 1$ if the parametric model is correct; $\lambda = 0$ otherwise. Hence, λ can be regarded as a parameter, the value of which indicates the correctness of the parametric model. The issue is to consistently estimate λ. We construct an estimator of λ in two steps: First, we replace $f(\beta, X_i)$ and $\theta(X_i)$ in (2.7) by $f(\hat{\beta}, X_i)$ and $\hat{g}^{(i)}(X_i)$ respectively, where $\hat{g}^{(i)}(X_i)$ is the leave-one-out version of the kernel estimate $\hat{g}(X_i)$; Second, we estimate λ from the resulting model,

$$Y_i - f(\hat{\beta}, X_i) = (\lambda - 1)[f(\hat{\beta}, X_i) - \hat{g}^{(i)}(X_i)] + \hat{u}_i, \tag{2.8}$$

where $\hat{u}_i = u_i + [f(\beta, X_i) - f(\hat{\beta}, X_i)] + (1-\lambda)[\theta(X_i) - \hat{g}^{(i)}(X_i)]$.

To overcome the random denominator problem, we multiply $\hat{h}^{(i)}(X_i)$ on both sides of Eq. (2.8) and obtain the OLS estimator of λ from the equation

$$\hat{h}^{(i)}(X_i)[Y_i - f(\hat{\beta}, X_i)] = \alpha[\hat{h}^{(i)}(X_i)\, f(\hat{\beta}, X_i) - \hat{r}^{(i)}(X_i)] + \hat{h}^{(i)}(X_i)\,\hat{u}_i, \tag{2.9}$$

where $\alpha = \lambda - 1$, $\hat{h}^{(i)}(X_i)$ and $\hat{r}^{(i)}(X_i)$ are leave-one-out versions of $\hat{h}(X_i)$ and $\hat{r}(X_i)$ given in (2.4) and (2.5), respectively.

The OLS estimate of a from (2.9) is

$$\begin{aligned}\hat{\alpha} &\equiv \hat{\lambda} - 1 \\ &= \frac{n^{-1}\sum_{i=1}^{n}\{\hat{h}^{(i)}(X_i)\, f(\hat{\beta}, X_i) - \hat{r}^{(i)}(X_i)\}\{\hat{h}^{(i)}(X_i)[Y_i - f(\hat{\beta}, X_i)]\}}{n^{-1}\sum_{i=1}^{n}\{\hat{h}^{(i)}(X_i)\, f(\hat{\beta}, X_i) - \hat{r}^{(i)}(X_i)\}^2}.\end{aligned} \tag{2.10}$$

Given $\hat{\lambda}$, we can now define our combined estimator of $\theta(x)$ by

$$\hat{\theta}(x) = \hat{\lambda} f(\hat{\beta}, x) + (1-\hat{\lambda})\,\hat{g}(x). \tag{2.11}$$

By comparing the definitions of $\hat{\theta}^*(x)$ and $\hat{\theta}(x)$, we note immediately that computationally $\hat{\theta}(x)$ may be less burdensome than $\hat{\theta}^*(x)$ especially for nonlinear regression models, because $\hat{\theta}(x)$ only requires one estimate of β, i.e., $\hat{\beta}$, while $\hat{\theta}^*(x)$ requires not only $\hat{\beta}$ but also $\hat{\beta}^{(i)}$ for each and every $i = 1, 2, ..., n$. Theoretically, $\hat{\theta}(x)$ has the same rate of convergence as $\hat{\theta}^*(x)$ and is easier to handle than $\hat{\theta}(x)$. Therefore, in this paper, we will derive the asymptotic distribution of $\hat{\theta}(x)$.

Throughout the rest of this paper, for any two functions $h_1(x)$ and $h_2(x)$, we define the norm and the inner product as $\|h_1\|^2 = \int h_1^2(x)\, h(x)\, dx$ and $\langle h_1, h_2 \rangle = \int h_1(x)\, h_2(x)\, h(x)\, dx$, where we recall that $h(\cdot)$ is the density function of X. We use K_{ij} to denote $K((X_i - X_j)/c)$ and K_{xi} to denote $K((x - X_i)/a)$. Unless otherwise stated, all the limits are taken as $n \to \infty$.

3. CORRECT PARAMETRIC SPECIFICATION

In this section, we will derive the asymptotic distribution of $\hat{\theta}(x)$ when the parametric model is correctly specified. We will accomplish this by first providing an asymptotic expansion for $\hat{\theta}(x)$ up to $O_p(n^{-1/2})$, inclusive, and then applying the Central Limit Theorem (CLT) for generalized multilinear forms in De Jong (1990) to the expansion.

We will make the following assumptions:

Assumption A. (A1) The error $\varepsilon \equiv Y - \theta(X)$ satisfies $E\,|\varepsilon^4| < \infty$. The conditional variance function $\sigma^2(x) \equiv E[\varepsilon^2 \mid X = x]$ and $m_4(x) \equiv E[\varepsilon^4 \mid X = x]$ are continuous. In addition, $h(x)\,\sigma^2(x)$ and $h(x)\, m_4(x)$ are bounded on R^d. The data $\{Y_i, X_i\}_{i=1}^n$ are i.i.d.

(A2) $h \in \mathscr{G}_2^\infty$ and $\theta \in \mathscr{G}_2^4$.

(A3) The kernel function $k(\cdot)$ satisfies $k \in \mathscr{K}_2$. Let $\Omega_K = \int uu' K(u)\, du$.

(A4) The window width $a = a_n$ satisfies $a \to 0$, $na^d \to \infty$, and $na^{d+4} \to \delta$ with $0 < \delta < \infty$.

(A5) (a) Let $\mathscr{B}$ be a compact subset of R^q. The function $f\colon \mathscr{B} \times R^d \to R$ is such that for each $\beta \in \mathscr{B}$, $f(\beta, \cdot)$ is measurable; for each $x \in R^d$, $f(\cdot, x)$ is continuous, and the absolute value of $f(\cdot, X_i)$ is dominated by a square integrable function. (b) For each $x \in R^d$, $f(\cdot, x)$ is continuously differentiable of order 2 on $\mathscr{B}$, and the absolute values of the elements of $Df(\cdot, X_i)$ and $D^2 f(\cdot, X_i)$ are dominated by square integrable functions.

(A6) The true parameter value $\beta_0 \in \operatorname{int} \mathscr{B}$ such that $E[Df(\beta_0, X)\, D'f(\beta_0, X)]$ and $E[D^2 f(\beta_0, X)]$ are nonsingular.

Assumption (A1) is a standard assumption on the error ε. The assumption of boundedness of $h(x)\,\sigma^2(x)$ and $h(x)\, m_4(x)$ is not very restrictive. It allows both $\sigma^2(x)$ and $m_4(x)$ to be unbounded functions of x. Assumption (A2) imposes smoothness and moment conditions on $h(x)$ and $\theta(x)$ that were first used by Robinson (1988). For more explanation of (A2), see Robinson (1988). (A3) states that the kernel function is a second order kernel. (A4) states that the smoothing parameter is optimal in terms of minimizing the asymptotic mean square error of the kernel density estimator or of the kernel regression estimator. (A5) and (A6) are standard regularity conditions commonly used in the literature on nonlinear regression models. They ensure that the nonlinear least squares estimator $\hat{\beta}$ converges to β_0 at rate $n^{-1/2}$ and is asymptotically normally distributed (see Jennrich (1969) and White (1981)).

Denote the numerator and the denominator of $(\hat{\lambda}-1)$ in (2.10) as $\hat{\lambda}_{N1}$ and $\hat{\lambda}_{D1}$ respectively. We first provide asymptotic expansions for $\hat{\lambda}_{N1}$, $\hat{\lambda}_{D1}$, and $\hat{\lambda}-1$.

LEMMA 3.1. *Define*

$$b(x)=tr[\Omega_K\{(\partial/\partial x)\, h(x)(\partial/\partial x')\, \theta(x)+\tfrac{1}{2}h(x)(\partial^2/\partial x\, \partial x')\, \theta(x)\}]$$

and

$$B(X_i)=h(X_i)\, b(X_i)-E[h(X_1)\, b(X_1)\, Df(\beta_0, X_1)]'\, \Sigma_X^{-1}\, Df(\beta_0, X_i),$$

where $\Sigma_X=E[Df(\beta_0, X_1)\, D'f(\beta_0, X_1)]$. *Suppose that Assumption A is satisfied. Then the following results hold*:

(a) $(na^d)\, \hat{\lambda}_{D1}=\sigma_D^2+o_p(1)$, *where*

$$\sigma_D^2=\delta\int b^2(u)\, h(u)\, du+\int K^2(u)\, du\int \sigma^2(u)\, h^2(u)\, du;$$

(b) $\hat{\lambda}_{N1}=U_1+U_2+o_P((na^{d/2})^{-1})$, *where*

$$U_1=\frac{a^2}{n}\sum_{i=1}^{n}\varepsilon_i B(X_i) \qquad \textit{and}$$

$$U_2=\frac{1}{n(n-1)\, a^d}\sum\sum_{1\leqslant i<j\leqslant n}\varepsilon_i\varepsilon_j K_{ij}[h(X_i)+h(X_j)];$$

(c) $na^{d/2}(U_1+U_2) \to N(0, \sigma_u^2)$ *in distribution, where*

$$\sigma_u^2 = \delta \int \sigma^2(x)\, B^2(x)\, h(x)\, dx + 2 \int K^2(u)\, du \left[\int \sigma^4(x)\, h^4(x)\, dx \right].$$

(d) $a^{-d/2}(\hat{\lambda}-1) = \sigma_D^{-2}[na^{d/2}(U_1+U_2)] + o_p(1) \to N(0, \sigma_\alpha^2)$ *in distribution, where* $\sigma_\alpha^2 = \sigma_D^{-4}\sigma_u^2$.

It is worth mentioning the following results from Lemma 3.1: First, Lemma 3.1(d) implies that $\hat{\lambda}$ approaches 1 in probability. Hence the combined estimate $\hat{\theta}(x)$ eventually places zero weight on the kernel estimate $\hat{g}(x)$ in the case where the parametric model is correct; second, Lemma 3.1(b) and (c) imply that $na^{d/2}\hat{\lambda}_{N1} \to N(0, \sigma_u^2)$ in distribution. This or Lemma 3.1(d) can be used to construct consistent tests for the parametric specification of the unknown regression function. For example, consider using Lemma 3.1(d) in this context: The null hypothesis is H_0: $P(\theta(X) = f(\beta_0, X)) = 1$ and the alternative hypothesis is $H_1 = P(\theta(X) = f(\beta, X)) < 1$ for all β. Obviously, $\lambda = 1$ under H_0 and $\lambda = 0$ under H_1. Hence, a consistent test for H_0 versus H_1 can be constructed based on $\hat{\lambda}$ for which the asymptotic distribution of $\hat{\lambda}$ given in Lemma 3.1(d) will be very useful. We shall not get into the details here in order not to distract the reader from the main theme of this paper.

We now use Lemma 3.1 to develop an asymptotic expansion for the combined estimator $\hat{\theta}(x)$ defined in (2.11). For this, we note that the following expression holds:

$$\hat{\theta}(x) - \theta(x) = \hat{\lambda}[f(\hat{\beta}, x) - f(\beta_0, x)] - \hat{\alpha}[\hat{g}(x) - \theta(x)]. \tag{3.1}$$

From Lemma 3.1, it follows that $\hat{\lambda} = O_p(1)$ and $\hat{\alpha} = O_p(a^{d/2})$. Since $f(\hat{\beta}, x) - f(\beta_0, x) = O_p(n^{-1/2})$ and $\hat{g}(x) - \theta(x) = O_p((na^d)^{-1/2}$, we know that both terms on the right hand side of (3.1) are of the same order $n^{-1/2}$. Thus, both terms will contribute to the asymptotic distribution of $\hat{\theta}(x)$. Below, we extract the dominant terms from each of these two terms and collect them in the following proposition.

Proposition 3.2. *Let*

$$H_n(X_i) = [D'f(\beta_0, X_i)\, \Sigma_X^{-1}\, Df(\beta_0, x) - \delta b(x)\, B(X_i)/\{\sigma_D^2 h(x)\}],$$

$$H_n(X_i, X_j) = \frac{\left[\begin{array}{r} B(X_i)\, K((x-X_j)/a) + B(X_j)\, K((x-X_i)/a) \\ + b(x)\, K_{ij}[h(X_i) + h(X_j)] \end{array}\right]}{\sigma_D^2 h(x)},$$

and

$$H_n(X_i, X_j, X_k) = \left\{ K_{ij}[h(X_i) + h(X_j)]\, K\left(\frac{x - X_k}{a}\right) + K_{kj}[h(X_k) + h(X_j)]\, K\left(\frac{x - X_i}{a}\right) + K_{ik}[h(X_i) + h(X_k)]\, K\left(\frac{x - X_j}{a}\right)\right\} \Big/ \{\sigma_D^2 h(x)\}.$$

Then under the assumptions of Lemma 3.1, *we get*

$$n^{1/2}[\hat{\theta}(x) - \theta(x)] = \frac{1}{n^{1/2}} \sum_i \varepsilon_i H_n(X_i) - \frac{a^2}{n^{1/2}} \sum \sum_{i<j} \varepsilon_i \varepsilon_j H_n(X_i, X_j) - \frac{1}{n^{1/2}(n-1)\, a^d} \sum \sum \sum_{i<j<k} \varepsilon_i \varepsilon_j \varepsilon_k H_n(X_i, X_j, X_k) + o_p(1).$$

From Proposition 3.2, one can see that the asymptotic distribution of $\hat{\theta}(x)$ will be more complicated than that of $f(\hat{\beta}, x)$ due to the influence of $\hat{g}(x)$ on $\hat{\theta}(x)$. The result is stated in the following theorem.

THEOREM 3.3. *Under the assumptions of Lemma* 3.1, *we have* $n^{1/2}[\hat{\theta}(x) - \theta(x)] \to N(0, \sigma_c^2)$ *in distribution, where*

$$\sigma_c^2 = \int \left[D'f(\beta_0, u)\, \Sigma_X^{-1}\, Df(\beta_0, x) - \frac{1}{\sigma_D^2 h(x)} \delta b(x)\, B(u) \right]^2 \sigma^2(u)\, h(u)\, du + \frac{2\delta}{\sigma_D^4 h^2(x)} \left[\sigma^2(x)\, h(x) \int \sigma^2(u)\, B^2(u)\, h(u)\, du + 2b^2(x) \int \sigma^4(u)\, h^4(u)\, du \right] \int K^2(v)\, dv + \frac{12}{\sigma_D^4 h^2(x)} \left\{ \sigma^2(x)\, h(x) \int \sigma^4(u)\, h^4(u)\, du \right\} \left[\int K^2(v)\, dv \right]^2.$$

Theorem 3.3 states that in the case where the parametric model is correct, the combined estimator $\hat{\theta}(x)$ converges to $\theta(x)$ at the parametric rate and is asymptotically normally distributed. Similar results were also observed in Gozalo and Linton (1995), Clad (1998), and other papers alluded to in Section 1. However, the parametric rate of convergence of our combined estimator $\hat{\theta}(x)$ is achieved by letting the smoothing parameter approach zero with respect to sample size n (see (A4)), while the same rate of convergence of other combined estimators is only achieved when the smoothing parameter is kept fixed.

4. INCORRECT PARAMETRIC SPECIFICATION

In this section, we will investigate the asymptotic distribution of $\hat{\theta}(x)$ when the parametric model is incorrectly specified in the sense that the distance between $\theta(x)$ and the family of parametric functions $f(\beta, x)$ is a positive constant. For this, we, replace (A6) with (A6′) given below.

(A6′) There exists a unique $\beta^* \in \operatorname{int} \mathscr{B}$ such that $\beta^* = \arg\min_{\beta \in \mathscr{B}} E[Y - f(\beta, X)]^2$. In addition, $E[Df(\beta^*, X)\, D'f(\beta^*, X)]$ and $E[D^2 f(\beta^*, X)]$ are nonsingular.

White (1981) showed that under (A5) and (A6′), $\hat{\beta}$ converges to β^* at rate $n^{-1/2}$ and is asymptotically normally distributed.

As in Section 3, we first examine $\hat{\lambda}$.

LEMMA 4.1. *Under* (A1)–(A5) *and* (A6′), *the following results hold*:

(a) $\hat{\lambda}_{D1} = \int [f(\beta^*, x) - \theta(x)]^2\, h^3(x)\, dx + o_p(1)$;

(b) *The numerator of* $\hat{\lambda}$, $\hat{\lambda}_N$ *(say), satisfies* $\hat{\lambda}_N = -a^2 \int [f(\beta^*, x) - \theta(x)]\, h^2(x)\, b(x)\, dx + o_p((na^d)^{-1/2})$;

(c) $\hat{\lambda} = -a^2 b_\lambda + o_p((na^d)^{-1/2})$, *where*

$$b_\lambda = \frac{\int [f(\beta^*, v) - \theta(v)]\, h^2(v)\, b(v)\, dv}{\int [f(\beta^*, v) - \theta(v)]^2\, h^3(v)\, dv}.$$

As opposed to Lemma 3.1, Lemma 4.1 states that if the parametric model is incorrect, then the combined estimate $\hat{\theta}(x)$ places zero weight on the parametric estimate in the limit. However, the parametric estimate $f(\hat{\beta}, x)$ will affect the asymptotic distribution of $\hat{\theta}(x)$ as is obvious from the expression

$$\begin{aligned}
&(na^d)^{1/2}\, [\hat{\theta}(x) - \theta(x)] \\
&\quad = (na^d)^{1/2}\, \hat{\lambda}[f(\hat{\beta}, x) - \theta(x)] + (na^d)^{1/2}\, (1 - \hat{\lambda})[\hat{g}(x) - \theta(x)] \\
&\quad = (na^d)^{1/2}\, [-a^2 b_\lambda + o_p((na^d)^{-1/2})][f(\hat{\beta}, x) - \theta(x)] \\
&\qquad + (na^d)^{1/2}\, (1 + o_p(1))[\hat{g}(x) - \theta(x)] \\
&\quad = -\delta^{1/2} b_\lambda [f(\hat{\beta}, x) - \theta(x)] + (na^d)^{1/2}\, [\hat{g}(x) - \theta(x)] + o_p(1), \qquad (4.1)
\end{aligned}$$

where we have used Lemma 4.1. The asymptotic distribution of $\hat{\theta}(x)$ follows immediately from (4.1) and the asymptotic distribution of the kernel estimate $\hat{g}(x)$ (see, e.g., Bierens (1987)).

THEOREM 4.2. *Under* (A1)–(A5) *and* (A6′), *we have*

$$(na^d)^{1/2}\,[\hat{\theta}(x)-\theta(x)]$$
$$\to N\left(-\delta^{1/2}b_\lambda[f(\beta^*,x)-\theta(x)]+\frac{\delta b(x)}{h(x)},\frac{\sigma^2(x)}{h(x)}\int K^2(u)\,du\right)$$

in distribution.

We point out that Theorem 4.2 or (4.1) imply that in the case where the parametric model is incorrect, the combined estimate $\hat{\theta}(x)$ not only converges to $\theta(x)$ at the same rate as the kernel estimate $\hat{g}(x)$ but also has the same asymptotic variance as $\hat{g}(x)$. The parametric estimate $f(\hat{\beta},x)$ only changes the center of the asymptotic distribution of $\hat{\theta}(x)$. In the case where $b_\lambda[f(\beta^*,x)-\theta(x)]>0$, the asymptotic bias of $\hat{\theta}(x)$ is smaller than that of the kernel estimate $\hat{g}(x)$.

5. APPROXIMATELY CORRECT PARAMETRIC SPECIFICATION

Theorems 3.3 and 4.2 imply that the combined estimator $\hat{\theta}(x)$ can never do worse than the kernel estimator $\hat{g}(x)$ in terms of convergence rate no matter how far apart the parametric model and the true model are, and in the case where the parametric model is the right one, $\hat{\theta}(x)$ has the same convergence rate as the parametric estimator $f(\hat{\beta},x)$. This certainly makes the combined estimator $\hat{\theta}(x)$ superior to the kernel estimator. In practice it is rarely the case that the chosen parametric model correctly describes the data, but often the parametric model is not too far from the true model. The purpose of this section is then to examine the asymptotic distribution of $\hat{\theta}(x)$ when the parametric model is approximately correct in the sense that $\|\theta(x)-f(\beta,x)\|\to 0$ as the sample size n approaches ∞.

To make the technical analysis feasible, we consider the case where

$$\theta(x)=f(\beta_0,x)+\delta_n\Delta(x), \tag{5.1}$$

where $\delta_n\to 0$ and $\Delta(x)$ is continuous with $\|\Delta(x)\|<\infty$.

Under (5.1), $\|\theta(x)-f(\beta_0,x)\|=\delta_n\,\|\Delta(x)\|$. The rate of convergence of $\hat{\theta}(x)$ depends crucially on the magnitude of δ_n in relation to $n^{-1/2}$ and $(na^d)^{-1/2}$. We will derive the asymptotic distribution of $\hat{\theta}(x)$ for three cases:

(a) $\delta_n = o(n^{-1/2})$; (b) $\delta_n (na^d)^{1/2} \to \infty$; and (c) δ_n is between $n^{-1/2}$ and $(na^d)^{-1/2}$.

We replace (A6) with the following assumption:

(A6″) Assume $\langle Df(\beta_0, x), \Delta(x) \rangle = 0$.

Assumption (A6″) is made for convenience. It ensures that under (5.1), $\hat{\beta}$ still converges to β_0 at rate $n^{-1/2}$ and is asymptotically normally distributed.

LEMMA 5.1. *Under* (5.1), *and* (A1)–(A5), *and* (A6″), *the following results hold*:

(a) *If* $\delta_n = o(n^{-1/2})$, *then*

$$(na^d)\,\hat{\lambda}_{D1} = \sigma_D^2 + o_p(1),$$

$$\hat{\lambda}_{N1} = U_1 + U_2 + o_p((na^{d/2})^{-1}),$$

$$a^{-d/2}(\hat{\lambda} - 1) = \sigma_D^{-2}[na^{d/2}(U_1 + U_2)] + o_p(1),$$

where σ_D^2, U_1, *and* U_2 *are defined in Lemma* 3.1;

(b) *If* $\delta_n (na^d)^{1/2} \to \infty$, *then*

$$\delta_n^{-2} \hat{\lambda}_{D1} = E[h^2(X)\,\Delta^2(X)] + o_p(1),$$

$$\delta_n^{-2} \hat{\lambda}_N = [\delta_n (na^d)^{1/2}]^{-1}\,\delta^{1/2} E[\Delta(X)\,h(X)\,b(X)] + o_p([\delta_n (na^d)^{1/2}]^{-1}),$$

$$\delta_n (na^d)^{1/2}\,\hat{\lambda} = \frac{\delta^{1/2} E[\Delta(X)\,h(X)\,b(X)]}{E[h^2(X)\,\Delta^2(X)]} + o_p(1);$$

(c) *If* $n^{-1/2} = o(\delta_n)$ *and* $\delta_n = o((na^d)^{-1/2})$, *then*

$$(na^d)\,\hat{\lambda}_{D1} = \sigma_D^2 + o_p(1),$$

$$\frac{(na^d)^{1/2}}{\delta_n}\,\hat{\lambda}_{N1} = \delta^{1/2} E[h(X)\,b(X)\,\Delta(X)] + o_p(1),$$

$$(\delta_n (na^d)^{1/2})^{-1}\,(\hat{\lambda} - 1) = \sigma_D^{-2}\,\delta^{1/2} E[h(X)\,b(X)\,\Delta(x)] + o_p(1).$$

The results in Lemma 5.1 are very interesting: If the parametric model is very close to the true regression model in the sense that $\delta_n = o(n^{-1/2})$, then the behavior of $\hat{\lambda}$ is exactly the same as that in the case where the parametric model is correct, in particular, it places zero weight on the kernel estimate asymptotically; If δ_n is relatively large in the sense that

$\delta_n(na^d)^{1/2} \to \infty$, then $\hat{\lambda}$ behaves in a similar way to the case where the parametric model is incorrect; If δ_n is in between $n^{-1/2}$ and $(na^d)^{-1/2}$, then $\hat{\lambda}$ places zero weight on the kernel estimate in the limit as in (a) but at a slower rate.

THEOREM 5.2. *Under the assumptions of Lemma* 5.1, *we have*

(a) *If* $\delta_n = o(n^{-1/2})$, *then*

$$n^{1/2}[\hat{\theta}(x) - \theta(x)] \to N(0, \sigma_c^2), \qquad \textit{in distribution},$$

where σ_c^2 *is defined in Theorem* 3.3;

(b) *If* $\delta_n(na^d)^{1/2} \to \infty$, *then*

$$(na^d)^{1/2}[\hat{\theta}(x) - \theta(x)] \to N\left(-\frac{\delta^{1/2} E[\Delta(X)\, h(X)\, b(X)]}{E[h^2(X)\Delta^2(X)]}\, \Delta(x) + \frac{\delta b(x)}{h(x)}, \frac{\sigma^2(x)}{h(x)} \int K^2(u)\, du\right);$$

in distribution;

(c) *If* $n^{-1/2} = o(\delta_n)$ *and* $\delta_n = o((na^d)^{-1/2})$, *then*

$$\delta_n^{-1}[\hat{\theta}(x) - \theta(x)] \to N\left(-\frac{\delta^{3/2} b(x)}{\sigma_D^2 h(x)}\, E[h(X)\, b(X)\, \Delta(X)] - \Delta(x), \right.$$
$$\left. \frac{\delta\sigma^2(x)}{\sigma_D^4 h(x)} \{E[h(X)\, b(X)\, \Delta(X)]\}^2 \int K^2(u)\, du\right),$$

in distribution.

Theorem 5.2(a) and (c) are very encouraging. They imply that when the true regression model is not too far from the parametric model used to construct $\hat{\theta}$ in the sense that $\delta_n = o((na^d)^{-1/2})$, the convergence rate of the combined estimator $\hat{\theta}(x)$ is faster than that of the kernel estimate and is dimension-free. Thus one would expect that $\hat{\theta}(x)$ perform much better than the kernel estimator especially in high dimensions. In addition, when the true model is close to the parametric model ($\delta_n = o(n^{-1/2})$), the combined estimator has the same asymptotic distribution as in the case where the parametric model is correct. We also note from Theorem 5.2(a) and (c) that although $\hat{\theta}(x)$ places zero weight on the kernel estimate in both cases, it has different asymptotic distribution due to different rate of convergence of $\hat{\lambda}$ or $\hat{\lambda} - 1$.

There are two other cases that are not covered by Lemma 5.1 and Theorem 5.2. These are: (a) $\delta_n = O(n^{-1/2})$; (b) $\delta_n = O(na^d)^{-1/2})$. Obviously, these two cases can be handled in exactly the same way as above. For completeness, we summarize the results in the following theorem without providing the proofs.

THEOREM 5.3. *Under the assumptions of Lemma* 5.1, *we have*

(a) *If* $\delta_n n^{1/2} \to \gamma_a$, *where* $0 < \gamma_a < \infty$, *then*

$$(na^d)\,\hat{\lambda}_{D1} = \sigma_D^2 + o_p(1),$$

$$\hat{\lambda}_{N1} = U_1 + U_2 + \frac{1}{na^{d/2}}\gamma_a\,\delta^{1/2}E[h(X)\,b(X)\,\Delta(X)] + o_p((na^{d/2})^{-1}),$$

$$a^{-d/2}(\hat{\lambda} - 1) = \sigma_D^{-2}\{na^{d/2}(U_1 + U_2) + \gamma_a\,\delta^{1/2}E[h(X)\,b(X)\,\Delta(X)]\} + o_p(1).$$

Consequently, $n^{1/2}[\hat{\theta}(x) - \theta(x)] \to$

$$N\left(-\frac{\gamma_a\,\delta E[h(X)\,b(X)\,\Delta(X)]\,b(x)}{\sigma_D^2 h(x)},\ \sigma_c^2 + \frac{\gamma_a^2\,\delta\{E[h(X)\,b(X)\,\Delta(X)]\}^2\,\sigma^2(x)}{\sigma_D^4 h(x)}\int K^2(u)\,du\right),$$

in distribution;

(b) *If* $\delta_n (na^d)^{1/2} \to \gamma_b$, *where* $0 < \gamma_b < \infty$, *then*

$$(na^d)\,\hat{\lambda}_{D1} = E[\delta^{1/2}b(X) - \gamma_b h(X)\,\Delta(X)]^2 + \int K^2(u)\,du \int \sigma^2(u)\,h^2(u)\,du + o_p(1),$$

$$\delta_n^{-2}\hat{\lambda}_N = \gamma_b\,\delta^{1/2}E[\Delta(X)\,h(X)\,b(X)] + \delta E[b^2(X)] + o_p(1),$$

$$\hat{\lambda} = \frac{\gamma_b^2\{\gamma_b\,\delta^{1/2}E[\Delta(X)\,h(X)\,b(X)] + \delta E[b^2(X)]\}}{E[\delta^{1/2}b(X) - \gamma_b h(X)\,\Delta(X)]^2 + \int K^2(u)\,du \int \sigma^2(u)\,h^2(u)\,du} + o_p(1) \equiv \mu_\lambda + o_p(1).$$

In addition, $(na^d)^{1/2}\,[\hat{\theta}(x) - \theta(x)] \to N(-\gamma_b\mu_\lambda\,\Delta(x) + (1 - \mu_\lambda)(\delta b(x)/h(x)), (1 - \mu_\lambda)^2\,(\sigma^2(x)/h(x))\int K^2(u)\,du)$ *in distribution.*

We mentioned in Section 3 that both $\hat{\lambda}_{N1}$ and $\hat{\lambda}$ can be used to construct consistent model specification tests. Theorem 5.3(a) can be used to analyze the local power properties of the resulting tests. In particular, under appropriate conditions, one can show that such tests may have non-trivial power against sequences of local alternatives that converge to the true model at rate $n^{-1/2}$ depending on the sign of $E[h(X)\, b(X)\, \Delta(X)]$. This result is similar to the corresponding result in Fan (1994) in the context of testing goodness-of-fit of a parametric density function. Theorem 5.3(a) also states that if $\delta_n = O(n^{-1/2})$, then $\hat{\theta}(x)$ places zero weight on the kernel estimate. This is consistent with our earlier observations, because in this case $\delta_n = o((na^d)^{-1/2})$. However, the asymptotic distribution of $\hat{\theta}(x)$ is different. More interestingly, if $\delta_n = O((na^d)^{-1/2})$, then $\hat{\theta}(x)$ places non-zero weight on both $f(\hat{\beta}, x)$ and $\hat{g}(x)$. However, the asymptotic variance of $\hat{\theta}(x)$ is determined by the kernel estimate $\hat{g}(x)$.

6. MONTE CARLO RESULTS

In this section, we report results from a small Monte Carlo simulation study to examine the finite sample performance of the proposed combined estimator of the regression function, $\hat{\theta}(x)$ in (2.11). This combined estimator is compared with the parametric estimator $f(\hat{\beta}, x)$, kernel estimator $\hat{g}(x)$, and an alternative combined estimator due to Glad (1998) as

$$\tilde{\theta}(x) = \frac{\sum_{i=1}^{n} (Y_i f(\hat{\beta}, x)/f(\hat{\beta}, X_i))\, K((X_i - x)/a)}{\sum_{i=1}^{n} K((X_i - x)/a)}. \tag{6.1}$$

Such a Monte Carlo comparison is useful since the analytical expressions for the asymptotic bias and variance of $\hat{\theta}(x)$ are too complicated to have meaningful comparisons with the corresponding analytical expressions of the parametric, kernel, and Glad's combined estimators.

Another Monte Carlo simulation is carried out to study the behavior of the test statistics

$$T_1 = a^{-d/2} \frac{(\hat{\lambda} - 1)}{\hat{\sigma}_\alpha}, \qquad T_2 = \frac{na^{d/2} \hat{\lambda}_{N1}}{\hat{\sigma}_u} \tag{6.2}$$

for testing the parametric specification of the unknown regression function, where

$$\hat{\sigma}_\alpha^2 = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_D^4}, \tag{6.3}$$

$$\hat{\sigma}_u^2 = \frac{1}{2a^d n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} e_i^2 e_i^2 K^2 \left(\frac{X_i - X_j}{a} \right) (\hat{h}(X_i) + \hat{h}(X_j))^2, \tag{6.4}$$

$$\hat{\sigma}_D^2 = \frac{1}{a^d (n-1)^2} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} e_j^2 K^2 \left(\frac{X_i - X_j}{a} \right), \tag{6.5}$$

and $e_i = Y_i - f(\hat{\beta}, X_i)$. Under certain conditions, both T_1 and T_2 tend to $N(0, 1)$ in distribution as $n \to \infty$ (see Lemma 3.1(d) and (c)) and they are described in Section 3 following Lemma 3.1.

6.1. *Performance of* $\hat{\theta}(x)$

To conduct a Monte Carlo simulation we assume the data generating process,

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \gamma[\gamma_1 \sin(\pi(X-1)\gamma_2)] + \varepsilon, \tag{6.6}$$

where $\beta_0 = 60.50$, $\beta_1 = -17$, $\beta_2 = 2$, $\gamma_1 = 10$, $\gamma_2 = 0.44$, $\pi = 3.1416$, and γ is the misspecification parameter which determines the deviation from the parametric specification $f(\beta, X) = \beta_0 + \beta_1 + \beta_2 X^2$. This parameter γ is chosen as 0, 0.3, and 1.0 in order to consider the cases of correct parametric specification ($\gamma = 0$), approximately correct parametric specification ($\gamma = 0.3$) and incorrect parametric specification ($\gamma = 1.0$). In addition to varying γ, the sample size n is varied as $n = 50$, 100, and 500. Both X and ε are generated from standard normal populations. Further the number of replications is 1000 in all cases. Finally, the normal kernel is used in all cases and the optimal bandwidth a is taken as $a = 1.06n^{-1/5}\hat{\sigma}_x$ is the sample standard deviation of x, see Härdle (1990) for details on the choice of kernel and bandwidth.

To compare the four techniques of obtaining the fitted value $\hat{y}$: $f(\hat{\beta}, X)$, $\hat{g}(X)$, $\tilde{\theta}(X)$ and $\hat{\theta}(x)$, we present in Table I the mean (m) and standard deviation (s) of the mean squared errors (MSE), that is the mean and standard deviation of $\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ over 1000 replications, in each case and see its closeness to the variance of ε which is one. It is seen that when the parametric model is correctly specified ($\gamma = 0$) our proposed combined estimator $\hat{\theta}(x)$ performs at least as well as the parametric estimator $f(\hat{\beta}, X)$ and it out performs the kernel estimator $\hat{g}(X)$. This continues to hold when $\gamma = 0.3$, that is the parametric model is approximately correct. However, when the parametric model is incorrectly specified ($\gamma = 1$) then the combined estimator performs better than the kernel estimator $\hat{g}(X)$ for small samples but for large samples their performances are similar. Further when $\gamma = 1$ the combined estimator performs better than the parametric regression for larger sample sizes. These results are consistent with the rates results for the combined estimator in Section 3. Similar findings can be seen for the Glad's combined estimator $\tilde{\theta}(x)$ compared to parametric and kernel estimators. With regard to the comparison of the combined estimators $\hat{\theta}(x)$ and $\tilde{\theta}(x)$ we note that while both have similar performances for large samples, $\hat{\theta}(x)$ tends to out perform $\tilde{\theta}(x)$ for smaller samples.

TABLE I

Mean (m) and Standard Deviation (s) of the MSE of Fitted Values

γ	n		$f(\hat{\beta}, x)$	$\hat{g}(x)$	$\tilde{\theta}(x)$	$\hat{\theta}(x)$
0.0	50	m	0.9365	10.7280	0.8908	0.9067
		s	0.1921	2.5773	0.1836	0.1796
	100	m	0.9737	7.5312	0.9406	0.9571
		s	0.1421	1.2578	0.1393	0.1386
	500	m	0.9935	3.2778	0.9894	0.9801
		s	0.0041	0.1970	0.0636	0.0643
0.3	50	m	1.4954	11.7159	0.9299	0.9501
		s	0.5454	3.1516	0.2162	0.1892
	100	m	1.6761	8.2164	0.9602	0.9720
		s	0.4811	1.5510	0.2001	0.1870
	500	m	1.8414	3.5335	1.0374	1.0412
		s	0.2818	0.2485	0.0652	0.0646
1.0	50	m	7.2502	16.9386	2.6089	1.9710
		s	5.2514	5.3785	1.0861	0.7925
	100	m	8.7423	11.8223	2.0001	1.8679
		s	4.8891	2.7239	0.6143	0.5306
	500	m	10.4350	4.8395	1.6320	1.6300
		s	2.9137	0.4689	0.1466	0.1479

6.2. *Performance of Test Statistics T_1 and T_2*

Here we conduct a Monte Carlo simulation to evaluate the size and power of the T-tests in (6.2). The null hypothesis we want to test is the linear regression model is correct,

$$H_0: \theta(x) = E(Y \mid X = x) = f(\beta, x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2, \tag{6.7}$$

where β's are parameters and x_1 and x_2 are the regressors. To investigate the size of the test we consider the model in which the dependent variable is generated by

$$Y_i = 1 + X_{1i} + X_{2i} + \varepsilon_i, \tag{6.8}$$

where the error term ε_j is drawn independently from the standard normal distribution, and the regressors X_1 and X_2 are defined as $X_1 = Z_1$, $X_2 = (Z_1 + Z_2)/\sqrt{2}$; Z_1 and Z_2 are vectors of independent random numbers of size n. To investigate the power of the test we consider an alternative model by adding the interaction term $X_{1i}X_{2i}$ into the null model (6.7),

$$Y_i = 1 + X_{1i} + X_{2i} + X_{1i}X_{2i} + \varepsilon_i. \tag{6.9}$$

In all experiments, we consider sample sizes of 100, 300, and 500 and we perform 1000 replications. The kernel function K is chosen to be the

TABLE II

Proportion of Rejections in the Model $Y = 1 + X_1 + X_2 + \varepsilon$

Parameter c	n	Tests	1%	5%	10%
0.5	100	T_1	0.8	6.5	11.7
		T_2	0.3	3.8	10.7
	300	T_1	1.5	5.0	10.0
		T_2	0.8	5.4	10.1
	500	T_1	0.5	4.0	10.1
		T_2	0.5	4.5	10.2
1.0	100	T_1	4.1	9.7	15.2
		T_2	0.4	3.4	8.5
	300	T_1	2.2	7.2	12.7
		T_2	1.0	5.5	10.5
	500	T_1	1.5	5.5	11.3
		T_2	1.0	4.4	9.8
2.0	100	T_1	9.1	17.3	21.6
		T_2	0.5	2.0	5.8
	300	T_1	5.7	11.7	16.9
		T_2	1.1	3.3	8.0
	500	T_1	4.7	11.4	16.1
		T_2	0.9	4.7	9.4

TABLE III

Proportion of Rejections in the Model $Y = 1 + X_1 + X_2 + X_1 X_2 + \varepsilon$

Parameter c	n	Tests	1%	5%	10%
0.5	100	T_1	0.6	10.6	24.7
		T_2	15.0	36.7	47.6
	300	T_1	30.5	67.6	80.0
		T_2	66.2	82.9	90.3
	500	T_1	91.1	93.0	96.4
		T_2	91.2	96.7	98.3
1.0	100	T_1	0.4	14.2	37.2
		T_2	62.2	75.9	83.3
	300	T_1	56.5	94.5	98.4
		T_2	99.5	99.7	99.9
	500	T_1	95.9	100.0	100.0
		T_2	100.0	100.0	100.0
2.0	100	T_1	0.3	13.0	27.9
		T_2	92.2	95.6	97.0
	300	T_1	0.1	66.7	97.8
		T_2	100.0	100.0	100.0
	500	T_1	27.9	99.4	100.0
		T_2	100.0	100.0	100.0

bivariate standard normal density function and the bandwidth a is chosen to be $cn^{-2/6}$ where c is a constant. The bandwidth satisfies $na^d \to \infty$ and $na^{d+4} \to 0$ where $d=2$ in our case. To analyze whether the tests are sensitive to the choice of bandwidth we calculate the test statistics for c equal to 0.5, 1.0 and 2.0. The critical values for the tests are from the standard normal table.

The results of the size study are shown in Table II. The T_2 test has adequate size in most cases. Further the sizes get closer to the limiting sizes when n becomes large. Also the size is not much sensitive to the choice of bandwidth. The size behavior of T_1-test is generally inferior to T_2-test and it is sensitive to the choice of bandwidth.

Table III gives the results of the power study. The powers of the T_2-test are very high in most cases and quickly converge to 1. Further the power increases with sample size in all cases for any chosen bandwidth. Finally the power behavior of T_1-test is inferior to T_2-test both across the bandwidth and sample size. In view of much better size and power performances of T_2 over T_1 we recommend using T_2 in practice.

7. CONCLUDING COMMENTS

This paper proposes a new estimator $\hat{\theta}(x)$ of the regression function $\theta(x)$ by combining a parametric estimator $f(\hat{\beta}, x)$ and a kernel estimator $\hat{g}(x)$. The weight λ given to the parametric estimator is estimated by OLS procedure applied to the compound model. Under general conditions, we show that the new estimator has desirable properties: It converges to $\theta(x)$ at a rate that is no slower than that of the kernel estimator; in cases where the parametric model is not too far from the true regression model, $\hat{\theta}(x)$ converges to $\theta(x)$ at a faster rate than the kernel estimator, and its rate of convergence is independent of the dimension of X, which makes the new estimator more appealing than the kernel especially in high dimensions; when the parametric model is correct, $\hat{\theta}(x)$ converges at the same rate as the parametric estimator.

There are several issues that deserve further investigation. First, we note that in constructing $\hat{\theta}(x)$, the same kernel estimator (smoothing parameter) is used to estimate both λ and $\theta(x)$ (see (2.10) and (2.11)). This has the advantage that $\hat{\theta}(x)$ only depends on one smoothing parameter a. It is quite legitimate to use one kernel estimate (with smoothing parameter a, say) to construct $\hat{\lambda}$ and then use another (with smoothing parameter b, say) to form $\hat{\theta}(x)$. In this case, since $\hat{\lambda}$ involves averages of kernel estimates, it may be beneficial to undersmooth the data, i.e., choose a such that $na^{d+4} \to 0$, see, e.g., Carroll and Härdle (1989), Härdle, Hart, Marron, and Tsybakov (1992), and Linton (1995). However, given $\hat{\lambda}$, $\hat{\theta}(x)$ depends on the smoothing parameter b only through kernel estimate $\hat{g}_b(x)$ (say). Thus it makes sense to choose b optimally. Second, it is also possible to use other nonparametric estimators in constructing the combined estimator. One interesting case is to consider using estimates based on local likelihood approach, the simplest one being the local linear smoother; see Fan and Gijbels (1992).

8. TECHNICAL LEMMAS AND PROOFS

We first provide four technical lemmas which will be used frequently in this section. The first two lemmas can be found in, e.g., Bierens (1987) and the last two lemmas are taken from Fan and Li (1996).

LEMMA 8.1. $(1/a^d)\, E[K((X-x)/a)] = h(x) + o(1)$.

LEMMA 8.2. $(1/a^d)\, E\{K((X-x)/a)[\theta(x)-\theta(X)]\} = a^2 b(x) + o(a^2)$, *where* $b(X) = \frac{1}{2} tr(\Omega_K(\partial/\partial X)(\partial/\partial X')[\theta(X)\, h(X)]) - \frac{1}{2}\theta(X)\, tr(\Omega_K(\partial/\partial X)(\partial/\partial X')\, h(X))$.

LEMMA 8.3. *For λ, μ satisfying $l-1<\lambda \leqslant l$, $m-1<\mu\leqslant m$, where $l \geqslant 1$, $m\geqslant 1$ are integers, and for $\delta \geqslant 1$, let $f\in \mathcal{G}_\lambda^\infty$, $g \in \mathcal{G}_\mu^\delta$, $k\in \mathcal{K}_{l+m-1}$. Then,*

$$|E[\{g(X_2)-g(X_1)\}\, K_{21} \mid X_1]| \leqslant D_g(X_1)(a^{(d+\eta)}),$$

where $D_g(\cdot)$ has finite δth moment and $\eta = \min(\lambda+1, \mu)$.

LEMMA 8.4. *For μ satisfying $m-1<\mu\leqslant m$, where $m\geqslant 1$ is an integer, and for $\delta \geqslant 1$, let $g \in \mathcal{G}_\mu^{2\delta}$ and $\zeta = \min(\mu, 1)$. Suppose* $\sup_u[|u|^{\delta\zeta+d} K^\delta(u)] < \infty$. *Then,*

$$|E[\{g(X_2)-g(X_1)\}^\delta K_{21}^\delta \mid X_1]| \leqslant M_g(X_1)(a^{(\delta\zeta+d)}),$$

where $M_g(\cdot)$ has finite second moment.

In the rest of this section, we will provide shortened proofs for the results stated in Sections 3–5 for the case where the parametric regression model is linear: $f(\beta, x) = \beta' x$. In this case, $Df(\beta_0, x) = x$. Detailed proofs can be found in Fan and Ullah (1996). The corresponding proofs for nonlinear $f(\beta, x)$ are essentially the same except that there will be more non-dominant terms.

Proof of Lemma 3.1. (a) It follows from (2.10) and the definition of $\hat{\lambda}_{D1}$ that

$$\hat{\lambda}_{D1} = n^{-1} \sum_{i=1}^{n} [\hat{h}^{(i)}(X_i)(\hat{\beta}-\beta^*)' X_i]^2 + n^{-1} \sum_{i=1}^{n} [\hat{h}^{(i)}(X_i)\, \beta^{*\prime} X_i - \hat{r}^{(i)}(X_i)]^2$$

$$+ 2n^{-1} \sum_{i=1}^{n} [\hat{h}^{(i)}(X_i)(\hat{\beta}-\beta^*)' X_i][\hat{h}^{(i)}(X_i)\, \beta^{*\prime} X_i - \hat{r}^{(i)}(X_i)]$$

$$\equiv I_D + II_D + 2III_D, \tag{8.1}$$

where $\beta^* \equiv plim \hat{\beta} = \beta_0$ under correct model specification and in this case $\beta^{*\prime} x = \theta(x)$. In the following, we will show that $(na^d)\, I_D = o_p(1)$, $(na^d)\, II_D = \sigma_D^2 + o_p(1)$, and $(na^d)\, III_D = o_p(1)$. These results and (8.1) lead to (a).

First we consider

$$I_D = (\hat{\beta}-\beta^*)' \left[n^{-1} \sum_{i=1}^{n} \{\hat{h}^{(i)}(X_i)\}^2 X_i X_i' \right] (\hat{\beta}-\beta^*)$$

$$\equiv (\hat{\beta}-\beta^*)' A_D (\hat{\beta}-\beta^*). \tag{8.2}$$

With slight abuse of notation, we have

$$E\,|A_D| \leqslant E\,|\{\hat{h}^{(1)}(X_1)\}^2 X_1 X_1'|$$

$$= \frac{1}{(n-1)^2 a^{2d}} \sum\sum_{j \neq k \neq 1} E(K_{1j} K_{1k}\, |X_1 X_1'|)$$

$$+ \frac{1}{(n-1)^2 a^{2d}} \sum_{j \neq 1} E(K_{1j}^2\, |X_1 X_1'|)$$

$$= \frac{(n-2)}{(n-1)\, a^{2d}} E[\{E(K_{12} \mid X_1)\}^2\, |X_1 X_1'|]$$

$$+ \frac{1}{(n-1)\, a^{2d}} E(K_{12}^2\, |X_1 X_1'|)$$

$$= O\left(\frac{1}{a^{2d}}\right) O(a^{2d}) + O\left(\frac{1}{na^{2d}}\right) O(a^d)$$

$$= O(1) + O\left(\frac{1}{na^d}\right) = O(1),$$

by Lemma 2 and Lemma 3 in Robinson (1988), where recall that $K_{ij} = K((X_i - X_j)/a)$. Thus, we get $A_D = O_p(1)$. It follows from (8.2), (A5), and (A6) that $I_D = O_p(n^{-1}) = o_p((na^d)^{-1})$.

Next, we examine II_D: $II_D = n^{-1} \sum_{i=1}^n [\hat{h}^{(i)}(X_i)\, \beta^{*\prime} X_i - \hat{r}^{(i)}(X_i)]^2$. We'll first show that

$$II_D = c_1(n) + c_2(n) + o_p((na^d)^{-1}), \tag{8.3}$$

where

$$\begin{aligned} c_1(n) &= \frac{1}{(n-1)\, a^{2d}} E[\varepsilon_2^2 K_{21}^2], \\ c_2(n) &= \frac{1}{(n-1)^2\, a^{2d}} E\left[\sum_{i=2}^n \{\theta(X_i) - \beta^{*\prime} X_1\}\, K_{i1} \right]^2. \end{aligned} \tag{8.4}$$

Then we shall show that

$$\begin{aligned} (na^d)\, c_1(n) &= \int K^2(u)\, du \int \sigma^2(u)\, h^2(u)\, du + o(1), \\ (na^d)\, c_2(n) &= \delta \int b^2(u)\, h(u)\, du + o(1). \end{aligned} \tag{8.5}$$

For convenience, we define $\bar{\varepsilon}^{(i)}(X_i)$ and $\bar{g}^{(i)}(X_i)$ in a similar way to $\hat{g}^{(i)}(X_i)$ with ε_j and $\theta(X_j)$ replacing Y_j in $\hat{g}^{(i)}(X_i)$ respectively. Noting that $\hat{g}^{(i)}(X_i) = \bar{\varepsilon}^{(i)}(X_i) + \bar{g}^{(i)}(X_i)$, we get

$$\begin{aligned} II_D &= \frac{1}{n} \sum_i [\bar{\varepsilon}^{(i)}(X_i)\, \hat{h}^{(i)}(X_i)]^2 + \frac{1}{n} \sum_i [\bar{g}^{(i)}(X_i) - \beta^{*\prime} X_i]^2\, [\hat{h}^{(i)}(X_i)]^2 \\ &\quad + \frac{2}{n} \sum_i [\bar{\varepsilon}^{(i)}(X_i)\, \hat{h}^{(i)}(X_i)][\bar{g}^{(i)}(X_i) - \beta^{*\prime} X_i]\, \hat{h}^{(i)}(X_i) \\ &\equiv IIF + IIT + 2IIS. \end{aligned} \tag{8.6}$$

We will complete the proof of (8.3) by showing that

$$IIF = c_1(n) + o_p((na^d)^{-1}), \tag{8.7}$$

$$IIS = o_p((na^d)^{-1}), \tag{8.8}$$

and

$$IIT = c_2(n) + o_p((na^d)^{-1}). \tag{8.9}$$

Proof of (8.7). It follows from (8.6) that

$$\begin{aligned} IIF &= \frac{1}{n(n-1)^2\, a^{2d}} \sum\sum \sum_{i\neq j\neq k} \varepsilon_j \varepsilon_k K_{ji} K_{ki} \\ &\quad + \frac{1}{n(n-1)^2\, a^{2d}} \sum \sum_{i\neq j} \varepsilon_j^2 K_{ji}^2 \\ &\equiv IIF_1 + IIF_2. \end{aligned} \tag{8.10}$$

We first show $(na^d)\, IIF_1 = o_p(1)$. It suffices to show that $Var[(na^d)\, IIF_1] = o(1)$, which follows immediately from

$$\begin{aligned} Var[(na^d)\, IIF_1] &= \frac{2}{a^{2d}} E[\sigma^2(X_2)\, \sigma^2(X_3)\, K_{21} K_{31} K_{24} K_{34}] \\ &\quad + \frac{2}{na^{2d}} E[\sigma^2(X_2)\, \sigma^2(X_3)\, K_{21}^2 K_{31}^2] \\ &= O(a^d) + O(n^{-1}), \end{aligned}$$

by change of variables or Lemmas 2 and 3 in Robinson (1988).

To analyze IIF_2, we let $S_{iD} = \sum_{j\neq i} \varepsilon_j^2 K_{ji}^2$. Then, we have from (8.10), $IIF_2 = [n(n-1)^2\, a^{2d}]^{-1} \sum_i S_{iD}$. We now show $IIF_2 = c_1(n) + o_p((na^d)^{-1})$. The result (8.7) will then follow from (8.10), this result, and $(na^d)\, IIF_1 = o_p(1)$. Noting that $E[IIF_2] = c_1(n)$, it suffices to show $Var[(na^d)\, IIF_2] = o(1)$,

$$Var[(na^d)\, IIF_2] \leqslant \frac{1}{n^3 a^{2d}} [E\{S_{1D}^2\} + nCov\{S_{1D}, S_{2D}\}]. \tag{8.11}$$

The first term on the right hand side of (8.11) equals $(n^3 a^{2d})^{-1} E[E\{S_{1D}^2 \mid X_1\}] = O(n^{-1}) = o(1)$ by Lemmas 2 and 3 in Robinson (1988), since

$$E[S_{1D}^2 \mid X_1] \leqslant nE(m_4(X_2)\, K_{21}^4 \mid X_1) + n^2 \{E(\sigma^2(X_2)\, K_{21}^2 \mid X_1)\}^2.$$

Now consider the second term on the right hand side of (8.11): It equals

$$\frac{1}{n^2a^{2d}}\,[E\{S_{1D}S_{2D}\}-E\{S_{1D}\}\,E\{S_{2D}\}]$$

$$=\frac{1}{n^2a^{2d}}\sum_{i=3}^{n}[E\{\varepsilon_i^4K_{i1}^2K_{i2}^2\}-E\{\varepsilon_i^2K_{i1}^2\}\,E\{\varepsilon_i^2K_{i2}^2\}]$$

$$+\frac{1}{n^2a^{2d}}\sum_{j=3}^{n}[E\{\varepsilon_2^2K_{21}^2\varepsilon_j^2K_{j2}^2\}-E\{\varepsilon_2^2K_{21}^2\}\,E\{\varepsilon_j^2K_{j2}^2\}]$$

$$+\frac{1}{n^2a^{2d}}\sum_{i=2}^{n}[E\{\varepsilon_i^2K_{i1}^2\varepsilon_1^2K_{12}^2\}-E\{\varepsilon_i^2K_{i1}^2\}\,E\{\varepsilon_1^2K_{12}^2\}]$$

$$=\frac{1}{n^2a^{2d}}(n-2)\,E\{\varepsilon_3^4K_{31}^2K_{32}^2\}-\frac{1}{n^2a^{2d}}(n-2)\,E\{\varepsilon_3^2K_{31}^2\}\,E\{\varepsilon_3^2K_{32}^2\}$$

$$+\frac{1}{n^2a^{2d}}[(n-2)\,E\{\varepsilon_2^2K_{21}^2\varepsilon_3^2K_{32}^2\}-(n-2)\,E\{\varepsilon_2^2K_{21}^2\}\,E\{\varepsilon_3^2K_{32}^2\}]$$

$$+\frac{1}{n^2a^{2d}}[E\{\varepsilon_2^2K_{21}^2\varepsilon_1^2K_{12}^2\}+(n-2)\,E\{\varepsilon_3^2K_{31}^2\varepsilon_1^2K_{12}^2\}]$$

$$-\frac{1}{n^2a^{2d}}[(n-1)\,E\{\varepsilon_2^2K_{21}^2\}\,E\{\varepsilon_1^2K_{12}^2\}]$$

$$=\frac{1}{n^2a^{2d}}[O(a^d)+O(na^{2d})]$$

$$=o(1)\qquad\text{by Lemmas 2 and 3 in Robinson (1988).}$$

Proof of (8.8). Now the third term on the right hand side of (8.6) is $IIS=[n(n-1)\,a^d]^{-1}\sum_{i=1}^{n}[\bar{\varepsilon}^{(i)}(X_i)\,\hat{h}^{(i)}(X_i)]T_i$, where $T_i=\sum_{j\neq i}[\theta(X_j)-\beta^{*\prime}X_i]\,K_{ji}\equiv\sum_{j\neq i}t_{ji}$ with $t_{ji}=[\theta(X_j)-\beta^{*\prime}X_i]\,K_{ji}$. So, we get

$$E[(na^dIIS)^2]$$

$$=\frac{1}{n^2}\left[E\left\{\sum_i[\bar{\varepsilon}^{(i)}(X_i)\,\hat{h}^{(i)}(X_i)]^2\,T_i^2\right\}\right.$$

$$\left.+E\left\{\sum\sum_{i\neq j}[\bar{\varepsilon}^{(i)}(X_i)\,\hat{h}^{(i)}(X_i)][\bar{\varepsilon}^{(j)}(X_j)\,\hat{h}^{(j)}(X_j)]\,T_iT_j\right\}\right]. \tag{8.12}$$

Let $t_i=E[t_{ji}\mid X_i]$. Then

$$E[(T_1-t_{21})^2\mid X_1]=\sum_{j=3}^{n}Var[t_{j1}\mid X_1]+(n-2)^2\,t_1^2$$

$$\leqslant nE[t_{31}^2\mid X_1]+n^2t_1^2. \tag{8.13}$$

Define $\mathscr{F}_1^n = \sigma(X_1, ..., X_n)$. Then, $E[\{\bar{\varepsilon}^{(i)}(X_i)\,\hat{h}^{(i)}(X_i)\}^2 \mid \mathscr{F}_1^n] = (na^d)^{-2} \sum_{j\neq i} \sigma^2(X_j)\,K_{ji}^2$. Thus, the first term on the right hand side of (8.12) equals

$$
\begin{aligned}
&\frac{1}{n^3 a^{2d}} E\left[T_1^2 \sum_{i\neq 1} \sigma^2(X_i)\,K_{i1}^2 \right] \\
&\quad \leqslant \frac{2}{n^2 a^{2d}} E[(T_1 - t_{21})^2\,\sigma^2(X_2)\,K_{21}^2] + \frac{2}{n^2 a^{2d}} E[t_{21}^2 \sigma^2(X_2)\,K_{21}^2] \\
&\quad \leqslant \frac{2}{n^2 a^{2d}} \{E[\{E[(T_1 - t_{21})^2 \mid X_1]\}^2]\}^{1/2} \\
&\qquad \times \{E\{[E(\sigma^2(X_2)\,K_{21}^2 \mid X_1)]^2\}\}^{1/2} \\
&\qquad + \frac{2}{n^2 a^{2d}} \{E(t_{21}^4)\}^{1/2} \{E[\sigma^4(X_1)\,K_{21}^4]\}^{1/2} \\
&\quad \leqslant \frac{2}{n^2 a^{2d}} \{E[\{nE(t_{21}^2 \mid X_1) + n^2 t_1^2\}^2]\}^{1/2} \{O(a^d)\} \\
&\qquad + \frac{2}{n^2 a^{2d}} \{O(a^{4+d})\}^{1/2} \{O(a^d)\}^{1/2} \\
&\quad = \frac{2}{n^2 a^{2d}} \{n^2[O(a^{2+d})]^2 + n^4[O(a^{2(d+2)})]^2\}^{1/2} O(a^d) + \frac{2}{n^2 a^{2d}} O(a^{2+d}) \\
&\quad = \left\{ O\left(\frac{a^4}{n^2 a^d}\right) + O(a^8) \right\}^{1/2} + O\left(\frac{a^2}{n^2 a^d}\right) = o(1),
\end{aligned} \tag{8.14}
$$

by (8.13), Lemmas 8.3 and 8.4, and Lemmas 2 and 3 in Robinson (1988). Using the fact that $E[\{\bar{\varepsilon}^{(1)}(X_1)\,\hat{h}^{(1)}(X_1)\}\{\bar{\varepsilon}^{(2)}(X_2)\,\hat{h}^{(2)}(X_2)\} \mid \mathscr{F}_1^n] = (na^d)^{-2} \sum_{i=3}^n \sigma^2(X_i)\,K_{i1} K_{i2}$, the second term on the right hand side of (8.12) is less than or equal to

$$
\begin{aligned}
&|E[\{\bar{\varepsilon}^{(1)}(X_1)\,\hat{h}^{(1)}(X_1)\}\{\bar{\varepsilon}^{(2)}(X_2)\,\hat{h}^{(2)}(X_2)\}\; T_1 T_2]| \\
&\quad = \left| E\left[\frac{1}{n^2 a^{2d}} \sum_{i=3}^{n} \sigma^2(X_i)\,K_{i1} K_{i2} T_1 T_2 \right] \right| \\
&\quad \leqslant \frac{C}{n a^{2d}} |E[\sigma^2(X_3)\,K_{31} K_{32} (T_1 - t_{21} - t_{31})^2]| \\
&\qquad + \frac{C}{n a^{2d}} |E[\sigma^2(X_3)\,K_{31} K_{32} t_{21}^2]| + \frac{C}{n a^{2d}} |E[\sigma^2(X_3)\,K_{31} K_{32} t_{31}^2]| \\
&\quad \equiv B_{1D} + B_{2D} + B_{3D}.
\end{aligned} \tag{8.15}
$$

To analyze the order of B_{1D}, note that

$$
\begin{aligned}
E(T_1^4 \mid X_1) &\leqslant CE\left[\left\{\sum_{i=2}^{n}(t_{i1}-t_1)\right\}^4 \mid X_1\right] + Cn^4 t_1^4 \\
&= CE\left\{\left[\sum_{i=2}^{n}(t_{i1}-t_1)^4 + \sum\sum_{i\neq j=2}^{n}(t_{i1}-t_1)^2(t_{j1}-t_1)^2\right] \mid X_1\right\} \\
&\quad + Cn^4 t_1^4 \\
&\leqslant CnE(t_{21}^4 \mid X_1) + Cn^2[E(t_{21}^2 \mid X_1)]^2 + Cn^4 t_1^4. \qquad (8.16)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
B_{1D} &\leqslant \frac{1}{na^{2d}}\{E[\sigma^4(X_3)\,|K_{31}K_{32}|]\,E[(T_1-t_{21}-t_{31})^4\,|K_{31}K_{32}|]\}^{1/2} \\
&= \frac{1}{na^{2d}}\{E[\sigma^4(X_3)[E(|K_{31}| \mid X_3)]^2]\}^{1/2} \\
&\quad \times \{E[E\{(T_1-t_{21}-t_{31})^4 \mid X_1\}\,E\{|K_{31}|\,E(|K_{32}| \mid X_3) \mid X_1\}]\}^{1/2} \\
&= \frac{1}{na^{2d}}\{O(a^d)[O(n^2a^{4(1+d)}) + O(n^4a^{6d+8})]^{1/2}\} \\
&= O(a^{2+d}) + O(na^{2d+4}) = o(1),
\end{aligned}
$$

where we have used (8.16), Lemmas 8.3 and 8.4, and Lemmas 2 and 3 in Robinson (1988). Similarly, we can show that $B_{2D}=o(1)$ and $B_{3D}=o(1)$. Hence, $IIS = o_p((na^d)^{-1})$.

Proof of (8.9). From (8.6), we have $IIT = [n(n-1)^2\,a^{2d}]^{-1}\sum_{i=1}^{n} T_i^2$. Thus,

$$
Var[na^d IIT] \leqslant \frac{1}{n^3a^{2d}}E[T_1^4] + \frac{1}{n^2a^{2d}}Cov[T_1^2, T_2^2] \equiv C_1 + C_2. \qquad (8.17)
$$

From (8.16) and (8.17), it follows that $C_1 = (n^3a^{2d})^{-1}\,E[E(T_1^4 \mid X_1)] = O(a^4/n) + O(na^{2d+8}) = o(1)$. It remains to consider C_2:

$$C_2 = \frac{1}{n^2 a^{2d}} \sum\sum\sum \sum_{i \neq j \neq k \neq l} [E(t_{i1} t_{j1} t_{k2} t_{l2}) - E(t_{i1} t_{j1}) E(t_{k2} t_{l2})]$$

$$+ \frac{1}{n^2 a^{2d}} \sum\sum \sum_{j \neq k \neq l} [E(t_{j1}^2 t_{k2} t_{l2}) - E(t_{j1}^2) E(t_{k2} t_{l2})]$$

$$+ \frac{1}{n^2 a^{2d}} \sum\sum \sum_{i \neq j \neq k} [E(t_{i1} t_{j1} t_{k2}^2) - E(t_{i1} t_{j1}) E(t_{k2}^2)]$$

$$+ \frac{4}{n^2 a^{2d}} \sum\sum \sum_{i \neq j \neq l} [E(t_{i1} t_{j1} t_{i2} t_{l2}) - E(t_{i1} t_{j1}) E(t_{i2} t_{l2})]$$

$$+ \frac{1}{n^2 a^{2d}} \sum \sum_{i \neq j} [E(t_{i1}^2 t_{j2}^2) - E(t_{i1}^2) E(t_{j2}^2)]$$

$$+ \frac{C}{n^2 a^{2d}} \sum \sum_{i \neq j} [E(t_{i1} t_{j1} t_{i2} t_{j2}) - E(t_{i1} t_{j1}) E(t_{i2} t_{j2})]$$

$$+ \frac{1}{n^2 a^{2d}} \sum_{i} [E(t_{i1}^2 t_{i2}^2) - E(t_{i1}^2) E(t_{i2}^2)]$$

$$\equiv D_1 + D_2 + D_3 + D_4 + D_5 + D_6 + D_7. \tag{8.18}$$

We shall show that $C_2 = o(1)$ by showing $D_i = o(1)$; $i = 1, 2, ..., 7$. To save space, we'll only provide the proof for $D_1 = o(1)$. Those for $D_2 ..., D_7$ can be found in Fan and Ullah (1996). By independence of the X_i' and the fact that $t_{ii} = 0$, it is easy to see that the non-vanishing part of D_1 are $i = 2$ $(j = 2)$ or $l = 1$ $(k = 1)$. Hence

$$D_1 = \frac{2}{n^2 a^{2d}} \sum\sum \sum_{j \neq k \neq l = 3}^{n} [E(t_{21} t_{j1} t_{k2} t_{l2}) - E(t_{21} t_{j1}) E(t_{k2} t_{l2})]$$

$$+ \frac{2}{n^2 a^{2d}} \sum\sum \sum_{j \neq k \neq l,\, k = 3}^{n} [E(t_{i1} t_{j1} t_{k2} t_{12}) - E(t_{i1} t_{j1}) E(t_{k2} t_{12})]$$

$$\equiv D_{11} + D_{12},$$

where $\sum\sum\sum_{j \neq k \neq l = 3}^{n}$ denotes triple summation over j, k, l all from 3 to n with $j \neq k \neq l$ and $j \neq l$, whereas $\sum\sum\sum_{i \neq j \neq k,\, k = 3}^{n}$ denotes triple summation over i, j, k with k from 3 to n; i, j from 1 to n with $i \neq j \neq k$ and $i \neq k$.

Using Lemma 8.3, Lemmas 2 and 3 in Robinson (1988), we have

$$|D_{11}| \leqslant \frac{2n}{a^{2d}} [|E\{E(t_{31} \mid X_1)\, E(t_{21} t_{42} t_{52} \mid X_1)\}| + |E\{[E(t_{21} \mid X_1)]^2\}|^2]$$

$$\leqslant \frac{2n}{a^{2d}} [[E\{(E(t_{31} \mid X_1))^2\}\, E\{[E(t_{21}(E(t_{42} \mid X_2))^2 \mid X_1)]^2\}]^{1/2}$$

$$+ O(h^{4(d+\eta)})]$$

$$= \frac{2n}{a^{2d}} [O(a^{d+2})\{O(a^{2d})\, O(a^{4(d+2)})\}^{1/2} + O(a^{4(d+2)})] = o(1).$$

Similarly, we can show that $D_{12} = o(1)$.

Equations (8.7), (8.8), and the above results imply $IIT = E[IIT] + o_p((na^d)^{-1})$, where $E[IIT] = [(n-1)\, a^d]^{-2}\, E[T_1^2] = c_2(n)$ by definition.

It remains to show (8.5). Note that $E[II_D] = c_1(n) + c_2(n)$. Hence (8.5) is equivalent to $(na^d)\, E[II_D] = \sigma_D^2 + o(1)$. First, we get from (8.4)

$$(na^d)\, c_1(n) = \frac{n}{(n-1)\, a^d} \int \int \left[K^2\left(\frac{u}{a}\right) \sigma^2(-u+v)\, h(-u+v) \right] h(v)\, du\, dv$$

$$= \int K^2(u)\, du \int \sigma^2(v)\, h^2(v)\, dv + o(1).$$

Next, we have

$$c_2(n) = \frac{1}{a^{2d}} E(E\{K_{21}[\beta^{*\prime} - \theta(X_2)] \mid X_1\})^2$$

$$+ \frac{1}{(n-1)\, a^{2d}} E\{K_{21}^2[\beta^{*\prime} X_1 - \theta(X_2)]^2\}$$

$$\equiv I + II. \tag{8.19}$$

For the first term, we have $(na^d)\, I = na^{-d} E\{E[K_{21}[\beta^{*\prime} X_1 - \theta(X_2)] \mid X_1]\}^2$, where by Lemma 8.2, we know

$$E\{K_{21}[\beta^{*\prime} X_1 - \theta(X_2)] \mid X_1 = x\} = \int K\left(\frac{u-x}{a}\right) [\theta(x) - \theta(u)]\, h(u)\, du$$

$$= a^{d+2}[b(x) + o(1)].$$

Thus, by (A4), we get

$$(na^d)\, I = (na^{4+d})\, E[b(X_1)]^2 + o(1) = \delta E[b(X)]^2 + o(1).$$

Now, consider the second term on the right hand side of (8.19). By Lemma 8.3,

$$(na^d)\, II = \frac{n}{(n-1)\, a^d} E\{K_{21}^2[\beta^{*\prime} X_1 - \theta(X_2)]^2\} = o(1).$$

Finally, we examine III_D. From (8.1), it is obvious that

$$\begin{aligned} III_D &= (\hat{\beta} - \beta^*)' \left\{ \frac{1}{n(n-1)^2\, a^{2d}} \sum_i \sum_{j \neq i} \sum_{k \neq i} K_{ji} X_i [\beta^{*\prime} X_i - \theta(X_k)]\, K_{ki} \right\} \\ &\quad - (\hat{\beta} - \beta^*)' \left\{ \frac{1}{n(n-1)^2\, a^{2d}} \sum_i \sum_{j \neq i} \sum_{k \neq i} K_{ji} X_i \varepsilon_k K_{ki} \right\} \\ &\equiv (\hat{\beta} - \beta^*)'\, [F_1 - F_2], \end{aligned} \tag{8.20}$$

where it is clear from (8.20) that

$$\begin{aligned} F_1 &= \frac{1}{n(n-1)^2\, a^{2d}} \sum \sum \sum_{i \neq j \neq k} k_{ji} X_i [\beta^{*\prime} X_i - \theta(X_k)]\, K_{ki} \\ &\quad + \frac{1}{n(n-1)^2\, a^{2d}} \sum \sum_{i \neq j} K_{ji}^2 [\beta^{*\prime} X_i - \theta(X_j)]\, X_i \\ &\equiv F_{11} + F_{12}, \end{aligned} \tag{8.21}$$

and

$$\begin{aligned} F_2 &= \frac{1}{n(n-1)^2\, a^{2d}} \sum \sum \sum_{i \neq j \neq k} K_{ji} X_i \varepsilon_k K_{ki} \\ &\quad + \frac{1}{n(n-1)^2\, a^{2d}} \sum \sum_{i \neq j} K_{ji}^2 X_i \varepsilon_j \\ &\equiv F_{21} + F_{22}. \end{aligned} \tag{8.22}$$

Since $\beta^{*\prime} X_1 = \theta(X_1)$ under correct model specification, we get from (8.21)

$$\begin{aligned} E[F_{11}] &= \frac{1}{a^{2d}} E[X_1 E(K_{21} \mid X_1)\, E\{(\beta^{*\prime} X_1 - \theta(X_3))\, K_{31} \mid X_1\}] \\ &\leqslant \frac{1}{a^{2d}} E[\,|X_1 E(K_{21} \mid X_1)|\, D_\theta(X_1)\, a^{d+2}] \\ &\qquad (\text{by Lemma 8.3}) = O(a^2), \end{aligned}$$

where $E[D_\theta(X_1)] < \infty$. Similarly, we get

$$E[F_{12}] = \frac{1}{na^{2d}} E[X_1 E\{(\beta^{*\prime} X_1 - \theta(X_2))\, K_{21}^2 \mid X_1\}]$$

$$= O\left(\frac{a}{na^d}\right) = o(a^2) \qquad \text{by (A4).}$$

Thus, we have from (8.21), $E[F_1] = O(a^2)$.

We now analyze $Var[F_1]$. For clarity, we rewrite F_1 by using T_i introduced before: $F_1 = -[n(n-1)\, a^d]^{-1} \sum_i \hat{h}^{(i)}(X_i)\, X_i T_i$. Hence

$$Var[(na^d)^{1/2} F_1] = \frac{1}{n^2 a^d} Var[\hat{h}^{(1)}(X_1)\, X_1 T_1]$$

$$+ \frac{1}{na^d} Cov[\hat{h}^{(1)}(X_1)\, X_1 T_1, \hat{h}^{(2)}(X_2)\, X_2 T_2]$$

$$\equiv VF_{11} + VF_{12}.$$

Similar to the proof of (8.8), one can show that $VF_{11} = o(1)$ and $VF_{12} = o(1)$. Therefore, $F_1 = E[F_1] + o_p((na^d)^{-1/2}) = E[F_1] + o_p(a^2) = O_p(a^2)$ by (A4).

Thus, we have shown that

$$F_1 = O_p(a^2). \tag{8.23}$$

We now derive the order of F_2. For this, note from (8.22) that

$$F_{21} = \frac{1}{n(n-1)^2\, a^{2d}} \sum_k \varepsilon_k \left[\sum \sum_{i \neq j \neq k} K_{ji} X_i K_{ki} \right]$$

$$\equiv \frac{1}{n(n-1)^2\, a^{2d}} \sum_k \varepsilon_k \bar{S}_k.$$

Then, $E[F_{21}] = 0$ and

$$Var(F_{21}) = \frac{1}{n^2(n-1)^4\, a^{4d}} \sum_k E(\varepsilon_k^2 \bar{S}_k^2)$$

$$= \frac{1}{n(n-1)^4\, a^{4d}} E[\sigma^2(X_1)\, \bar{S}_1^2] = O(n^{-1}),$$

because

$$E[\sigma^2(X_1)\,\bar{S}_1^2] = \sum_{i\neq j\neq 1}\sum \sum_{k\neq l\neq 1}\sum E[K_{ji}X_iK_{1i}K_{lk}X_kK_{1k}\sigma^2(X_1)]$$
$$= O(n^4a^{4d}).$$

Thus, $F_{21} = O_p(n^{-1/2})$. Similarly, one can show that $E[F_{22}] = 0$ and

$$Var(F_{22}) = \frac{1}{n^2(n-1)^4\,a^{4d}}\sum \sum_{i\neq j}\sum_{k\neq j} E[K_{ji}^2X_iK_{jk}^2X_k\varepsilon_j^2]$$
$$= O\left(\frac{1}{n^3a^{2d}}\right) = o\left(\frac{1}{n}\right).$$

Thus, we have shown that $F_2 = O_p(n^{-1/2})$, which together with (8.20) and (8.23) imply

$$III_D = O_p(n^{-1/2}[a^2 + n^{-1/2}]) = O_p\left(\frac{a^2}{n^{1/2}} + \frac{1}{n}\right).$$

Thus, by (A4), we get

$$na^dIII_D = O_p(n^{1/2}a^{d+2} + a^d) = O_p((na^{2(d+4)})^{1/2}) + o_p(1) = o_p(1).$$

(b) From the proof of (a), one sees that it is extremely tedious to write out all the details of the derivations. Since most of the basic techniques for the proof of (b) are the same as those for (a), we'll omit some of the details and only focus our attention on the new arguments that will arise from the proof of (b).

From (2.10), it follows that

$$\hat{\lambda}_{N1} = n^{-1}\sum_{i=1}^{n} \hat{h}^{(i)}(X_i)\{f(\hat{\beta}, X_i) - f(\beta^*, X_i)\}\{\hat{h}^{(i)}(X_i)[Y_i - f(\hat{\beta}, X_i)]\}$$
$$+ n^{-1}\sum_{i=1}^{n} \{\hat{h}^{(i)}(X_i)\, f(\beta^*, X_i) - \hat{r}^{(i)}(X_i)\}\{\hat{h}^{(i)}(X_i)[Y_i - f(\hat{\beta}, X_i)]\}$$
$$\equiv F + S. \tag{8.24}$$

It is not difficult to show that under correct model specification: $F = O_p(n^{-1})$. We now examine S.

From (8.24), we have

$$
\begin{aligned}
S &= n^{-1} \sum_{i=1}^{n} [\hat{h}^{(i)}(X_i)]^2 [f(\beta^*, X_i) - \theta(X_i)][\varepsilon_i + (\beta_0 - \hat{\beta})' X_i] \\
&\quad + n^{-1} \sum_{i=1}^{n} [\hat{h}^{(i)}(X_i)]^2 [\theta(X_i) - \hat{g}^{(i)}(X_i)][\varepsilon_i + (\beta_0 - \hat{\beta})' X_i] \\
&\equiv S_1 + S_2.
\end{aligned}
\tag{8.25}
$$

Note that $S_1 = 0$ under correct model specification and

$$
\begin{aligned}
S_2 &= \frac{1}{n(n-1)^2 a^{2d}} \sum_i \sum_{j \neq i} \sum_{k \neq i} \varepsilon_i K_{ji} [\theta(X_i) - \theta(X_k)] K_{ki} \\
&\quad - \frac{1}{n(n-1)^2 a^{2d}} \sum_i \sum_{j \neq i} \sum_{k \neq i} K_{ji} [\theta(X_i) - \theta(X_k)] K_{ki} [K_i'(\hat{\beta} - \beta_0)] \\
&\quad + \frac{1}{n(n-1)^2 a^{2d}} \sum_i \sum_{j \neq i} \sum_{k \neq i} K_{ji} K_{ki} \varepsilon_k \varepsilon_i \\
&\quad - \frac{1}{n(n-1)^2 a^{2d}} \sum_i \sum_{j \neq i} \sum_{k \neq i} K_{ji} \varepsilon_k K_{ki} [X_i'(\hat{\beta} - \beta_0)] \\
&= B_1 - B_2'(\hat{\beta} - \beta_0) + B_3 - B_4'(\hat{\beta} - \beta_0).
\end{aligned}
\tag{8.26}
$$

We'll show that

$$
\begin{aligned}
B_1 &= \frac{a^2}{n} \sum_{i=1}^{n} \varepsilon_i h(X_i)\, b(X_i) + o_p((na^{d/2})^{-1}), \\
B_2 &= a^2 E\{h(X_1)\, b(X_1)\, X_1\} + o_p((na^{d/2})^{-1}), \\
B_3 &= \frac{1}{n(n-1)\, a^d} \sum_{i<j} \sum \varepsilon_i \varepsilon_j K_{ij} [h(X_i) + h(X_j)] + o_p((na^{d/2})^{-1}), \quad \text{and} \\
B_4 &= O_p(n^{-1/2}).
\end{aligned}
\tag{8.27}
$$

To save space, we shall only provide the proof for B_1. The proofs for B_2, B_3, and B_4 can be found in Fan and Ullah (1996). From (8.26), we get

$$
\begin{aligned}
B_1 &= \frac{1}{n(n-1)^2 a^{2d}} \sum \sum \sum_{i \neq j \neq k} \varepsilon_i K_{ji} [\theta(X_i) - \theta(X_k)] K_{ki} \\
&\quad + \frac{1}{n(n-1)^2 a^{2d}} \sum \sum_{i \neq j} \varepsilon_i K_{ji}^2 [\theta(X_i) - \theta(X_k)] \\
&\equiv B_{11} + B_{12}.
\end{aligned}
\tag{8.28}
$$

It is obvious from (8.28) that $B_{12}=o_p(B_{11})$. Below we show that $B_{11}=O_p((na^{d/2})^{-1})$. Hence $B_{12}=o_p((na^{d/2})^{-1})$.

$$
\begin{aligned}
B_{11} =& \frac{1}{6a^{2d}}\left[\binom{n}{3}^{-1} \sum\sum\sum_{i<j<k} [\varepsilon_i K_{ji}K_{ki}[\theta(X_i)-\theta(X_k)]\right.\\
&+\varepsilon_i K_{ki}K_{ji}[\theta(X_i)-\theta(X_j)]\\
&+\varepsilon_j K_{ij}K_{kj}[\theta(X_j)-\theta(X_k)]+\varepsilon_j K_{kj}K_{ij}[\theta(X_j)-\theta(X_i)]\\
&\left.+\varepsilon_k K_{ik}K_{jk}[\theta(X_k)-\theta(X_i)]+\varepsilon_k K_{jk}K_{ik}[\theta(X_k)-\theta(X_j)]\right]\\
=&\frac{a^2}{6}\sum\sum\sum_{i<j<k} P_n(Z_i, Z_j, Z_k).
\end{aligned}
$$

Note that $6a^{-2}B_{11}$ is a non-degenerate U-statistic with zero mean. Let $P_{n1}=E[P_n(Z_1, Z_2, Z_3)\,|\,Z_1]$ and $P_n^{(2)}=E[P_n(Z_1, Z_2, Z_3)\,|\,Z_1, Z_2]$. By Lemma 3 in Lavergne and Vuong (1996), we get: if $Var[P_n^{(2)}]=o(n)$ and $Var[P_n]=o(n^2)$ (the proofs of these two results are straightforward applications of Lemmas 8.3 and 8.4 and are thus omitted), then

$$
\begin{aligned}
B_{11} =& \frac{a^2}{6}\left\{\frac{3}{n}\sum_{i=1}^{n} P_{ni}+o_p(n^{-1/2})\right\}\\
=&\frac{1}{6a^{2d}}\left[\frac{3}{n}\sum_{i=1}^{n}\varepsilon_i E\{K_{ji}K_{ki}([\theta(X_i)-\theta(X_k)]+[\theta(X_i)-\theta(X_j)])\,|\,X_i\}\right]\\
&+o_p(n^{-1/2}a^2)\\
=&\frac{1}{6a^{2d}}\left[\frac{6}{n}\sum_{i=1}^{n}\varepsilon_i E\{K_{ji}K_{ki}[\theta(X_i)-\theta(X_k)]\,|\,X_i\}\right]+o_p(n^{-1/2}a^2)\\
=&\frac{1}{na^{2d}}\sum_{i=1}^{n}\varepsilon_i E[K_{ji}\,|\,X_i]\,E\{K_{ki}[\theta(X_i)-\theta(X_k)]\,|\,X_i\}+o_p((na^{d/2})^{-1})\\
=&\frac{a^2}{n}\sum_{i=1}^{n}\varepsilon_i h(X_i)\,b(X_i)+o_p((na^{d/2})^{-1}),
\end{aligned}
\tag{8.29}
$$

where we have used (A4), Lemmas 8.1 and 8.2. From (A4) and (8.29), it follows that $B_{11}=O_p((na^{d/2})^{-1})$.

To complete the proof of (b), we note that

$$\hat{\beta}-\beta_0=\Sigma_X^{-1}\frac{\sum_{i=1}^n X_i\varepsilon_i}{n}+o_p(n^{-1/2}).$$

Hence, (8.24), $F = O_p(n^{-1})$, (8.25), $S_1 = 0$, (8.26) and (8.27) lead to

$$\hat{\lambda}_{N1} = \frac{a^2}{n} \sum_i \varepsilon_i h(X_i)\, b(X_i) - a^2 E\{h(X_1)\, b(X_1)\, X_1\}' (\hat{\beta} - \beta_0)$$

$$+ \frac{1}{n(n-1)\, a^d} \sum_{i<j} \sum \varepsilon_i \varepsilon_j K_{ij} [h(X_i) + h(X_j)] + o_p((na^{d/2})^{-1})$$

$$= \frac{a^2}{n} \sum_i \varepsilon_i \{h(X_i)\, b(X_i) - E[h(X_i)\, b(X_i)\, X_1]' \Sigma_X^{-1} X_i\}$$

$$+ \frac{1}{n(n-1)\, a^d} \sum_{i<j} \sum \varepsilon_i \varepsilon_j K_{ij} [h(X_i) + h(X_j)] + o_p((na^{d/2})^{-1})$$

$$= U_1 + U_2 + o_p((na^{d/2})^{-1}). \tag{8.30}$$

(c) By the Liapounov CLT, we get

$$na^{d/2} U_1 = (na^{d+4})^{1/2} \left[n^{-1/2} \sum_i \varepsilon_i B(X_i) \right]$$

$$\to \delta^{1/2} N(0, E[\varepsilon_1^2 B^2(X_1)]) = N\left(0, \delta \int \sigma^2(x)\, B^2(x)\, h(x)\, dx\right). \tag{8.31}$$

To analyze U_2, we rewrite U_2 as: $U_2 \equiv [n(n-1)\, a^d]^{-1} \sum \sum_{1 \leqslant i < j \leqslant n} \bar{H}_n(Z_i, Z_j)$, where $Z_i' = (\varepsilon_i, X_i')$ and $\bar{H}_n(Z_i, Z_j) = \varepsilon_i \varepsilon_j K_{ij} [h(X_i) + h(X_j)]$. It is obvious that $\bar{H}_n(\cdot, \cdot)$ is symmetric and satisfies $E[\bar{H}_n(Z_i, Z_j) \mid Z_i] = 0$ for $i \neq j$. Further, if we let $G_n(Z_1, Z_2) = E[\bar{H}_n(Z_3, Z_1)\, \bar{H}_n(Z_3, Z_2) \mid Z_1, Z_2]$, then

$$E[G_n^2(Z_1, Z_2)] \leqslant CE\{\varepsilon_1^2 \varepsilon_1^2 E[\varepsilon_3^2 |K_{31}| \mid Z_1]\, E[\varepsilon_3^2 |K_{31}|\, K_{32}^2 \mid Z_1, Z_2]\}$$

$$= CE\{\varepsilon_1^2 \varepsilon_2^2 \varepsilon_3^2 |K_{31}|\, K_{32}^2 E[\varepsilon_3^2 |K_{31}| \mid Z_1]\}$$

$$\leqslant O(a^d)\, E\{\varepsilon_1^2 \varepsilon_2^2 \varepsilon_3^2 |K_{31}|\, K_{32}^2\} = O(a^{3d}).$$

It follows from Theorem 1 in Hall (1984) that $\sum \sum_{1 \leqslant i < j \leqslant n} \bar{H}_n(Z_i, Z_j)$ is asymptotically normally distributed with zero mean and variance given by $2^{-1} n^2 E\{\bar{H}_n^2(Z_1, Z_2)\}$, if

$$\frac{[E\{G_n^2(Z_1, Z_2)\} + n^{-1} E\{\bar{H}_n^4(Z_1, Z_2)\}]}{[E\{\bar{H}_n^2(Z_1, Z_2)\}]^2} \to 0. \tag{8.32}$$

From Lemma 2 in Robinson (1988), it follows that $E\{\bar{H}_n^4(Z_1, Z_2)\} = O(a^d)$ and $E\{\bar{H}_n^2(Z_1, Z_2)\} = O(a^d)$. Thus, the left hand side of (8.32) is of order $O(a^d) + O((na^d)^{-1}) = o(1)$. Therefore, $(na^{d/2})U_2 \to N(0, (2a^d)^{-1} E\{\bar{H}_n^2(Z_1, Z_2)\})$, where

$$E\{\bar{H}_n^2(Z_1, Z_2)\} = E\{\varepsilon_1^2 \varepsilon_2^2 K_{12}^2 [h(X_1) + h(X_2)]^2\}$$

$$= 4a^d \int K^2(u)\, du \int \sigma^4(x)\, h^4(x)\, dx + o(a^d).$$

The result in (c) then follows from the observation that U_1 and U_2 are uncorrelated, and Delta method.

(d) The result in (d) follows from Eq. (2.10) and Lemma 3.1(a), (b), (c). ∎

Proof of Proposition 3.2. From Lemma 3.1(c) and (d), we get $\hat{\alpha} = \sigma_D^{-2} a^{d/2} [na^{d/2}(U_1 + U_2)] + o_p(a^{d/2})$ and $\hat{\lambda} = 1 + O_p(a^{d/2}) = 1 + o_p(1)$. Hence, we have by (2.11),

$$\hat{\theta}(x) - \theta(x) = [f(\hat{\beta}, x) - \theta(x)] - \frac{a^{d/2}}{\sigma_D^2 \hat{h}(x)} [na^{d/2}(U_1 + U_2)]$$

$$\times \{\hat{h}(x)[\hat{g}(x) - \theta(x)]\} + o_p(n^{-1/2}). \tag{8.33}$$

Note that under correct model specification,

$$f(\hat{\beta}, x) - \theta(x) = (\hat{\beta} - \beta_0)'\, x = \frac{1}{n} \sum_{i=1}^{n} [\varepsilon_i X_i' \Sigma_X^{-1} x] + o_p(n^{-1/2}). \tag{8.34}$$

In addition, from Bierens (1987), we have by (A4)

$$\hat{h}(x)[\hat{g}(x) - \theta(x)]$$

$$= \frac{1}{na^d} \sum_i \varepsilon_i K\left(\frac{x - X_i}{a}\right) + \frac{1}{na^d} \sum_i [\theta(x_i) - \theta(x)]\, K\left(\frac{x - X_i}{a}\right)$$

$$= \frac{1}{na^d} \sum_i \varepsilon_i K\left(\frac{x - X_i}{a}\right) + a^2 b(x) + o_p((na^d)^{-1/2}). \tag{8.35}$$

Let $Z_i = \{\varepsilon_i, X_i'\}'$ and $\bar{P}_n(Z_i, Z_j) = \varepsilon_i \varepsilon_j K_{ij}[h(X_i) + h(X_j)]$. We have from (8.35)

$$
\begin{aligned}
&(U_1 + U_2)\{\hat{h}(x)[\hat{g}(x) - \theta(x)]\} \\
&\quad = \frac{a^4}{n} b(x) \sum_i \varepsilon_i B(X_i) + \frac{a^2}{n^2 a^d} \sum \sum_{i \neq j} \varepsilon_i \varepsilon_j B(X_i)\, K\left(\frac{x - X_j}{a}\right) \\
&\qquad + \frac{a^2}{n^2 a^d} \sum_i \varepsilon_i^2 B(X_i)\, K\left(\frac{x - X_i}{a}\right) \\
&\qquad + \frac{a^2 b(x)}{n(n-1)\, a^d} \sum \sum_{i<j} \varepsilon_i \varepsilon_j K_{ij}[h(X_i) + h(X_j)] \\
&\qquad + \frac{1}{n^2 (n-1)\, a^{2d}} \sum \sum_{i<j} \sum_k \varepsilon_i \varepsilon_j \varepsilon_k K_{ij}[h(X_i) + h(X_j)]\, K\left(\frac{x - X_k}{a}\right) \\
&\qquad + o_p(n^{-3/2} a^{-d}) \\
&\quad \equiv A_1 + A_2 + A_3 + A_4 + A_5 + o_p(n^{-3/2} a^{-d}).
\end{aligned} \tag{8.36}
$$

Note that from (8.36), we know

$$
A_2 = \frac{a^2}{n^2 a^d} \sum \sum_{i<j} \varepsilon_i \varepsilon_j \left[B(X_i)\, K\left(\frac{x - X_j}{a}\right) + B(X_j)\, K\left(\frac{x - X_i}{a}\right) \right]. \tag{8.37}
$$

It is easy to show that $A_3 = O_p(a^2 n^{-1}) = o_p(n^{-3/2} a^{-d})$. Now consider A_5. From (8.36),

$$
\begin{aligned}
A_5 &= \frac{1}{n^2 (n-1)\, a^{2d}} \sum \sum_{i<j} \sum_{k \neq i,\, k \neq j} \varepsilon_i \varepsilon_j \varepsilon_k K_{ij}[h(X_i) + h(X_j)]\, K\left(\frac{x - X_k}{a}\right) \\
&\quad + \frac{1}{n^2 (n-1)\, a^{2d}} \sum \sum_{i<j} \varepsilon_i \varepsilon_j^2 K_{ij}[h(X_i) + h(X_j)]\, K\left(\frac{x - X_j}{a}\right) \\
&\quad + \frac{1}{n^2 (n-1)\, a^{2d}} \sum \sum_{i<j} \varepsilon_j \varepsilon_i^2 K_{ij}[h(X_i) + h(X_j)]\, K\left(\frac{x - X_i}{a}\right) \\
&\equiv A_{51} + A_{52} + A_{53}.
\end{aligned} \tag{8.38}
$$

One can show that $(n^{3/2} a^d)\, A_{52} = o_p(1)$ and $(n^{3/2} a^d)\, A_{53} = o_p(1)$; see Fan and Ullah (1996). Now we consider A_{51}. From (8.38), we get

$$A_{51} = \frac{1}{n^2(n-1)\,a^{2d}} \sum\sum\sum_{i<j<k} \varepsilon_i\varepsilon_j\varepsilon_k \left\{K_{ij}[h(X_i)+h(X_j)]\,K\left(\frac{x-X_k}{a}\right)\right.$$

$$+K_{kj}[h(X_k)+h(X_j)]\,K\left(\frac{x-X_i}{a}\right)$$

$$\left.+K_{ik}[h(X_i)+h(X_k)]\,K\left(\frac{x-X_j}{a}\right)\right\}. \tag{8.39}$$

Therefore, $A_5 = A_{51} + o_p(n^{-3/2}a^{-d})$ and (8.36)–(8.39) imply

$$[(na^{d/2})(U_1+U_2)]\{(na^d)^{1/2}\,\hat{h}(x)[\hat{g}(x)-\theta(x)]\}$$

$$= n^{1/2}a^{d+4}b(x)\sum_i \varepsilon_i B(X_i)$$

$$+\frac{a^2}{n^{1/2}}\sum\sum_{i<j}\varepsilon_i\varepsilon_j\left\{B(X_i)\,K\left(\frac{x-X_j}{a}\right)\right.$$

$$\left.+B(X_j)\,K\left(\frac{x-X_i}{a}\right)+b(x)\,K_{ij}[h(X_i)+h(X_j)]\right\}$$

$$+\frac{1}{n^{1/2}(n-1)\,a^d}\sum\sum\sum_{i<j<k}\varepsilon_i\varepsilon_j\varepsilon_k$$

$$\times\left\{K_{ij}[h(X_i)+h(X_j)]\,K\left(\frac{x-X_k}{a}\right)\right.$$

$$+K_{kj}[h(X_k)+h(X_j)]\,K\left(\frac{x-X_i}{a}\right)$$

$$\left.+K_{ik}[h(X_i)+h(X_k)]\,K\left(\frac{x-X_j}{a}\right)\right\}+o_p(1)$$

$$\equiv n^{1/2}a^{d+4}b(x)\sum_i \varepsilon_i B(X_i)+\frac{a^2}{n^{1/2}}\sum\sum_{i<j}\varepsilon_i\varepsilon_j[\sigma_D^2 h(x)\,H_n(X_i,X_j)]$$

$$+\frac{1}{n^{1/2}(n-1)\,a^d}\sum\sum\sum_{i<j<k}\varepsilon_i\varepsilon_j\varepsilon_k[\sigma_D^2 h(x)\,H_n(X_i,X_j,X_k)]$$

$$+o_p(1). \tag{8.40}$$

Hence, (8.33), (8.34), and (8.40) yield

$$
\begin{aligned}
n^{1/2}[\hat{\theta}(x)-\theta(x)]
&=\frac{1}{n^{1/2}}\sum_i[\varepsilon_i X_i'\Sigma_X^{-1}x]\\
&\quad-\frac{n^{3/2}a^d}{\sigma_D^2\hat{h}(x)}\{(U_1+U_2)\,\hat{h}(x)[\hat{g}(x)-\theta(x)]\}+o_p(1)\\
&=\frac{1}{n^{1/2}}\sum_{i=1}^n\varepsilon_i H_n(X_i)-\frac{a^2}{n^{1/2}}\sum\sum_{i<j}\varepsilon_i\varepsilon_j H_n(X_i,X_j)\\
&\quad-\frac{1}{n^{1/2}(n-1)\,a^d}\sum\sum\sum_{i<j<k}\varepsilon_i\varepsilon_j\varepsilon_k H_n(X_i,X_j,X_k)+o_p(1). \quad \blacksquare
\end{aligned}
$$

Proof of Theorem 3.3. From Proposition 3.2, it follows that $n^{1/2}[\hat{\theta}(x)-\theta(x)]$ has the same limiting distribution as $n^{1/2}U_n$, where

$$
\begin{aligned}
U_n&=\frac{1}{n}\sum_{i=1}^n\varepsilon_i H_n(X_i)-\frac{a^2}{n}\sum\sum_{i<j}\varepsilon_i\varepsilon_j H_n(X_i,X_j)\\
&\quad-\frac{1}{n(n-1)\,a^d}\sum\sum\sum_{i<j<k}\varepsilon_i\varepsilon_j\varepsilon_k H_n(X_i,X_j,X_k).
\end{aligned}
\tag{8.41}
$$

Noting that the three terms on the right hand side of (8.41) are uncorrelated, one can show that $n^{1/2}U_n$ or $n^{1/2}[\hat{\theta}(x)-\theta(x)]$ is asymptotically normally distributed with zero mean and variance $Avar[n^{1/2}U_n]$ by first applying the CLTs in De Jong (1990) and Hall (1984), and then using the Delta method. Since the proof is similar to that of Lemma 3.1(c), it is omitted.

It remains to evaluate the asymptotic variance of U_n. From (8.41), we have by noting that the three terms on the right hand side of (8.41) are uncorrelated,

$$
\begin{aligned}
Var[n^{1/2}U_n]&=Var[\varepsilon_1 H_n(X_1)]+\frac{a^4}{n}Var\left[\sum\sum_{i<j}\varepsilon_i\varepsilon_j H_n(X_i,X_j)\right]\\
&\quad+\frac{1}{n(n-1)^2\,a^{2d}}Var\left[\sum\sum\sum_{i<j<k}\varepsilon_i\varepsilon_j\varepsilon_k H_n(X_i,X_j,X_k)\right]\\
&\equiv v1+v2+v3.
\end{aligned}
$$

One can easily show that

$$v2 = \frac{na^{d+4}}{\sigma_D^4 h^2(x)} \left\{ 2\sigma^2(x)\, h(x) \int \sigma^2(u)\, B^2(u)\, h(u)\, du \right.$$
$$\left. + 4b^2(x) \int \sigma^4(u)\, h^4(u)\, du \right\} \int K^2(v)\, dv + o(1), \tag{8.43}$$

and

$$[\sigma_D^4 h^2(x)]\, v3 = 12\sigma^2(x)\, h(x) \left[\int \sigma^4(x_1)\, h^4(x_1)\, dx_1 \right] \left[\int K^2(x_2)\, dx_2 \right]^2$$
$$+ o(1). \tag{8.44}$$

Therefore, we have from (8.42), (8.43), and (8.44), $Var[n^{1/2}(\hat{\theta}(x) - \theta(x))] = \sigma_c^2 + o(1)$. ∎

Proof of Lemma 4.1. Because the proof is very similar to that of Lemma 3.1, we will only provide a sketch.

(a) Obviously, (8.1) still holds. It is also true that $(na^d)\, I_D = o_p(1)$. Hence, we only need to re-evaluate II_D and III_D on the right hand side of (8.1).

For II_D, we'll only provide details for evaluating $E[II_D]$. Similar but more tedious derivations show that $Var[II_D] = o(1)$. Consequently, $II_D = E[II_D] + o_p(1)$. Note that it is still true that $E[II_D] = c_1(n) + c_2(n) = c_1(n) + I + II$ from (8.1) and (8.19). Obviously, the result that $na^d c_1(n) = O(1)$ is still correct so that $c_1(n) = o(1)$. Now consider I and II. First

$$I = \frac{1}{a^{2d}} E[(E\{K_{21}[\beta^{*\prime} X_1 - \theta(X_2)] \mid X_1\})^2]$$
$$= E([\beta^{*\prime} X_1 - \theta(X_1)]\, h(X_1) + o(1))^2$$
$$= \int [\beta^{*\prime} x - \theta(x)]^2\, h^3(x)\, dx + o(1).$$

It is easy to see that $II = O((na^d)^{-1}) = o(1)$. Thus, we have shown that $E[II_D] = \int [\beta^{*\prime} x - \theta(x)]^2\, h^3(x)\, dx + o(1)$.

Finally we consider III_D: Note that (8.20)–(8.22) still hold and $F_2 = O_p(n^{-1/2})$ is correct, but we need to re-evaluate F_1: From (8.21), it is evident that $E\,|F_{11}| \leqslant a^{-2d} E\,|K_{21} X_1 [\beta^{*\prime} X_1 - \theta(X_3)]\, K_{31}| = O(1)$. Hence, $F_{11} = O_p(1)$. Similarly, one can easily show that $F_{12} = O_p((na^d)^{-1}) = o_p(1)$. Thus, it follows from (8.21) that $F_1 = O_p(1)$. (8.20), (A5), and (A6′) imply $III_D = O_p(n^{-1/2})$.

The result follows immediately from (8.1) and the above discussion.

(b) From (2.10), it is easy to see that

$$\hat{\lambda}_N = n^{-1} \sum_{i=1}^{n} \{\hat{h}^{(i)}(X_i)\}^2 \{f(\hat{\beta}, X_i) - f(\beta^*, X_i)\}[Y_i - \hat{g}^{(i)}(X_i)]$$
$$+ n^{-1} \sum_{i=1}^{n} \{\hat{h}^{(i)}(X_i)\}^2 \{f(\beta^*, X_i) - \hat{g}^{(i)}(X_i)\}[Y_i - \hat{g}^{(i)}(X_i)]$$
$$\equiv F_N + S_N. \tag{8.45}$$

It is straightforward to show that

$$F_N = n^{-1} \sum_{i=1}^{n} \{\hat{h}^{(i)}(X_i)\}^2 (\hat{\beta} - \beta^*)' X_i \varepsilon_i$$
$$+ n^{-1} \sum_{i=1}^{n} \{\hat{h}^{(i)}(X_i)\}^2 (\hat{\beta} - \beta^*)' X_i [\theta(X_i) - \hat{g}^{(i)}(X_i)]$$
$$= o_p(n^{-1/2}) = o_p((na^d)^{-1/2}). \tag{8.46}$$

Now we evaluate S_N. Rewrite S_N as

$$S_N = n^{-1} \sum_{i=1}^{n} \{\hat{h}^{(i)}(X_i)\}^2 [\beta^{*\prime} X_i - \theta(X_i)][Y_i - \hat{g}^{(i)}(X_i)]$$
$$+ n^{-1} \sum_{i=1}^{n} \{\hat{h}^{(i)}(X_i)\}^2 [\theta(X_i) - \hat{g}^{(i)}(X_i)][Y_i - \hat{g}^{(i)}(X_i)]$$
$$\equiv S_{N1} + S_{N2}, \tag{8.47}$$

where

$$S_{n1} = n^{-1} \sum_{i=1}^{n} \{\hat{h}^{(i)}(X_i)\}^2 [\beta^{*\prime} X_i - \theta(X_i)]\, \varepsilon_i$$
$$+ n^{-1} \sum_{i=1}^{n} \{\hat{h}^{(i)}(X_i)\}^2 [\beta^{*\prime} X_i - \theta(X_i)][\theta(X_i) - \hat{g}^{(i)}(X_i)] \tag{8.48}$$

and

$$S_{N2} = n^{-1} \sum_{i=1}^{n} \{\hat{h}^{(i)}(X_i)\}^2 [\theta(X_i) - \hat{g}^{(i)}(X_i)]\, \varepsilon_i$$
$$+ n^{-1} \sum_{i=1}^{n} \{\hat{h}^{(i)}(X_i)\}^2 [\theta(X_i) - \hat{g}^{(i)}(X_i)]^2. \tag{8.49}$$

Noting that

$$\hat{h}^{(i)}(X_i)[\theta(X_i) - \hat{g}^{(i)}(X_i)]$$
$$= \frac{1}{(n-1)\,a^d} \sum_{j\neq i} [\theta(X_i) - \theta(X_j)]\, K_{ji} - \frac{1}{(n-1)\,a^d} \sum_{j\neq i} \varepsilon_j K_{ji},$$

we get from (8.48) and (8.49),

$$\begin{aligned}
S_{N1} &= \frac{1}{n(n-1)^2\,a^{2d}} \sum_i \sum_{j\neq i} \sum_{k\neq i} K_{ji}K_{ki}[\beta^{*\prime} X_i - \theta(X_i)]\,\varepsilon_i \\
&\quad + \frac{1}{n(n-1)^2\,a^{2d}} \sum_i \sum_{j\neq i} \sum_{k\neq i} K_{ji}K_{ki}[\beta^{*\prime} X_i - \theta(X_i)][\theta(X_i) - \theta(X_k)] \\
&\quad - \frac{1}{n(n-1)^2\,a^{2d}} \sum_i \sum_{j\neq i} \sum_{k\neq i} K_{ji}K_{ki}[\beta^{*\prime} X_i - \theta(X_i)]\,\varepsilon_k \\
&= a^{-2d} E\{K_{21}[\beta^{*\prime} X_1 - \theta(X_1)][\theta(X_1) - \theta(X_3)]\,K_{31}\} \\
&\quad + o_p((na^d)^{-1/2}). \qquad (8.50)
\end{aligned}$$

Similar to the analysis of S_2 in (8.26), one can show that

$$\begin{aligned}
S_{N2} &= \frac{1}{n(n-1)^2\,a^{2d}} \sum_i \sum_{j\neq i} \sum_{k\neq i} K_{ji}K_{ki}[\theta(X_i) - \theta(X_k)]\,\varepsilon_i \\
&\quad - \frac{1}{n(n-1)^2\,a^{2d}} \sum_i \sum_{j\neq i} \sum_{k\neq i} K_{ji}K_{ki}\varepsilon_k\varepsilon_i \\
&\quad + \frac{1}{n(n-1)^2\,a^{2d}} \sum_i \sum_{j\neq i} \sum_{k\neq i} K_{ji}K_{ki}[\theta(X_i) - \theta(X_j)][\theta(X_i) - \theta(X_k)] \\
&\quad + \frac{1}{n(n-1)^2\,a^{2d}} \sum_i \sum_{j\neq i} \sum_{k\neq i} K_{ji}K_{ki}\varepsilon_k\varepsilon_j \\
&\quad - \frac{2}{n(n-1)^2\,a^{2d}} \sum_i \sum_{j\neq i} \sum_{k\neq i} K_{ji}K_{ki}[\theta(X_i) - \theta(X_j)]\,\varepsilon_k \\
&= O_p\left(\frac{1}{na^{d/2}}\right) + O_p\left(\frac{1}{na^{d/2}}\right) + O(a^4) + O_p\left(\frac{1}{na^{d/2}}\right) + O_p\left(\frac{1}{na^{d/2}}\right) \\
&= O_p\left(\frac{1}{na^{d/2}}\right) + O(a^4) = o_p((na^d)^{-1/2}) \qquad \text{by (A4).} \qquad (8.51)
\end{aligned}$$

Consequently, we have from (8.45)–(8.51) that

$$\hat{\lambda}_N = a^{-2d} E\{K_{21}[\beta^{*\prime} X_1 - \theta(X_1)][\theta(X_1) - \theta(X_3)]\, K_{31}\} + o_p((na^d)^{-1/2})$$

$$= a^{-2d} E\{[\beta^{*\prime} X_1 - \theta(X_1)]\, E[K_{21} \mid X_1]\, E[(\theta(X_1) - \theta(X_3))\, K_{31} \mid X_1]\}$$

$$+ o_p((na^d)^{-1/2})$$

$$= -a^2 E\{[\beta^{*\prime} X_1 - \theta(X_1)]\, h(X_1)\, b(X_1)\} + o_p((na^d)^{-1/2}),$$

where we have used Lemma 8.1 and Lemma 8.2.

(c) Noting that $\hat{\lambda} = \hat{\lambda}_N / \hat{\lambda}_{D1}$, we obtain immediately (c) from (a) and (b). ∎

Proof of Lemma 5.1. We first analyze $\hat{\lambda}_{D1}$. Because the argument follows that of (a) in Lemma 3.1, we only point out the necessary modifications.

Obviously, (8.1) still holds, where $\beta^* = \beta_0$ under (A6″), but $\theta(x) - f(\beta, x) = \delta_n \Delta(x)$. It is easy to see that the derivation of the order of I_D is still correct so that

$$I_D = O_p\left(\frac{1}{n}\right) = o_p\left(\frac{1}{na^d}\right). \tag{8.52}$$

We now examine II_D. As before, we only consider $E[II_D]$. For this, we note that $E[II_D] = c_1(n) + c_2(n) = c_1(n) + I + II$ from (8.19). Also, the orders of II and $c_1(n)$ are still correct, i.e., $II = o((na^d)^{-1})$ and $c_1(n) = (na^d)^{-1}\, [\int K^2(u)\, du \int \sigma^2(v)\, h^2(v)\, dv + o(1)]$. It remains to consider I: From (8.19), we have $I = a^{-2d} E(E\{K_{21}[\beta^{*\prime} X_1 - \theta(X_2)] \mid X_1\})^2$, where as in the proof of Lemma 3.1, we obtain

$$E\{K_{21}[\beta^{*\prime} X_1 - \theta(X_2)] \mid X_1 = x\}$$

$$= E\{K_{21}[\theta(X_1) - \theta(X_2)] \mid X_1 = x\} = \delta_n \Delta(x)\, E\{K_{21} \mid X_1 = x\}$$

$$= a^{d+2}[b(x) + o(1)] - \delta_n a^d \Delta(x)[h(x) + o(1)].$$

Hence,

$$I = E(a^2[b(X_1) + o(1)] - \delta_n \Delta(X_1)[h(X_1) + o(1)])^2$$

$$= a^4 E[b^2(X_1)] + \delta_n^2 E[h^2(X_1)\, \Delta^2(X_1)] - 2a^2 \delta_n E[b(X_1)\, \Delta(X_1)\, h(X_1)]$$

$$+ o(a^4 + \delta_n^2 + a^2 \delta_n).$$

Consequently, we have shown that

$$II_D = \{a^4 E[b^2(X_1)] + \delta_n^2 E[h^2(X_1)\, \Delta^2(X_1)] - 2a^2\delta_n E[b(X_1)\, \Delta(X_1)\, h(X_1)]$$
$$+ o_p(a^4 + \delta_n^2 + a_n^2)\} + o_p\left(\frac{1}{na^d}\right)$$
$$+ \frac{1}{na^d}\left[\int K^2(u)\, du \int \sigma^2(u)\, h^2(u)\, du + o_p(1)\right]$$
$$= \frac{1}{na^d}\sigma_D^2 \delta_n^2 E[h^2(W_1)\, \Delta^2(X_1)] - 2a^2\delta_n E[b(X_1)\, \Delta(X_1)\, h(X_1)]$$
$$+ o_p(\delta_n^2 + a^2\delta_n) + o_p\left(\frac{1}{na^d}\right). \tag{8.53}$$

For III_D, a brief review of the derivation reveals that (8.20) and (8.21) are still valid, so is the order of F_2: $F_2 = O_p(n^{-1/2})$, but we need to modify the analysis of F_1 in the following way: To save space, we only examine $E[F_1]$ or equivalently $E[F_{11}]$ and $E[F_{12}]$. First, from (8.21) it follows that

$$E[F_{11}] = \frac{1}{a^{2d}} E[K_{21} X_1\{\beta^{*\prime} X_1 - \theta(X_3)\}\, K_{31}]$$
$$= \frac{1}{a^{2d}} E[K_{21} X_1\{\theta(X_1) - \theta(X_3) - \delta_n \Delta(X_1)\}\, K_{31}]$$
$$= \frac{1}{a^{2d}} E[K_{21} X_1\{\theta(X_1) - \theta(X_3)\}\, K_{31}] - \frac{\delta_n}{a^{2d}} E[K_{21} X_1 \Delta(X_1)\, K_{31}].$$

As in the proof of Lemma 3.1, one can easily show by using the above expression that $E[F_{11}] = O(a^2) + O(\delta_n)$. Similarly, one can show that $E[F_{12}] = O(a/(na^d)) + O(\delta_n/(na^d))$. Hence, we have $F_1 = O_p(a^2 + \delta_n + a(na^d)^{-1} + \delta_n(na^d)^{-1})$. From (8.20), we have

$$II_D = O_p\left(\frac{a^2}{n^{1/2}} + \frac{\delta_n}{n^{1/2}} + \frac{a}{n^{3/2}a^d} + \frac{\delta_n}{n^{3/2}a^d} + \frac{1}{n}\right)$$
$$= o_p\left(\frac{1}{na^d}\right) + O_p\left(\frac{\delta_n}{n^{1/2}}\right). \tag{8.54}$$

Equations (8.1), (8.52), (8.53), and (8.54) imply

$$\hat{\lambda}_{D1} = \frac{1}{na^d}\sigma_D^2 + \delta_n^2 E[h^2(X_1)\, \Delta^2(X_1)] - 2a^2\delta_n E[b(X_1)\, \Delta(X_1)\, h(X_1)]$$
$$+ o_p(\delta_n^2 + a^2\delta_n) + o_p\left(\frac{1}{na^d}\right) + O_p\left(\frac{\delta_n}{n^{1/2}}\right). \tag{8.55}$$

(a) and (c) If $\delta_n = o((na^d)^{-1/2})$, then one can see that $(na^d)[\delta_n n^{-1/2} + \delta_n^2 + a^2\delta_n] = o(1)$. Hence, we get from (8.55) the desired result.

(b) If $\delta_n (na^d)^{1/2} \to \infty$, then it is obvious that $\delta_n^{-2}[(na^d)^{-1} + a^2\delta_n + \delta_n n^{-1/2}] = o(1)$. Hence, $\delta_n^{-2}\hat{\lambda}_{D1} = E[h^2(X_1)\,\Delta^2(X_1)] + o_p(1)$.

We now analyze $\hat{\lambda}_{N1}$ by modifying the proof of Lemma 3.1(b). Obviously, (8.24) is still correct. We first consider F. From (8.24), it follows that

$$
\begin{aligned}
F &= \frac{1}{n} \sum_i [\hat{h}^{(i)}(X_i)]^2 (\hat{\beta} - \beta^*)' X_i \varepsilon_i \\
&\quad + \frac{1}{n} \sum_i [\hat{h}^{(i)}(X_i)]^2 (\hat{\beta} - \beta^*)' X_i X_i' (\hat{\beta} - \beta^*) \\
&\quad + \frac{\delta_n}{n} \sum_i [\hat{h}^{(i)}(X_i)]^2 (\hat{\beta} - \beta^*)' X_i \Delta(X_i) \\
&= O_p\left(\frac{1}{n} + \frac{\delta_n}{n^{1/2}}\right).
\end{aligned} \tag{8.56}
$$

Now consider S. For simplicity, we let S_Δ denote S under (5.1) and reserve S for the case where $\theta(x) = f(\beta, x)$. From (8.24) and (8.25), it follows that

$$
\begin{aligned}
S_\Delta &= \left\{ S_1 + \frac{\delta_n}{n} \sum_i [\hat{h}^{(i)}(X_i)]^2 [f(\beta^*, X_i) - \theta(X_i)]\, \Delta(X_i) \right\} \\
&\quad + \left\{ S_2 + \frac{\delta_n}{n} \sum_i [\hat{h}^{(i)}(X_i)]^2 [\theta(X_i) - \hat{g}^{(i)}(X_i)]\, \Delta(X_i) \right\} \\
&\equiv S_{1\Delta} + S_{2\Delta}.
\end{aligned} \tag{8.57}
$$

From (8.25) and (8.57), we get

$$
\begin{aligned}
S_{1\Delta} &= -\frac{\delta_n}{n} \sum_i [\hat{h}^{(i)}(X_i)]^2\, \Delta(X_i)\, \varepsilon_i \\
&\quad + \frac{\delta_n}{n} \sum_i [\hat{h}^{(i)}(X_i)]^2\, \Delta(X_i)(\hat{\beta} - \beta_0)' X_i - \frac{\delta_n}{n} \sum_i [\hat{h}^{(i)}(X_i)]^2\, \Delta^2(X_i) \\
&= O_p\left(\frac{\delta_n}{n^{1/2}} + \delta_n^2\right).
\end{aligned} \tag{8.58}
$$

On inspecting S_2 in (8.25), it is obvious that

$$
S_2 = U_1 + U_2 + o_p\left(\frac{1}{na^{d/2}}\right). \tag{8.59}
$$

It remains to consider $S_{2\Delta} - S_2$. From (8.57), we have

$$\begin{aligned} S_{2\Delta} - S_2 &= \frac{\delta_n}{n} \sum_i [\hat{h}^{(i)}(X_i)]^2 \, [\theta(X_i) - \hat{g}^{(i)}(X_i)] \, \Delta(X_i) \\ &= \frac{\delta_n}{n(n-1)^2 \, a^{2d}} \sum_i \sum_{j \neq i} \sum_{k \neq i} K_{ji} K_{ki} [\theta(X_i) - \theta(X_k)] \, \Delta(X_i) \\ &\quad + \frac{\delta_n}{n(n-1)^2 \, a^{2d}} \sum_i \sum_{j \neq i} \sum_{k \neq i} K_{ji} K_{ki} \varepsilon_k \Delta(X_i) \\ &\equiv B_5 + B_7. \end{aligned} \tag{8.60}$$

Comparing B_5 with B_2, one gets immediately from (8.30) that

$$B_5 = \delta_n a^2 E[h(X) \, b(X) \, \Delta(X)] + o_p(\delta_n a^2) + O_p\left(\frac{\delta_n}{n^{1/2}}\right).$$

Similar to B_4, one can show that $B_6 = O_p(n^{-1/2} \delta_n)$. Hence we have

$$S_{2\Delta} - S_2 = \delta_n a^2 E[h(X) \, b(X) \, \Delta(X)] + o_p(\delta_n a^2) + O_p\left(\frac{\delta_n}{n^{1/2}}\right). \tag{8.61}$$

Combining (8.60) and (8.61) yields

$$\begin{aligned} S_{2\Delta} &= U_1 + U_2 + \delta_n a^2 E[h(X) \, b(X) \, \Delta(X)] \\ &\quad + o_p(\delta_n a^2) + O_p\left(\frac{\delta_n}{n^{1/2}}\right) + o_p\left(\frac{1}{n a^{d/2}}\right). \end{aligned} \tag{8.62}$$

It follows from (8.57), (8.58), and (8.62) that

$$\begin{aligned} S_{\Delta} &= U_1 + U_2 + \delta_n a^2 E[h(X) \, b(X) \, \Delta(X)] \\ &\quad + o_p(\delta_n a^2) + O_p\left(\frac{\delta_n}{n^{1/2}}\right) + o_p\left(\frac{1}{n a^{d/2}}\right) + O_p(\delta_n^2). \end{aligned} \tag{8.63}$$

Finally, we obtain from (8.24), (8.56), and (8.63),

$$\begin{aligned} \hat{\lambda}_{N1} &= U_1 + U_2 + \delta_n a^2 E[h(X) \, b(X) \, \Delta(X)] \\ &\quad + o_p(\delta_n a^2) + O_p\left(\frac{\delta_n}{n^{1/2}}\right) + o_p\left(\frac{1}{n a^{d/2}}\right) + O_p(\delta_n^2). \end{aligned} \tag{8.64}$$

(a) If $\delta_n = o(n^{-1/2})$, then $(n a^{d/2})[\delta_n a^2 + (\delta_n / n^{1/2}) + \delta_n^2] = o(1)$. Hence, we obtain from (8.64) the result $\hat{\lambda}_{N1} = U_1 + U_2 + o_p((n a^{d/2})^{-1})$.

(c) If $n^{-1/2} = o(\delta_n)$ and $\delta_n = o((na^d)^{-1/2})$, then one can show that

$$\begin{aligned}(na^d)^{1/2}\,\delta_n^{-1}\,\hat{\lambda}_{N1} &= O_p(1/\delta_n n^{1/2}) + \delta^{1/2} E[h(X)\,b(X)\,\Delta(X)] + o_p(1)\\ &\quad + o_p((na^{d+4})^{1/2}) + O_p(a^{d/2}) + O_p(\delta_n (na^d)^{1/2})\\ &= \delta^{1/2} E[h(X)\,b(X)\,\Delta(X)] + o_p(1).\end{aligned}$$

It remains to show (b). For this, we follow the corresponding proof of Lemma 4.1. It is evident that (8.45) still holds. Using (8.46) and similar arguments in analyzing III_D in (8.1), one can show that

$$F_N = O_p\left(\frac{a^2}{n^{1/2}} + \frac{1}{n}\right). \tag{8.65}$$

We now consider S_N. Note that (8.47), (8.48), and (8.49) are still correct. The first equality in (8.50) still holds which implies

$$\begin{aligned}S_{N1} &= -\frac{\delta_n}{n(n-1)^2\,a^{2d}} \sum_i \sum_{j\neq i} \sum_{k\neq i} K_{ji} K_{ki} \Delta(X_i)\,\varepsilon_i\\ &\quad -\frac{\delta_n}{n(n-1)^2\,a^{2d}} \sum_i \sum_{j\neq i} \sum_{k\neq i} K_{ji} K_{ki} \Delta(X_i)[\theta(X_i) - \theta(X_k)]\\ &\quad +\frac{\delta_n}{n(n-1)^2\,a^{2d}} \sum_i \sum_{j\neq i} \sum_{k\neq i} K_{ji} K_{ki} \varepsilon_k\\ &= O_p\left(\frac{\delta_n}{n^{1/2}}\right) - \delta_n\left[\frac{1}{a^{2d}} E\{K_{21}K_{31}\Delta(X_1)[\theta(X_1) - \theta(X_3)]\} + o_p(a^2)\right]\\ &\quad + O_p\left(\frac{\delta_n}{n^{1/2}}\right)\\ &= \delta_n a^2 E[\Delta(X_1)\,h(X_1)\,b(X_1)] + O_p\left(\frac{\delta_n}{n^{1/2}}\right) + o_p(\delta_n a^2).\end{aligned} \tag{8.66}$$

The order of S_{N2} given in (8.51) is still correct, i.e., $S_{N2} = O_p((na^{d/2})^{-1}) + O_p(a^4)$. Hence,

$$\begin{aligned}S_N &= \delta_n a^2 E[\Delta(X_1)\,h(X_1)\,b(X_1)]\\ &\quad + O_p\left(\frac{\delta_n}{n^{1/2}}\right) + o_p(\delta_n a^2) + O_p\left(\frac{1}{na^{d/2}}\right) + O_p(a^4).\end{aligned} \tag{8.67}$$

Equations (8.46), (8.65), and (8.67) imply

$$\hat{\lambda}_N = \delta_n a^2 E[\Delta(X_1)\, h(X_1)\, b(X_1)] + O_p\left(\frac{\delta_n}{n^{1/2}}\right) + o_p(\delta_n a^2)$$

$$+ O_p\left(\frac{1}{na^{d/2}}\right) + O_p(a^4) + O_p\left(\frac{a^2}{n^{1/2}}\right) + O_p\left(\frac{1}{n}\right). \tag{8.68}$$

Let $\tau_n = [\delta_n (na^d)^{1/2}]^{-1} \to 0$. Then we have the above expression that

$$\tau_n^{-1} \delta_n^{-2} \hat{\lambda}_N = (na^{d+4})^{1/2}\, E[\Delta(X_1)\, h(X_1)\, b(X_1)] + O_p(a^{d/2}) + o_p((na^{d+4})^{1/2})$$

$$+ O_p\left(\frac{a^{d/2}}{\delta_n (na^d)^{1/2}}\right) + O_p\left(\frac{na^{d+4}}{\delta_n (na^d)^{1/2}}\right)$$

$$+ O_p\left(\frac{(na^{d+4})^{1/2}\, a^{d/2}}{(na^d)^{1/2}\, \delta_n}\right) + O_p\left(\frac{a^d}{(na^d)^{1/2}\, \delta_n}\right)$$

$$= \delta^{1/2} E[\Delta(X_1)\, h(X_1)\, b(X_1)] + o_p(1). \quad \blacksquare$$

Proof of Theorem 5.2. From (2.11), we get

$$\hat{\theta}(x) - \theta(x) = \hat{\lambda}(\hat{\beta} - \beta^*)'\, x - \delta_n \hat{\lambda} \Delta(x) - \hat{\alpha}[\, \hat{g}(x) - \theta(x)]. \tag{8.69}$$

(a) If $\delta_n = o(n^{-1/2})$, then $\delta_n \hat{\lambda} = o_p(n^{-1/2})$. The result follows from (8.69), Proposition 3.2, and Theorem 3.3;

(b) If $\delta_n (na^d)^{1/2} \to \infty$, then from Lemma 5.1, we obtain

$$(na^d)^{1/2}\, [\hat{\theta}(x) - \theta(x)]$$

$$= (na^d)^{1/2}\, \hat{\lambda}(\hat{\beta} - \beta^*)'\, x - (na^d)^{1/2}\, \delta_n \hat{\lambda} \Delta(x) - (na^d)^{1/2}\, \hat{\alpha}[\, \hat{g}(x) - \theta(x)]$$

$$= O_p\left((na^d)^{1/2} \frac{1}{\delta_n (na^d)^{1/2}}\, n^{-1/2}\right)$$

$$- \left[\frac{\delta^{1/2} E[\Delta(X)\, h(X)\, b(X)]}{E[h^2(X)\, \Delta^2(X)]} + o_p(1)\right] \Delta(x)$$

$$- [-1 + o_p(1)](na^d)^{1/2}\, [\, \hat{g}(x) - \theta(x)]$$

$$= -\frac{\delta^{1/2} E[\Delta(X)\, h(X)\, b(X)]}{E[h^2(X)\, \Delta^2(X)]} + (na^d)^{1/2}\, [\, \hat{g}(x) - \theta(x)] + o_p(1).$$

(c) If $n^{-1/2}=o(\delta_n)$ and $\delta_n=o((na^d)^{-1/2})$, then

$$\begin{aligned}\delta_n^{-1}[\hat{\theta}(x)-\theta(x)] &= \delta_n^{-1}\hat{\lambda}(\hat{\beta}-\beta^*)'\,x-\hat{\lambda}\Delta(x)-\delta_n^{-1}\hat{\alpha}[\hat{g}(x)-\theta(x)]\\ &= o_p(1)-\Delta(x)-\left[\frac{\delta^{1/2}E[h(X)\,b(X)\,\Delta(X)]}{\sigma_D^2}+o_p(1)\right]\\ &\quad\times(na^d)^{1/2}\,[\hat{g}(x)-\theta(x)]. \quad \blacksquare\end{aligned}$$

REFERENCES

1. H. J. Bierens, Kernel estimators of regression functions, *in* "Advances in Econometrics," Cambridge Univ. Press, Cambridge, UK, 1987.
2. P. Burman and P. Chaudhuri, A hybrid approach to parametric and nonparametric regression, manuscript University of California, Davis, 1994.
3. R. J. Caroll and W. Härdle, Second order effects in semiparametric weighted least squares regression, *Statistics* **2** (1989), 179–187.
4. J. Fan and I. Gijbels, Variable bandwidth and local linear regression smoothers, *Ann. Statist.* **20** (1992), 2008–2037.
5. Y. Fan, Testing the goodness-of-fit of a parametric density function by kernel method, *Econom. Theory* **10**, No. 2 (1994), 316–357.
6. Y. Fan and Q. Li, Consistent model specification tests: Omitted variables and semi-parametric functional forms, *Econometrica* **64** (1996), 865–890.
7. Y. Fan and A. Ullah, "Asymptotic Normality of a Combined Regression Estimator," Technical Report, University of Windsor, 1996.
8. P. De Jong, A central limit theorem for generalized quadratic forms, *Probab. Theory Related Fields* **75** (1987), 261–277.
9. P. De Jong, A central limit theorem for generalized multilinear forms, *J. Multivariate Anal.* **34** (1990), 275–289.
10. I. K. Glad, Nonparametric correction of parametric regression estimates, *Scand. J. Statist.* **25** (1998), 649–668.
11. P. Gozalo and O. Linton, Using parametric information in nonparametric regression, manuscript Brown University, 1995.
12. P. Hall, Central limit theorem for integrated square error of multivariate nonparametric density estimators, *J. Multivariate Anal.* **14** (1984), 1–17.
13. W. Härdle, "Applied Nonparametric Regression," Cambridge Univ. Press, Cambridge, UK, 1990.
14. W. Härdle, J. Hart, J. S. Marron, and A. B. Tsybakov, Bandwidth choice for average derivative estimation, *J. Amer. Statist. Assoc.* **87** (1992), 218–227.
15. N. L. Hjort, Dynamic likelihood hazard estimation, *Biometrika*, in press.
16. N. L. Hjort and I. K. Glad, Nonparametric density estimation with a parametric start, *Ann. Statist.* **23** (1995), 882–904.
17. N. L. Hjort and M. C. Jones, Locally parametric nonparametric density estimation, *Ann. Statist.* **24** (1996), 1619–1647.
18. R. I. Jennrich, Asymptotic properties of nonlinear least squares estimators, *Ann. Math. Statist.* **40** (1969), 633–643.
19. P. Lavergne and Q. Vuong, Nonparametric selection of regressors: The nonnested case, *Econometrica* **64** (1996), 207–219.

20. O. Linton, Second order approximation in the partially linear regression model, *Econometrica* **63** (1995), 1079–1112.
21. I. Olkin and C. H. Speigelman, A semiparametric approach to density estimation, *J. Amer. Statist. Assoc.* **82** (1987), 858–865.
22. P. M. Robinson, Root-N consistent semiparametric regression, *Econometrica* **56**, No. 4 (1988), 931–954.
23. A. Ullah and H. D. Vinod, General nonparametric regression estimation and testing in econometrics, *in* "Handbook of Statistics" (G. S. Maddala, C. R. Rao, and H. D. Vinod, Eds.), Vol. 11, pp. 85–117, 1993.
24. H. White, Consequences and detection of misspecified nonlinear regression models, *Amer. Statist. Assoc.* **76**, No. 314 (1981), 419–433.