6-2014

Adaptive Nonparametric Regression with Conditional Heteroskedasticity

Sainan JIN
Singapore Management University, snjin@smu.edu.sg

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Zhijie XIAO
Boston University

Adaptive Nonparametric Regression with Conditional Heteroskedasticity*

Sainan Jin[a], Liangjun Su[a], Zhijie Xiao[b]

[a] *School of Economics, Singapore Management University*

[b] *Department of Economics, Boston College*

March 12, 2014

Abstract

In this paper, we study adaptive nonparametric regression estimation in the presence of conditional heteroskedastic error terms. We demonstrate that both the conditional mean and conditional variance functions in a nonparametric regression model can be estimated adaptively based on the local profile likelihood principle. Both the one-step Newton-Raphson estimator and the local profile likelihood estimator are investigated. We show that the proposed estimators are asymptotically equivalent to the infeasible local likelihood estimators (e.g., Aerts and Claeskens, 1997), which require knowledge of the error distribution. Simulation evidence suggests that when the distribution of the error term is different from Gaussian, the adaptive estimators of both conditional mean and variance can often achieve significant efficiency over the conventional local polynomial estimators.

JEL classifications: C13, C14

Key Words: Adaptive Estimation, Conditional Heteroskedasticity, Local Profile Likelihood Estimation, Local Polynomial Estimation, Nonparametric Regression, One-step Estimator.

1 Introduction

We consider the following regression model:

$$Y_i = m(X_i) + \sigma(X_i)\,\varepsilon_i,\ \ i = 1, ..., n, \tag{1.1}$$

where ε_i is independent and identically distributed (IID hereafter) with mean zero and variance one, X_i is a $d \times 1$ IID independent variable, $m(\cdot)$ and $\sigma^2(\cdot)$ are assumed to be unknown smooth functions. For

*We would like to thank Oliver Linton, Wolfgang Härdle, Peter Phillips and two referees for helpful comments on early versions of the paper. Jin gratefully acknowledges the financial support from the SMU research grant (#09-C244-SMU-011). Su gratefully acknowledges the Singapore Ministry of Education for Academic Research Fund under grant number MOE2012-T2-2-021. Xiao thanks Boston College for research support. Address correspondence to: Sainan Jin, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903. Phone: (65)6828-0431. E-mail: snjin@smu.edu.sg.

simplicity, we will assume that ε_i is independent of X_i and has an unknown density function. We are interested in estimating the infinite dimensional parameters $m(\cdot)$ and $\sigma^2(\cdot)$ adaptively in the sense that they are asymptotically equivalent to the infeasible likelihood estimators which require knowledge of the error distribution.

Model (1.1) has attracted a lot of research attention in the last two decades; see Härdle and Tsybakov (1997), Ruppert et al. (1997), Fan and Yao (1998), Akritas and van Keilegom (2001), and Ziegelmann (2002), among others. Härdle and Tsybakov (1997) consider local polynomial estimation of the volatility function in a nonparametric autoregression model; Ruppert et al. (1997) study the conditional bias and variance of the local polynomial estimates of variance functions. Fan and Yao (1998) suggest estimating the conditional mean function first and then the conditional variance function and they show that their estimator of the conditional variance function is asymptotically adaptive to the unknown conditional mean function. Akritas and van Keilegom (2001) are interested in estimating the distribution of ε_i after estimating the conditional mean and variance functions. Ziegelmann (2002) propose a local linear exponential tilting estimator of the conditional variance function to ensure its positivity. Nevertheless, all estimators of the conditional variance functions reviewed here are based on the least squares principle and none of them takes into account the error distribution.

Motivated by efficiency considerations, Linton and Xiao (2007) study adaptive estimation for model (1.1) in the case where $\sigma^2(X_i) = \sigma^2$ almost surely (a.s.). They propose an adaptive estimator in the sense that it is asymptotically equivalent to the infeasible local likelihood estimator of Staniswalis (1989) and Fan et al. (1998), which requires the knowledge of the error distribution. In the case where conditional heteroskedasticity is present, i.e., $\sigma^2(\cdot)$ is not a constant function, the density of $u_i \equiv \sigma(X_i)\varepsilon_i$ also has mean zero but is a multiplicative convolution of the two terms and such that an estimator based on a direct application of Linton and Xiao (2007) may not be adaptive unless ε_i is symmetric about zero. In the current paper we propose jointly estimating the location and scale parameters $\left(m(x), \sigma^2(x)\right)$ efficiently by a feasible multiparameter local likelihood method.

There are several advantages associated with our approach. First, our estimator takes into account the useful information in the error distribution and is adaptive to the unknown error distribution. Second, by estimating the conditional mean and variance functions jointly, we relax the symmetry assumption on the error density, which is very helpful for applications in empirical finance and economics, since both conditional heteroskedasticity and asymmetric error distributions have frequently been detected in practice. Third, compared to Linton and Xiao (2007), additional issues arise due to the estimation of conditional variance and its nonnegativity. We resolve these issues by using a link function for the variance parameter.

In this paper, we consider the model (1.1) where the regression function $m(\cdot)$ is assumed to be a general smooth function of the regressor x. It is well-known that such models, although general, suffer from the curse-of-dimensionality problem and have a slow rate of convergence when the dimension of x is high. In the case of high dimensional covariates, other types of models such as the additive functions may be considered to avoid the curse of dimensionality. See, e.g., Claeskens and Aerts (2000) for nonparametric estimation of additive multiparameter models. But to conserve space, we limit our attention to the model (1.1).

The rest of the paper is organized as follows. We introduce the model and estimators in Section 2 and the asymptotic properties of the proposed estimators in Section 3. In section 4 we provide results from a small Monte Carlo experiment evaluating the finite sample performance of the adaptive estimators. Section 5 concludes. All proofs are relegated to the appendix.

Throughout the paper, we use $f^{(j)}$ to denote the jth derivatives of a function f. When $j = 1, 2$, we also use f' and f'' to denote the first and the second order derivatives, respectively. For a matrix A, we use $\|A\|$ to denote its Frobenius norm $\{\mathrm{tr}(A^\top A)\}^{1/2}$ and $A^\top$ its transpose. Let I_n denote the $n \times n$ identity matrix. For a $d \times 1$ multi-index vector $\mathbf{j} = (j_1, ..., j_d)'$ and a general $d \times 1$ vector $x = (x_1, ..., x_d)'$, we follow Masry (1996a, 1996b) and Linton and Xiao (2007) and use the following notation

$$\mathbf{j}! = \prod_{s=1}^{d} j_s!, \ \ |\mathbf{j}| = \sum_{s=1}^{d} j_s, \ \ x^{\mathbf{j}} = \prod_{s=1}^{d} x_s^{j_s}, \ \text{and} \ \sum_{0 \le |\mathbf{j}| \le p} = \sum_{k=0}^{p} \sum_{\substack{j_1=0 \\ j_1+...+j_d=k}}^{k} \cdots \sum_{j_d=0}^{k}.$$

2 The Model and Estimator

We introduce the multi-parameter likelihood model in Section 2.1 where the likelihood function is assumed to be known. This infeasible estimator serves as an efficiency standard with which we can compare the proposed adaptive estimator. We propose an adaptive estimator in Section 2.2, where we study the efficient estimation in the case when the error density is unknown and has to be estimated from the data.

2.1 The Multi-parameter Likelihood Model

Suppose that we have a random sample $\{(X_1, Y_1), ..., (X_n, Y_n)\}$, where $X_i \in \mathbb{R}^d$ and $Y_i \in \mathbb{R}$, from the nonparametric regression model (1.1). We consider the regression model (1.1) where ε_i is independent of X_i and has mean zero and variance one. Assume that ε_i admits a Lebesgue density f. We are interested in estimating $\left(m(x), \sigma^2(x)\right)$ at some interior point x.

Although kernel estimation or other types of methods can be used, in this paper we give asymptotic analysis based on the local polynomial procedure. See Fan (1992, 1993) and Fan and Gijbels (1996) for discussions on the attractive properties of local polynomial estimators.

For a kernel function K and a bandwidth parameter h, let $K_h(\cdot) \equiv K(\cdot/h)/h^d$. Following the notation of Masry (1996a, b), let $N_l \equiv (l + d - 1)!/(l!(d-1)!)$ be the number of distinct d-tuples $\mathbf{j}$ with $|\mathbf{j}| = l$. Arrange the N_l d-tuples as a sequence in a lexicographical order (with highest priority to last position so that $(0, 0, ..., l)$ is the first element in the sequence and $(l, 0, ..., 0)$ is the last element), and let ϕ_l^{-1} denote this one-to-one map. For each $\mathbf{j}$ with $0 \le |\mathbf{j}| \le 2p$, let $\mu_{\mathbf{j}}(K) \equiv \int_{\mathbb{R}^d} u^{\mathbf{j}} K(u) du$, $\nu_{\mathbf{j}}(K) \equiv \int_{\mathbb{R}^d} u^{\mathbf{j}} K^2(u) du$, and define the $N \times N$ dimensional matrices M and Γ and $N \times N_{p+1}$ matrix B, where $N \equiv \sum_{l=0}^{p} N_l$, by

$$M \equiv \begin{bmatrix} M_{0,0} & M_{0,1} & ... & M_{0,p} \\ M_{1,0} & M_{1,1} & ... & M_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ M_{p,0} & M_{p,1} & ... & M_{p,p} \end{bmatrix}, \ \Gamma \equiv \begin{bmatrix} \Gamma_{0,0} & \Gamma_{0,1} & ... & \Gamma_{0,p} \\ \Gamma_{1,0} & \Gamma_{1,1} & ... & \Gamma_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{p,0} & \Gamma_{p,1} & ... & \Gamma_{p,p} \end{bmatrix}, \ B \equiv \begin{bmatrix} M_{0,p+1} \\ M_{1,p+1} \\ \vdots \\ M_{p,p+1} \end{bmatrix}, \tag{2.1}$$

where $M_{i,j}$ and $\Gamma_{i,j}$ are $N_i \times N_j$ dimensional matrices whose (l, r) elements are, respectively, $\mu_{\phi_i(l)+\phi_j(r)}$ and $\nu_{\phi_i(l)+\phi_j(r)}$. Note that the elements of the matrices $M = M(K, p)$ and $\Gamma = \Gamma(K, p)$ are simply multivariate moments of the kernel K and K^2, respectively; and the matrix $B = B(K, p)$ depends on the kernel and the order of the local polynomial in use. In addition, we arrange the N_s elements of the derivatives

$$\left(D^{\mathbf{s}}m\right)(x) \equiv \frac{1}{s_1! \cdots s_d!} \frac{\partial^{|\mathbf{s}|} m(x)}{\partial^{s_1} x_1 ... \partial^{s_d} x_d}, \quad |\mathbf{s}| = s,$$

as an $N_s \times 1$ column vector $m^{(s)}(x)$ in the lexicographical order. $\left(D^{\mathbf{s}}\sigma^2\right)(x)$ and $\boldsymbol{\sigma}^{2(s)}(x)$ are similarly defined.

Let $f(\cdot)$ denote the probability density function (PDF) of ε_i. Then we can write the density function of Y_i given $X_i = x$ as

$$f(y; \beta_1(x), \beta_2(x)) \equiv f\left((y - \beta_1(x)) / \sqrt{\varphi(\beta_2(x))}\right) / \sqrt{\varphi(\beta_2(x))}, \tag{2.2}$$

where $\varphi(\cdot)$ is a "link" function that is strictly monotonic and positive, and the true value $(\beta_1^0(x), \beta_2^0(x))$ of $(\beta_1(x), \beta_2(x))$ satisfies $\beta_1^0(x) = m(x)$ and $\varphi(\beta_2^0(x)) = \sigma^2(x)$. A simple choice for $\varphi(\cdot)$ is the identity function, i.e., $\varphi(u) = u$, but this parametrization generally does not ensure the positivity of the variance function estimate. Another choice is $\varphi(u) = \exp(u)$, ensuring that the estimate of $\sigma^2(x)$ is always positive; Ziegelmann (2002) uses this function to obtain a local linear exponential tilting estimate. It is worth mentioning that both link functions yield the same asymptotic variance but different asymptotic biases for the local polynomial estimates. We will consider both link functions below.

For the ease of presentation, we denote for $r = 1, 2$,

$$\boldsymbol{\beta}_r(x) \equiv \left(\boldsymbol{\beta}_{r0}(x), \boldsymbol{\beta}_{r1}(x)^\top, ..., \boldsymbol{\beta}_{rp}(x)^\top\right)^\top, \ \boldsymbol{\beta}_r^0(x) \equiv \left(\boldsymbol{\beta}_{r0}^0(x), \boldsymbol{\beta}_{r1}^0(x)^\top, ..., \boldsymbol{\beta}_{rp}^0(x)^\top\right)^\top,$$

where $\boldsymbol{\beta}_{1s}^0(x) = m^{(s)}(x)$, and $\boldsymbol{\beta}_{2s}^0(x) = a^{(s)}(x)$ for $s = 0, 1, ..., p$ where $a(x) = \varphi^{-1}\left(\sigma^2(x)\right)$, $\varphi^{-1}(\cdot)$ is the inverse function of $\varphi(\cdot)$, and $a^{(s)}$ is analogously defined as $m^{(s)}(x)$ with $a(x)$ replacing $m(x)$. In particular, $\boldsymbol{\beta}_{10}^0(x) = m(x)$ and $\boldsymbol{\beta}_{20}^0(x) = \varphi^{-1}(\sigma^2(x))$. We will frequently suppress the dependence of $\boldsymbol{\beta}_r(x)$ and $\boldsymbol{\beta}_r^0(x)$ on x for $r = 1, 2$. Let $P_i(\boldsymbol{\beta}_r) \equiv \sum_{0 \le |\mathbf{j}| \le p} \beta_{r\mathbf{j}}(x) (X_i - x)^{\mathbf{j}}$. Apparently, $P_i\left(\boldsymbol{\beta}_1^0(x)\right)$ and $P_i\left(\boldsymbol{\beta}_2^0(x)\right)$ are p-th order Taylor expansions of $m(X_i)$ and $\varphi^{-1}(\sigma^2(X_i))$ around x, respectively.

Given $\{(Y_i, X_i)\}_{i=1}^n$, the local polynomial maximum likelihood estimator (MLE) $\bar{\boldsymbol{\beta}}^\top = (\bar{\boldsymbol{\beta}}_1^\top, \bar{\boldsymbol{\beta}}_2^\top) = (\bar{\boldsymbol{\beta}}_{10}, \bar{\boldsymbol{\beta}}_{11}^\top, ..., \bar{\boldsymbol{\beta}}_{1p}^\top, \bar{\boldsymbol{\beta}}_{20}, \bar{\boldsymbol{\beta}}_{21}^\top, ..., \bar{\boldsymbol{\beta}}_{2p}^\top)$ maximizes the kernel-weighted log-likelihood function

$$\mathcal{L}_n(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \equiv \frac{1}{n} \sum_{i=1}^n \log f(Y_i; P_i(\boldsymbol{\beta}_1), P_i(\boldsymbol{\beta}_2)) K_h(x - X_i), \tag{2.3}$$

with respect to $(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top) = (\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{11}^\top, ..., \boldsymbol{\beta}_{1p}^\top, \boldsymbol{\beta}_{20}, \boldsymbol{\beta}_{21}^\top, ..., \boldsymbol{\beta}_{2p}^\top)$.

Let $\psi(\varepsilon) \equiv f'(\varepsilon) / f(\varepsilon)$ and $\rho(\varepsilon) \equiv \psi(\varepsilon)\varepsilon + 1$. Let $\varepsilon(\beta) \equiv (y - \beta_1) / \sqrt{\varphi(\beta_2)}$ and $q^{(s)}(y; \beta_1, \beta_2) \equiv \frac{\partial^s}{\partial \varepsilon^s} \log f(\varepsilon)|_{\varepsilon = \varepsilon(\beta)}$ for $s = 1, 2, 3$. To study the asymptotic properties of $\bar{\boldsymbol{\beta}}$, we make the following assumptions.

A1. ε_i and X_i are IID and are mutually independent with $E(\varepsilon_i) = 0$ and $E\left(\varepsilon_i^2\right) = 1$. The PDF $f(\cdot)$ of ε_i has support $\mathbb{R}$, and uniformly bounded continuous derivatives of up to the order $p + 2$. Furthermore, $f^{(p+2)}(\cdot)$ is Lipschitz continuous of order 1, i.e., there exists $C_1 < \infty$ such that for all u and v on the support of f, we have $\left|f^{(p+2)}(u) - f^{(p+2)}(v)\right| \le C_1 |u - v|$.

A2. (i) $E[\psi(\varepsilon)] = 0$, $E[\rho(\varepsilon)] = 0$, $E\left[\psi^{2\gamma}(\varepsilon)\right] < \infty$ and $E\left[\rho^{2\gamma}(\varepsilon)\right] < \infty$ for some $\gamma > 1$.

(ii) $\begin{bmatrix} E\left[\psi^2(\varepsilon)\right] & E[\psi(\varepsilon)\rho(\varepsilon)] \\ E[\psi(\varepsilon)\rho(\varepsilon)] & E\left[\rho^2(\varepsilon)\right] \end{bmatrix}$ is positive definite (p.d.).

(*iii*) $E\left|\psi^{(r)}(\varepsilon)\right| < \infty$ and $E\left|\rho^{(r)}(\varepsilon)\right| < \infty$ for $r = 1, 2$.

(*iv*) There exists a function $J(y)$, $\left|q^{(s)}(y;\beta_1,\beta_2)\varepsilon(\beta)^t\right| \leq J(y)$ for all $\beta \in \mathcal{B}_0$, $0 \leq t \leq s$, and $s = 1,2,3$, and $E\left[J^2(Y)\right] < \infty$.

A3. The PDF $f_X(\cdot)$ of X_i is differentiable, bounded, and bounded away from zero on its compact support $\mathcal{X}$. $\beta_r^0(x)$ $(r = 1,2)$ have $(p+1)$th order derivatives, and $\left(D^{\mathbf{j}}\beta_r^0\right)(x)$ are bounded and Lipschitz continuous on $\mathcal{X}$ for all $|\mathbf{j}| = p+1$.

A4. K is a product kernel of a univariate kernel function $k : K(u) \equiv \Pi_{i=1}^d k(u_i)$, where k is a symmetric PDF that has compact support and bounded variation. For each d-tuple $\mathbf{j}$ with $0 \leq |\mathbf{j}| \leq 2p+1$, $H_{\mathbf{j}}(u) \equiv u^{\mathbf{j}}K(u)$ is Lipschitz continuous.

A5. As $n \to \infty$, $h \to 0$, $nh^d \to \infty$, and $nh^{2(p+1)+d} \to C_2 \in [0,\infty)$.

Assumptions A1-A2 are parallel to the Assumptions A1-A2 in Linton and Xiao (2007). The main difference is that our Assumption A2 is stronger than theirs. In addition to the conditions on the score function $\psi(\cdot)$ for the conditional mean parameter, we also impose conditions on the score function for the conditional variance parameter that is associated with $\rho(\cdot)$. It is easy to verify that the normal or student t distributions with degrees of freedom larger than three will satisfy A2(i). A2(ii) ensures the positive definiteness of certain information matrix. A3 mainly specifies conditions on the density of X_i and the smoothness of functions of interest. A4 and A5 impose conditions on the kernel function and bandwidth parameter, respectively.

Let $H \equiv \text{diag}\left(I_{N_0}, hI_{N_1}, ..., h^pI_{N_p}\right)$ and $\bar{H} \equiv \text{diag}(H, H)$. Define

$$\mathcal{I}_{\beta^0}(x) = f_X(x)\begin{bmatrix} \frac{E\left[\psi^2(\varepsilon)\right]}{\sigma^2(x)} & E\left[\psi(\varepsilon)\rho(\varepsilon)\right]\frac{\varphi'\left(\beta_2^0(x)\right)}{2\sigma^3(x)} \\ E\left[\psi(\varepsilon)\rho(\varepsilon)\right]\frac{\varphi'\left(\beta_2^0(x)\right)}{2\sigma^3(x)} & E\left[\rho(\varepsilon)\right]^2\frac{\left[\varphi'\left(\beta_2^0(x)\right)\right]^2}{4\sigma^4(x)} \end{bmatrix}. \tag{2.4}$$

The following proposition reports the asymptotic distribution of $\bar{\boldsymbol{\beta}}(x)$.

Proposition 2.1 *Suppose that Assumptions A1-A5 hold. Then*

$$\sqrt{nh^d}\left[\bar{H}\left(\bar{\boldsymbol{\beta}}(x) - \boldsymbol{\beta}^0(x)\right) - h^{p+1}\begin{pmatrix} M^{-1}B\mathbf{m}^{(p+1)}(x) \\ M^{-1}B\boldsymbol{\beta}_2^{0(p+1)}(x)\end{pmatrix}\right] \xrightarrow{d} N\left(0, \mathcal{I}_{\beta^0}^{-1}(x) \otimes \left[M^{-1}\Gamma M^{-1}\right]\right), \tag{2.5}$$

where $\boldsymbol{\beta}_2^{0(p+1)}(x)$ *is analogously defined as* $\mathbf{m}^{(p+1)}(x)$.

Proposition 2.1 complements existing results in the literature. Staniswalis (1989) studies the local constant $(p = 0)$ estimation of a single location parameter by maximizing a kernel-weighted likelihood function. Fan et al. (1995) discuss the local polynomial estimator of a single location parameter when the error density belongs to a one-parameter exponential family. Aerts and Claeskens (1997) study the local polynomial estimation in multiparameter-likelihood models where $d = 1$ and X_i is nonrandom. Claeskens and Van Keilegom (2003) study the construction of confidence bands via local polynomial estimation. One can apply their method to construct the confidence bands for both the conditional mean and variance function estimates in our framework. In addition, under appropriate conditions, the local MLE is equivalent to nonparametric estimation by treating the locally weighted score functions as

Table 1: Relative efficiency of the local likelihood estimators over the conventional local polynomial estimators

Estimator	$t(\gamma)$	$t(2)$	$t(3)$	$t(4)$	$t(5)$	$t(10)$	$t(20)$	$t(50)$	$t(100)$	$t(\infty)$
$\overline{m}(x)$	$\frac{(\gamma-2)((\gamma+3)}{\gamma(\gamma+1)}$	-	0.5	0.7	0.8	0.945	0.986	0.998	0.999	1
$\overline{\sigma}^2(x)$	$\frac{(\gamma+3)(\gamma-4)}{\gamma(\gamma+2)}$	-	-	-	0.229	0.650	0.836	0.938	0.969	1

estimating equations. So we can obtain the solution to the local MLE through solving the estimating equations. See Claeskens and Aerts (2000) for such a local polynomial estimation setup.

If $\varphi(u) = u$, then $\varphi'\left(\beta_2^0(x)\right) = 1$ and Proposition 2.1 holds with $\varphi'\left(\beta_2^0(x)\right)$ being replaced by 1 in the definition of $\mathcal{I}_{\beta^0}(x)$. If, in addition, the density f of ε_i is symmetric about zero, then $\psi(\cdot)$ is an odd equation, implying that $E[\psi(\varepsilon)\rho(\varepsilon)] = 0$ under Assumption A2. In this case, $\mathcal{I}_{\beta^0}^{-1}(x)$ is a diagonal matrix and thus the estimation of conditional variance is not affected by the estimation of conditional mean. In particular, we have

$$\sqrt{nh^d}\left(\overline{m}(x) - m(x) - h^{p+1}\left[M^{-1}B\mathbf{m}^{(p+1)}(x)\right]_1\right) \xrightarrow{d} N\left(0, \frac{\sigma^2(x)}{f_X(x)E\left[\psi^2(\varepsilon)\right]}\left[M^{-1}\Gamma M^{-1}\right]_{1,1}\right),$$

and

$$\sqrt{nh^d}\left(\overline{\sigma}^2(x) - \sigma^2(x) - h^{p+1}\left[M^{-1}B\boldsymbol{\sigma}^{\mathbf{2}^{(p+1)}}(x)\right]_1\right) \xrightarrow{d} N\left(0, \frac{4\sigma^4(x)}{f_X(x)E\left[\rho^2(\varepsilon)\right]}\left[M^{-1}\Gamma M^{-1}\right]_{1,1}\right),$$

where $[C]_1$ denotes the first element of C, and $[A]_{i,i}$ signifies the (i,i)th element of A. This indicates that in the case of symmetric error density, the asymptotic biases of $\overline{m}(x)$ and $\overline{\sigma}^2(x)$ are the same as those of the local polynomial estimators of $m(x)$ and $\sigma^2(x)$ when they are estimated separately (e.g., Ruppert et al., 1997, Fan and Yao, 1998). The asymptotic variance of $\overline{m}(x)$ is smaller than that of the conventional local polynomial estimator of $m(x)$, which is $\sigma^2(x)\left[M^{-1}\Gamma M^{-1}\right]_{1,1}/f_X(x)$. Similarly, the asymptotic variance of $\overline{\sigma}^2(x)$ is smaller than that of the conventional local polynomial estimator of $\sigma^2(x)$, which is $\sigma^4(x)E\left(\varepsilon_1^2 - 1\right)^2\left[M^{-1}\Gamma M^{-1}\right]_{1,1}/f_X(x)$. In the special case where the error term ε is proportional to a student-t random variable, let ε^* have the student t distribution with $\gamma > 2$ degrees of freedom, we can normalize ε^* so that $\varepsilon = \sqrt{(\gamma-2)/\gamma}\varepsilon^*$ has variance one. Table 1 lists the relative asymptotic efficiency ratio of $\left(\overline{m}(x), \overline{\sigma}^2(x)\right)$ over the conventional local polynomial estimators of $\left(m(x), \sigma^2(x)\right)$ in terms of asymptotic variance for the approximately $t(\gamma)$ distributed ε. Smaller number means that larger asymptotic gain can be achieved by using the local likelihood approach. Table 1 indicates that a large efficiency gain can be achieved by using the local likelihood approach when the error distribution is far away from normality.

If $\varphi(u) = \exp(u)$, then $\varphi'\left(\beta_2^0(x)\right) = \exp(\beta_2^0(x)) = \sigma^2(x)$, and Proposition 2.1 holds with $\varphi'\left(\beta_2^0(x)\right)$ being replaced by $\sigma^2(x)$ in the definition of $\mathcal{I}_{\beta^0}(x)$. In this case, Proposition 2.1 implies that

$$\sqrt{nh^d}\left[\overline{m}(x) - m(x)\right] - h^{p+1}\left[M^{-1}B\mathbf{m}^{(p+1)}(x)\right]_1 \xrightarrow{d} N\left(0, \left[\overline{\mathcal{I}}_{\beta^0}^{-1}(x) \otimes \left(M^{-1}\Gamma M^{-1}\right)\right]_{1,1}\right), \quad (2.6)$$

and that, by a simple application of the delta method,

$$\sqrt{nh^d}\left[\overline{\sigma}^2(x) - \sigma^2(x)\right] - h^{p+1}\sigma^2(x)\left[M^{-1}B\boldsymbol{\beta}_2^{0(p+1)}(x)\right]_1 \xrightarrow{d} N\left(0, \sigma^4(x)\left[\overline{\mathcal{I}}_{\beta^0}^{-1}(x) \otimes \left(M^{-1}\Gamma M^{-1}\right)\right]_{N+1,N+1}\right),$$

where

$$\overline{\mathcal{I}}_{\beta^0}(x) = f_X(x) \begin{bmatrix} \frac{E\left[\psi^2(\varepsilon)\right]}{\sigma^2(x)} & \frac{E[\psi(\varepsilon)\rho(\varepsilon)]}{2\sigma(x)} \\ \frac{E[\psi(\varepsilon)\rho(\varepsilon)]}{2\sigma(x)} & \frac{E\left[\rho^2(\varepsilon)\right]}{4} \end{bmatrix}, \tag{2.7}$$

and $\boldsymbol{\beta}_2^{0(p+1)}(x)$ is analogously defined as $\mathbf{m}^{(p+1)}$ by stacking all $(p+1)$th derivatives of $\log \sigma^2(x)$ into a column vector. Compared with the case of $\varphi(u) = u$, the bias of $\overline{m}(x)$ remains the same as before whereas the bias of $\overline{\sigma}^2(x)$ differs from that of the latter case; the asymptotic variances of these two estimators also remain the same as before. Furthermore, if f is symmetric about zero, then one can readily verify that the asymptotic variance of $\overline{\sigma}^2(x)$ is the same as that of the local linear exponential-tilting (ET) estimator of $\sigma^2(x)$ as obtained by Ziegelmann (2002).[1]

It is worth mentioning that in general, the estimate $\overline{\boldsymbol{\beta}}(x)$ is only implicitly defined as a nonlinear function of the random sample. In practice, one may resort to a numerical algorithm to compute it. We may work with the one-step Newton-Raphson (NR hereafter) estimator from a preliminary consistent estimator $\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\beta}}_1(x)^\top, \tilde{\boldsymbol{\beta}}_2(x)^\top)^\top$, where $\tilde{\boldsymbol{\beta}}_1(x)$ and $\tilde{\boldsymbol{\beta}}_2(x)$ can be the local polynomial least squares estimator for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, respectively. Let $\tilde{\mathbf{X}}_i \equiv \tilde{\mathbf{X}}_i(x) = (\tilde{\mathbf{X}}_{i,0}(x)^\top, ..., \tilde{\mathbf{X}}_{i,p}(x)^\top)^\top$ where $\tilde{\mathbf{X}}_{i,|\mathbf{j}|}(x)$ $(0 \le |\mathbf{j}| \le p)$ is an $N_{|\mathbf{j}|} \times 1$ subvector whose r-th element is given by $\left[\tilde{\mathbf{X}}_{i,|\mathbf{j}|}\right]_r = (X_i - x)^{\phi_{|\mathbf{j}|}(r)}$. Let

$$s_i(\boldsymbol{\beta}) \equiv \begin{pmatrix} -\frac{\psi(\varepsilon_i(\boldsymbol{\beta}))}{\varphi(P_i(\boldsymbol{\beta}_2))^{1/2}} \\ -\frac{\varphi'(P_i(\boldsymbol{\beta}_2))\rho(\varepsilon_i(\boldsymbol{\beta}))}{2\varphi(P_i(\boldsymbol{\beta}_2))} \end{pmatrix}, \tag{2.8}$$

where $\varepsilon_i(\boldsymbol{\beta}) \equiv (Y_i - P_i(\boldsymbol{\beta}_1)) / \sqrt{\varphi(P_i(\boldsymbol{\beta}_2))}$. Define the smoothed score function

$$S_n(\boldsymbol{\beta}; f) \equiv \frac{1}{n}\sum_{i=1}^n K_h(x - X_i)\, s_i(\boldsymbol{\beta}) \otimes \tilde{\mathbf{X}}_i, \tag{2.9}$$

and the smoothed information matrix

$$I_n(\boldsymbol{\beta}; f) \equiv \frac{1}{n}\sum_{i=1}^n K_h(x - X_i)\left[s_i(\boldsymbol{\beta})\, s_i(\boldsymbol{\beta})^\top\right] \otimes \left(\tilde{\mathbf{X}}_i \tilde{\mathbf{X}}_i^\top\right). \tag{2.10}$$

Then the one-step NR estimator from a preliminary consistent estimator $\tilde{\boldsymbol{\beta}}$ is given by

$$\overline{\boldsymbol{\beta}}_{NR}(x) \equiv \tilde{\boldsymbol{\beta}}(x) + I_n\left(\tilde{\boldsymbol{\beta}}(x); f\right)^{-1} S_n\left(\tilde{\boldsymbol{\beta}}(x); f\right). \tag{2.11}$$

It can be shown that $\overline{\boldsymbol{\beta}}_{NR}(x)$ shares the same asymptotic distribution as $\overline{\boldsymbol{\beta}}(x)$ under some regularity conditions.

We call $\overline{\boldsymbol{\beta}}(x)$ in Proposition 2.1 an oracle estimator because its definition uses knowledge that only an oracle could have. In practice, the density f of ε_i is generally unknown and so both $\overline{\boldsymbol{\beta}}(x)$ and $\overline{\boldsymbol{\beta}}_{NR}(x)$ are infeasible.

2.2 The Adaptive Nonparametric Regression Estimators

To obtain a feasible analogue of $\overline{\boldsymbol{\beta}}(x)$, we need to replace f in (2.11) by a nonparametric estimate, say, $\tilde{f}$. Since $\tilde{f}$ appears as a random denominator, it causes technical difficulty when it is small. For this reason, we propose to trim out small values of $\tilde{f}$ as do Bickel (1982), Kreiss (1987), and Linton and Xiao (2007).

In particular, we consider the following smoothed trimming, which has been used by Andrews (1995), Ai (1997), and Linton and Xiao (2007). Let $g(\cdot)$ be a density function with support [0,1], $g(0) = g(1) = 0$. Let b be the trimming parameter that goes to zero at a certain rate as $n \to \infty$. Define $g_b(z) = \frac{1}{b} g\left(\frac{z}{b} - 1\right)$. Clearly, $g_b(z)$ has support on $[b, 2b]$. Defining $G_b(z) = \int_{-\infty}^{z} g_b(t)\, dt$, we have

$$G_b(z) = \begin{cases} 0 & \text{if } z < b \\ \int_{-\infty}^{z} g_b(t)\, dt & \text{if } b \leq z \leq 2b \\ 1 & \text{if } z > 2b \end{cases} \tag{2.12}$$

We assume that g is second order differentiable and its derivatives are uniformly bounded.

In this section we propose a feasible estimator by substituting a suitable pilot estimator of f in (2.11). The proposed three-step estimation procedure is as follows:

1. Obtain a preliminary consistent estimate of $\left(\beta_1^0, \beta_2^0\right) \equiv \left(m(x), \sigma^2(x)\right)$ and its derivatives by the p-th order local polynomial smoothing with kernel K and bandwidth h_1. Denote the preliminary estimate as $\tilde{\boldsymbol{\beta}}(x) = (\tilde{\boldsymbol{\beta}}_1^\top(x), \tilde{\boldsymbol{\beta}}_2^\top(x))^\top$, where $\tilde{\boldsymbol{\beta}}_r(x)$ estimates $\boldsymbol{\beta}_r^0(x)$ for $r = 1, 2$. Define the residuals $\tilde{u}_i \equiv Y_i - \tilde{m}(X_i)$ and its standardized version $\tilde{\varepsilon}_i \equiv \tilde{u}_i / \tilde{\sigma}(X_i)$ for $i = 1, ..., n$.

2. Obtain a consistent estimator for the error density and its derivatives by the leave-one-out kernel method:
$$\tilde{f}_i(e_i) \equiv \frac{1}{nh_0} \sum_{j \neq i} k_0\left(\frac{e_i - \tilde{\varepsilon}_j}{h_0}\right), \text{ and } \tilde{f}_i^{\prime}(e_i) \equiv \frac{1}{nh_0^2} \sum_{j \neq i} k_0^{\prime}\left(\frac{e_i - \tilde{\varepsilon}_j}{h_0}\right), \tag{2.13}$$
where k_0 and h_0 are the univariate kernel and bandwidth parameter, respectively.

3. Define the trimmed and re-centered local score function
$$\tilde{S}_n\left(\boldsymbol{\beta}; \tilde{f}\right) \equiv \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tilde{s}_{n,i}(\boldsymbol{\beta}) \otimes \tilde{\mathbf{X}}_i, \tag{2.14}$$
and the trimmed local information matrix
$$\tilde{I}_n\left(\boldsymbol{\beta}; \tilde{f}\right) \equiv \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) G_b\left(\tilde{f}_i(\varepsilon_i(\boldsymbol{\beta}))\right) \left[\tilde{s}_i(\boldsymbol{\beta})\, \tilde{s}_i(\boldsymbol{\beta})^\top\right] \otimes (\tilde{\mathbf{X}}_i \tilde{\mathbf{X}}_i^\top), \tag{2.15}$$
where $\tilde{s}_{n,i}(\boldsymbol{\beta}) \equiv G_b(\tilde{f}_i(\varepsilon_i(\boldsymbol{\beta})))\tilde{s}_i(\boldsymbol{\beta}) + \tilde{s}_i^*(\tilde{\boldsymbol{\beta}})$,
$$\tilde{s}_i(\boldsymbol{\beta}) \equiv \begin{pmatrix} \tilde{q}_{1,i}(\boldsymbol{\beta}) \\ \tilde{q}_{2,i}(\boldsymbol{\beta}) \end{pmatrix} \equiv - \begin{pmatrix} \frac{\tilde{\psi}_i(\varepsilon_i(\boldsymbol{\beta}))}{\varphi(P_i(\boldsymbol{\beta}_2))^{1/2}} \\ \frac{\varphi'(P_i(\boldsymbol{\beta}_2))\tilde{\rho}(\varepsilon_i(\boldsymbol{\beta}))}{2\varphi(P_i(\boldsymbol{\beta}_2))} \end{pmatrix}, \tag{2.16}$$
$$\begin{aligned} \tilde{s}_i^*(\boldsymbol{\beta}) &\equiv \begin{pmatrix} \tilde{q}_{1,i}^*(\boldsymbol{\beta}) \\ \tilde{q}_{2,i}^*(\boldsymbol{\beta}) \end{pmatrix} \\ &\equiv -\log\left(\frac{\tilde{f}_i(\varepsilon_i(\boldsymbol{\beta}))}{\varphi(P_i(\boldsymbol{\beta}_2))^{1/2}}\right) g_b\left(\tilde{f}_i(\varepsilon_i(\boldsymbol{\beta}))\right) \tilde{f}_i'(\varepsilon_i(\boldsymbol{\beta})) \begin{pmatrix} \frac{1}{\varphi(P_i(\boldsymbol{\beta}_2))^{1/2}} \\ \frac{\varphi'(P_i(\boldsymbol{\beta}_2))\varepsilon_i(\boldsymbol{\beta})}{2\varphi(P_i(\boldsymbol{\beta}_2))} \end{pmatrix}, \end{aligned} \tag{2.17}$$
and $\tilde{\psi}_i(\cdot) \equiv \tilde{f}_i'(\cdot) / \tilde{f}_i(\cdot)$, $\tilde{\rho}(\varepsilon_i(\boldsymbol{\beta})) = \left[\tilde{\psi}_i(\varepsilon_i(\boldsymbol{\beta}))\, \varepsilon_i(\boldsymbol{\beta}) + 1\right]$. The proposed one-step adaptive estimator can then be calculated by
$$\hat{\boldsymbol{\beta}}(x) \equiv \tilde{\boldsymbol{\beta}}(x) + \tilde{I}_n\left(\tilde{\boldsymbol{\beta}}(x); \tilde{f}\right)^{-1} \tilde{S}_n\left(\tilde{\boldsymbol{\beta}}(x); \tilde{f}\right). \tag{2.18}$$

We shall show below that, under appropriate assumptions, the proposed estimator $\hat{\boldsymbol{\beta}}(x)$ is asymptotically equivalent to the infeasible estimator $\bar{\boldsymbol{\beta}}(x)$.

Note that Linton and Xiao (2007) only resort to the trimming but not the re-centering technique. There the score function is given by $S_{1n} \equiv \frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)\psi(\varepsilon_i)\tilde{\mathbf{X}}_i$ and can be estimated by $\tilde{S}_{1n} \equiv S_{1n}(\tilde{\boldsymbol{\beta}})$, where

$$S_{1n}(\boldsymbol{\beta}) \equiv \frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) G_b\left(\tilde{f}_i(\varepsilon_i(\boldsymbol{\beta}))\right)\tilde{\psi}_i(\varepsilon_i(\boldsymbol{\beta}))\tilde{\mathbf{X}}_i.$$

An important step in their paper is justifying the adaptivity of the conditional mean function by showing that $\sqrt{nh^d}\tilde{S}_{1n}$ and $\sqrt{nh^d}S_{1n}$ share the same asymptotic distribution. The latter is based on the demonstration that

$$\frac{1}{\sqrt{nh^d}}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)\tilde{\mathbf{X}}_i\psi(\varepsilon_i)\left[1-G_b(f(\varepsilon_i))\right] = o_p(1)$$

and that

$$E\left\{\psi(\varepsilon_i)\left[1-G_b(f(\varepsilon_i))\right]\right\} = 0. \tag{2.19}$$

By the construction of G_b and the nature of ψ, the last condition can be ensured no matter whether f is symmetric about zero or not. In contrast, to prove the adaptivity of both the conditional mean and variance functions, we also require the asymptotic equivalence of $\sqrt{nh^d}\tilde{S}_{2n}$ and $\sqrt{nh^d}S_{2n}$, where $S_{2n} \equiv \frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)\rho(\varepsilon_i)\tilde{\mathbf{X}}_i$, $\tilde{S}_{2n} \equiv \tilde{S}_{2n}(\boldsymbol{\beta})$, and

$$\tilde{S}_{2n}(\boldsymbol{\beta}) \equiv \frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) G_b\left(\tilde{f}_i(\varepsilon_i(\boldsymbol{\beta}))\right)\frac{\varphi'(P_i(\boldsymbol{\beta}_2))\,\tilde{\rho}(\varepsilon_i(\boldsymbol{\beta}))}{2\varphi(P_i(\boldsymbol{\beta}_2))}\tilde{\mathbf{X}}_i.$$

Analogously to (2.19), a key step toward the establishment of the above claim is to demonstrate that

$$E\left\{\rho(\varepsilon_i)\left[1-G_b(f(\varepsilon_i))\right]\right\} = 0, \tag{2.20}$$

where recall that $\rho(\varepsilon) = \psi(\varepsilon)\varepsilon + 1$. Unfortunately, (2.20) does not hold generally even if we assume the symmetry of f. And there is no obvious way to design another trimming function for the estimation of S_{2n} such that we can ensure the above asymptotic equivalence. An intuitive explanation is that even though the score function for the conditional variance has zero mean, this does not ensure that the weighted population score (i.e., after being multiplied by $G_b(f(\varepsilon_i))$) has zero mean. When the error density is symmetric, the score function for the conditional mean function is antisymmetric, which is still antisymmetric after being multiplied by $G_b(f(\varepsilon_i))$. This ensures that its weighted population score has zero mean. In contrast, the score function for the conditional variance function is symmetric and it is still symmetric after being multiplied by $G_b(f(\varepsilon_i))$, which cannot have zero mean.

To avoid the non-zero asymptotic mean of the estimated score function, we rely on the re-centering term $\tilde{s}_i^*$ in the above definition of $\tilde{S}_n(\boldsymbol{\beta};\tilde{f})$ or $\tilde{s}_{n,i}(\boldsymbol{\beta})$. This is motivated by maximizing the following local profile log-likelihood function

$$\tilde{Q}_n(x,\boldsymbol{\beta}) \equiv \frac{1}{n}\sum_{i=1}^n \log\left(\tilde{f}_i(\varepsilon_i(\boldsymbol{\beta}))/\sqrt{\varphi(P_i(\boldsymbol{\beta}_2))}\right) G_b\left(\tilde{f}_i(\varepsilon_i(\boldsymbol{\beta}))\right) K_h(x-X_i) \tag{2.21}$$

with respect to $\boldsymbol{\beta}$. The previously defined score function results from the first order condition to this maximization problem and hence one expects that its population analog has asymptotic zero mean. Below we will study the asymptotic properties of both the one-step NR adaptive estimator and the local profile likelihood (LPL) estimator obtained from maximizing a local profile log-likelihood function of the above type.

3 The Main Results

In this section we first study the asymptotic properties of the one-step Newton-Raphson (NR) estimator, and then the LPL estimator.

3.1 The One-Step Newton-Raphson Estimator

To proceed, we add the following assumptions.

A6. (i) The kernel k_0 has compact support and is symmetric about zero and satisfies $\int k_0(u)\,du = 1$, $\int u^j k_0(u)\,du = 0$ for $j = 1, ..., p$, and $\int u^{p+1} k_0(u)\,du \neq 0$. (ii) k_0 is three times differentiable on its support with $k'(0) = 0$. In addition, $k_0^{(j)}(u)$, $j = 1, 2, 3$, is Lipschitz continuous and $|k_0'(u)\,u|$ is uniformly bounded.

A7. As $n \to \infty$, the trimming parameter b and bandwidth sequences h_0 and h_1 satisfy (i) $b \propto h^{\tau}$ for $\tau \in (0, 1/3)$, (ii) $h_0 \propto h/\log n$, (iii) $h_1 \propto h/\log n$.

Assumptions A6-A7 are similar to the Assumptions A6 and A7 in Linton and Xiao (2007). We just mention two main differences. First, our assumption on the kernel k_0 in A6 is slightly different from theirs in that we need a restriction on the tail thickness of the derivative of k_0 (a similar assumption is made by Andrews (1995, Assumption NP4)), and we require that the Lipschitz condition hold instead of the fourth order differentiability. A6 requires that k_0 be a $(p+1)$th order kernel, but the compactness condition can be relaxed at the cost of lengthy arguments. Second, our assumption on the bandwidth sequences in A7 is weaker than theirs whereas our requirement on the trimming parameter is stricter. This is due to the differences in proving that the higher order terms are asymptotically negligible.

In this paper we only focus on the case where the error density f has an unbounded support in order to apply some uniform convergence results for the kernel estimates of density function and its derivatives (e.g., Hansen, 2008) and avoid the well-known boundary bias problems for kernel density estimates in the case of compact support. Linton and Xiao (2007) also consider the case of bounded support for the error density where special attention is needed. In particular, they assume that the density f vanishes at the boundary at a sufficiently fast rate so that the properties of regular density estimation can hold.

The following theorem states the asymptotic property of the one-step NR estimator.

Theorem 3.1 *Suppose that Assumptions A1-A7 hold. Then*

$$\sqrt{nh^d}\left[\bar{H}\left(\hat{\boldsymbol{\beta}}(x) - \boldsymbol{\beta}^0(x)\right) - h^{p+1}\begin{pmatrix} M^{-1}B\mathbf{m}^{(p+1)}(x) \\ M^{-1}B\boldsymbol{\beta}_2^{0(p+1)}(x) \end{pmatrix}\right] \xrightarrow{d} N\left(0, \mathcal{I}_{\beta^0}^{-1}(x) \otimes \left[M^{-1}\Gamma M^{-1}\right]\right). \quad (3.1)$$

We denote the first and $(N+1)$th elements of $\hat{\boldsymbol{\beta}}(x)$ as $\hat{m}(x)$ and $\varphi^{-1}(\hat{\sigma}^2(x))$, which are the one-step NR adaptive estimator of $m(x)$ and $\varphi^{-1}(\sigma^2(x))$, respectively. Theorem 3.1 shows that the one-step Newton-Raphson estimator is "oracle": the feasible estimator $\hat{\boldsymbol{\beta}}(x)$ is asymptotically equivalent to $\bar{\boldsymbol{\beta}}(x)$ and hence is more efficient than conventional local polynomial estimator in the case of $\varphi(u) = u$ and the local polynomial ET estimator in the case of $\varphi(u) = \exp(u)$. Following Remarks 2 and 3 we can readily obtain the asymptotic normal distributions for $\hat{m}(x)$ and $\hat{\sigma}^2(x)$, based on which one can also construct the pointwise confidence intervals. To do this we require an estimation of the asymptotic variance. The procedure is standard and we omit it for brevity.

3.2 The LPL Estimator

The one-step NR estimates $\hat{m}(x)$ and $\hat{\sigma}^2(x))$ are easy to obtain in general. Nevertheless, we have to estimate both the error density and its derivative in order to construct these estimates. It is well known that precise estimation of the density's derivatives can be difficult for some distributions. For this reason, we now propose another adaptive estimator that avoids estimation of density derivatives.

We can obtain the adaptive estimator by maximizing the local profile likelihood in (2.21) by taking $\varphi(\cdot) = \exp(\cdot)$. Let $\boldsymbol{\beta}^+(x) = (\boldsymbol{\beta}_1^{+\top}, \boldsymbol{\beta}_2^{+\top})^\top$ denote the solution. The local profile likelihood estimators $m^+(x)$ and $\sigma^{+2}(x)$ for $m(x)$ and $\sigma^2(x)$ are given respectively by $\boldsymbol{\beta}_{10}^+$ and $\exp(\boldsymbol{\beta}_{20}^+)$, where $\boldsymbol{\beta}_{r0}^+$ is the first element of $\boldsymbol{\beta}_r^+$, $r = 1, 2$. The corresponding infeasible local likelihood estimator of $\boldsymbol{\beta}$ can be obtained by maximizing the following criterion function

$$Q_n(x, \boldsymbol{\beta}) \equiv \frac{1}{n}\sum_{i=1}^{n} \log\left(f(\varepsilon_i(\boldsymbol{\beta})) / \sqrt{\exp(P_i(\boldsymbol{\beta}_2))}\right) G_b(f(\varepsilon_i(\boldsymbol{\beta}))) K_h(x - X_i), \tag{3.2}$$

where $G_b(f(\varepsilon_i(\boldsymbol{\beta})))$ can be absent as in Aerts and Claeskens (1997). Replacing f in (3.2) by $\tilde{f}_i$ gives the local profile likelihood function (2.21). We will show that such a replacement does not affect the asymptotic properties of the resulting estimator.

To derive the uniform consistency and pointwise asymptotic normality of $\boldsymbol{\beta}^+$, we shall show that under certain conditions that $\tilde{Q}_n(x, \boldsymbol{\beta})$ and $Q_n(x, \boldsymbol{\beta})$ converge uniformly in $(x, \boldsymbol{\beta})$ to the non-random function

$$Q(x, \boldsymbol{\beta}) \equiv E\left[\log\left(f(\varepsilon_i(\boldsymbol{\beta})) / \sqrt{\exp(P_i(\boldsymbol{\beta}_2))}\right) K_h(x - X_i)\right], \tag{3.3}$$

where we suppress the dependence of $Q(x, \boldsymbol{\beta})$ on n through h. By the theory on local likelihood estimation (e.g., Aerts and Claeskens, 1997), the maximizer of the limit of $Q(x, \boldsymbol{\beta})$ is given by $\boldsymbol{\beta}^0 \equiv \boldsymbol{\beta}^0(x)$, which is composed of $m(x)$, $\log\sigma^2(x)$, and their derivatives of up to the p-th order. This corresponds to the identifiable uniqueness condition of White (1994, p. 28). Consequently we can establish the following uniform consistency result.

Theorem 3.2 *Suppose that Assumptions A1-A7 hold, and for all* $\boldsymbol{\beta} \in \mathcal{B}_0$, $|\log f(\varepsilon_i(\boldsymbol{\beta}))| \leq D(Y_i)$ *such that* $E[D(Y_i)^\gamma] < \infty$ *for some* $\gamma > 1$. *Then*

$$\sup_{x \in \mathcal{X}} \left|\boldsymbol{\beta}^+(x) - \boldsymbol{\beta}^0(x)\right| \to 0 \ a.s. \tag{3.4}$$

The asymptotic normality of $\boldsymbol{\beta}^+(x)$ can be established in several ways. One way is to apply and modify the results of Andrews (1994a) for semiparametric estimators. See also Andrews (1994b) for a review of the theoretical literature and results for more general sampling schemes using stochastic equicontinuity concepts and empirical process techniques. Here, we follow the traditional approach to establish the asymptotic normality of $\boldsymbol{\beta}^{+}(x)$ by expanding the score function around the population truth. The result is stated in the next theorem.

Theorem 3.3 *Under the conditions in Theorem 3.2, we have*

$$\sqrt{nh^d}\left[\bar{H}\left(\boldsymbol{\beta}^{+}(x)-\boldsymbol{\beta}^{0}(x)\right)-h^{p+1}\left(\begin{array}{c} M^{-1}B\mathbf{m}^{(p+1)}(x) \\ M^{-1}B\boldsymbol{\beta}_2^{0(p+1)}(x)\end{array}\right)\right]\xrightarrow{d} N\left(0,\overline{\mathcal{I}}_{\beta^0}^{-1}(x)\otimes\left[M^{-1}\Gamma M^{-1}\right]\right), \quad (3.5)$$

where $\overline{\mathcal{I}}_{\beta^0}(x)$ *is as defined in (2.7).*

Note that $\overline{\mathcal{I}}_{\beta^0}(x)$ is given by (2.7). So Theorem 3.3 indicates the asymptotic equivalence between $\boldsymbol{\beta}^{+}(x)$ and $\hat{\boldsymbol{\beta}}(x)$ when $\varphi(u)=\exp(u)$ is used to obtain the latter estimate. As expected, the local profile likelihood estimators $m^{+}(x)$ and $\sigma^{+2}(x)$ share the same asymptotic properties of the one-step NR adaptive estimators $\hat{m}(x)$ and $\hat{\sigma}^2(x)$ when $\varphi(u)=\exp(u)$. Despite the need to estimate certain density derivatives, the computation for the NR adaptive estimators is not very heavy. By contrast, even though the local profile likelihood estimates $m^{+}(x)$ and $\sigma^{+2}(x)$ only require estimation of the density function, they are computationally more demanding because certain optimization routine is needed. In the next section, we shall evaluate the finite sample performance of these estimators in Matlab.

4 Monte Carlo Simulations

In this section we conduct a small set of Monte Carlo simulations to evaluate the finite sample performance of the proposed estimators and compare them with the conventional local polynomial estimators for the conditional mean and variance functions.

4.1 Data Generating Processes

We generate data from (1.1) with different specifications of the conditional mean and variance functions and different choices of the error distributions. In all cases, the regressor X_i is independently and uniformly distributed on [-2, 2]. The conditional mean and variance functions are specified as follows:
DGP 1: $m(x)=1+x+x^2$, $\sigma^2(x)=0.1+x^2$;
DGP 2: $m(x)=1+x+x^2$, $\sigma^2(x)=\exp(2x)$;
DGP 3: $m(x)=4\sin(x)$, $\sigma^2(x)=0.1+x^2$;
DGP 4: $m(x)=4\sin(x)$, $\sigma^2(x)=\exp(2x)$.
For the error term ε_i, we consider two distributions: $t(4.1)$ and Beta$(2,3)$. Note that the $t(4.1)$ distribution is symmetric around zero with variance $41/21$ and the Beta$(2,3)$ distribution is asymmetric with mean $2/5$ and variance $1/25$. For each case, we first generate ε_i^* independently according to the specified distribution and then normalize it to have mean zero and variance one (e.g., let $\varepsilon_i=5(\varepsilon_i^*-2/5)$ for the Beta(2,3) case). In DGPs 1-4, we consider cases where the error term ε_i^* are IID $t(4.1)$. DGPs 5-8 are specified as DGPs 1-4, respectively, but with ε_i^* being generated from Beta$(2,3)$.[2]

In addition, we also consider bivariate regressions where the conditional mean and variance functions are specified as follows:

DGP 9: $m(x_1, x_2) = 1 + (x_1 + x_2)^2$, $\sigma^2(x_1, x_2) = 0.5\left(1 + x_1^2 + x_2^2\right)$;
DGP 10: $m(x_1, x_2) = 1 + (x_1 + x_2)^2$, $\sigma^2(x_1, x_2) = \exp(x_1 + x_2 - 0.5)$;
DGP 11: $m(x_1, x_2) = \cos(x_1 + x_2) + x_1^2 + 2x_2^2$, $\sigma^2(x_1, x_2) = 0.5\left(1 + x_1^2 + x_2^2\right)$;
DGP 12: $m(x_1, x_2) = \cos(x_1 + x_2) + x_1^2 + 2x_2^2$, $\sigma^2(x_1, x_2) = \exp(x_1 + x_2 - 0.5)$.
We generate the IID error terms ε_i from normalized $\chi^2(6)$ distribution, namely, $\varepsilon_i = (\varepsilon_i^* - 6)/\sqrt{12}$ where ε_i^* are IID $\chi^2(6)$.

4.2 Implementation

We investigate the local linear estimation ($p = 1$) with the normalized Epanechnikov kernel $K(u) = \frac{3}{4\sqrt{5}}\left(1 - \frac{1}{5}u^2\right)1\{|u| \leq \sqrt{5}\}$. For the estimation of the error density and its first derivative, we use the second order Gaussian kernel $k_0(u) = \frac{1}{\sqrt{2\pi}}e^{-u^2/2}$.

For comparison purpose, we examine the finite sample performance of the local linear (LL) estimators and the two adaptive estimators. For the conditional variance function, we also report the local linear ET estimator of Ziegelmann (2002). The preliminary estimator that is used for our adaptive estimation is composed of the LL estimator for the conditional mean function and the ET estimator for the conditional variance function (to achieve nonnegativity). We calculate the empirical variances and mean squared errors (MSEs) of the estimators of $m(x)$ and $\sigma^2(x)$ at selected values of x.

For the conventional LL estimators, the bandwidth sequences are chosen by Silverman's rule of thumb (ROT): $h_{LL} = s_X n^{-1/(d+4)}$ where s_X is the sample standard deviation of X_i.[3] The ET estimator also adopts the same bandwidth used for estimating the conditional variance. For our estimator, we have not designed a data-driven procedure for choosing the bandwidth. Instead, we set $h = h_{LL}$, $h_0 = 1.5n^{-1/(d+4)}/\log n$, and $h_1 = (d+3)h/\log n$. For the trimming parameter, we set $b = 0.01h^{1/4}$ when $d = 1$ and $b = 0.01\bar{h}^{1/4}$ when $d = 2$ with $\bar{h}$ being the average of the two elements in h. We choose the trimming function g to be the Beta(5,5) density function. The number of replications is 500 in each case.

4.3 Results

Tables 2 and 3 report the results for estimating the regression mean and variance functions at x =-1.2, -0.6, 0, 0.6, and 1.2, respectively for DGPs 1-8. Table 2 suggests that both the one-step NR adaptive estimator and the local profile likelihood (LPL) estimator generally have lower MSE than the local linear (LL) estimators for the regression mean. Somewhat surprisingly, in terms of MSE the efficiency gains for the LPL estimators are not as large as the case of the one-step NR adaptive estimators. For the estimation of the variance functions, Table 3 suggests that the conventional LL estimator is typically outperformed by the ET estimator, which is in turn outperformed by the NR and LPL estimators. Exceptions may occur when x moves toward the boundary points.

Tables 4 and 5 report the results for estimating the regression mean and variance functions at x =(-1.2,-1.2), (-0.6, -0.6), (0,0), (0.6, 0.6), and (1.2,1.2), respectively, for DGPs 9-12. As Table 4 suggests, the performance of the one-step NR estimator of the regression mean function is still good for most DGPs at most data evaluation points. But this is not the case for the LPL estimator. We find the performance of this estimator is not stable when $d = 2$ because of the need to utilize certain numerical optimization routine. Interestingly, for the estimates of the conditional variance function, the performance of the LPL estimator is comparable with that of the NR estimator and both tend to outperform the conventional LL estimator at most data evaluation points.

5 Conclusions

In this paper we propose adaptive estimators for nonparametric regression models with conditional heteroskedasticity. Consistency and asymptotic normality for the proposed estimators are studied. Our simulations confirm our theoretical results and suggest that significant gains can often be achieved by adopting our approach. The methodology can be extended to a general multi-parameter model by using the local likelihood method. It can also be extended to regression models where both ε_i and X_i are stationary time series, or autoregression models with lagged dependent variables in the regressors.

Note to readers. In the Appendices that follow we provide the proofs of the main results in the paper which further require some technical lemmas. The proofs of these lemmas are rather long and can be found in the Supplementary Material at Cambridge Journals Online (journals.cambridge.org/ect).

Notes

[1]Let $\tilde{m}(\cdot)$ be the p-th order local polynomial estimator of $m(\cdot)$ by using the kernel K and bandwidth h_1. We regress $\tilde{u}_i^2 = [Y_i - \tilde{m}(X_i)]^2$ on X_i by using the p-th order local polynomial (LP) ET technique based on the following minimization problem:

$$\left(\breve{\boldsymbol{\beta}}_{20}, \breve{\boldsymbol{\beta}}_{21}, ..., \breve{\boldsymbol{\beta}}_{2p}\right) = \arg\min_{\{\beta_{2\mathbf{j}}\}} \frac{1}{nh^d} \sum_{i=1}^n \left\{ \tilde{u}_i^2 - \exp\left(\sum_{0\leq|\mathbf{j}|\leq p} \beta_{2\mathbf{j}}(x)(X_i - x)^{\mathbf{j}} \right) \right\}^2 K\left(\frac{x - X_i}{h_1}\right).$$

The LP ET estimator $\tilde{\sigma}^2(x)$ of $\sigma^2(x)$ is then given by $\exp(\breve{\boldsymbol{\beta}}_{20})$. In addition, $\exp(\breve{\boldsymbol{\beta}}_{20})\breve{\boldsymbol{\beta}}_{21}$ estimates the first derivatives of $\sigma^2(x)$ and the estimates of other derivatives of $\sigma^2(x)$ can also be recovered. Ziegelmann (2002) shows that $\tilde{\sigma}^2(x)$ obtained this way is also adaptive to the unknown conditional mean function and it shares the same asymptotic variance as the two-step local polynomial least squares estimator of Fan and Yao (1998) but has different asymptotic biases. One can establish the uniform convergence rate for this type of estimators by following the arguments of Masry (1996a, 1996b).

[2]Note that the Beta distribution is compactedly supported and Assumption A1 is not satisfied in this case. We use this distribution simply to check whether our estimators are robust to errors with compact support.

[3]We also tried to choose the the bandwidth sequences by least squares cross-validation (LSCV) for estimating the conditional mean and variance functions separately, and the results were qualitatively similar.

REFERENCES

Aerts, M., and G. Claeskens (1997) Local polynomial estimation in multiparameter likelihood models. *Journal of the American Statistical Association* 92, 1536–1545.

Ai, C. (1997) A semiparametric maximum likelihood estimator. *Econometrica* 65, 933-963.

Akritas, M., and I. Van Keilegom (2001) Nonparametric estimation of the residual distribution. *Scandinavian Journal of Statistics* 28, 549-568.

Table 2: Comparison of our NR adaptive estimator and local profile likelihood (LPL) estimator with the conventional local linear (LL) estimator for regression mean, d=1, n=100

DGP	x	Variance			MSE			Efficiency ratio	
		LL (1)	NR (2)	LPL (3)	LL (4)	NR (5)	LPL (6)	$\frac{(5)}{(4)}$	$\frac{(6)}{(4)}$
1	-1.2	0.049	0.050	0.053	0.073	0.067	0.072	0.924	0.989
	-0.6	0.015	0.017	0.017	0.055	0.036	0.045	0.648	0.816
	0	0.007	0.007	0.007	0.047	0.024	0.029	0.519	0.626
	0.6	0.015	0.017	0.017	0.052	0.032	0.039	0.618	0.754
	1.2	0.051	0.050	0.054	0.071	0.065	0.073	0.914	1.031
2	-1.2	0.004	0.004	0.007	0.033	0.021	0.027	0.623	0.819
	-0.6	0.010	0.011	0.011	0.052	0.027	0.038	0.517	0.737
	0	0.030	0.034	0.034	0.070	0.051	0.057	0.727	0.814
	0.6	0.098	0.117	0.117	0.136	0.132	0.133	0.976	0.983
	1.2	0.443	0.403	0.443	0.462	0.416	0.456	0.900	0.987
3	-1.2	0.051	0.054	0.057	0.175	0.118	0.153	0.674	0.873
	-0.6	0.015	0.019	0.023	0.076	0.045	0.063	0.589	0.830
	0	0.006	0.005	0.005	0.006	0.005	0.005	0.964	0.982
	0.6	0.015	0.017	0.017	0.076	0.043	0.057	0.569	0.751
	1.2	0.052	0.052	0.061	0.186	0.119	0.160	0.636	0.858
4	-1.2	0.006	0.008	0.007	0.141	0.074	0.109	0.526	0.768
	-0.6	0.010	0.012	0.013	0.073	0.041	0.060	0.556	0.818
	0	0.029	0.031	0.036	0.029	0.032	0.037	1.098	1.267
	0.6	0.099	0.122	0.118	0.160	0.133	0.141	0.835	0.882
	1.2	0.443	0.423	0.473	0.581	0.449	0.534	0.772	0.918
5	-1.2	0.044	0.050	0.051	0.076	0.072	0.076	0.940	0.994
	-0.6	0.015	0.018	0.018	0.059	0.040	0.048	0.677	0.814
	0	0.007	0.007	0.007	0.048	0.028	0.031	0.587	0.637
	0.6	0.016	0.019	0.019	0.053	0.036	0.040	0.684	0.747
	1.2	0.052	0.058	0.060	0.071	0.070	0.071	0.993	1.003
6	-1.2	0.004	0.004	0.005	0.035	0.023	0.025	0.667	0.710
	-0.6	0.010	0.012	0.013	0.053	0.031	0.040	0.585	0.755
	0	0.029	0.035	0.037	0.071	0.054	0.060	0.758	0.849
	0.6	0.110	0.126	0.144	0.146	0.143	0.164	0.980	1.126
	1.2	0.442	0.442	0.446	0.457	0.450	0.453	0.985	0.992
7	-1.2	0.046	0.053	0.056	0.187	0.137	0.155	0.729	0.828
	-0.6	0.016	0.019	0.022	0.081	0.054	0.066	0.658	0.817
	0	0.006	0.006	0.006	0.006	0.006	0.006	1.058	1.105
	0.6	0.016	0.019	0.021	0.078	0.044	0.060	0.561	0.768
	1.2	0.051	0.058	0.058	0.191	0.139	0.161	0.729	0.842
8	-1.2	0.006	0.007	0.007	0.145	0.085	0.110	0.588	0.762
	-0.6	0.011	0.013	0.014	0.075	0.047	0.058	0.630	0.781
	0	0.029	0.032	0.035	0.029	0.033	0.035	1.158	1.219
	0.6	0.109	0.130	0.128	0.172	0.140	0.154	0.811	0.891
	1.2	0.437	0.428	0.476	0.587	0.470	0.559	0.801	0.952

Table 3: Comparison of our NR adaptive estimator and local profile likelihood (LPL) estimator with the conventional local linear (LL) and local linear exponential-tilting (ET) estimators for conditional variance, d=1, n=100

DGPs	x	Variance				MSE				Efficiency ratio		
		LL(1)	ET(2)	NR(3)	LPL(4)	LL(5)	ET(6)	NR(7)	LPL(8)	$\frac{(6)}{(5)}$	$\frac{(7)}{(5)}$	$\frac{(8)}{(5)}$
1	-1.2	2.178	2.233	0.590	0.663	2.198	2.264	0.695	0.725	1.030	0.316	0.330
	-0.6	0.466	0.135	0.055	0.061	0.513	0.135	0.064	0.064	0.263	0.124	0.124
	0	0.036	0.037	0.012	0.029	0.083	0.051	0.015	0.036	0.608	0.177	0.426
	0.6	0.098	0.091	0.062	0.070	0.130	0.091	0.070	0.072	0.700	0.542	0.558
	1.2	1.048	0.741	0.490	0.474	1.074	0.786	0.607	0.547	0.732	0.565	0.509
2	-1.2	0.011	0.007	0.002	0.004	0.015	0.007	0.002	0.004	0.478	0.147	0.235
	-0.6	0.090	0.087	0.030	0.049	0.117	0.087	0.032	0.050	0.743	0.272	0.424
	0	0.694	0.687	0.345	0.420	0.905	0.687	0.366	0.424	0.760	0.405	0.469
	0.6	5.185	4.808	3.366	3.544	6.986	4.890	3.796	3.742	0.700	0.543	0.536
	1.2	101.78	56.50	36.08	34.72	117.19	57.00	39.49	36.49	0.486	0.337	0.311
3	-1.2	2.188	2.235	0.476	0.562	2.227	2.244	0.564	0.599	1.008	0.253	0.269
	-0.6	0.469	0.136	0.056	0.088	0.541	0.140	0.061	0.089	0.259	0.113	0.165
	0	0.038	0.041	0.012	0.026	0.094	0.055	0.015	0.033	0.590	0.160	0.350
	0.6	0.099	0.092	0.059	0.075	0.151	0.093	0.065	0.076	0.617	0.430	0.502
	1.2	1.060	0.728	0.526	0.548	1.099	0.744	0.606	0.587	0.677	0.552	0.534
4	-1.2	0.013	0.009	0.003	0.006	0.028	0.017	0.004	0.010	0.612	0.149	0.342
	-0.6	0.092	0.090	0.028	0.042	0.138	0.094	0.029	0.043	0.681	0.208	0.309
	0	0.709	0.714	0.312	0.397	0.935	0.714	0.339	0.404	0.763	0.363	0.432
	0.6	5.192	4.817	2.864	3.238	7.129	4.876	3.406	3.487	0.684	0.478	0.489
	1.2	102.19	56.67	34.60	32.97	117.95	57.04	38.32	34.86	0.484	0.325	0.296
5	-1.2	0.135	0.174	0.163	0.170	0.143	0.214	0.215	0.201	1.492	1.501	1.402
	-0.6	0.021	0.024	0.022	0.022	0.060	0.024	0.026	0.022	0.402	0.423	0.370
	0	0.005	0.011	0.006	0.007	0.051	0.027	0.011	0.017	0.529	0.220	0.328
	0.6	0.024	0.024	0.020	0.021	0.060	0.024	0.024	0.022	0.396	0.399	0.356
	1.2	0.147	0.201	0.196	0.200	0.157	0.249	0.262	0.244	1.590	1.676	1.556
6	-1.2	0.001	0.001	0.001	0.001	0.005	0.002	0.001	0.001	0.327	0.227	0.250
	-0.6	0.009	0.010	0.009	0.011	0.031	0.010	0.010	0.011	0.343	0.341	0.351
	0	0.101	0.124	0.091	0.101	0.308	0.125	0.106	0.103	0.408	0.344	0.334
	0.6	1.260	1.379	1.243	1.202	3.281	1.420	1.472	1.268	0.433	0.449	0.387
	1.2	13.27	15.71	14.32	14.93	23.63	16.73	16.90	15.66	0.708	0.715	0.663
7	-1.2	0.138	0.147	0.154	0.168	0.159	0.157	0.180	0.176	0.983	1.128	1.107
	-0.6	0.022	0.025	0.021	0.023	0.083	0.029	0.022	0.023	0.348	0.265	0.279
	0	0.005	0.013	0.007	0.009	0.061	0.027	0.010	0.016	0.441	0.162	0.258
	0.6	0.026	0.029	0.022	0.028	0.085	0.032	0.026	0.028	0.380	0.309	0.326
	1.2	0.143	0.164	0.155	0.165	0.160	0.177	0.177	0.181	1.106	1.108	1.132
8	-1.2	0.002	0.003	0.002	0.002	0.016	0.011	0.004	0.007	0.687	0.242	0.434
	-0.6	0.010	0.015	0.011	0.012	0.049	0.017	0.011	0.013	0.349	0.229	0.260
	0	0.105	0.133	0.097	0.102	0.332	0.135	0.123	0.107	0.407	0.370	0.323
	0.6	1.277	1.485	1.130	1.194	3.453	1.511	1.501	1.275	0.438	0.435	0.369
	1.2	12.97	15.21	13.15	14.00	23.58	16.00	16.66	15.25	0.679	0.707	0.647

Table 4: Comparison of our NR adaptive estimator and local profile likelihood (LPL) estimator with the conventional local linear (LL) estimator for regression mean, d=2, n=200

DGP	x	Variance			MSE			Efficiency ratio	
		LL (1)	NR (2)	LPL (3)	LL (4)	NR (5)	LPL (6)	$\frac{(5)}{(4)}$	$\frac{(6)}{(4)}$
9	$(-1.2, -1.2)$	0.090	0.114	0.199	0.099	0.115	0.200	1.163	2.019
	$(-0.6, 0.6)$	0.032	0.048	0.042	0.125	0.077	0.138	0.616	1.108
	$(0, 0)$	0.019	0.037	0.026	0.163	0.076	0.200	0.466	1.224
	$(0.6, 0.6)$	0.031	0.050	0.041	0.136	0.086	0.154	0.633	1.135
	$(1.2, 1.2)$	0.095	0.130	0.217	0.101	0.130	0.218	1.279	2.150
10	$(-1.2, -1.2)$	0.036	0.033	0.051	0.044	0.045	0.061	1.003	1.373
	$(-0.6, 0.6)$	0.013	0.023	0.022	0.109	0.033	0.132	0.303	1.219
	$(0, 0)$	0.021	0.039	0.029	0.166	0.074	0.203	0.446	1.226
	$(0.6, 0.6)$	0.062	0.092	0.088	0.170	0.135	0.211	0.796	1.242
	$(1.2, 1.2)$	0.268	0.364	0.398	0.273	0.369	0.399	1.351	1.458
11	$(-1.2, -1.2)$	0.075	0.095	0.133	0.257	0.246	0.356	0.958	1.388
	$(-0.6, 0.6)$	0.029	0.051	0.044	0.276	0.123	0.344	0.445	1.243
	$(0, 0)$	0.017	0.039	0.025	0.175	0.064	0.255	0.366	1.453
	$(0.6, 0.6)$	0.031	0.054	0.050	0.297	0.140	0.378	0.470	1.270
	$(1.2, 1.2)$	0.083	0.110	0.151	0.276	0.282	0.376	1.023	1.365
12	$(-1.2, -1.2)$	0.021	0.022	0.035	0.206	0.151	0.198	0.733	0.961
	$(-0.6, 0.6)$	0.010	0.030	0.023	0.262	0.050	0.390	0.192	1.488
	$(0, 0)$	0.019	0.039	0.030	0.178	0.064	0.249	0.359	1.397
	$(0.6, 0.6)$	0.062	0.098	0.086	0.334	0.200	0.402	0.598	1.202
	$(1.2, 1.2)$	0.255	0.335	0.335	0.453	0.517	0.582	1.142	1.285

Table 5: Comparison of our NR adaptive estimator and local profile likelihood (LPL) estimator with the conventional local linear (LL) and local linear exponential-tilting (ET) estimators for conditional variance, d=2, n=200

DGPs	x	Variance				MSE				Efficiency ratio		
		LL(1)	ET(2)	NR(3)	LPL(4)	LL(5)	ET(6)	NR(7)	LPL(8)	$\frac{(6)}{(5)}$	$\frac{(7)}{(5)}$	$\frac{(8)}{(5)}$
9	x_1	0.739	1.471	0.879	1.017	2.663	1.922	1.820	1.739	0.722	0.684	0.653
	x_2	0.095	0.146	0.136	0.110	0.208	0.192	0.137	0.119	0.923	0.657	0.575
	x_3	0.035	0.041	0.058	0.039	0.256	0.258	0.120	0.137	1.005	0.470	0.534
	x_4	0.098	0.144	0.142	0.123	0.229	0.202	0.142	0.136	0.882	0.621	0.594
	x_5	0.757	1.525	0.837	1.195	2.915	1.888	1.726	1.776	0.648	0.592	0.609
10	x_1	0.181	0.052	0.028	0.040	2.682	0.091	0.040	0.061	0.034	0.015	0.023
	x_2	0.006	0.007	0.009	0.008	0.072	0.056	0.013	0.031	0.773	0.186	0.424
	x_3	0.057	0.066	0.065	0.051	0.209	0.136	0.069	0.068	0.651	0.330	0.328
	x_4	0.659	0.941	0.694	0.669	0.937	0.960	0.928	0.789	1.024	0.990	0.841
	x_5	9.276	13.970	8.765	9.814	12.175	17.978	18.142	16.755	1.477	1.490	1.376
11	x_1	0.556	1.322	0.939	0.951	0.941	1.503	1.379	1.319	1.597	1.466	1.402
	x_2	0.090	0.109	0.133	0.086	0.383	0.343	0.179	0.141	0.895	0.466	0.369
	x_3	0.035	0.039	0.060	0.038	0.360	0.372	0.144	0.148	1.031	0.400	0.412
	x_4	0.093	0.120	0.148	0.099	0.406	0.365	0.195	0.159	0.898	0.480	0.391
	x_5	0.561	1.371	1.025	1.038	1.052	1.486	1.364	1.320	1.412	1.297	1.255
12	x_1	0.021	0.006	0.003	0.004	0.679	0.006	0.003	0.004	0.008	0.005	0.006
	x_2	0.007	0.008	0.016	0.012	0.220	0.228	0.060	0.108	1.036	0.275	0.493
	x_3	0.055	0.069	0.071	0.051	0.294	0.193	0.083	0.080	0.655	0.282	0.272
	x_4	0.613	0.806	0.760	0.581	1.135	0.832	0.819	0.614	0.733	0.721	0.541
	x_5	8.659	11.262	8.230	8.211	9.529	14.075	14.426	13.459	1.477	1.514	1.413

Note. $x_1, ..., x_5$ represents points $(-1.2, -1.2)$, $(-0.6, 0.6)$, $(0, 0)$, $(0.6, 0.6)$, and $(1.2, 1.2)$, respectively.

Andrews, D. W. K. (1994a) Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica* 62, 43-72.

Andrews, D. W. K. (1994b) Empirical process methods in econometrics. In: R. F. Engle and D. McFadden (eds.). *Handbook of Econometrics,* Vol. IV. North-Holland, Amsterdam.

Andrews, D. W. K. (1995) Nonparametric kernel estimation for semiparametric models. *Econometric Theory* 11, 560-596.

Bernstein, D. S. (2005) *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*. Princeton University Press, Princeton.

Bickel, P. J. (1982) On adaptive estimation. *Annals of Statistics* 10, 647-671.

Claeskens, G., and M. Aerts (2000) On local estimating equations in additive multiparameter models. *Statistics and Probability Letters* 49, 139-148.

Claeskens, G., and I. Van Keilegom (2003) Bootstrap Confidence Bands for Regression Curves and their Derivatives. *Annals of Statistics* 31, 1852-1884.

Fan, J. (1992) Design-adaptive nonparametric regression. *Journal of the American Statistical Association* 87, 998-1004.

Fan, J. (1993) Local linear regression smoothers and their minimax efficiencies. *Annals of Statistics* 21, 196-216.

Fan, J., M. Farmen, and I. Gijbels (1998) Local maximum likelihood estimation and inference. *Journal of the Royal Statistics Society, Series B* 60*,* 591-608.

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

Fan, J., N. E. Heckman, and M. P. Wand (1995) Local polynomial kernel regression for generalized linear models and quasi-likelihood functions, *Journal of the American Statistical Association* 90, 141-150.

Fan, J. and Q. Yao (1998) Efficient estimation of conditional variance functions in stochastic regression, *Biometrika* **85**, 645-660.

Hansen, B. E. (2008) Uniform convergence rates for kernel regression. *Econometric Theory* 24, 726-748.

Härdle, W. and A. Tsybakov (1997) Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of Econometrics* 81, 233-242.

Kreiss, J. (1987) On adaptive estimation in stationary ARMA process. *Annals of Statistics* 15, 112-133.

Lee, A. J. (1990) *U-statistics: Theory and Practice*. Marcel Dekker, New York and Basel.

Linton, O., and Z. Xiao (2007) A nonparametric regression estimator that adapts to error distribution of unknown form. *Econometric Theory* 23, 371-413.

Masry, E. (1996a) Multivariate regression estimation: local polynomial fitting for time series. *Stochastic Processes and Their Applications* 65, 81-101.

Masry, E. (1996b) Multivariate local polynomial regression for time series: uniform strong consistency rates. *Journal of Time Series Analysis* 17, 571-599.

Rudin, W. (1976) *Principles of Mathematical Analysis*. 3rd ed., McGraw-Hill, London.

Ruppert, D., M. P. Wand, U. Holst, and O. Hössjer (1997) Local polynomial variance function estimation. *Technometrics* 39, 262-273.

Staniswalis, J. E. (1989) On the kernel estimate of a regression function in likelihood-based models. *Journal of the American Statistical Association* 84, 276-283.

White, H. (1994) *Estimation, Inference and Specification Analysis*. Cambridge University Press, New York.

Ziegelmann, F. (2002) Nonparametric estimation of volatility functions: the local exponential estimator. *Econometric Theory* 18, 985-991.

APPENDIX

In this appendix we first define some notation and provide some technical lemmas that are used in the proof of the main results in the text. The proofs of all technical lemmas are provided in the supplementary material.

A Notation and Some Preliminary Results

We use C to signify a generic constant whose exact value may vary from case to case. Let $e_1 \equiv (1, 0, ..., 0)^\top$ denote an $N \times 1$ vector with one in the first position and zero elsewhere. Let $\tilde{\mathbf{Z}}_i = H_1^{-1} \tilde{\mathbf{X}}_i$ and $\mathbf{Z}_i = H^{-1} \tilde{\mathbf{X}}_i$, where H_1 is defined as H but with h_1 in place of h. Let

$$K_{ix} \equiv K\left((x - X_i)/h\right) \text{ and } K_{ix,\mathbf{j}} \equiv \left((X_i - x)/h\right)^{\mathbf{j}} K_{ix}. \tag{A.1}$$

We write $A_n \simeq B_n$ to signify that $A_n = B_n (1 + o_p(1))$ as $n \to \infty$. Let

$$\bar{\varepsilon}_i \equiv \varepsilon_i\left(\boldsymbol{\beta}^0\right) = \frac{Y_i - P_i\left(\boldsymbol{\beta}_1^0\right)}{\sqrt{\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)}} = \frac{u_i + \delta_i}{\sqrt{\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)}}, \text{ and } \overrightarrow{\varepsilon}_i \equiv \frac{Y_i - P_i(\tilde{\boldsymbol{\beta}}_1)}{\sqrt{\varphi\left(P_i(\tilde{\boldsymbol{\beta}}_2)\right)}} = \frac{u_i + \delta_i - \upsilon_{1i}(x)}{\sqrt{\varphi\left(P_i(\tilde{\boldsymbol{\beta}}_2)\right)}}, \tag{A.2}$$

where $\delta_i \equiv m(X_i) - P_i\left(\boldsymbol{\beta}_1^0\right)$, $\upsilon_{1i}(x) \equiv P_i(\tilde{\boldsymbol{\beta}}_1) - P_i\left(\boldsymbol{\beta}_1^0\right)$, and $\varphi(u) = u$ or $\exp(u)$. Further, define

$$\upsilon_{0n} \equiv h_1 + n^{-1/2} h_1^{-d/2} \sqrt{\log n}, \ \upsilon_{1n} \equiv h_1^{p+1} + n^{-1/2} h_1^{-d/2} \sqrt{\log n}, \text{ and } \upsilon_{2n} \equiv \upsilon_{1n}\left(1 + (h/h_1)^p\right). \tag{A.3}$$

Let $\tilde{f}_i^{(s)}(e_i) = \frac{1}{n h_0^{s+1}} \sum_{j \neq i} k_0^{(s)}\left(\frac{e_i - \tilde{\varepsilon}_j}{h_0}\right)$, $s = 0, 1, 2, 3$, where $k_0^{(0)} = k_0$. We first study the uniform consistency of $\tilde{f}_i^{(s)}(\overrightarrow{\varepsilon}_i)$ with $f^{(s)}(\varepsilon_i)$ by the following two lemmas.

Lemma A.1 *Suppose that Assumptions A1-A7 hold. Then*

$$\max_{\{K_{ix} > 0\}} |\tilde{f}_i^{(s)}(\overrightarrow{\varepsilon}_i) - f^{(s)}(\bar{\varepsilon}_i)| = O_p(\upsilon_{3n,s}) \text{ for } s = 0, 1, 2, 3, \tag{A.4}$$

where $\upsilon_{3n,s} \equiv \upsilon_{2n} + (\upsilon_{0n} h_0^{-1-s} + \alpha_{n,s}) n^{-1/2} h_1^{-d/2} \sqrt{\log n} + \upsilon_{2n}^2 h_0^{-2-s}$ *and* $\alpha_{n,s} = h^{[(2p+d)/4 - (s+1)]} (\log n)^{s+1}$.

Lemma A.2 *Suppose that Assumptions A1-A7 hold. Then*

$$\max_{1 \le i \le n} \left|\tilde{f}_i^{(s)}(\bar{\varepsilon}_i) - f^{(s)}(\bar{\varepsilon}_i)\right| = O_p\left(h_0^{p+1} + n^{-1/2} h_0^{-1/2-s} \sqrt{\log n} + \upsilon_{1n}\right) \text{ for } s = 0, 1, 2, \tag{A.5}$$

and the above result is also true if one replaces $\bar{\varepsilon}_i \equiv \varepsilon_i\left(\boldsymbol{\beta}^0\right)$ *by* ε_i.

To proceed, we use linear functional notation and write $P\xi = \int \xi dP$ for any probability measure P and random variable $\xi(Z)$, where $Z = \left(X^\top, Y\right)^\top$. P_n denotes the empirical probability measure of the observations $\{Z_1, ..., Z_n\}$ sampled randomly from P.

Lemma A.3 (USLLN) *Let* $\theta = (x^\top, \boldsymbol{\beta}^\top)^\top$ *be an element of* $\Theta = \mathcal{X} \times \mathcal{B}$. *Let* $q_{n,1}(Z, \theta) = \log(f(\varepsilon(\boldsymbol{\beta})))$ $G_b(f(\varepsilon(\boldsymbol{\beta}))) K((x - X)/h)$, *and* $q_{n,2}(Z, \theta) = P(\boldsymbol{\beta}_2) G_b(f(\varepsilon(\boldsymbol{\beta}))) K((x - X)/h)$, *where* $Z = \left(X^\top, Y\right)^\top$, $P(\boldsymbol{\beta}_r)$ *(*$r = 1, 2$*) is defined as* $P_i(\boldsymbol{\beta}_r)$ *with* X_i *replaced by* X, *and* $\varepsilon(\boldsymbol{\beta}) = (Y - P(\boldsymbol{\beta}_1)) / \sqrt{\exp(P(\boldsymbol{\beta}_2))}$. *Under the conditions in Theorem 3.2, we have*

$$\sup_{\theta \in \Theta} \left|h^{-d}\left[P_n q_{n,r}(Z, \theta) - P q_{n,r}(Z_i, \theta)\right]\right| = O_{a.s.}\left(n^{-1/2} h^{-d/2} \sqrt{\log n}\right), \ r = 1, 2.$$

Lemma A.4 (Equicontinuity) *Let* $\theta = (x^\top, \boldsymbol{\beta}^\top)^\top$ *be an element of* $\Theta \equiv \mathcal{X} \times \mathcal{B}$. *For* $r = 1, 2$, *let* $p_1(Z_i, \theta) = \log(f(\varepsilon_i(\boldsymbol{\beta})))(1 - G_b(f(\varepsilon_i(\boldsymbol{\beta}))))K_h(x - X_i)$, $p_2(Z_i, \theta) = P_i(\boldsymbol{\beta}_2)(1 - G_b(f(\varepsilon_i(\boldsymbol{\beta}))))$ $K_h(x - X_i)$, *and* $Z_i = \left(X_i^\top, Y_i\right)^\top$. *Then under the conditions in Theorem 3.2,* $\bar{P}_{n,r}(\theta) \equiv E[P_n p_r(Z, \theta)]$ *is equicontinuous,* $r = 1, 2$.

B Proof of the Main Results

B.1 Proof of Proposition 2.1

Let $\alpha_n = 1/\sqrt{nh^d}$ and $\boldsymbol{\beta}^* \equiv (\boldsymbol{\beta}_1^{*\top}, \boldsymbol{\beta}_2^{*\top})^\top$ with $\boldsymbol{\beta}_r^* \equiv \sqrt{nh^d} H\left(\boldsymbol{\beta}_r - \boldsymbol{\beta}_r^0\right)$ for $r = 1, 2$. If $\bar{\boldsymbol{\beta}} \equiv (\bar{\boldsymbol{\beta}}_1^\top, \bar{\boldsymbol{\beta}}_2^\top)^\top$ maximizes (2.3), then $\bar{\boldsymbol{\beta}}^* \equiv (\bar{\boldsymbol{\beta}}_1^{*\top}, \bar{\boldsymbol{\beta}}_2^{*\top})^\top$ maximizes

$$\mathcal{L}_n(\boldsymbol{\beta}^*) = \frac{1}{n}\sum_{i=1}^n \log f\left(Y_i; P_i\left(\boldsymbol{\beta}_1^0\right) + \alpha_n \boldsymbol{\beta}_1^{*\top}\mathbf{Z}_i, P_i\left(\boldsymbol{\beta}_2^0\right) + \alpha_n \boldsymbol{\beta}_2^{*\top}\mathbf{Z}_i\right) K_{ix}.$$

To study the asymptotic properties of $\bar{\boldsymbol{\beta}}^*$, we resort to the quadratic approximation lemma of Fan et al. (1995) to the maximization of

$$L_n(\boldsymbol{\beta}^*) = \sum_{i=1}^n \{\log f(Y_i; P_i\left(\boldsymbol{\beta}_1^0\right) + \alpha_n \boldsymbol{\beta}_1^{*\top}\mathbf{Z}_i, P_i\left(\boldsymbol{\beta}_2^0\right) + \alpha_n \boldsymbol{\beta}_2^{*\top}\mathbf{Z}_i) - \log f\left(Y_i; P_i\left(\boldsymbol{\beta}_1^0\right), P_i\left(\boldsymbol{\beta}_2^0\right)\right)\} K_{ix}.$$

Notice that $\bar{\boldsymbol{\beta}}^*$ also maximizes $L_n(\boldsymbol{\beta}^*)$.

Let $q_r(y; \beta_1, \beta_2) \equiv \frac{\partial}{\partial \beta_r}\log\left(f(y; \beta_1, \beta_2)\right)$, $q_{rs}(y; \beta_1, \beta_2) \equiv \frac{\partial^2}{\partial\beta_r\partial\beta_s}\log\left(f(y; \beta_1, \beta_2)\right)$, and $q_{rst}(y; \beta_1, \beta_2) \equiv \frac{\partial^3}{\partial\beta_r\partial\beta_s\partial\beta_t}\log\left(f(y; \beta_1, \beta_2)\right)$, $r, s, t = 1, 2$. One can verify that

$$q_1(y; \beta_1, \beta_2) = \frac{-1}{\sqrt{\varphi(\beta_2)}}\frac{f'(\varepsilon(\beta))}{f(\varepsilon(\beta))}, \text{ and } q_2(y; \beta_1, \beta_2) = \frac{-\varphi'(\beta_2)}{2\varphi(\beta_2)}\left[\frac{f'(\varepsilon(\beta))}{f(\varepsilon(\beta))}\varepsilon(\beta) + 1\right],$$

where $\varepsilon(\beta) \equiv (y - \beta_1)/\sqrt{\varphi(\beta_2)}$. The expressions for q_{rs} and q_{rst} are complicated and are given in supplementary Appendix D. In addition, let $I(x) \equiv \begin{pmatrix} I_{11}(x) & I_{12}(x) \\ I_{21}(x) & I_{22}(x)\end{pmatrix}$, where $I_{rs}(x) \equiv E[q_r(Y_i; \beta_1^0(X_i), \beta_2^0(X_i))q_s(Y_i; \beta_1^0(X_i), \beta_2^0(X_i))|X_i = x]$. It is easy to verify that

$$I(x) = \begin{bmatrix} \frac{E\left[\psi^2(\varepsilon)\right]}{\sigma^2(x)} & \frac{E[\psi(\varepsilon)(\psi(\varepsilon)\varepsilon+1)]}{2\sigma^3(x)} \\ \frac{E[\psi(\varepsilon)(\psi(\varepsilon)\varepsilon+1)]}{2\sigma^3(x)} & \frac{E[\psi(\varepsilon)\varepsilon+1]^2}{4\sigma^4(x)} \end{bmatrix} = \begin{bmatrix} 1/\sigma(x) & 0 \\ 0 & 1/[2\sigma^2(x)] \end{bmatrix} A \begin{bmatrix} 1/\sigma(x) & 0 \\ 0 & 1/[2\sigma^2(x)] \end{bmatrix}$$

where $A = \begin{bmatrix} E\left[\psi^2(\varepsilon)\right] & E\left[\psi(\varepsilon)(\psi(\varepsilon)\varepsilon + 1)\right] \\ E\left[\psi(\varepsilon)(\psi(\varepsilon)\varepsilon + 1)\right] & E\left[\psi(\varepsilon)\varepsilon + 1\right]^2 \end{bmatrix}$ is p.d. by Assumption A2(ii). As a result, $I(x)$ is p.d. as $0 < \sigma(x) < \infty$.

By Taylor series expansions,

$$\begin{aligned} L_n(\boldsymbol{\beta}^*) &= \alpha_n \sum_{r=1}^2\sum_{i=1}^n q_r\left(Y_i; P_i\left(\boldsymbol{\beta}_1^0\right), P_i\left(\boldsymbol{\beta}_2^0\right)\right) K_{ix}\mathbf{Z}_i^\top \boldsymbol{\beta}_r^* \\ &\quad + \frac{\alpha_n^2}{2}\sum_{r=1}^2\sum_{s=1}^2\sum_{i=1}^n q_{rs}\left(Y_i; P_i\left(\boldsymbol{\beta}_1^0\right), P_i\left(\boldsymbol{\beta}_2^0\right)\right) K_{ix}\boldsymbol{\beta}_r^{*\top}\mathbf{Z}_i\mathbf{Z}_i^\top\boldsymbol{\beta}_s^* \\ &\quad + \frac{\alpha_n^3}{6}\sum_{r=1}^2\sum_{s=1}^2\sum_{t=1}^2\sum_{i=1}^n q_{rst}\left(Y_i; P_{1i}^\dagger, P_{2i}^\dagger\right) K_{ix} Z_i^\top \boldsymbol{\beta}_r^*\mathbf{Z}_i^\top\boldsymbol{\beta}_s^*\mathbf{Z}_i^\top\boldsymbol{\beta}_t^* \\ &\equiv L_{n1}(\boldsymbol{\beta}^*) + L_{n2}(\boldsymbol{\beta}^*) + L_{n3}(\boldsymbol{\beta}^*), \text{ say,} \end{aligned}$$

where $P_{ri}^\dagger$ lies between $P_i\left(\boldsymbol{\beta}_r^0\right)$ and $P_i\left(\boldsymbol{\beta}_r^0\right) + \alpha_n\boldsymbol{\beta}_r^{*\top}\mathbf{Z}_i$, $r = 1, 2$. Let

$$W_n \equiv \frac{1}{\sqrt{nh^d}}\sum_{i=1}^n \begin{pmatrix} q_1\left(Y_i; P_i\left(\boldsymbol{\beta}_1^0\right), P_i\left(\boldsymbol{\beta}_2^0\right)\right) K_{ix}\mathbf{Z}_i \\ q_2\left(Y_i; P_i\left(\boldsymbol{\beta}_1^0\right), P_i\left(\boldsymbol{\beta}_2^0\right)\right) K_{ix}\mathbf{Z}_i \end{pmatrix}, \text{ and } A_n \equiv \begin{pmatrix} A_{n,11} & A_{n,12} \\ A_{n,21} & A_{n,22} \end{pmatrix},$$

where $A_{n,rs} \equiv \alpha_n^2 \sum_{i=1}^n q_{rs}\left(Y_i; P_i\left(\boldsymbol{\beta}_1^0\right), P_i\left(\boldsymbol{\beta}_2^0\right)\right) K_{ix} \mathbf{Z}_i \mathbf{Z}_i^\top$. Then $L_{n1} = W_n^\top \boldsymbol{\beta}^*$ and $L_{n2} = \frac{1}{2}\boldsymbol{\beta}^{*\top} A_n \boldsymbol{\beta}^*$. By Assumption A2(iv) and the explicit expressions for $q_{rst}(y;\beta_1,\beta_2)$, $r,s,t = 1,2$, in Appendix D, $|L_{n3}(\boldsymbol{\beta}^*)| \le C^3 \bar{L}_{n3}$ for $\|\boldsymbol{\beta}^*\| \le C$, where $\bar{L}_{n3} = \frac{\alpha_n^3}{6}\sum_{r=1}^2\sum_{s=1}^2\sum_{t=1}^2\sum_{i=1}^n J(Y_i) K_{ix}\|\mathbf{Z}_i\|^3$. Noting that $E(\bar{L}_{n3}) = O(nh^d\alpha_n^3) = O((nh^d)^{-1/2}) = o(1)$ by Assumption A5, $\bar{L}_{n3} = o_p(1)$ by Markov inequality. Consequently $L_{n3}(\boldsymbol{\beta}^*) = o_p(1)$ uniformly in $\boldsymbol{\beta}^*$ in a compact set. Consequently, $L_n(\boldsymbol{\beta}^*) = W_n^\top \boldsymbol{\beta}^* + \frac{1}{2}\boldsymbol{\beta}^{*\top} A_n \boldsymbol{\beta}^* + o_p(1)$, which implies that $\bar{\boldsymbol{\beta}}^* = -A_n^{-1}W_n + o_p(1)$ provided that A_n is asymptotically non-singular.

By Taylor series expansions,

$$
\begin{aligned}
A_{n,rs} &= \alpha_n^2 \sum_{i=1}^n q_{rs}\left(Y_i; \beta_1^0(X_i), \beta_2^0(X_i)\right) K_{ix}\mathbf{Z}_i\mathbf{Z}_i^\top \\
&\quad -\alpha_n^2 \sum_{i=1}^n \sum_{t=1}^2 q_{rst}\left(Y_i; \beta_1^\dagger(X_i,x), \beta_2^\dagger(X_i x)\right) K_{ix} \sum_{|\mathbf{j}|=p+1}\left[\frac{1}{\mathbf{j}!}\frac{\partial^{|\mathbf{j}|}\beta_t^0(x)}{\partial^{j_1}x_1 \ldots \partial^{j_d}x_d}(X_i - x)^{\mathbf{j}}\right]\mathbf{Z}_i\mathbf{Z}_i^\top \\
&\equiv A_{n,rs,1} + A_{n,rs,2}, \text{ say,}
\end{aligned}
$$

where $\beta_l^\dagger(X_i,x)$ lies between $P_i\left(\boldsymbol{\beta}_l^0\right)$ and $\beta_l^0(X_i)$ for $l = 1,2$. By the weak law of large numbers (WLLN) and the information matrix inequality (e.g., White (1994, Ch. 4),

$$
A_{n,rs,1} = h^{-d}E\left[q_{rs}\left(Y_i; \beta_1^0(X_i), \beta_2^0(X_i)\right) K_{ix}\mathbf{Z}_i\mathbf{Z}_i^\top\right] + o_p(1) = -f_X(x) I_{rs}(x) M + o_p(1).
$$

By Assumptions A1-A5 and Markov inequality, we can show that $A_{n,rs,2} = O_p\left(h^{p+1}\right)$. It follows that $A_{n,rs} = -f_X(x) I_{rs}(x) M + o_p(1)$ and that $A_n = -\mathcal{I}_{\beta^0}(x) \otimes M + o_p(1)$ is asymptotically non-singular. By the Liapounov central limit theorem, we can readily show that $W_n - E(W_n) \xrightarrow{d} N\left(0, \mathcal{I}_{\beta^0}(x) \otimes \Gamma\right)$. Combining these results, we have

$$
\bar{\boldsymbol{\beta}}^* + A_n^{-1}E(W_n) = \bar{\boldsymbol{\beta}}^* - \left[\mathcal{I}_{\beta^0}^{-1}(x) \otimes M^{-1} + o_p(1)\right] E(W_n) \xrightarrow{d} N\left(0, \mathcal{I}_{\beta^0}^{-1}(x) \otimes \left[M^{-1}\Gamma M^{-1}\right]\right).
$$

In addition, by Taylor expansions and Assumption A2,

$$
\begin{aligned}
E(W_n) &= \sqrt{nh^d} E\begin{pmatrix} q_1\left(Y_i; \beta_1^0(X_i), \beta_2^0(X_i)\right) K_{ix} Z_i \\ q_2\left(Y_i; \beta_1^0(X_i), \beta_2^0(X_i)\right) K_{ix} Z_i \end{pmatrix} \\
&\quad -\sqrt{nh^d} E\begin{pmatrix} \sum_{r=1}^2 q_{1r}\left(Y_i; \beta_1^0(X_i), \beta_2^0(X_i)\right) K_{ix}\gamma_{ri}(x) Z_i \\ \sum_{r=1}^2 q_{2r}\left(Y_i; \beta_1^0(X_i), \beta_2^0(X_i)\right) K_{ix}\gamma_{ri}(x) Z_i \end{pmatrix} + \sqrt{nh^d} O\left(h^{p+2}\right) \\
&= \sqrt{nh^d} h^{p+1}\left(\mathcal{I}_{\beta^0}(x) \otimes B\right)\begin{pmatrix} \boldsymbol{\beta}_1^{0(p+1)}(x) \\ \boldsymbol{\beta}_2^{0(p+1)}(x) \end{pmatrix} + o(1),
\end{aligned}
$$

where $\gamma_{ri}(x) \equiv \sum_{|\mathbf{j}|=p+1}\frac{1}{\mathbf{j}!}\frac{\partial^{|\mathbf{j}|}\beta_r^0(x)}{\partial^{j_1}x_1 \ldots \partial^{j_d}x_d}(X_i - x)^{\mathbf{j}}$ and we use the fact that $E\left[q_1\left(Y_i; \beta_1^0(X_i), \beta_2^0(X_i)\right)|X_i\right]$ $= -E[\psi(\varepsilon_i)]/\sigma(X_i) = 0$ and $E\left[q_2\left(Y_i; \beta_1^0(X_i), \beta_2^0(X_i)\right)|X_i\right] = -E[\rho(\varepsilon_i)]\varphi'\left(\beta_2^0(X_i)\right)/[2\sigma^2(X_i)] = 0$ by Assumption A2. Then Proposition 2.1 follows. ■

B.2 Proof of Theorem 3.1

For notational simplicity, denote $\tilde{I}_n(\boldsymbol{\beta}(x); \tilde{f})$ as $\tilde{I}_n(\boldsymbol{\beta})$ and $\tilde{S}_n(\boldsymbol{\beta}(x); \tilde{f})$ as $\tilde{S}_n(\boldsymbol{\beta})$. We frequently suppress the dependence of $\hat{\boldsymbol{\beta}}(x)$, $\tilde{\boldsymbol{\beta}}(x)$, $\boldsymbol{\beta}(x)$, etc., on x. Denote $G_b(\tilde{f}_i(\varepsilon_i(\tilde{\boldsymbol{\beta}})))$ as $\tilde{G}_i$ and $G_b(f(\bar{\varepsilon}_i))$ as G_i, where $\bar{\varepsilon}_i \equiv \varepsilon_i\left(\boldsymbol{\beta}^0\right)$. Write

$$
\sqrt{nh^d}\begin{pmatrix} H(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0) \\ H(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^0) \end{pmatrix} = \sqrt{nh^d}\begin{pmatrix} H(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0) \\ H(\tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^0) \end{pmatrix} + \sqrt{nh^d}\bar{H}\tilde{I}_n(\tilde{\boldsymbol{\beta}})^{-1}\tilde{S}_n(\tilde{\boldsymbol{\beta}}).
$$

Expanding $\tilde{s}_i(\tilde{\boldsymbol{\beta}})$ in (2.16) around $\boldsymbol{\beta}^0$, we obtain $\tilde{s}_i(\tilde{\boldsymbol{\beta}}) = \tilde{s}_i\left(\boldsymbol{\beta}^0\right) + \frac{\partial \tilde{s}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) + R_i\left(\boldsymbol{\beta}^*\right)$, where β^* is the element-by-element intermediate value between $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^0$, and the jth element of $R_i\left(\boldsymbol{\beta}\right)$ is given by $R_{i,j}\left(\boldsymbol{\beta}\right) \equiv \frac{1}{2}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)^\top \frac{\partial^2 \tilde{s}_{i(j)}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)$ with $\tilde{s}_{i(j)}\left(\boldsymbol{\beta}\right)$ being the jth element of $\tilde{s}_i\left(\boldsymbol{\beta}\right)$. Let

$$
\begin{aligned}
\bar{S}_n\left(\boldsymbol{\beta}, \tilde{\boldsymbol{\beta}}\right) &\equiv \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)\tilde{s}_{n,i}\left(\boldsymbol{\beta}\right) \otimes \tilde{\mathbf{X}}_i, \\
\bar{R}_{1n}\left(\boldsymbol{\beta}, \tilde{\boldsymbol{\beta}}\right) &\equiv \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)\left(\tilde{G}_i \frac{\partial \tilde{s}_i\left(\boldsymbol{\beta}\right)}{\partial \boldsymbol{\beta}^\top} \otimes \tilde{\mathbf{X}}_i\right) + \tilde{I}_n\left(\boldsymbol{\beta}\right), \\
\bar{R}_{2n}\left(\boldsymbol{\beta}^*, \tilde{\boldsymbol{\beta}}\right) &\equiv \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)\tilde{G}_i R_i\left(\boldsymbol{\beta}^*\right) \otimes \tilde{\mathbf{X}}_i,
\end{aligned}
$$

where recall $\tilde{s}_{n,i}\left(\boldsymbol{\beta}\right) \equiv G_b(\tilde{f}_i\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right))\tilde{s}_i\left(\boldsymbol{\beta}\right) + \tilde{s}_i^*(\tilde{\boldsymbol{\beta}})$, and $\tilde{s}_i^*\left(\boldsymbol{\beta}\right)$ is defined in (2.17). Then $\tilde{S}_n(\tilde{\boldsymbol{\beta}}) = \bar{S}_n(\boldsymbol{\beta}^0, \tilde{\boldsymbol{\beta}}) - \tilde{I}_n\left(\boldsymbol{\beta}^0\right)(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) + \bar{R}_{1n}(\boldsymbol{\beta}^0, \tilde{\boldsymbol{\beta}})(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) + \bar{R}_{2n}(\boldsymbol{\beta}^*, \tilde{\boldsymbol{\beta}})$.

It follows that

$$
\begin{aligned}
&\sqrt{nh^d}\begin{pmatrix} H\left(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0\right) \\ H\left(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^0\right) \end{pmatrix} \\
= &\sqrt{nh^d}\left[\bar{H}^{-1}\tilde{I}_n(\tilde{\boldsymbol{\beta}})\bar{H}^{-1}\right]^{-1}\bar{H}^{-1}\bar{S}_n(\boldsymbol{\beta}^0, \tilde{\boldsymbol{\beta}}) \\
&+\left\{\sqrt{nh^d}\begin{pmatrix} H(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0) \\ H(\tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^0) \end{pmatrix} - \left[\bar{H}^{-1}\tilde{I}_n(\tilde{\boldsymbol{\beta}})\bar{H}^{-1}\right]^{-1}\left[\bar{H}^{-1}\tilde{I}_n\left(\boldsymbol{\beta}^0\right)\bar{H}^{-1}\right]\sqrt{nh^d}\begin{pmatrix} H(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0) \\ H(\tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^0) \end{pmatrix}\right\} \\
&+\sqrt{nh^d}\left[\bar{H}^{-1}\tilde{I}_n(\tilde{\boldsymbol{\beta}})\bar{H}^{-1}\right]^{-1}\bar{H}^{-1}\bar{R}_{1n}(\boldsymbol{\beta}^0, \tilde{\boldsymbol{\beta}})\bar{H}^{-1}\bar{H}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) \\
&+\left[\bar{H}^{-1}\tilde{I}_n(\tilde{\boldsymbol{\beta}})\bar{H}^{-1}\right]^{-1}\sqrt{nh^d}\bar{H}^{-1}\bar{R}_{2n}(\boldsymbol{\beta}^*, \tilde{\boldsymbol{\beta}}).
\end{aligned}
$$

It suffices to prove the theorem by showing that

$$
\begin{aligned}
\bar{H}^{-1}\left[\tilde{I}_n(\tilde{\boldsymbol{\beta}}) - I_n\left(\boldsymbol{\beta}^0, f\right)\right]\bar{H}^{-1} &= O_p\left(h^\epsilon\right) \text{ for some } \epsilon > 0, &\text{(B.1)} \\
\sqrt{nh^d}\left[\bar{H}^{-1}\bar{S}_n(\boldsymbol{\beta}^0; \tilde{\boldsymbol{\beta}}) - S_n\left(\boldsymbol{\beta}^0, f\right)\right] &= o_p\left(1\right), &\text{(B.2)} \\
\bar{H}^{-1}\bar{R}_{1n}(\boldsymbol{\beta}^0, \tilde{\boldsymbol{\beta}})\bar{H}^{-1} &= o_p\left(1\right), &\text{(B.3)} \\
\sqrt{nh^d}\bar{H}^{-1}\bar{R}_{2n}(\boldsymbol{\beta}^*, \tilde{\boldsymbol{\beta}}) &= o_p\left(1\right). &\text{(B.4)}
\end{aligned}
$$

B.2.1 Proof of (B.1)

Recall $\bar{\varepsilon}_i \equiv \varepsilon_i\left(\boldsymbol{\beta}^0\right)$, $\overrightarrow{\varepsilon}_i \equiv \varepsilon_i(\tilde{\boldsymbol{\beta}})$, $\tilde{G}_i \equiv G_b(\tilde{f}_i\left(\overrightarrow{\varepsilon}_i\right))$ and $G_i \equiv G_b\left(f\left(\bar{\varepsilon}_i\right)\right)$. Note that

$$
\bar{H}^{-1}\tilde{I}_n(\tilde{\boldsymbol{\beta}})\bar{H}^{-1} = \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)\tilde{G}_i\left[\tilde{s}_i(\tilde{\boldsymbol{\beta}})\tilde{s}_i(\tilde{\boldsymbol{\beta}})^\top\right] \otimes \mathbf{Z}_i\mathbf{Z}_i^\top
$$

and a typical element of $\bar{H}^{-1}\tilde{I}_n(\tilde{\boldsymbol{\beta}})\bar{H}^{-1}$ is $\frac{1}{nh^d}\sum_{i=1}^{n} K_{ix,\mathbf{j}}\tilde{G}_i\tilde{q}_{r,i}(\tilde{\boldsymbol{\beta}})\tilde{q}_{s,i}(\tilde{\boldsymbol{\beta}})$, where $K_{ix,\mathbf{j}}$ is defined in (A.1), $r, s = 1, 2$, and $0 \leq |\mathbf{j}| \leq 2p$. It suffices to show $T_{n\mathbf{j}}\left(r, s\right) \equiv \frac{1}{nh^d}\sum_{i=1}^{n} K_{ix,\mathbf{j}}\{\tilde{G}_i\tilde{q}_{r,i}(\tilde{\boldsymbol{\beta}})\tilde{q}_{s,i}(\tilde{\boldsymbol{\beta}}) - q_r\left(Y_i; P_i\left(\boldsymbol{\beta}_1^0\right), P_i\left(\boldsymbol{\beta}_2^0\right)\right)q_s\left(Y_i; P_i\left(\boldsymbol{\beta}_1^0\right), P_i\left(\boldsymbol{\beta}_2^0\right)\right)\} = O_p\left(h^\epsilon\right)$ for $r, s = 1, 2$, and $0 \leq |\mathbf{j}| \leq 2p$. Noting that $\tilde{q}_r\tilde{q}_s - q_rq_s = q_r\left(\tilde{q}_s - q_s\right) + q_s\left(\tilde{q}_r - q_r\right) + \left(\tilde{q}_r - q_r\right)\left(\tilde{q}_s - q_s\right)$, we have $T_{n\mathbf{j}}\left(r, s\right) = T_{1n\mathbf{j}}\left(r, s\right) + T_{2n\mathbf{j}}\left(r, s\right) +$

$T_{3n\mathbf{j}}(r,s)$, where

$$T_{1n\mathbf{j}}(r,s) \equiv \frac{1}{nh^d}\sum_{i=1}^{n} K_{ix,\mathbf{j}}\tilde{G}_i q_{r,i}(\boldsymbol{\beta})\left[\tilde{q}_{s,i}(\tilde{\boldsymbol{\beta}}) - q_{s,i}(\boldsymbol{\beta}^0)\right],$$

$$T_{2n\mathbf{j}}(r,s) \equiv \frac{1}{nh^d}\sum_{i=1}^{n} K_{ix,\mathbf{j}}\tilde{G}_i \left\{\tilde{q}_{r,i}(\tilde{\boldsymbol{\beta}}) - q_{r,i}(\boldsymbol{\beta}^0)\right\}\left[\tilde{q}_{s,i}(\tilde{\boldsymbol{\beta}}) - q_{s,i}(\boldsymbol{\beta}^0)\right],$$

$$T_{3n\mathbf{j}}(r,s) \equiv \frac{1}{nh^d}\sum_{i=1}^{n} K_{ix,\mathbf{j}}\left(1-\tilde{G}_i\right) q_{r,i}(\boldsymbol{\beta}^0)\, q_{s,i}(\boldsymbol{\beta}^0),$$

and $q_{s,i}(\boldsymbol{\beta}) \equiv q_s(Y_i; P_i(\boldsymbol{\beta}_1), P_i(\boldsymbol{\beta}_2))$ for $s = 1,2$. We complete the proof of (B.1) by showing $T_{ln\mathbf{j}}(r,s) = O_p(h^{\epsilon})$ for $l = 1,2,3$, $r,s = 1,2$, $0 \le |\mathbf{j}| \le 2p$ in Lemmas B.1-B.3 below.

Lemma B.1 *Suppose that the conditions in Theorem 3.1 hold. Then* $T_{1n\mathbf{j}}(r,s) = O_p(h^{\epsilon})$ *for* $r,s = 1,2$, $0 \le |\mathbf{j}| \le 2p$.

Lemma B.2 *Suppose that the conditions in Theorem 3.1 hold. Then* $T_{2n\mathbf{j}}(r,s) = O_p(h^{\epsilon})$ *for* $r,s = 1,2$, $0 \le |\mathbf{j}| \le 2p$.

Lemma B.3 *Suppose that the conditions in Theorem 3.1 hold. Then* $T_{3n\mathbf{j}}(r,s) = O_p(h^{\epsilon})$ *for* $r,s = 1,2$, $0 \le |\mathbf{j}| \le 2p$.

B.2.2 The proof of (B.2)

Note that $\sqrt{nh^d}\bar{H}^{-1}\bar{S}_n(\boldsymbol{\beta}^0,\tilde{\boldsymbol{\beta}}) - \sqrt{nh^d}\bar{H}^{-1}S_n(\boldsymbol{\beta}^0; f) = (nh^d)^{-1/2}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)[(\tilde{G}_i\tilde{s}_i(\boldsymbol{\beta}^0) + \tilde{s}_i^*(\tilde{\boldsymbol{\beta}})) - s_i(\boldsymbol{\beta}^0)] \otimes \mathbf{Z}_i$, of which a typical element is

$$S_{n\mathbf{j},r} \equiv \frac{1}{\sqrt{nh^d}}\sum_{i=1}^{n} K_{ix,\mathbf{j}}\left[\left(\tilde{G}_i\tilde{q}_{r,i}(\boldsymbol{\beta}^0) + \tilde{q}^*_{r,i}(\tilde{\boldsymbol{\beta}})\right) - q_{r,i}(\boldsymbol{\beta}^0)\right], \tag{B.5}$$

where $r = 1,2$, and $0 \le |\mathbf{j}| \le p$. It suffices to show that $S_{n\mathbf{j},r} = o_p(1)$, $r = 1,2$. We only consider the $r = 2$ case, since the $r = 1$ case is similar but simpler. (Without bias correction, the proof for the case $r = 1$ would be analogous to that in Appendix A.2.2 of Linton and Xiao (2007)). We make the following decomposition:

$$\begin{aligned}
S_{n\mathbf{j},2} &= \frac{1}{\sqrt{nh^d}}\sum_{i=1}^{n} K_{ix,\mathbf{j}}\left\{q_{2,i}(\boldsymbol{\beta}^0)[\tilde{G}_i - 1] - \log\left(f(\bar{\varepsilon}_i)\varphi\left(P_i(\boldsymbol{\beta}_2^0)\right)^{-1/2}\right) g_b(f(\bar{\varepsilon}_i)) f'(\bar{\varepsilon}_i)\frac{\bar{\varepsilon}_i\varphi'\left(P_i(\boldsymbol{\beta}_2^0)\right)}{2\varphi\left(P_i(\boldsymbol{\beta}_2^0)\right)}\right\} \\
&\quad + \frac{1}{\sqrt{nh^d}}\sum_{i=1}^{n} K_{ix,\mathbf{j}}\left\{\tilde{G}_i\left[\tilde{q}_{2,i}(\boldsymbol{\beta}^0) - q_{2,i}(\boldsymbol{\beta}^0)\right]\right\} \\
&\quad - \frac{1}{\sqrt{nh^d}}\sum_{i=1}^{n} K_{ix,\mathbf{j}}\left\{\log\left(\tilde{f}_i(\vec{\varepsilon}_i)\varphi\left(P_i(\tilde{\boldsymbol{\beta}}_2)\right)^{-1/2}\right) g_b\left(\tilde{f}_i(\vec{\varepsilon}_i)\right)\tilde{f}_i'(\vec{\varepsilon}_i)\frac{\vec{\varepsilon}_i\varphi'\left(P_i(\tilde{\boldsymbol{\beta}}_2)\right)}{2\varphi\left(P_i(\tilde{\boldsymbol{\beta}}_2)\right)}\right. \\
&\qquad\qquad \left. - \log\left(f(\bar{\varepsilon}_i)\varphi\left(P_i(\boldsymbol{\beta}_2^0)\right)^{-1/2}\right) g_b(f(\bar{\varepsilon}_i)) f'(\bar{\varepsilon}_i)\frac{\bar{\varepsilon}_i\varphi'\left(P_i(\boldsymbol{\beta}_2^0)\right)}{2\varphi\left(P_i(\boldsymbol{\beta}_2^0)\right)}\right\} \\
&\equiv \mathcal{S}_{1n\mathbf{j}} + \mathcal{S}_{2n\mathbf{j}} - \mathcal{S}_{3n\mathbf{j}}, \text{ say.}
\end{aligned}$$

By Lemmas B.4-B.6 below, $S_{n\mathbf{j},2} = o_p(1)$.

Lemma B.4 *Suppose that the conditions in Theorem 3.1 hold. Then* $\mathcal{S}_{1n\mathbf{j}} = o_p(1)$ *for* $0 \le |\mathbf{j}| \le p$.

Lemma B.5 *Suppose that the conditions in Theorem 3.1 hold. Then* $\mathcal{S}_{2n\mathbf{j}} = o_p(1)$ *for* $0 \le |\mathbf{j}| \le p$.

Lemma B.6 *Suppose that the conditions in Theorem 3.1 hold. Then* $\mathcal{S}_{3n\mathbf{j}} = o_p(1)$ *for* $0 \le |\mathbf{j}| \le p$.

B.2.3 The proof of (B.3)

To show (B.3), we first decompose $\bar{H}^{-1}\bar{R}_{1n}(\boldsymbol{\beta}^0,\tilde{\boldsymbol{\beta}})\bar{H}^{-1}$ as follows

$$\begin{aligned}
&\bar{H}^{-1}\bar{R}_{1n}(\boldsymbol{\beta}^0,\tilde{\boldsymbol{\beta}})\bar{H}^{-1} \\
=\;& \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\bar{H}^{-1}\tilde{G}_i\left[\frac{\partial \tilde{s}_i\left(\boldsymbol{\beta}^0\right)}{\partial \boldsymbol{\beta}^{\top}}\otimes \tilde{\mathbf{X}}_i + \left(\tilde{s}_i\left(\boldsymbol{\beta}^0\right)\tilde{s}_i\left(\boldsymbol{\beta}^0\right)^{\top}\right)\otimes\left(\tilde{\mathbf{X}}_i\tilde{\mathbf{X}}_i^{\top}\right)\right]\bar{H}^{-1} \\
=\;& \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\bar{H}^{-1}[\tilde{G}_i\tilde{s}_i\left(\boldsymbol{\beta}^0\right)\tilde{s}_i\left(\boldsymbol{\beta}^0\right)^{\top} - G_i s_i\left(\boldsymbol{\beta}^0\right)s_i\left(\boldsymbol{\beta}^0\right)^{\top}]\otimes\left(\tilde{\mathbf{X}}_i\tilde{\mathbf{X}}_i^{\top}\right)\bar{H}^{-1} \\
&+\frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\bar{H}^{-1}\left\{\left[\tilde{G}_i\frac{\partial \tilde{s}_i\left(\boldsymbol{\beta}^0\right)}{\partial \boldsymbol{\beta}^{\top}} - G_i\frac{\partial s_i\left(\boldsymbol{\beta}^0\right)}{\partial \boldsymbol{\beta}^{\top}}\right]\otimes\tilde{\mathbf{X}}_i\right\}\bar{H}^{-1} \\
&+\frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)G_i\bar{H}^{-1}\left[\frac{\partial s_i\left(\boldsymbol{\beta}^0\right)}{\partial \boldsymbol{\beta}^{\top}}\otimes\tilde{\mathbf{X}}_i + \left(s_i\left(\boldsymbol{\beta}^0\right)s_i\left(\boldsymbol{\beta}^0\right)^{\top}\right)\otimes\left(\tilde{\mathbf{X}}_i\tilde{\mathbf{X}}_i^{\top}\right)\right]\bar{H}^{-1} \\
\equiv\;& \mathcal{R}_{1n}+\mathcal{R}_{2n}+\mathcal{R}_{3n},\ \text{say.}
\end{aligned}$$

By Lemmas B.7-B.8 below, $\mathcal{R}_{1n}=o_p(1)$ and $\mathcal{R}_{2n}=o_p(1)$. We are left to show that $\mathcal{R}_{3n}=o_p(1)$. Using (2.8), we can readily obtain

$$\frac{\partial s_i\left(\boldsymbol{\beta}^0\right)}{\partial\boldsymbol{\beta}^{\top}}=\begin{pmatrix} \frac{\psi'(\bar{\varepsilon}_i)}{\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)} & \frac{\varphi'\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)\left[\psi'(\bar{\varepsilon}_i)\bar{\varepsilon}_i+\psi(\bar{\varepsilon}_i)\right]}{2\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{3/2}} \\ \frac{\varphi'\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)\left[\psi'(\bar{\varepsilon}_i)\bar{\varepsilon}_i+\psi(\bar{\varepsilon}_i)\right]}{2\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{3/2}} & \frac{2c_{i\varphi}[\psi(\bar{\varepsilon}_i)\bar{\varepsilon}_i+1]+\varphi'\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^2\bar{\varepsilon}_i\left[\psi'(\bar{\varepsilon}_i)\bar{\varepsilon}_i+\psi(\bar{\varepsilon}_i)\right]}{4\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^2}\end{pmatrix}\otimes\tilde{\mathbf{X}}_i^{\top},$$

where $c_{i\varphi}=\varphi'\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^2-\varphi''\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right).$

For notational simplicity, we focus on the case where $\varphi(u)=\exp(u)$. In this case, we have $\frac{\partial s_i(\boldsymbol{\beta}^0)}{\partial\boldsymbol{\beta}^{\top}}=\bar{A}_i\otimes\tilde{\mathbf{X}}_i^{\top}$ and $s_i\left(\boldsymbol{\beta}^0\right)s_i\left(\boldsymbol{\beta}^0\right)^{\top}=\bar{B}_i$, where

$$\bar{A}_i\equiv\begin{pmatrix}\frac{\psi'(\bar{\varepsilon}_i)}{\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)} & \frac{\psi'(\bar{\varepsilon}_i)\bar{\varepsilon}_i+\psi(\bar{\varepsilon}_i)}{2[\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)]^{1/2}} \\ \frac{\psi'(\bar{\varepsilon}_i)\bar{\varepsilon}_i+\psi(\bar{\varepsilon}_i)}{2[\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)]^{1/2}} & \frac{\bar{\varepsilon}_i\left[\psi'(\bar{\varepsilon}_i)\bar{\varepsilon}_i+\psi(\bar{\varepsilon}_i)\right]}{4}\end{pmatrix}\ \text{and}\ \bar{B}_i\equiv\begin{bmatrix}\frac{\psi^2(\bar{\varepsilon}_i)}{\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)} & \frac{\psi(\bar{\varepsilon}_i)[\psi(\bar{\varepsilon}_i)\bar{\varepsilon}_i+1]}{2[\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)]^{1/2}} \\ \frac{\psi(\bar{\varepsilon}_i)[\psi(\bar{\varepsilon}_i)\bar{\varepsilon}_i+1]}{2[\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)]^{1/2}} & \frac{[\psi(\bar{\varepsilon}_i)\bar{\varepsilon}_i+1]^2}{4}\end{bmatrix}.$$

Using these notation and $H^{-1}\tilde{\mathbf{X}}_i=\mathbf{Z}_i$, we have

$$\begin{aligned}
\mathcal{R}_{3n} &= \frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right)G_i(\bar{A}_i+\bar{B}_i)\otimes(\mathbf{Z}_i\mathbf{Z}_i^{\top}) \\
&= \frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right)(\bar{A}_i+\bar{B}_i)\otimes(\mathbf{Z}_i\mathbf{Z}_i^{\top})+\frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right)(G_i-1)(\bar{A}_i+\bar{B}_i)\otimes(\mathbf{Z}_i\mathbf{Z}_i^{\top}) \\
&\equiv \mathcal{R}_{3n,1}+\mathcal{R}_{3n,2},\ \text{say.}
\end{aligned}$$

We want to show that $\mathcal{R}_{3n,1}=o_p(1)$ and $\mathcal{R}_{3n,2}=o_p(1)$. Consider $\mathcal{R}_{3n,1}$ first. Using the definition of $\bar{\varepsilon}_i$ in (A.2), we can readily show that $\mathcal{R}_{3n,1}=\overline{\mathcal{R}}_{3n,1}+o_p(1)$, where $\overline{\mathcal{R}}_{3n,1}=\frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right)(A_i+B_i)\otimes(\mathbf{Z}_i\mathbf{Z}_i^{\top})$, and A_i and B_i are analogously defined as $\bar{A}_i$ and $\bar{B}_i$ respectively with $\bar{\varepsilon}_i$ being replaced by ε_i. Consider the following auxiliary location-scale regression model

$$Y_i^{\dagger}=P_i\left(\boldsymbol{\beta}_1^0\right)+\sqrt{\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)}\varepsilon_i.$$

Recall that ε_i is independent of X_i and has PDF $f(\cdot)$. So the conditional density of $Y_i^{\dagger}$ given X_i is

$f_x^\dagger(\boldsymbol{\beta};Y_i,X_i) \equiv f\left(\varepsilon_i^\dagger(\boldsymbol{\beta})\right)/\sqrt{\exp(P_i(\boldsymbol{\beta}_2))}$, where $\varepsilon_i^\dagger(\boldsymbol{\beta}) \equiv \left(Y_i^\dagger - P_i(\boldsymbol{\beta}_1)\right)/\sqrt{\exp(P_i(\boldsymbol{\beta}_2))}$. The corresponding conditional-log-likelihood function is given by

$$Q_{nx}(\boldsymbol{\beta}) \equiv \frac{1}{n}\sum_{i=1}^{n}\log\left(f\left(\varepsilon_i^\dagger(\boldsymbol{\beta})\right)/\sqrt{\exp(P_i(\boldsymbol{\beta}_2))}\right) = \frac{1}{n}\sum_{i=1}^{n}\left[\log f\left(\varepsilon_i^\dagger(\boldsymbol{\beta})\right) - \frac{1}{2}P_i(\boldsymbol{\beta}_2)\right].$$

It is easy to show that for this auxiliary maximum likelihood estimation problem, the Hessian and information matrices for the ith observation, when evaluated at the true parameter values $\boldsymbol{\beta}_1^0$ and $\boldsymbol{\beta}_2^0$, are given by $A_i \otimes \tilde{\mathbf{X}}_i\tilde{\mathbf{X}}_i^\top$ and $B_i \otimes \tilde{\mathbf{X}}_i\tilde{\mathbf{X}}_i^\top$, respectively. By the information matrix equality, we have

$$E\left[(A_i + B_i)\otimes \tilde{\mathbf{X}}_i\tilde{\mathbf{X}}_i^\top | X_i\right] = 0.$$

It follows that $E\left(\overline{\mathcal{R}}_{3n,1}\right) = 0$. In addition, we can show that $\mathrm{Var}\left(\overline{\mathcal{R}}_{3n,1}\right) = O(\left(nh^d\right)^{-1}) = o(1)$. Hence $\mathcal{R}_{3n,1} = o_p(1)$ by Chebyshev inequality.

We are left to show that $\mathcal{R}_{3n,2} = o_p(1)$. By Minkowski inequality, the fact that $\|A\otimes B\| = \|A\|\,\|B\|$ (e.g., Bernstein (2005, p. 398)), the compact support of K and Assumption A1, we have

$$E\|\mathcal{R}_{3n,2}\| \le h^{-d}E\left[K_{ix}(1-G_i)\left\|\bar{A}_i + \bar{B}_i\right\|\left\|\mathbf{Z}_i\mathbf{Z}_i^\top\right\|\right] \le C\ ER_{3n,2}.$$

where $ER_{3n,2} = h^{-d}E\left[K_{ix}(1-G_i)\left\|\bar{A}_i+\bar{B}_i\right\|\right]$. By the fact that $0 \le 1 - G_i \le \mathbf{1}\{f(\bar{\varepsilon}_i) \le 2b\}$ and Hölder inequality,

$$\begin{aligned} ER_{3n,2} &= h^{-d}E\left[K_{ix}(1-G_i)\left\|\bar{A}_i+\bar{B}_i\right\|\right] \\ &\le \left\{h^{-d}E\left[\left\|\bar{A}_i+\bar{B}_i\right\|^\gamma K_{ix}\right]\right\}^{1/\gamma}\left\{h^{-d}E\left[P(f(\bar{\varepsilon}_i)\le 2b|X_i)K_{ix}\right]\right\}^{(\gamma-1)/\gamma}. \end{aligned}$$

Under Assumption A2 and as in the proof of Lemma B.2, we can readily show that $h^{-d}E\left[\left\|\bar{A}_i+\bar{B}_i\right\|^\gamma K_{ix}\right]$ $= O(1)$ and $h^{-d}E\left[P(f(\bar{\varepsilon}_i)\le 2b|X_i)K_{ix}\right] = O\left(b^{1/2}\right)$. It follows that $ER_{3n,2} = O\left(b^{(\gamma-1)/2\gamma}\right) = o(1)$. Then $\mathcal{R}_{3n,2} = o_p(1)$ by the Markov inequality.

Lemma B.7 *Suppose that the conditions in Theorem 3.1 hold. Then* $\mathcal{R}_{1n} = o_p(1)$.

Lemma B.8 *Suppose that the conditions in Theorem 3.1 hold. Then* $\mathcal{R}_{2n} = o_p(1)$.

B.2.4 The proof of (B.4)

To show (B.4), note that a typical element of $\sqrt{nh^d}\bar{H}^{-1}\bar{R}_{2n}(\boldsymbol{\beta}^*,\tilde{\boldsymbol{\beta}})$ is given by

$$\frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n}K_{ix,\mathbf{j}}\tilde{G}_i\frac{\partial^2\tilde{q}_{r,i}(\boldsymbol{\beta}^*)}{\partial\boldsymbol{\beta}_{rj}\partial\boldsymbol{\beta}_{sl}}(\tilde{\boldsymbol{\beta}}_{rj}-\boldsymbol{\beta}_{rj}^0)(\tilde{\boldsymbol{\beta}}_{sl}-\boldsymbol{\beta}_{sl}^0).$$

So it suffices to show $||\tilde{G}_i\partial^2\tilde{q}_{r,i}(\boldsymbol{\beta})/(\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top)|| = o_p(\sqrt{nh^d})$ for any $\boldsymbol{\beta}\in\mathcal{B}_0 = \{\boldsymbol{\beta}: \left\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\right\| \le Cn^{-1/2}h^{-d/2}\}$. This is true because, by Lemma A.1 and Assumption A7 we can show that uniformly in $\boldsymbol{\beta}\in\mathcal{B}_0$, $||\tilde{G}_i\partial^2\tilde{q}_{r,i}(\boldsymbol{\beta})$ $/(\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top)|| = O_p(1 + b^{-1}\upsilon_{3n,3} + b^{-2}\upsilon_{3n,2} + b^{-3}\upsilon_{3n,1}) = o_p(\sqrt{nh^d})$. ■

B.3 Proof of Theorem 3.2

Let $\tilde{Q}_n(x,\boldsymbol{\beta})$ and $Q(x,\boldsymbol{\beta})$ be as defined in (2.21) and (3.3). By White (1994, Theorem 3.4), it suffices to show

$$\sup_{(x,\beta)\in\mathcal{X}\times\mathcal{B}}\left|\tilde{Q}_n(x,\boldsymbol{\beta}) - Q(x,\boldsymbol{\beta})\right| \to 0 \text{ wp} \to 1, \quad \text{(B.6)}$$

and for every neighborhood $\mathcal{B}_0$ of $\boldsymbol{\beta}^0$,

$$\limsup_{n\to\infty} \sup_{\beta\in\mathcal{B}/\mathcal{B}^0} \left[Q\left(x,\boldsymbol{\beta}\right) - Q\left(x,\boldsymbol{\beta}^0\right)\right] < 0 \text{ uniformly in } x. \tag{B.7}$$

We first show (B.6). Write

$$\begin{aligned}\tilde{Q}_n\left(x,\boldsymbol{\beta}\right) &= \frac{1}{n}\sum_{i=1}^n \log\left(\tilde{f}_i\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)\right) G_b\left(\tilde{f}_i\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)\right) K_h\left(x-X_i\right) - \frac{1}{2n}\sum_{i=1}^n P_i\left(\boldsymbol{\beta}_2\right) G_b\left(\tilde{f}_i\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)\right) K_h\left(x-X_i\right)\\ &\equiv \tilde{Q}_{n1}\left(x,\boldsymbol{\beta}\right) - \tilde{Q}_{n2}\left(x,\boldsymbol{\beta}\right).\end{aligned}$$

Let $\theta \equiv \left(x^\top, \boldsymbol{\beta}^\top\right)^\top$. We expand $\tilde{Q}_{n1}\left(\theta\right)$ about $f\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)$:

$$\begin{aligned}\tilde{Q}_{n1}\left(\theta\right) &= \frac{1}{n}\sum_{i=1}^n \log\left(f\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)\right) G_b\left(f\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)\right) K_h\left(x-X_i\right)\\ &\quad+\frac{1}{n}\sum_{i=1}^n \frac{1}{\tilde{f}_i^*\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)} G_b\left(\tilde{f}_i^*\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)\right) K_h\left(x-X_i\right)\left(\tilde{f}_i\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right) - f\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)\right)\\ &\quad+\frac{1}{n}\sum_{i=1}^n \log\left(\tilde{f}_i^*\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)\right) g_b\left(\tilde{f}_i^*\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)\right) K_h\left(x-X_i\right)\left(\tilde{f}_i\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right) - f\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)\right)\\ &\equiv Q_{n1}\left(\theta\right) + R_{n1a}\left(\theta\right) + R_{n1b}\left(\theta\right),\end{aligned}$$

where $\tilde{f}_i^*\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)$ lies between $\tilde{f}_i\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)$ and $f\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)$. First, uniformly in β,

$$\sup_\theta \left|R_{n1a}\left(\theta\right)\right| \le \frac{1}{b}\sup_\varepsilon \left|\tilde{f}_i\left(\varepsilon\right) - f\left(\varepsilon\right)\right| \sup_{x\in\mathcal{X}} \frac{1}{n}\sum_{i=1}^n \left|K_h\left(x-X_i\right)\right| \to 0 \text{ wp} \to 1,$$

where the last line follows from the fact that $\sup_\varepsilon \left|\tilde{f}_i\left(\varepsilon\right) - f\left(\varepsilon\right)\right|/b \to 0$ wp$\to$ 1 by Lemma A.2 and

$$\sup_{x\in\mathcal{X}} \frac{1}{n}\sum_{i=1}^n \left|K_h\left(x-X_i\right)\right| \simeq \sup_{x\in\mathcal{X}} \int K\left(u\right) f_X\left(x-hu\right) du \le C\int K\left(u\right) du = C.$$

Similarly,

$$\sup_\theta \left|R_{n1b}\left(\theta\right)\right| \le \left|\frac{-\log(b)}{b}\right| \sup_\varepsilon \left|\tilde{f}_i\left(\varepsilon\right) - f\left(\varepsilon\right)\right| \sup_{x\in\mathcal{X}} \frac{1}{n}\sum_{i=1}^n \left|K_h\left(x-X_i\right)\right| \to 0 \text{ wp} \to 1.$$

Thus uniformly in θ, $\tilde{Q}_{n1}\left(\theta\right) \to Q_{n1}\left(\theta\right)$ wp$\to$ 1. Similarly, we can show that uniformly in θ, $\tilde{Q}_{n2}\left(\theta\right) \to Q_{n2}\left(\theta\right)$ wp$\to$ 1, where $Q_{n2}\left(\theta\right) \equiv \left(2n\right)^{-1}\sum_{i=1}^n P_i\left(\boldsymbol{\beta}_2\right) G_b\left(f\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)\right) K_h\left(x-X_i\right)$. Now by Lemma A.3, uniformly in θ and wp$\to$ 1,

$$\tilde{Q}_n\left(\theta\right) \to E\left\{\left[\log\left(f\left(\varepsilon\left(\boldsymbol{\beta}\right)\right)\right) - \frac{1}{2}P\left(\boldsymbol{\beta}_2\right)\right] G_b\left(f\left(\varepsilon\left(\boldsymbol{\beta}\right)\right)\right) K_h\left(x-X\right)\right\}.$$

Now, let $P_n\left(\theta\right) \equiv \left[\log\left(f\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)\right) - \frac{1}{2}P_i\left(\boldsymbol{\beta}_2\right)\right]\left[1 - G_b\left(f\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)\right)\right] K_h\left(x-X_i\right)$. By Lemma A.4, $\bar{P}_n\left(\theta\right) \equiv E\left[P_n\left(\theta\right)\right]$ is equicontinuous. Notice that $E\left|K_h\left(x-X_i\right)\right| < \infty$ and

$$\begin{aligned}&E\left|\left[\log\left(f\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)\right) - \frac{1}{2}P_i\left(\boldsymbol{\beta}_2\right)\right]\left[1 - G_b\left(f\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)\right)\right] K_h\left(x-X_i\right)\right|\\ \le\ &E\left|\left[\log\left(f\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)\right) - \frac{1}{2}P_i\left(\boldsymbol{\beta}_2\right)\right] \cdot 1\left\{f\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right) \le 2b\right\} K_h\left(x-X_i\right)\right| = o\left(1\right)\end{aligned}$$

by the dominated convergence theorem. So $\bar{P}_n(\theta) = o(1)$. It follows from Rudin (1976, Exercise 7.16) that $\sup_\theta \bar{P}_n(\theta) = o(1)$. Consequently, $\tilde{Q}_n(\theta) \to Q(\theta)$ uniformly in θ wp$\to$ 1.

Now, note that the elements of $\boldsymbol{\beta}^0(x)$ correspond to $m(x)$ and $\log \sigma^2(x)$ and their derivatives of up to order p, which are uniquely defined for each x in the interior of $\mathcal{X}$ (c.f., Aerts and Claeskens, 1997). Hence (B.7) holds. ■

B.4 Proof of Theorem 3.3

Given Assumption A3, an element-by-element mean value expansion of $\partial\tilde{Q}_n(x,\boldsymbol{\beta}^+)/\partial\boldsymbol{\beta}$ about $\boldsymbol{\beta}^0$ gives

$$0 = \frac{\partial \tilde{Q}_n}{\partial \boldsymbol{\beta}}(x,\boldsymbol{\beta}^+) = \frac{\partial \tilde{Q}_n}{\partial \boldsymbol{\beta}}\left(x,\boldsymbol{\beta}^0\right) + \frac{\partial^2 \tilde{Q}_n}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top}(x,\boldsymbol{\beta}^*)\left(\boldsymbol{\beta}^+ - \boldsymbol{\beta}^0\right), \tag{B.8}$$

where $\boldsymbol{\beta}^*$ is a random variable such that elements of $\boldsymbol{\beta}^*$ lies on the segment joining the corresponding elements of $\boldsymbol{\beta}^+$ and $\boldsymbol{\beta}_0$, and hence $\boldsymbol{\beta}^+ \to \boldsymbol{\beta}_0$ a.s. From (B.8), we have

$$\begin{aligned}\sqrt{nh^d}\bar{H}\left(\boldsymbol{\beta}^+ - \boldsymbol{\beta}^0\right) &= -\left\{\bar{H}^{-1}\frac{\partial^2 \tilde{Q}_n}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top}(x,\boldsymbol{\beta}^*)\bar{H}^{-1}\right\}^{-1}\sqrt{nh^d}\bar{H}^{-1}\frac{\partial \tilde{Q}_n}{\partial \boldsymbol{\beta}}\left(x,\boldsymbol{\beta}^0\right) \\ &\equiv B_n(x,\boldsymbol{\beta}^*)^{-1} S_n\left(x,\boldsymbol{\beta}^0\right). \end{aligned} \tag{B.9}$$

The proof is completed by showing that

$$S_n\left(x,\boldsymbol{\beta}^0\right) - \sqrt{nh^d}h^{p+1}\overline{\mathcal{I}}_{\beta^0}(x)\otimes B \xrightarrow{d} N\left(0,\overline{\mathcal{I}}_{\beta^0}(x)\otimes M\right), \tag{B.10}$$

$$\left\|\bar{H}^{-1}\frac{\partial^2 \tilde{Q}_n}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top}\left(x,\boldsymbol{\beta}^0\right)\bar{H}^{-1} - \Gamma\right\| = o_p(1), \tag{B.11}$$

$$\left\|\bar{H}^{-1}\left[\frac{\partial^2 \tilde{Q}_n}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top}(x,\boldsymbol{\beta}^*) - \frac{\partial^2 \tilde{Q}_n}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top}\left(x,\boldsymbol{\beta}^0\right)\right]\bar{H}^{-1}\right\| = o_p(1). \tag{B.12}$$

The study of (B.11) parallels that of (B.1). We can prove (B.12) by showing that

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\|\le Cn^{-1/2}h^{-d/2}}\left\|\bar{H}^{-1}\left[\frac{\partial^2 \tilde{Q}_n}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top}(x,\boldsymbol{\beta}) - \frac{\partial^2 \tilde{Q}_n}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top}\left(x,\boldsymbol{\beta}^0\right)\right]\bar{H}^{-1}\right\| = o_p(1)$$

by standard uniform consistency arguments and applying Lemmas A.1-A.2 repeatedly; see also the proof of (B.4). Below, we focus on the proof of (B.10).

By a geometric expansion, we can write

$$S_n\left(x,\boldsymbol{\beta}^0\right) = \frac{1}{\sqrt{nh^d}}\bar{H}^{-1}\sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\beta}}\left[\left(\log \tilde{f}_i\left(\varepsilon_i\left(\boldsymbol{\beta}^0\right)\right) - \frac{1}{2}P_i\left(\boldsymbol{\beta}_2^0\right)\right)G_b\left(\tilde{f}_i\left(\varepsilon_i\left(\boldsymbol{\beta}^0\right)\right)\right)\right]K_{ix} = \sum_{r=1}^6 \mathcal{J}_r,$$

where

$$\begin{aligned}\mathcal{J}_1 = -\tfrac{1}{\sqrt{nh^d}}\textstyle\sum_{i=1}^n &\left\{\begin{pmatrix}\psi(\bar{\varepsilon}_i)/\sqrt{\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)} \\ \frac{1}{2}\left(\psi(\bar{\varepsilon}_i)\bar{\varepsilon}_i + 1\right)\end{pmatrix} G_b\left(\tilde{f}_i(\bar{\varepsilon}_i)\right)\right. \\ &\left.+\begin{pmatrix}1 \\ \bar{\varepsilon}_i\end{pmatrix}\log\left(f(\bar{\varepsilon}_i)/\sqrt{\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)}\right)g_b\left(f(\bar{\varepsilon}_i)\right)f'(\bar{\varepsilon}_i)\right\}\otimes \mathbf{Z}_i K_{ix}, \\ \mathcal{J}_2 = \tfrac{1}{\sqrt{nh^d}}\textstyle\sum_{i=1}^n &\begin{pmatrix} f^{-1}(\bar{\varepsilon}_i)\left(\tilde{f}_i'(\bar{\varepsilon}_i) - f'(\bar{\varepsilon}_i)\right)/\sqrt{\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)} \\ \frac{1}{2}f^{-1}(\bar{\varepsilon}_i)\left(\tilde{f}_i'(\bar{\varepsilon}_i) - f'(\bar{\varepsilon}_i)\right)\bar{\varepsilon}_i\end{pmatrix}\otimes \mathbf{Z}_i G_b\left(\tilde{f}_i(\bar{\varepsilon}_i)\right)K_{ix},\end{aligned}$$

$$\mathcal{J}_3 = -\tfrac{1}{\sqrt{nh^d}} \sum_{i=1}^n \begin{pmatrix} f^{-1}(\bar{\varepsilon}_i) \tilde{f}_i'(\bar{\varepsilon}_i) \left(\tilde{f}_i(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)\right) / \sqrt{\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)} \\ \tfrac{1}{2} f^{-1}(\bar{\varepsilon}_i) \tilde{f}_i'(\bar{\varepsilon}_i) \left(\tilde{f}_i(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)\right) \bar{\varepsilon}_i \end{pmatrix} \otimes \mathbf{Z}_i G_b\left(\tilde{f}_i(\bar{\varepsilon}_i)\right) K_{ix},$$

$$\mathcal{J}_4 = \tfrac{1}{\sqrt{nh^d}} \sum_{i=1}^n \begin{pmatrix} f^{-2}(\bar{\varepsilon}_i) \tilde{f}_i^{-1}(\bar{\varepsilon}_i) \tilde{f}_i'(\bar{\varepsilon}_i) \left(\tilde{f}_i(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)\right)^2 / \exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{1/2} \\ \tfrac{1}{2} f^{-2}(\bar{\varepsilon}_i) \tilde{f}_i^{-1}(\bar{\varepsilon}_i) \tilde{f}_i'(\bar{\varepsilon}_i) \left(\tilde{f}_i(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)\right)^2 \bar{\varepsilon}_i \end{pmatrix} \otimes \mathbf{Z}_i G_b\left(\tilde{f}_i(\bar{\varepsilon}_i)\right) K_{ix},$$

$$\mathcal{J}_5 = \tfrac{1}{\sqrt{nh^d}} \sum_{i=1}^n \begin{pmatrix} 1 \\ \bar{\varepsilon}_i \end{pmatrix} \otimes \mathbf{Z}_i K_h(x - X_i) \left[\log \tilde{f}_i(\bar{\varepsilon}_i) g_b\left(\tilde{f}_i(\bar{\varepsilon}_i)\right) \tilde{f}_i'(\bar{\varepsilon}_i) - \log f(\bar{\varepsilon}_i) g_b(f(\bar{\varepsilon}_i)) f'(\bar{\varepsilon}_i)\right],$$

and

$$\mathcal{J}_6 = -\tfrac{1}{2\sqrt{nh^d}} \sum_{i=1}^n \begin{pmatrix} 1 \\ \bar{\varepsilon}_i \end{pmatrix} \otimes \mathbf{Z}_i P_i\left(\boldsymbol{\beta}_2^0\right) \left(g_b\left(\tilde{f}_i(\bar{\varepsilon}_i)\right) \tilde{f}_i'(\bar{\varepsilon}_i) - g_b(f(\bar{\varepsilon}_i)) f'(\bar{\varepsilon}_i)\right) K_{ix}.$$

It suffices to show $\mathcal{J}_1 - \sqrt{nh^d} h^{p+1} \overline{\mathcal{I}}_{\beta^0}(x) \otimes B \xrightarrow{d} N(0, \overline{\mathcal{I}}_{\beta^0}(x) \otimes M)$ and each of other terms is $o_p(1)$.

To analyze $\mathcal{J}_1$, let $\tilde{f}_i = \tilde{f}_i(\bar{\varepsilon}_i)$, $f_i = f(\bar{\varepsilon}_i)$, and

$$\begin{aligned} a_i(\boldsymbol{\beta}) &= \bar{H}^{-1} \left\{ \begin{pmatrix} \psi(\varepsilon_i(\boldsymbol{\beta})) / \sqrt{\exp(P_i(\boldsymbol{\beta}_2))} \\ \tfrac{1}{2}[\psi(\varepsilon_i(\boldsymbol{\beta})) \varepsilon_i(\boldsymbol{\beta}) + 1] \end{pmatrix} G_b(f(\varepsilon_i(\boldsymbol{\beta}))) \right. \\ &\quad \left. + \begin{pmatrix} 1 \\ \varepsilon_i(\boldsymbol{\beta}) \end{pmatrix} \log\left(f(\varepsilon_i(\boldsymbol{\beta})) / \sqrt{\exp(P_i(\boldsymbol{\beta}_2))}\right) g_b(f(\varepsilon_i(\boldsymbol{\beta}))) f'(\varepsilon_i(\boldsymbol{\beta})) \right\} \otimes \tilde{\mathbf{X}}_i K_{ix}. \end{aligned}$$

Noticing that $G_b(\tilde{f}_i) - G_b(f_i) = g_b(f_i)(\tilde{f}_i - f_i) + \frac{1}{2} g_b'(f_i^*)(\tilde{f}_i - f_i)^2$, where f_i^* lies between $\tilde{f}_i$ and f_i, we have

$$\begin{aligned} \mathcal{J}_1 &= -\frac{1}{\sqrt{nh^d}} \sum_{i=1}^n a_i\left(\boldsymbol{\beta}^0\right) + \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n \begin{pmatrix} \psi(\bar{\varepsilon}_i) / \sqrt{\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)} \\ \tfrac{1}{2}[\psi(\bar{\varepsilon}_i) \bar{\varepsilon}_i + 1] \end{pmatrix} \otimes \mathbf{Z}_i g_b(f_i) \left(\tilde{f}_i - f_i\right) K_{ix} \\ &\quad + \frac{1}{2\sqrt{nh^d}} \sum_{i=1}^n \begin{pmatrix} \psi(\bar{\varepsilon}_i) / \sqrt{\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)} \\ \tfrac{1}{2}[\psi(\bar{\varepsilon}_i) \bar{\varepsilon}_i + 1] \end{pmatrix} \otimes \mathbf{Z}_i g_b'(f_i^*) \left(\tilde{f}_i - f_i\right)^2 K_{ix} \\ &\equiv -\mathcal{J}_{11} + \mathcal{J}_{12} + \mathcal{J}_{13}, \text{ say.} \end{aligned}$$

Using a rough bound on the last term, we have by Lemma A.2 and Assumption A7

$$\mathcal{J}_{13} = \sqrt{nh^d} O_p\left(\left(h_0^{p+1} + n^{-1/2} h_0^{-1/2} \sqrt{\log n} + \upsilon_{1n}\right)^2\right) = o_p(1).$$

Let $\overline{Q}_{in}(\boldsymbol{\beta}) \equiv \left[\log(f(\varepsilon_i(\boldsymbol{\beta}))) - \frac{1}{2} P_i(\boldsymbol{\beta}_2)\right] G_b(f(\varepsilon_i(\boldsymbol{\beta}))) K_{ix}$, $Q_{in}(\boldsymbol{\beta}) \equiv \left[\log f(\varepsilon_i(\boldsymbol{\beta})) - \frac{1}{2} P_i(\boldsymbol{\beta}_2)\right] K_{ix}$, and $\varsigma_n(\boldsymbol{\beta}) \equiv h^{-d}\{E\left[\overline{Q}_{in}(\boldsymbol{\beta})\right] - E[Q_{in}(\boldsymbol{\beta})]\} \to 0$. We can verify that i) $\varsigma_n\left(\boldsymbol{\beta}^0\right) \to 0$, (ii) $\varsigma_n(\boldsymbol{\beta})$ is differentiable in a small ϵ_0-neighborhood $N_{\epsilon_0}\left(\boldsymbol{\beta}^0\right)$ of $\boldsymbol{\beta}^0$, (iii) $\varsigma_n'(\boldsymbol{\beta})$ converges uniformly on $N_{\epsilon_0}\left(\boldsymbol{\beta}^0\right)$. Then by Theorem 7.17 of Rudin (1976), we have

$$\begin{aligned} E\mathcal{J}_{11} &= \sqrt{nh^{-d}} E\left[a_i\left(\boldsymbol{\beta}^0\right)\right] = \sqrt{nh^{-d}} E\left[\frac{\partial}{\partial \boldsymbol{\beta}} \overline{Q}_{in}\left(\boldsymbol{\beta}^0\right)\right] \\ &= \sqrt{nh^{-d}} E\left[\frac{\partial}{\partial \boldsymbol{\beta}} Q_{in}\left(\boldsymbol{\beta}^0\right)\right] \{1 + o(1)\} \\ &= -\sqrt{nh^{-d}} E\left(\begin{pmatrix} \psi(\bar{\varepsilon}_i) / \sqrt{\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)} \\ \tfrac{1}{2}[\psi(\bar{\varepsilon}_i) \bar{\varepsilon}_i + 1] \end{pmatrix} \otimes \mathbf{Z}_i K_{ix}\right) \{1 + o(1)\} \\ &= \sqrt{nh^d} h^{p+1} \left(\overline{\mathcal{I}}_{\beta^0}(x) \otimes B\right) \{1 + o(1)\}. \end{aligned}$$

Similarly, $\mathrm{Var}(\mathcal{J}_{11}) = h^{-d}\mathrm{Var}\big(a_i\left(\boldsymbol{\beta}^0\right)\big) \simeq h^{-d}\mathrm{Var}\left(\begin{pmatrix} \psi\left(\varepsilon_i\right)/\sqrt{\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)} \\ \frac{1}{2}\left[\psi\left(\varepsilon_i\right)\varepsilon_i+1\right] \end{pmatrix} \otimes \mathbf{Z}_i K_{ix}\right) \to \overline{\mathcal{I}}_{\beta^0}\left(x\right) \otimes M$. By the Liapounov central limit theorem,

$$\mathcal{J}_{11} \xrightarrow{d} N\left(\overline{\mathcal{I}}_{\beta^0}\left(x\right) \otimes B, \overline{\mathcal{I}}_{\beta^0}\left(x\right) \otimes M\right). \tag{B.13}$$

For $\mathcal{J}_{12}$, by Lemma A.2, $\tilde{f}_i\left(\bar{\varepsilon}_i\right) - \overline{f}\left(\bar{\varepsilon}_i\right) = \frac{1}{nh_0}\sum_{j\neq i}\left\{k_0\left(\frac{\bar{\varepsilon}_i-\tilde{\varepsilon}_j}{h_0}\right) - k_0\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\right\}$ can be written as

$$\frac{1}{nh_0^2}\sum_{j\neq i} k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\frac{m\left(X_j\right)-\tilde{m}\left(X_j\right)}{\sigma\left(X_j\right)} + \frac{1}{nh_0^2}\sum_{j\neq i} k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\varepsilon_j\frac{\sigma\left(X_j\right)-\tilde{\sigma}\left(X_j\right)}{\sigma\left(X_j\right)} + o_p\left(n^{-1/2}h^{-d/2}\right),$$

where $o_p\left(n^{-1/2}h^{-d/2}\right)$ holds uniformly in $\{K_{ix}>0\}$. Using $\tilde{f}_i - f_i = (\tilde{f}_i - \overline{f}_i) + (\overline{f}_i - f_i)$ and the expressions for $\mathcal{V}\left(\bar{\varepsilon}_i\right)$ and $\mathcal{B}\left(\bar{\varepsilon}_i\right)$ in (C.30) and (C.31), we can decompose $\mathcal{J}_{12}$ as follows

$$\begin{aligned}
& \frac{1}{\sqrt{nh^d}}\sum_{i=1}^n \begin{pmatrix} \psi\left(\bar{\varepsilon}_i\right)/\sqrt{\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)} \\ \frac{1}{2}\left[\psi\left(\bar{\varepsilon}_i\right)\bar{\varepsilon}_i+1\right] \end{pmatrix} \otimes \mathbf{Z}_i g_b\left(f_i\right)\frac{1}{nh_0^2}\sum_{j\neq i} k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\frac{m\left(X_j\right)-\tilde{m}\left(X_j\right)}{\sigma\left(X_j\right)}K_{ix} \\
& + \frac{1}{\sqrt{nh^d}}\sum_{i=1}^n \begin{pmatrix} \psi\left(\bar{\varepsilon}_i\right)/\sqrt{\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)} \\ \frac{1}{2}\left[\psi\left(\bar{\varepsilon}_i\right)\bar{\varepsilon}_i+1\right] \end{pmatrix} \otimes \mathbf{Z}_i g_b\left(f_i\right)\frac{1}{nh_0^2}\sum_{j\neq i} k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\varepsilon_j\frac{\sigma\left(X_j\right)-\tilde{\sigma}\left(X_j\right)}{\sigma\left(X_j\right)}K_{ix} \\
& + \frac{1}{\sqrt{nh^d}}\sum_{i=1}^n \begin{pmatrix} \psi\left(\bar{\varepsilon}_i\right)/\sqrt{\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)} \\ \frac{1}{2}\left[\psi\left(\bar{\varepsilon}_i\right)\bar{\varepsilon}_i+1\right] \end{pmatrix} \otimes \mathbf{Z}_i g_b\left(f_i\right)\mathcal{V}\left(\bar{\varepsilon}_i\right)K_{ix} \\
& + \frac{1}{\sqrt{nh^d}}\bar{H}^{-1}\sum_{i=1}^n \begin{pmatrix} \psi\left(\bar{\varepsilon}_i\right)/\sqrt{\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)} \\ \frac{1}{2}\left[\psi\left(\bar{\varepsilon}_i\right)\bar{\varepsilon}_i+1\right] \end{pmatrix} \otimes \tilde{\mathbf{X}}_i g_b\left(f_i\right)\mathcal{B}\left(\bar{\varepsilon}_i\right)K_{ix} + o_p\left(1\right) \\
\equiv\ & \mathcal{J}_{12a} + \mathcal{J}_{12b} + \mathcal{J}_{12c} + \mathcal{J}_{12d} + o_p\left(1\right).
\end{aligned}$$

For $\mathcal{J}_{12a}$ and $\mathcal{J}_{12b}$, we can write them as the sum of a third order U-statistic and a term that is asymptotically negligible. Using the standard theory for third order U-statistics (e.g., Lee, 1990), we can show $\mathcal{J}_{12a}$ and $\mathcal{J}_{12b}$ are each $o_p\left(1\right)$. Writing $\mathcal{J}_{12c}$ as a second order U-statistic we can verify that $E\left[\mathcal{J}_{12c}\right]^2 = o\left(1\right)$ and thus $\mathcal{J}_{12c} = o_p\left(1\right)$. For $\mathcal{J}_{12d}$, we verify that $\mathcal{J}_{12d} = o_p\left(n^{1/2}h^{d/2}h^{p+1}\right) = o_p\left(1\right)$. Consequently $\mathcal{J}_{12} = o_p\left(1\right)$ and $\mathcal{J}_1 \xrightarrow{d} N\left(\overline{\mathcal{I}}_{\beta}\left(x\right)\otimes B, \overline{\mathcal{I}}_{\beta}\left(x\right)\otimes M\right)$.

For $\mathcal{J}_2$, a typical element of $\mathcal{J}_2$ is

$$\begin{aligned}
& \frac{1}{2^r\sqrt{nh^d}}\bar{H}^{-1}\sum_{i=1}^n K_{ix,\mathbf{j}}f^{-1}\left(\bar{\varepsilon}_i\right)\left(\tilde{f}_i'\left(\bar{\varepsilon}_i\right) - \overline{f}'\left(\bar{\varepsilon}_i\right)\right)\bar{\varepsilon}_i^r\left[\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)\right]^{-\frac{2-r}{2}}G_b\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right) \\
\simeq\ & \frac{1}{2^r\sqrt{nh^d}}\bar{H}^{-1}\sum_{i=1}^n K_{ix,\mathbf{j}}f^{-1}\left(\bar{\varepsilon}_i\right)\left(\tilde{f}_i'\left(\bar{\varepsilon}_i\right) - \overline{f}'\left(\bar{\varepsilon}_i\right)\right)\bar{\varepsilon}_i^r\left[\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)\right]^{-\frac{2-r}{2}}G_i \\
& + \frac{1}{2^r\sqrt{nh^d}}\bar{H}^{-1}\sum_{i=1}^n K_{ix,\mathbf{j}}f^{-1}\left(\bar{\varepsilon}_i\right)\left(\overline{f}'\left(\bar{\varepsilon}_i\right) - f'\left(\bar{\varepsilon}_i\right)\right)\bar{\varepsilon}_i^r\left[\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)\right]^{-\frac{2-r}{2}}G_i \\
\equiv\ & \mathcal{J}_{21} + \mathcal{J}_{22},
\end{aligned}$$

where $0 \le |\mathbf{j}| \le p$, $r = 0, 1$, and recall $G_i \equiv G_b(f_i)$. A typical element of $\mathcal{J}_{21}$ is

$$
\begin{aligned}
& \frac{1}{2^r\sqrt{nh^d}} \bar{H}^{-1} \sum_{i=1}^n K_{ix,\mathbf{j}} f^{-1}(\bar{\varepsilon}_i) \frac{1}{nh_0^2} \sum_{j \neq i} \left[k_0'\left(\frac{\bar{\varepsilon}_i - \tilde{\varepsilon}_j}{h_0}\right) - k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right) \right] \bar{\varepsilon}_i^r \left[\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)\right]^{-\frac{2-r}{2}} G_i \\
\simeq\ & \frac{1}{2^r\sqrt{nh^d}} \bar{H}^{-1} \sum_{i=1}^n K_{ix,\mathbf{j}} f^{-1}(\bar{\varepsilon}_i) \frac{1}{nh_0^3} \sum_{j \neq i} k_0''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right) (\tilde{\varepsilon}_j - \varepsilon_j) \bar{\varepsilon}_i^r \left[\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)\right]^{-\frac{2-r}{2}} G_i \\
=\ & -\frac{1}{2^r\sqrt{nh^d}} \bar{H}^{-1} \sum_{i=1}^n K_{ix,\mathbf{j}} f^{-1}(\bar{\varepsilon}_i) \frac{1}{nh_0^3} \sum_{j \neq i} k_0''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right) \frac{\tilde{m}(X_j) - m(X_j)}{\sigma(X_j)} \bar{\varepsilon}_i^r \left[\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)\right]^{-\frac{2-r}{2}} G_i \\
& -\frac{1}{2^r\sqrt{nh^d}} \bar{H}^{-1} \sum_{i=1}^n K_{ix,\mathbf{j}} f^{-1}(\bar{\varepsilon}_i) \frac{1}{nh_0^3} \sum_{j \neq i} k_0''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right) \varepsilon_j \frac{\tilde{\sigma}(X_j) - \sigma(X_j)}{\sigma(X_j)} \bar{\varepsilon}_i^r \left[\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)\right]^{-\frac{2-r}{2}} G_i \\
& + o_p(1) \\
\equiv\ & -\mathcal{J}_{21a} - \mathcal{J}_{21b} + o_p(1).
\end{aligned}
$$

Decompose $\mathcal{J}_{21a}$ as

$$
\begin{aligned}
& \frac{1}{2^r\sqrt{nh^d}} \bar{H}^{-1} \sum_{i=1}^n K_{ix,\mathbf{j}} f^{-1}(\bar{\varepsilon}_i) \frac{1}{nh_0^3} \sum_{j \neq i} k_0''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right) \frac{e_1^\top M_n^{-1}(X_j) U_{1,n}(X_j)}{\sigma(X_j)} \bar{\varepsilon}_i^r \left[\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)\right]^{-\frac{2-r}{2}} G_i \\
& + \frac{1}{2^r\sqrt{nh^d}} \bar{H}^{-1} \sum_{i=1}^n K_{ix,\mathbf{j}} f^{-1}(\bar{\varepsilon}_i) \frac{1}{nh_0^3} \sum_{j \neq i} k_0''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right) \frac{e_1^\top M_n^{-1}(X_j) B_{1,n}(X_j)}{\sigma(X_j)} \bar{\varepsilon}_i^r \left[\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)\right]^{-\frac{2-r}{2}} G_i.
\end{aligned}
$$

The analysis of these two terms is similar to the analysis of $\mathcal{S}_{1n\mathbf{j},212}$ in the proof of Lemma B.4. In particular, the first term is $o_p(1)$ by the replacement of $M_n^{-1}(X_j)$ by $[Mf_X(X_j)]^{-1}$ and moment calculations and the second term is $O_p(\sqrt{nh^d} h_1^{p+1}) = o_p(1)$. Similarly, we can verify that $\mathcal{J}_{21b} = o_p(1)$. Next, decompose $\mathcal{J}_{22}$ as follows

$$
\begin{aligned}
& \frac{1}{2^r\sqrt{nh^d}} \bar{H}^{-1} \sum_{i=1}^n K_{ix,\mathbf{j}} f^{-1}(\bar{\varepsilon}_i) \left\{ \frac{1}{nh_0^2} \sum_{j \neq i} \left\{ k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right) - E_j\left[k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\right] \right\} \right\} \bar{\varepsilon}_i^r \left[\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)\right]^{-\frac{2-r}{2}} G_i \\
& + \frac{1}{2^r\sqrt{nh^d}} \bar{H}^{-1} \sum_{i=1}^n K_{ix,\mathbf{j}} f^{-1}(\bar{\varepsilon}_i) \left\{ \frac{1}{nh_0^2} \sum_{j \neq i} E_j\left[k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\right] - f'(\bar{\varepsilon}_i) \right\} \bar{\varepsilon}_i^r \left[\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)\right]^{-\frac{2-r}{2}} G_i.
\end{aligned}
$$

Analogous to the study of $\mathcal{S}_{2n\mathbf{j},12}$ in the proof of Lemma B.5, we can show that $\mathcal{J}_{22} = o_p(1)$.

For $\mathcal{J}_3$, a typical element of $\mathcal{J}_3$ is

$$
\begin{aligned}
& -\frac{1}{2^r\sqrt{nh^d}} \bar{H}^{-1} \sum_{i=1}^n K_{ix,\mathbf{j}} f^{-1}(\bar{\varepsilon}_i) \tilde{f}_i'(\bar{\varepsilon}_i) \left(\tilde{f}_i(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)\right) \bar{\varepsilon}_i^r \left[\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)\right]^{-\frac{2-r}{2}} G_b\left(\tilde{f}_i(\bar{\varepsilon}_i)\right) \\
\simeq\ & -\frac{1}{2^r\sqrt{nh^d}} \bar{H}^{-1} \sum_{i=1}^n K_{ix,\mathbf{j}} f^{-1}(\bar{\varepsilon}_i) \tilde{f}_i'(\bar{\varepsilon}_i) \left(\tilde{f}_i(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)\right) \bar{\varepsilon}_i^r \left[\exp\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)\right]^{-\frac{2-r}{2}} G_b(f_i),
\end{aligned}
$$

where $0 \le |\mathbf{j}| \le p$, and $r = 0, 1$. The rest of the proof is similar to that of $\mathcal{J}_2$ and thus omitted.

For $\mathcal{J}_4$, we apply Lemma A.2 and the remark after it:

$$
\mathcal{J}_4 = \sqrt{nh^d} b^{-2} O_p\left(\left(h_0^{p+1} + n^{-1/2} h_0^{-d/2} \log n + \upsilon_{1n}\right)^2\right) = o_p(1).
$$

$\mathcal{J}_5$ and $\mathcal{J}_6$ can be analyzed by similar techniques to the above. ■

Supplementary Material on
"Adaptive Nonparametric Regression with Conditional Heteroskedasticity"

Sainan Jin[a], Liangjun Su[a], Zhijie Xiao[b]
[a] *School of Economics, Singapore Management University*
[b] *Department of Economics, Boston College*

THIS APPENDIX PROVIDES PROOFS FOR TECHNICAL LEMMAS IN THE ABOVE PAPER.

C Proofs of the Technical Lemmas

To facilitate the proof, we define an $N \times N$ matrix $M_n(x)$ and $N \times 1$ vectors $\Psi_{s,n}(x)$ $(s = 1, 2)$ as:

$$M_n(x) \equiv \begin{bmatrix} M_{n,0,0}(x) & M_{n,0,1}(x) & \ldots & M_{n,0,p}(x) \\ M_{n,1,0}(x) & M_{n,1,1}(x) & \ldots & M_{n,1,p}(x) \\ \vdots & \vdots & \ddots & \vdots \\ M_{n,p,0}(x) & M_{n,p,1}(x) & \ldots & M_{t,p,p}(x) \end{bmatrix},\ \Psi_{s,n}(x) \equiv \begin{bmatrix} \Psi_{s,n,0}(x) \\ \Psi_{s,n,1}(x) \\ \vdots \\ \Psi_{s,n,p}(x) \end{bmatrix}, \quad \text{(C.1)}$$

where $M_{n,|j|,|k|}(x)$ is an $N_{|j|} \times N_{|k|}$ submatrix with the (l, r) element given by

$$\left[M_{n,|j|,|k|}(x)\right]_{l,r} \equiv \frac{1}{nh^d} \sum_{i=1}^{n} \left(\frac{X_i - x}{h_1}\right)^{\phi_{|j|}(l)+\phi_{|k|}(r)} K\left(\frac{X_i - x}{h_1}\right),$$

$\Psi_{1,n,|j|}(x)$ is an $N_{|j|} \times 1$ subvector whose r-th element is given by

$$\left[\Psi_{1,n,|j|}(x)\right]_r \equiv \frac{1}{nh^d} \sum_{i=1}^{n} \left(\frac{X_i - x}{h_1}\right)^{\phi_{|j|}(r)} K\left(\frac{X_i - x}{h_1}\right) Y_i,$$

and $\Psi_{2,n,|j|}(x)$ is an $N_{|j|} \times 1$ subvector whose r-th element is given by

$$\left[\Psi_{2,n,|j|}(x)\right]_r \equiv \frac{1}{nh^d} \sum_{i=1}^{n} \left(\frac{X_i - x}{h_1}\right)^{\phi_{|j|}(r)} K\left(\frac{X_i - x}{h_1}\right) u_i^2.$$

Define $\widetilde{\Psi}_{2,n}(x)$ analogously as $\Psi_{2,n}(x)$ with u_i^2 being replaced by $\tilde{u}_i^2$, where $\tilde{u}_i \equiv Y_i - \tilde{m}(X_i)$. The p-th order local polynomial estimates of $m(x)$ and $\sigma^2(x)$ are given respectively by

$$\tilde{m}(x) = e_1^\top M_n^{-1}(x) \Psi_{1,n}(x) \text{ and } \tilde{\sigma}^2(x) = e_1^\top M_n^{-1}(x) \widetilde{\Psi}_{2,n}(x).$$

For $s = 1, 2$, let

$$U_{s,n}(x) \equiv \begin{bmatrix} U_{s,n,0}(x) \\ U_{s,n,1}(x) \\ \vdots \\ U_{1,n,p}(x) \end{bmatrix},\ B_{s,n}(x) \equiv \begin{bmatrix} B_{s,n,0}(x) \\ B_{s,n,1}(x) \\ \vdots \\ B_{s,n,p}(x) \end{bmatrix},$$

where $U_{s,n,l}(x)$ and $B_{s,n,l}(x)$ are defined analogously as $\Psi_{s,n,l}(x)$ so that $U_{s,n,|j|}(x)$ and $B_{s,n,|j|}(x)$ are $N_{|j|} \times 1$ subvectors whose r-th elements are given by

$$\left[U_{s,n,|j|}(x)\right]_r = \frac{1}{nh_1^d} \sum_{i=1}^{n} \left(\frac{X_i - x}{h_1}\right)^{\phi_{|j|}(r)} K\left(\frac{X_i - x}{h_1}\right) u_{s,i},$$

$$\left[B_{s,n,|j|}(x)\right]_r = \frac{1}{nh_1^d} \sum_{i=1}^{n} \left(\frac{X_i - x}{h_1}\right)^{\phi_{|j|}(r)} K\left(\frac{X_i - x}{h_1}\right) \Delta_{s,i}(x),$$

where $u_{1,i} \equiv u_i$, $u_{2,i} \equiv u_i^2 - E(u_i^2|X_i) = \sigma^2(X_i)(\varepsilon_i^2 - 1)$, and $\Delta_{s,i}(x) \equiv \beta_s(X_i) - \sum_{0\le|\mathbf{j}|\le p}\beta_{s\mathbf{j}}(x) \times (X_i - x)^{\mathbf{j}}$. We further define $\tilde{U}_{2,n}(x)$ analogously as $U_{2,n}(x)$ but with $u_{2,i}$ being replaced by $\tilde{u}_{2,i} \equiv \tilde{u}_i^2 - E(u_i^2|X_i)$. Then

$$\begin{aligned} \tilde{m}(x) - m(x) &= e_1^\top M_n^{-1}(x)U_{1,n}(x) + e_1^\top M_n^{-1}(x)B_{1,n}(x), \text{ and} \\ \tilde{\sigma}^2(x) - \sigma^2(x) &= e_1^\top M_n^{-1}(x)\tilde{U}_{2,n}(x) + e_1^\top M_n^{-1}(x)B_{2,n}(x). \end{aligned} \tag{C.2}$$

By Masry 1996(a), we can readily show that

$$\tilde{m}(x) - m(x) = e_1^\top [f_X(x)M]^{-1}\frac{1}{n}\sum_{i=1}^n K_{h_1}(x - X_i)\mathbf{Z}_i u_i + h_1^{p+1}e_1^\top M^{-1}B\mathbf{m}^{(p+1)}(x) + o_p(h_1^{p+1}) \tag{C.3}$$

uniformly in x. Furthermore,

$$\sup_{x\in\mathcal{X}}|M_n(x) - f_X(x)M| = O_p(\upsilon_{0n}) \text{ and } \sup_{x\in\mathcal{X}}|\tilde{m}(x) - m(x)| = O_p(\upsilon_{1n}). \tag{C.4}$$

The following lemma studies the asymptotic property of the local polynomial estimator $\tilde{\sigma}^2(x)$ of $\sigma^2(x)$.

Lemma C.1 *Suppose Assumptions A1-A5 hold. Then* $\tilde{\sigma}^2(x) - \sigma^2(x) = e_1^\top M_n^{-1}(x)U_{2,n}(x) + e_1^\top M_n^{-1}(x) \times B_{2,n}(x) + O_p((\upsilon_{0n} + \upsilon_{1n})\upsilon_{1n})$ *uniformly in* x.

Proof of Lemma C.1. Let $K^*(X_i, x) \equiv e_1^\top M_n^{-1}(x)K((X_i - x)/h_1)\tilde{\mathbf{Z}}_i$. Then $\tilde{\sigma}^2(x) = (nh_1^d)^{-1}\sum_{i=1}^n K^*(X_i, x)\tilde{u}_i^2$. It follows from $M_n^{-1}(x)M_n(x) = I_N$ that

$$\frac{1}{nh_1^d}\sum_{i=1}^n K^*(X_i, x) = \frac{1}{nh_1^d}\sum_{i=1}^n e_1^\top M_n^{-1}(x)K((X_i - x)/h_1)\tilde{\mathbf{Z}}_i = 1,$$

and

$$\frac{1}{nh_1^d}\sum_{i=1}^n K^*(X_i, x)(X_i - x)^{\mathbf{j}} = \frac{1}{nh_1^d}\sum_{i=1}^n e_1^\top M_n^{-1}(x)K((X_i - x)/h_1)\tilde{\mathbf{Z}}_i(X_i - x)^{\mathbf{j}} = 0,$$

for $1 \le |\mathbf{j}| \le p$. Consequently,

$$\tilde{\sigma}^2(x) - \sigma^2(x) = e_1^\top M_n^{-1}(x)\widetilde{\Psi}_{2,n}(x) = \frac{1}{nh_1^d}\sum_{i=1}^n K^*(X_i, x)\left\{\tilde{u}_i^2 - \overline{\sigma}^2(x, X_i)\right\},$$

where $\overline{\sigma}^2(x, X_i) \equiv \sum_{0\le|\mathbf{j}|\le p}\left(D^{(\mathbf{j})}\sigma^2\right)(x)(X_i - x)^{\mathbf{j}}$. Noting that $\tilde{u}_i^2 = [Y_i - \tilde{m}(X_i)]^2 = [\sigma(X_i)\varepsilon_i + m(X_i) - \tilde{m}(X_i)]^2 = \sigma^2(X_i)\varepsilon_i^2 + 2\sigma(X_i)\varepsilon_i[m(X_i) - \tilde{m}(X_i)] + [m(X_i) - \tilde{m}(X_i)]^2$, we have

$$\begin{aligned} \tilde{\sigma}^2(x) - \sigma^2(x) &= \frac{1}{nh_1^d}\sum_{i=1}^n K^*(X_i, x)\left\{\sigma^2(X_i) - \overline{\sigma}^2(x, X_i)\right\} \\ &\quad + \frac{1}{nh_1^d}\sum_{i=1}^n K^*(X_i, x)\sigma^2(X_i)\left(\varepsilon_i^2 - 1\right) \\ &\quad + \frac{2}{nh_1^d}\sum_{i=1}^n K^*(X_i, x)\sigma(X_i)\varepsilon_i[m(X_i) - \tilde{m}(X_i)] \\ &\quad + \frac{1}{nh_1^d}\sum_{i=1}^n K^*(X_i, x)[m(X_i) - \tilde{m}(X_i)]^2 \\ &\equiv A_1(x) + A_2(x) + 2A_3(x) + A_4(x), \text{ say.} \end{aligned}$$

Noting that $\Delta_{2,i}(x)=\sigma^2(X_i)-\overline{\sigma}^2(x,X_i)$, we have $A_1(x)=e_1^\top M_n^{-1}(x)B_{2,n}(x)$. In addition $A_2(x)=e_1^\top M_n^{-1}(x)U_{2,n}(x)$ by the definition of $u_{2,i}$, and $\sup_{x\in\mathcal{X}}|A_4(x)|=\upsilon_{1n}^2$ by (C.4). For $A_3(x)$, write $-A_3(x)=A_{31}(x)+A_{32}(x)$, where

$$\begin{aligned}A_{31}(x) &\equiv \frac{1}{nh_1^d}\sum_{i=1}^n K^*(X_i,x)u_ie_1^\top M_n^{-1}(X_i)U_{1,n}(X_i), \text{ and}\\ A_{32}(x) &\equiv \frac{1}{nh_1^d}\sum_{i=1}^n K^*(X_i,x)u_ie_1^\top M_n^{-1}(X_i)B_{1,n}(X_i).\end{aligned}$$

Note that

$$\begin{aligned}A_{31}(x) &= \frac{1}{nh_1^d}\sum_{i=1}^n K^*(X_i,x)u_ie_1^\top[Mf_X(X_i)]^{-1}U_{1,n}(X_i)\\ &\quad -\frac{1}{nh_1^d}\sum_{i=1}^n K^*(X_i,x)u_ie_1^\top\left\{M_n(x)^{-1}-[Mf_X(X_i)]^{-1}\right\}U_{1,n}(X_i)\\ &\equiv A_{31,1}(x)-A_{31,2}(x), \text{ say.}\end{aligned}$$

We dispose $A_{31,2}(x)$ first. By (C.4), the facts that $\sup_{x\in\mathcal{X}}\|U_{1,n}(x)\|=O_p(n^{-1/2}h_1^{-d/2}\sqrt{\log n})$ and $\sup_{x\in\mathcal{X}}\frac{1}{nh_1^d}\sum_{i=1}^n|K^*(X_i,x)u_i|=O_p(1)$, we have

$$\begin{aligned}\sup_{x\in\mathcal{X}}|A_{31,2}(x)| &\le \sup_{x\in\mathcal{X}}\left\|M_n(x)^{-1}-[Mf_X(X_i)]^{-1}\right\|\sup_{x\in\mathcal{X}}\|U_{1,n}(x)\|\sup_{x\in\mathcal{X}}\frac{1}{nh_1^d}\sum_{i=1}^n|K^*(X_i,x)u_i|\\ &= O_p(\upsilon_{0n})O_p(n^{-1/2}h_1^{-d/2}\sqrt{\log n})O_p(1)=O_p(\upsilon_{0n}n^{-1/2}h_1^{-d/2}\sqrt{\log n}).\end{aligned}$$

Using $U_{1,n}(x)=\frac{1}{nh_1^d}\sum_{j=1}^n K((X_j-x)/h_1)\tilde{\mathbf{Z}}_ju_j$ and $K^*(X_i,x)=e_1^\top M_n^{-1}(x)K((X_i-x)/h_1)\tilde{\mathbf{Z}}_i$, we have

$$\begin{aligned}A_{31,1}(x) &= \frac{1}{n^2h_1^{2d}}e_1^\top M_n^{-1}(x)\sum_{i=1}^n\sum_{j=1}^n K((X_i-x)/h_1)\tilde{\mathbf{Z}}_ie_1^\top[Mf_X(X_i)]^{-1}K((X_j-X_j)/h_1)\tilde{\mathbf{Z}}_ju_iu_j\\ &= \frac{1}{n^2h_1^{2d}}e_1^\top M_n^{-1}(x)\sum_{1\le i\ne j\le n}K((X_i-x)/h_1)\tilde{\mathbf{Z}}_ie_1^\top[Mf_X(X_i)]^{-1}K((X_j-X_j)/h_1)\tilde{\mathbf{Z}}_ju_iu_j\\ &\quad +\frac{1}{n^2h_1^{2d}}e_1^\top M_n^{-1}(x)\sum_{i=1}^n K((X_i-x)/h_1)\tilde{\mathbf{Z}}_ie_1^\top[Mf_X(X_i)]^{-1}K(0)\tilde{\mathbf{Z}}_iu_i^2\\ &\equiv A_{31,1a}(x)+A_{31,1b}(x), \text{ say.}\end{aligned}$$

Let $\varsigma_{ij}(x)\equiv\{e_1^\top[Mf_X(x)]^{-1}K((X_i-x)/h_1)\tilde{\mathbf{Z}}_i\}\{e_1^\top[Mf_X(X_i)]^{-1}K((X_j-X_i)/h_1)\tilde{\mathbf{Z}}_j\}u_iu_j$. Then by (C.4), $A_{31,1a}(x)=[1+O_p(\upsilon_{0n})]\bar{A}_{31,1a}(x)$, where

$$\bar{A}_{31,1a}(x)=\frac{1}{n^2h_1^{2d}}\sum_{1\le i\ne j\le n}\varsigma_{ij}(x).$$

is a second order degenerate U-statistic. We can readily show that $\bar{A}_{31,1a}(x)=O_p\left(n^{-1}h_1^{-d}\right)$ for each x by Chebyshev inequality. By using Bickel's (1975) standard chaining argument, we can show $\sup_{x\in\mathcal{X}}\left|\bar{A}_{31,1a}(x)\right|=O_p\left(n^{-1}h_1^{-d}\log n\right)$. For $A_{31,1b}(x)$, we have

$$\begin{aligned}\sup_{x\in\mathcal{X}}|A_{31,1b}(x)| &\le \frac{1}{nh_1^d}\sup_{x\in\mathcal{X}}\left\|M_n^{-1}(x)\right\|\sup_{x\in\mathcal{X}}\left\|\frac{1}{nh_1^d}\sum_{i=1}^n K((X_i-x)/h_1)\tilde{\mathbf{Z}}_ie_1^\top[Mf_X(X_i)]^{-1}K(0)\tilde{\mathbf{Z}}_iu_i^2\right\|\\ &= O_p\left(n^{-1}h_1^{-d}\right)O_p(1)O_p(1)=O_p\left(n^{-1}h_1^{-d}\right).\end{aligned}$$

It follows that $\sup_{x\in\mathcal{X}}|A_{31,1}(x)| = O_p\left(n^{-1}h_1^{-d}\log n\right)$. Consequently, we have shown that $\sup_{x\in\mathcal{X}}|A_{31}(x)| = O_p(\upsilon_{0n}n^{-1/2}h_1^{-d/2}\sqrt{\log n})$.

Note that

$$\begin{aligned} A_{32}(x) &= \frac{1}{nh_1^d}\sum_{i=1}^{n}K^*(X_i,x)\,u_i e_1^\top\left[Mf_X(X_i)\right]^{-1}B_{1,n}(X_i) \\ &\quad -\frac{1}{nh_1^d}\sum_{i=1}^{n}K^*(X_i,x)\,u_i e_1^\top\left\{M_n(x)^{-1}-\left[Mf_X(X_i)\right]^{-1}\right\}B_{1,n}(X_i) \\ &\equiv A_{32,1}(x)-A_{32,2}(x)\,,\ \text{say.} \end{aligned}$$

As in the study of $A_{31}(x)$, using (C.4) and the fact that $\sup_{x\in\mathcal{X}}|B_{1,n}(x)| = O_p(h_1^{p+1})$ we can readily show that $\sup_{x\in\mathcal{X}}|A_{32,2}(x)| = O_p(\upsilon_{0n}h_1^{p+1})$ and that $\sup_{x\in\mathcal{X}}|A_{32,2}(x)| = O_p(n^{-1/2}h_1^{-d/2}\sqrt{\log n}h_1^{p+1})$. Hence $\sup_{x\in\mathcal{X}}|A_{32}(x)| = O_p(\upsilon_{0n}h_1^{p+1})$. Consequently, $\sup_{x\in\mathcal{X}}|A_3(x)| = O_p(\upsilon_{0n}\upsilon_{1n})$. This completes the proof. ■

Remark C1. Using the notation defined in the proof of Lemma C.1, we can also show that $A_1(x) = h_1^{p+1}e_1^\top M^{-1}B\boldsymbol{\sigma}^{2(p+1)}(x)+o_p(h_1^{p+1})$, and $\sqrt{nh_1^d}A_2(x)\xrightarrow{d}N(0,\left(\sigma^4(x)/f_X(x)\right)E\left(\varepsilon_1^2-1\right)^2e_1^\top M^{-1}\Gamma M^{-1}e_1^\top)$. By standard results on local polynomial estimators, Lemma A.1 implies

$$\sup_{x\in\mathcal{X}}\left|\tilde{\sigma}^2(x)-\sigma^2(x)\right| = O_p(\upsilon_{1n})\,, \tag{C.5}$$

where υ_{1n} is the rate we can obtain even if the conditional mean function $m(x)$ is known.

Let δ_i and $v_{ri}(x)$ be as defined in Appendix A. To prove Lemmas A.1-A.2, we will frequently use the facts that

$$\delta_i = O_p\left(h^{p+1}\right)\ \text{uniformly on the set }\{K_{ix}>0\}\,, \tag{C.6}$$

$$v_{ri}(x) = O_p\left((h_1^{p+1}+n^{-1/2}h_1^{-d/2})(1+(h/h_1)^p)\right)\ \text{on the set }\{K_{ix}>0\}\,,\ r=1,2, \tag{C.7}$$

$$\max_{\{K_{ix}>0\}}|v_{ri}(x)| = O_p(\upsilon_{2n})\,,\ r=1,2. \tag{C.8}$$

To facilitate the asymptotic analysis, we also define the kernel density and derivative estimator based on the unobserved errors $\{\varepsilon_j\}$:

$$\overline{f}_i(e_i) = \frac{1}{nh_0}\sum_{j\neq i}k_0\left(\frac{e_i-\varepsilon_j}{h_0}\right),\ \text{and}\ \overline{f}_i^{(s)}(e_i) = \frac{1}{nh_0^{1+s}}\sum_{j\neq i}k_0^{(s)}\left(\frac{e_i-\varepsilon_j}{h_0}\right)\ \text{for } s=1,2,3.$$

We will need the result in the following lemma which is adopted from Hansen (2008).

Lemma C.2 *Let ε_i, $i=1,...,n$, be IID. Assume that (i) the PDF of ε_i, $f(\cdot)$, is uniformly bounded, and the $(p+1)$th derivative of $f^{(s)}(\varepsilon)$ is uniformly continuous; (ii) there exists $q>0$ such that $\sup_\varepsilon|\varepsilon|^q f(\varepsilon)<\infty$ and $|k_0^{(s)}(e)|\leq C|e|^{-q}$ for $|e|$ large; (iii) $k_0(\cdot)$ is a (p+1)th order kernel and $\int|e|^{p+s+1}|k_0(e)|de<\infty$; (iv) $h_0\to 0$ and $nh_0^{1+2s}/\log n\to\infty$ as $n\to\infty$. Then*

$$\max_{1\leq i\leq n}\left|\overline{f}_i^{(s)}(\bar{\varepsilon}_i)-f^{(s)}(\bar{\varepsilon}_i)\right| = O_p(h_0^{p+1}+n^{-1/2}h_0^{-1/2-s}\sqrt{\log n}).$$

Proof of Lemma C.2. The above result is essentially a special case of Theorem 6 in Hansen (2008) who allows for strong mixing processes. For an IID sequence, the parameters β and θ in Hansen (2008) correspond to ∞ and one, respectively. Another noticeable difference is that Hansen considers the usual kernel estimates whereas we consider the leave-one-out kernel estimates here. The difference between these two kernel estimates is uniformly $(nh_0^{1+s})^{-1}k_0^{(s)}(0)$, which is $o(n^{-1/2}h_0^{-1/2-s}\sqrt{\log n})$ under condition (iv) and thus does not contribute to the uniform convergence rate of $\overline{f}_i^{(s)}(\bar{\varepsilon}_i) - f^{(s)}(\bar{\varepsilon}_i)$ to 0. ■

Proof of Lemma A.1. We only prove the lemma with $s = 0$ as the other cases can be treated analogously. Write $\tilde{f}_i(\overrightarrow{\varepsilon}_i) - f(\bar{\varepsilon}_i) = [\overline{f}(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)] + [\tilde{f}_i(\overrightarrow{\varepsilon}_i) - \overline{f}(\bar{\varepsilon}_i)]$. Noting that k_0 is a $(p+1)$-th order kernel with compact support by Assumption A6, the conditions on the kernel in Lemma C.2 are satisfied. One can readily check that the other conditions in that lemma are also satisfied under Assumptions A1, A2, and A7. So we can apply Lemma C.2 to obtain $\max_{1\leq i\leq n}\left|\overline{f}_i(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)\right| = O_p(h_0^{p+1} + n^{-1/2}h_0^{-1/2}\sqrt{\log n})$. Let

$$\begin{aligned} r_{1ij} \equiv\; & \frac{\bar{\varepsilon}_i\left[\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{1/2} - \varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2}\right]}{\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2}} - \frac{v_{1i}(x)}{\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2}} + \frac{\tilde{m}(X_j) - m(X_j)}{\sigma(X_j)} \\ & + \left[\varepsilon_j + \frac{m(X_j) - \tilde{m}(X_j)}{\sigma(X_j)}\right]\frac{\tilde{\sigma}(X_j) - \sigma(X_j)}{\tilde{\sigma}(X_j)}. \end{aligned} \tag{C.9}$$

Then

$$\overrightarrow{\varepsilon}_i - \tilde{\varepsilon}_j = (\bar{\varepsilon}_i - \varepsilon_j) + r_{1ij}. \tag{C.10}$$

By a first order Taylor expansion with an integral remainder, we have

$$\begin{aligned} \tilde{f}_i(\overrightarrow{\varepsilon}_i) - \overline{f}(\bar{\varepsilon}_i) &= \frac{1}{nh_0}\sum_{j\neq i}\left[k_0\left(\frac{\overrightarrow{\varepsilon}_i - \tilde{\varepsilon}_j}{h_0}\right) - k_0\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\right] \\ &= \frac{-1}{nh_0^2}\sum_{j\neq i}k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\frac{v_{1i}(x)}{\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2}} \\ &\quad + \frac{1}{nh_0^2}\sum_{j\neq i}k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\bar{\varepsilon}_i\left[\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{1/2} - \varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2}\right]\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{-1/2} \\ &\quad + \frac{1}{nh_0^2}\sum_{j\neq i}k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\frac{\tilde{m}(X_j) - m(X_j)}{\sigma(X_j)} \\ &\quad + \frac{1}{nh_0^2}\sum_{j\neq i}k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\left[\varepsilon_j + \frac{m(X_j) - \tilde{m}(X_j)}{\sigma(X_j)}\right]\frac{\tilde{\sigma}(X_j) - \sigma(X_j)}{\tilde{\sigma}(X_j)} \\ &\quad + \frac{1}{nh_0^2}\sum_{j\neq i}\int_0^1\left[k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j + wr_{1ij}}{h_0}\right) - k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\right]dwr_{1ij} \\ &\equiv -B_{1i}(x) + B_{2i}(x) + B_{3i}(x) + B_{4i}(x) + B_{5i}(x), \text{ say.} \end{aligned} \tag{C.11}$$

We will establish the uniform probability order for $B_{ji}(x)$, $j = 1, 2, ..., 5$, in order.

For $B_{1i}(x)$, we apply Lemma C.2 to obtain that, uniformly in i,

$$\frac{1}{nh_0^2}\sum_{j\neq i}k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right) = f'(\bar{\varepsilon}_i) + O_p\left(n^{-1/2}h_0^{-3/2}\sqrt{\log n} + h_0^{p+1}\right). \tag{C.12}$$

Then by (C.8) and the uniform boundedness of $f'(\varepsilon)$, we have

$$\max_{\{K_{ix}>0\}}|B_{1i}(x)| = O_p(\upsilon_{2n}). \tag{C.13}$$

Similarly, by (C.12), (C.8), and the uniform boundedness of $f'(\varepsilon)\varepsilon$, we have

$$\max_{\{K_{ix}>0\}}|B_{2i}(x)| = O_p(\upsilon_{2n}). \tag{C.14}$$

Expanding $M_n^{-1}(x)$ around its probability limit $[Mf_X(x)]^{-1}$, we have

$$\begin{aligned} B_{3i}(x) &= \frac{1}{nh_0^2}\sum_{j\neq i} k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\sigma^{-1}(X_j)\, e_1^\top [Mf_X(X_j)]^{-1} U_{1,n}(X_j) \\ &\quad -\frac{1}{nh_0^2}\sum_{j\neq i} k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\sigma^{-1}(X_j)\, e_1^\top \alpha_n(X_j) M_n^{-1}(X_j) U_{1,n}(X_j) \\ &\quad +\frac{1}{nh_0^2}\sum_{j\neq i} k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\sigma^{-1}(X_j)\, e_1^\top [Mf_X(X_j)]^{-1} B_{1,n}(X_j) \\ &\quad -\frac{1}{nh_0^2}\sum_{j\neq i} k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\sigma^{-1}(X_j)\, e_1^\top \alpha_n(X_j) M_n^{-1}(X_j) B_{1,n}(X_j) \\ &\equiv B_{31i}(x) - B_{32i}(x) + B_{33i}(x) - B_{34i}(x), \end{aligned}$$

where $\alpha_n(x) \equiv [Mf_X(x)]^{-1}[M_n(x) - Mf_X(x)]$. Write

$$\begin{aligned} B_{31i}(x) &= \frac{1}{nh_0^2}\sum_{j\neq i} k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\sigma^{-1}(X_j)\, e_1^\top [Mf_X(X_j)]^{-1} U_{1,n}(X_j) \\ &= \frac{1}{nh_0^2}\sum_{j\neq i} E_j\left[k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\right]\sigma^{-1}(X_j)\, e_1^\top [Mf_X(X_j)]^{-1} U_{1,n}(X_j) \\ &\quad +\frac{1}{nh_0^2}\sum_{j\neq i}\left\{k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right) - E_j\left[k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\right]\right\}\sigma^{-1}(X_j)\, e_1^\top [Mf_X(X_j)]^{-1} U_{1,n}(X_j) \\ &\equiv B_{31i,1}(x) + B_{31i,2}(x), \text{ say.} \end{aligned}$$

For $B_{31i,1}(x)$, we have

$$\begin{aligned} \max_{1\le i\le n}|B_{31i,1}(x)| &\le \max_{1\le i\le n}\left|\frac{n-1}{nh_0^2}E_j\left[k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\right]\right| \times \sup_{x\in\mathcal{X}}\left\|\sigma^{-1}(x)\, e_1^\top [Mf_X(x)]^{-1}\right\| \sup_{x\in\mathcal{X}}\|U_{1,n}(x)\| \\ &= O_p(1)\, O_p(1)\, O_p\left(n^{-1/2}h_1^{-d/2}\sqrt{\log n}\right) = O_p\left(n^{-1/2}h_1^{-d/2}\sqrt{\log n}\right), \end{aligned}$$

where we use the facts that $\sup_{x\in\mathcal{X}}\|U_{1,n}(x)\| = O_p(n^{-1/2}h_1^{-d/2}\sqrt{\log n})$ by Masry (1996a), $\max_{1\le i\le n}|h_0^{-2} \times E_j[k_0'((\bar{\varepsilon}_i - \varepsilon_j)/h_0)] - f'(\bar{\varepsilon}_i)| = O(h_0^{p+1})$ by standard bias calculation for kernel estimates and $\max_{1\le i\le n} |f'(\bar{\varepsilon}_i)| \le \sup_\varepsilon |f'(\varepsilon)| \le C < \infty$.

Let $v_j(\bar{\varepsilon}_i) = k_0'((\bar{\varepsilon}_i - \varepsilon_j)/h_0) - E[k_0'((\bar{\varepsilon}_i - \varepsilon_j)/h_0)]$. Then

$$\begin{aligned} B_{31i,2}(x) &= \frac{1}{n^2h_1^dh_0^2}\sum_{j\neq i}\sum_{l} v_j(\bar{\varepsilon}_i)\,\sigma^{-1}(X_j)\, e_1^\top [Mf_X(X_j)]^{-1}\tilde{\mathbf{Z}}_l K\left(\frac{X_l-X_j}{h_1}\right)u_l \\ &= \frac{1}{n^2h_1^dh_0^2}\sum_{j\neq i}\sum_{l\neq j,i} v_j(\bar{\varepsilon}_i)\,\sigma^{-1}(X_j)\, e_1^\top [Mf_X(X_j)]^{-1}\tilde{\mathbf{Z}}_l K\left(\frac{X_l-X_j}{h_1}\right)u_l \\ &\quad +\frac{1}{n^2h_1^dh_0^2}\sum_{j\neq i} v_j(\bar{\varepsilon}_i)\,\sigma^{-1}(X_j)\, e_1^\top [Mf_X(X_j)]^{-1}\tilde{\mathbf{Z}}_j K(0)\, u_j \\ &\quad +\frac{1}{n^2h_1^dh_0^2}\sum_{j\neq i} v_j(\bar{\varepsilon}_i)\,\sigma^{-1}(X_j)\, e_1^\top [Mf_X(X_j)]^{-1}\tilde{\mathbf{Z}}_i K\left(\frac{X_i-X_j}{h_1}\right)u_i \\ &\equiv B_{31i,2a}(x) + B_{31i,2b}(x) + B_{31i,2c}(x), \text{ say.} \end{aligned}$$

By construction, $B_{31i,2a}(x)$ is a second order degenerate U-statistic (see, e.g., Lee (1990)) and we can bound it by straightforward moment calculations. Let $\epsilon_n \equiv Cn^{-1/2}h_1^{-d/2}\sqrt{\log n}$ for some $C > 0$. By the

Boole and Markov inequalities,

$$P\left(\max_{1\le i\le n}|B_{31i,2a}(x)|\ge\epsilon_n\right)\le\sum_{i=1}^{n}P\left(|B_{31i,2a}(x)|\ge\epsilon_n\right)\le\sum_{i=1}^{n}\frac{E\left[|B_{31i,2a}(x)|^4\right]}{\epsilon_n^4}.$$

Let $a_{lj}=e_1^{\top}\left[Mf_X(X_j)\right]^{-1}\tilde{\mathbf{Z}}_lK((X_l-X_j)/h_1)$. Note that

$$\begin{aligned}E\left[|B_{31i,2}(x)|^4\right] &= \frac{1}{\left(n^2h_1^dh_0^2\right)^4}\sum_{j_s\ne l_s\ne i\text{ for }s=1,2,3,4}\\&\quad\times E\left\{a_{l_1j_1}a_{l_2j_2}a_{l_3j_3}a_{l_4j_4}v_{j_1}(\bar{\varepsilon}_i)u_{l_1}v_{j_2}(\bar{\varepsilon}_i)u_{l_2}v_{j_3}(\bar{\varepsilon}_i)u_{l_3}v_{j_4}(\bar{\varepsilon}_i)u_{l_4}\right\},\end{aligned}$$

where the summations are only taken with respect to j and l's. Consider the index set $S\equiv\{j_s,l_s,s=1,2,3,4\}$. If the number of distinct elements in S is larger than 4, then the expectation in the last expression is zero by the IID condition in Assumption A1. We can readily show that $E\left[|B_{31i,2}(x)|^4\right]=O(n^{-4}h_1^{-2d}h_0^{-6})$. It follows that

$$\begin{aligned}P\left(\max_{1\le i\le n}|B_{31i,2a}(x)|\ge\epsilon_n\alpha_{n,0}\right) &\le \frac{nO(n^{-4}h_1^{-2d}h_0^{-6})}{Cn^{-2}h_1^{-2d}(\log n)^2\alpha_{n,0}^4}=\frac{O\left(n^{-1}h_0^{-6}(\log n)^{-2}\right)}{\alpha_n^4}\\&= O\left(n^{-1}h^{-(2p+1)-d}\right)=O(1).\end{aligned}$$

where recall $\alpha_{n,s}=h^{[(2p+d)/4-(s+1)]}(\log n)^{s+1}$. Then $\max_{1\le i\le n}|B_{31i,2a}(x)|=O_p(\alpha_{n,0}n^{-1/2}h_1^{-d/2}\sqrt{\log n})$ by the Markov inequality. Analogously, we can show that $\max_{1\le i\le n}|B_{31i,2c}(x)|=o(n^{-1/2}h_1^{-d/2}\sqrt{\log n})$. For $B_{31i,2b}(x)$, we continue to decompose it as follows

$$\begin{aligned}B_{31i,2b}(x) &= \frac{K(0)}{n^2h_1^dh_0^2}\sum_{j\ne i}\sigma^{-1}(X_j)e_1^{\top}\left[Mf_X(X_j)\right]^{-1}\tilde{\mathbf{Z}}_j\left\{v_j(\bar{\varepsilon}_i)u_j-E_j\left[v_j(\bar{\varepsilon}_i)u_j\right]\right\}\\&\quad+\frac{K(0)}{n^2h_1^dh_0^2}\sum_{j\ne i}\sigma^{-1}(X_j)e_1^{\top}\left[Mf_X(X_j)\right]^{-1}\tilde{\mathbf{Z}}_jE_j\left[v_j(\bar{\varepsilon}_i)u_j\right]\\&\equiv B_{31i,2b1}(x)+B_{31i,2b2}(x),\end{aligned}$$

where E_j denotes expectation with respect to the variable indexed by j. We bound the second term first:

$$\begin{aligned}\max_{1\le i\le n}|B_{31i,2b2}(x)| &\le \max_{1\le i\le n}\left|h_0^{-1}E_j\left[v_j(\bar{\varepsilon}_i)u_j\right]\right|\frac{K(0)}{n^2h_1^dh_0}\sum_{j=1}^{n}\sigma^{-1}(X_j)\left|e_1^{\top}\left[Mf_X(X_j)\right]^{-1}\tilde{\mathbf{Z}}_j\right|\\&= O_p(1)O(n^{-1}h_1^{-d}h_0^{-1})=O_p(n^{-1}h_1^{-d}h_0^{-1}).\end{aligned}$$

By the Boole and Markov inequalities,

$$\begin{aligned}P\left(\max_{1\le i\le n}|B_{31i,2b1}(x)|\ge\epsilon_n\right) &\le \sum_{i=1}^{n}\frac{E\left[|B_{31i,2b1}(x)|^4\right]}{\epsilon_n^4}=\frac{nO(n^{-6}h_1^{-4d}h_0^{-6})}{Cn^{-2}h_1^{-2d}(\log n)^2}\\&= O(n^{-3}h_1^{-2d}h_0^{-6}(\log n)^{-2})=o(1),\end{aligned}$$

implying that $\max_{1\le i\le n}|B_{31i,2b1}(x)|=o_p(n^{-1/2}h_1^{-d/2}\sqrt{\log n})$. Hence $\max_{1\le i\le n}|B_{31i,2}(x)|=O_p(n^{-1}h_1^{-d}h_0^{-1})$ $+o_p(n^{-1/2}h_1^{-d/2}\sqrt{\log n})$. Consequently, we have shown that

$$\max_{1\le i\le n}|B_{31i}(x)|=O_p(n^{-1}h_1^{-d}h_0^{-1})+(\alpha_{n,0}+o(1))O_p(n^{-1/2}h_1^{-d/2}\sqrt{\log n}).$$

By (C.4), the fact that $\sup_{x\in\mathcal{X}} \|U_{1,n}(x)\| = O_p(n^{-1/2}h_1^{-d/2}\sqrt{\log n})$, and the fact that $\max_{1\le i\le n} \frac{1}{nh_0^2}\sum_{j\ne i}$ $|k_0'((\bar{\varepsilon}_i-\varepsilon_j)/h_0)| = O(h_0^{-1})$, we can readily show that $\max_{1\le i\le n}|B_{32i}(x)| = O_p(\upsilon_{0n}n^{-1/2}h_1^{-d/2}\sqrt{\log n}h_0^{-1})$. For the other terms, we have $\max_{1\le i\le n}|B_{33i}(x)| = O_p(h_1^{p+1})$, and $\max_{1\le i\le n}|B_{34i}(x)| = O_p(h_1^{p+1})O_p(\upsilon_{0n})O_p(h_0^{-1}) = O_p(\upsilon_{0n}h_1^{p+1}h_0^{-1})$. Consequently,

$$\max_{1\le i\le n}|B_{3i}(x)| = O_p\left(n^{-1}h_1^{-d}h_0^{-1} + \upsilon_{1n} + \upsilon_{0n}\upsilon_{1n}h_0^{-1} + \alpha_{n,0}n^{-1/2}h_1^{-d/2}\sqrt{\log n}\right). \tag{C.15}$$

Now write

$$\begin{aligned} B_{4i}(x) &= \frac{1}{nh_0^2}\sum_{j\ne i}k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\varepsilon_j\frac{\tilde{\sigma}(X_j)-\sigma(X_j)}{\tilde{\sigma}(X_j)} \\ &\quad + \frac{1}{nh_0^2}\sum_{j\ne i}k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\frac{m(X_j)-\tilde{m}(X_j)}{\sigma(X_j)}\frac{\tilde{\sigma}(X_j)-\sigma(X_j)}{\tilde{\sigma}(X_j)} \\ &\equiv B_{41i}(x) + B_{42i}(x). \end{aligned}$$

By (C.4) and Lemma C.1, it is easy to show that $\max_{1\le i\le n}|B_{42i}(x)| = O_p\left(\upsilon_{1n}^2h_0^{-1}\right)$. Using analogous arguments as used in the analysis of $B_{3i}(x)$ and Lemma C.1, we can show that $\max_{1\le i\le n}|B_{41i}(x)|$ $= O_p(n^{-1}h_1^{-d}h_0^{-1} + \upsilon_{0n}\upsilon_{1n}h_0^{-1} + h_1^{p+1})$. Consequently,

$$\max_{1\le i\le n}|B_{4i}(x)| = O_p(n^{-1}h_1^{-d}h_0^{-1} + \upsilon_{0n}\upsilon_{1n}h_0^{-1} + h_1^{p+1}). \tag{C.16}$$

where we use the fact that $\upsilon_{1n}^2h_0^{-1} = o_p(n^{-1}h_1^{-d}h_0^{-1} + h_1^{p+1})$. As argued by Hansen (2008, pp.740-741), under Assumption A6 there exists an integral function k_0^* such that

$$\left|k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j+wr_{1ij}}{h_0}\right) - k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\right| \le wh_0^{-1}k_0^*\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)|r_{1ij}|.$$

It follows that

$$\begin{aligned} \max_{1\le i\le n}|B_{5i}(x)| &\le \frac{1}{nh_0^3}\sum_{j\ne i}k_0^*\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)r_{1ij}^2 = \frac{O_p\left(\upsilon_{2n}^2\right)}{nh_0^3}\sum_{j\ne i}k_0^*\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\left(\bar{\varepsilon}_i^2+\varepsilon_j^2\right) \\ &= O_p\left(\upsilon_{2n}^2h_0^{-2}\right). \end{aligned} \tag{C.17}$$

Combining (C.11), (C.13), (C.14), (C.15), (C.16), and (C.17) and using the facts that $n^{-1}h_1^{-d}h_0^{-1} = o\left(\upsilon_{1n}^2h_0^{-2}\right)$ and that $h_1^{p+1} = o(\upsilon_{2n})$ yield the desired result for $s=0$.

When $s>0$, we can decompose $\tilde{f}_i^{(s)}(\overrightarrow{\varepsilon}_i) - \overline{f}^{(s)}(\bar{\varepsilon}_i)$ as in (C.11) with the corresponding terms denoted as $B_{ri}^{(s)}(x)$ for $r=1,2,...,5$. The probability orders of $B_{1i}^{(s)}(x)$ and $B_{2i}^{(s)}(x)$ are the same as those of $B_{1i}(x)$ and $B_{2i}(x)$, those of $B_{3i}^{(s)}(x)$ and $B_{4i}^{(s)}(x)$ become $O_p(n^{-1}h_1^{-d}h_0^{-1-s} + (\upsilon_{0n}h_0^{-1-s} + \alpha_{n,s})n^{-1/2}h_1^{-d/2}\sqrt{\log n} + h_1^{p+1})$, and the probability order of $B_{5i}^{(s)}(x)$ is $O_p(\upsilon_{2n}^2h_0^{-2-s})$. Consequently, $\max_{1\le i\le n}|\tilde{f}_i^{(s)}(\overrightarrow{\varepsilon}_i) - \overline{f}^{(s)}(\bar{\varepsilon}_i)| = O_p(\upsilon_{2n} + (\upsilon_{0n}h_0^{-1-s} + \alpha_{n,s})n^{-1/2}h_1^{-d/2}\sqrt{\log n} + \upsilon_{2n}^2h_0^{-2-s})$. ■

Proof of Lemma A.2. The proof is similar to but much simpler than that of Lemma A.1 and thus omitted. ■

Proof of Lemma A.3. The proof is analogous to that of Lemma USSLN in Gozalo and Linton (2000) and thus we only sketch the proof for the $r=1$ case. Let $\mathcal{C}_n = \{q_{1n}(\cdot,\theta) : \theta\in\Theta\}$. Under the permissibility and envelope integrability of $\mathcal{C}_n$, the almost sure convergence of $\sup_{\theta\in\Theta}\left|h^{-d}\left[P_nq_{n,1}(Z,\theta) - Pq_{n,1}(Z_i,\theta)\right]\right|$ is equivalent to its convergence in probability. By the boundedness of Θ and measurability of the $q_{n,1}$, the class $\mathcal{C}_n$ is permissible in the sense of Pollard (1984, p196). We now show the envelope integrability of $\mathcal{C}_n$.

By Assumption A1 and the compactness of K, $|\log\left(f\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)\right)| \leq D(Y_i)$ on the set $K_{ix} > 0$. Consequently, we can take the dominance function $\overline{q}_n = D\left(Y\right) K\left(\left(x - X\right)/h\right)$. Let $E\left[D\left(Y\right)|X\right] = \bar{D}\left(X\right)$. Assumptions A1 and A3 ensure that

$$P\overline{q}_n = E\left[\bar{D}\left(X\right) K\left(\left(x - X\right)/h\right)\right] = h^d \int \bar{D}\left(x - hu\right) f\left(x - hu\right) K\left(u\right) du = O\left(h^d\right).$$

The envelope integrability allows us to truncate the functions to a finite range. Let $\alpha_n > 1$ be a sequence of constants such that $\alpha_n \to \infty$ as $n \to \infty$. Define

$$\mathcal{C}^*_{\alpha_n} = \left\{q^*_{\alpha_n} = \alpha_n^{-1} q_{n,1} \mathbf{1}\left\{\overline{q}_n \leq \alpha_n\right\} : q_n \in \mathcal{C}_n\right\}.$$

Let b_n be a non-increasing sequence of positive numbers for which $nh^d b_n^2 \gg \log n$. By analysis similar to that of Gozalo and Linton (2000) and Theorem II.37 of Pollard (1984, p.34), to show that $\sup_{\mathcal{C}_n} |P_n q_{n,1} - P q_{n,1}| = o_p\left(h^d b_n\right)$, it suffices to show

$$\sup_{\mathcal{C}^*_{\alpha_n}} \left|P_n q^*_{\alpha_n} - P q^*_{\alpha_n}\right| = o_p\left(h^d b_n\right), \tag{C.18}$$

which holds provided

$$\sup_{\mathcal{C}^*_{\alpha_n}} \left\{P\left[q^*_{\alpha_n}\right]^2\right\}^{1/2} < h^{d/2} \tag{C.19}$$

and

$$\sup N_1\left(\epsilon, G, \mathcal{C}^*_{\alpha_n}\right) \leq C_1 \epsilon^{-C_2} \text{ for } 0 < \epsilon \leq 1, \tag{C.20}$$

where $N_1\left(\epsilon, G, \mathcal{C}^*_{\alpha_n}\right)$ is the covering number of $\mathcal{C}^*_{\alpha_n}$, i.e., the smallest value J for which there exists functions $g_1, ..., g_J$ such that $\min_{j \leq J} G\left|q - g_j\right| \leq \epsilon$ for each $q \in \mathcal{C}^*_{\alpha_n}$, the supremum is taken over all probability measures G, and C_1 and C_2 are positive constants independent of n.

(C.19) holds by construction. For (C.20), we need to show that $\mathcal{C}^*_{\alpha_n}$ is a Euclidean class (Nolan and Pollard, 1987, p.789). Since the functions in $\mathcal{C}^*_{\alpha_n}$, $q^*_{\alpha_n} = \alpha_n^{-1} \log\left(f\left(\varepsilon\left(\boldsymbol{\beta}\right)\right)\right) G_b\left(f\left(\varepsilon\left(\boldsymbol{\beta}\right)\right)\right) K\left(\left(x - X\right)/h\right)$ $\mathbf{1}\left\{\overline{q}_n \leq \alpha_n\right\}$, are composed from the classes of functions

$$\begin{aligned}
\mathcal{C}_1 &= \left\{c_1 \log f\left(\frac{y - P\left(\boldsymbol{\beta}_1\right)}{\sqrt{\exp\left(P\left(\boldsymbol{\beta}_2\right)\right)}}\right) : \left(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top\right)^\top \in \mathcal{B},\ c_1 \leq 1\right\}, \\
\mathcal{C}_2 &= \left\{c_2 G_b\left(f\left(\frac{y - P\left(\boldsymbol{\beta}_1\right)}{\sqrt{\exp\left(P\left(\boldsymbol{\beta}_2\right)\right)}}\right)\right) : \left(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top\right)^\top \in \mathcal{B},\ c_2 \leq 1\right\}, \\
\mathcal{C}_3 &= \left\{K\left(x^\top c_3 + c_4\right) : c_3 \in \mathbb{R}^d,\ c_4 \in \mathbb{R}\right\}, \text{ and } \mathcal{C}_4 = \left\{\mathbf{1}\left\{c_5 \overline{q}_n \leq 1\right\} : c_5 \in \mathbb{R}\right\},
\end{aligned}$$

it suffices to show that the $\mathcal{C}_j's$ form Euclidean classes by Nolan and Pollard (1987, pp. 796-797) and Pakes and Pollard (1989, Lemmas 2.14 and 2.15).

First, for $j = 1, 2$, $\{P(\boldsymbol{\beta}_j)\}$ forms a polynomial class of functions and is Euclidean by Lemma 2.12 of Pakes and Pollard (1989). By Example 2.10 of Pakes and Pollard (1989) and the bounded variation assumption on f, the class $\{f\left(\frac{y-m}{s}\right) : m \in \mathbb{R},\ s > 0\}$ is Euclidean for the constant envelope $\sup_\varepsilon |f\left(\varepsilon\right)|$. It follows from Pakes and Pollard (1989, Lemmas 2.15) that $\mathcal{C}_1$ is also Euclidean. Similarly, $\mathcal{C}_2$ is Euclidean. By Nolan and Pollard (1987, Lemma 22) and the bounded variation of K, $\mathcal{C}_3$ forms a Euclidean class with constant envelope $\sup_x |K\left(x\right)|$. Finally, by Pollard (1984, Lemma II.25) and the Euclidean property of $\mathcal{C}_j$, $j = 1, 2, 3$, $\mathcal{C}_4$ is Euclidean. Consequently

$$\sup_\theta \left|\frac{1}{nh^d} \sum_{i=1}^n q_{1n}\left(Z_i, \theta\right) - E q_{1n}\left(Z_i, \theta\right)\right| = o_{a.s.}\left(b_n\right).$$

Since Pollard's Theorem requires that $b_n \gg n^{-1/2}h^{-d/2}\sqrt{\log n}$, we can take $b_n = n^{-1/2}h^{-d/2}\sqrt{\log n}$ to obtain the desired result. ■

Proof of Lemma A.4. The proof is analogous to that of Newey (1991, Corollary 3.2). We first show $\bar{P}_{n,1}(\theta)$ is equicontinuous. Let $D_{n,i}(S) = \mathbf{1}\{Y_i \notin S\} D(Y_i) K_h(x - X_i)$ for a compact set S on $\mathbb{R}$. By the Hölder inequality and the law of iterated expectations,

$$\begin{aligned} ED_{n,i}(S) &= EE[D_{n,i}(S)|X_i] \\ &\leq E\left[\{P(Y_i \notin S|X_i)\}^{(\gamma-1)/\gamma}\{E[D^\gamma(Y_i)|X_i]\}^{1/\gamma} K_h(x - X_i)\right] \\ &= E\left[\{P(Y_i \notin S|X_i)\}^{(\gamma-1)/\gamma}\left[\bar{D}(X_i)\right]^{1/\gamma} K_h(x - X_i)\right]. \end{aligned} \tag{C.21}$$

Note that

$$E\left[\left[\bar{D}(X_i)\right]^{1/\gamma} K_h(x - X_i)\right] = \int \left[\bar{D}(x - hv)\right]^{1/\gamma} f(x - hv) K(v)\, dv \leq C \int K(v)\, dv. \tag{C.22}$$

Consider $\epsilon, \eta > 0$. By Assumption A2, we can choose S large enough such that $P(Y_i \notin S|X_i)$ is arbitrary small to ensure $ED_{n,i}(S) < \epsilon\eta/4$. Also, $q_n(z, \theta)$ is uniformly continuous on $(\mathcal{X} \times S) \times \Theta$ for each compact set $\mathcal{X} \times S$, implying that for any $\theta \in \Theta$ there exists $\mathcal{N} \equiv \mathcal{N}(\theta)$ such that $\sup_{(z,\theta') \in (\mathcal{X} \times S) \times \mathcal{N}} |p_1(z, \theta') - p_1(z, \theta)| < \epsilon/2$. Consequently

$$\sup_{\theta' \in \mathcal{N}} \left|p_1(Z_i, \theta') - p_1(Z_i, \theta)\right| < \epsilon/2 + 2 \cdot \mathbf{1}\{Y_i \notin S\} D(Y_i) K_h(x - X_i). \tag{C.23}$$

Let $\triangle_n(\epsilon, \eta) = \epsilon/2 + 2\bar{D}_n(S)$, where $\bar{D}_n(S) = n^{-1}\sum_{i=1}^n D_{n,i}(S)$. By (C.23) and the triangle inequality

$$\sup_{\theta' \in \mathcal{N}} \left|P_n p_1(Z, \theta') - P_n p_1(Z, \theta)\right| < \triangle_n(\epsilon, \eta).$$

Also,

$$P(\triangle_n(\epsilon, \eta) > \epsilon) = P\left(\bar{D}_n(S) > \epsilon/4\right) \leq \frac{E[D_{n,i}(S)]}{\epsilon/4} < \eta.$$

Consequently

$$\begin{aligned} \sup_{\theta' \in \mathcal{N}} \left|\bar{P}_{n,1}(\theta') - \bar{P}_{n,1}(\theta)\right| &= \sup_{\theta' \in \mathcal{N}} \left|E\left[P_n p_1(Z, \theta') - P_n p_1(Z, \theta)\right]\right| \\ &\leq E\left[\sup_{\theta' \in \mathcal{N}} \left|P_n p_1(Z, \theta') - P_n p_1(Z, \theta)\right|\right] \leq E[\triangle_n(\epsilon, \eta)] < \eta. \end{aligned}$$

That is, $\{\bar{P}_{n,1}(\theta)\}$ is equicontinuous.

Notice that under our assumption on the compactness of $\mathcal{B}$ and the support of K, $P_i(\boldsymbol{\beta}_2)$ is bounded. So the proof for the equicontinuity of $\bar{P}_{n,2}(\theta)$ is simpler than that of $\bar{P}_{n,1}(\theta)$ and thus omitted. ■

Proof of Lemma B.1. We only prove the case $(r, s) = (1, 1)$ as the other cases are similar. For notational simplicity, write $T_{1n\mathbf{j}} = T_{1n\mathbf{j}}(1, 1)$. By the fact that $\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{-1/2} - \varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2} = O_p(\upsilon_{2n})$ uniformly in i on the set $\{K_{ix} > 0\}$, we can write

$$\begin{aligned} \tilde{q}_{1,i}(\tilde{\boldsymbol{\beta}}) - q_{1,i}(\boldsymbol{\beta}^0) &= \frac{\tilde{f}_i'(\overrightarrow{\varepsilon}_i)}{\tilde{f}_i(\overrightarrow{\varepsilon}_i)}\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{-1/2} - \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2} \\ &= \left[\frac{\tilde{f}_i'(\overrightarrow{\varepsilon}_i)}{\tilde{f}_i(\overrightarrow{\varepsilon}_i)} - \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\right]\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{-1/2} + \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\left[\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{-1/2} - \varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}\right] \\ &= \left[\frac{\tilde{f}_i'(\overrightarrow{\varepsilon}_i)}{\tilde{f}_i(\overrightarrow{\varepsilon}_i)} - \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\right]\left[\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2} + O_p(\upsilon_{2n})\right] + \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}O_p(\upsilon_{2n}). \end{aligned} \tag{C.24}$$

Thus

$$\begin{aligned}
|T_{1n\mathbf{j}}| &\leq \frac{1}{nh^d}\sum_{i=1}^{n}\left|K_{ix,\mathbf{j}}\tilde{G}_i q_{1,i}\left(\boldsymbol{\beta}^0\right)\left[\frac{\tilde{f}_i'\left(\overrightarrow{\varepsilon}_i\right)}{\tilde{f}_i\left(\overrightarrow{\varepsilon}_i\right)}-\frac{f'\left(\bar{\varepsilon}_i\right)}{f\left(\bar{\varepsilon}_i\right)}\right]\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1/2}\right| \\
&\quad+\frac{O_p\left(\upsilon_{2n}\right)}{nh^d}\sum_{i=1}^{n}\left|K_{ix,\mathbf{j}}\tilde{G}_i q_{1,i}\left(\boldsymbol{\beta}^0\right)\left[\frac{\tilde{f}_i'\left(\overrightarrow{\varepsilon}_i\right)}{\tilde{f}_i\left(\overrightarrow{\varepsilon}_i\right)}-\frac{f'\left(\bar{\varepsilon}_i\right)}{f\left(\bar{\varepsilon}_i\right)}\right]\right| \\
&\quad+\frac{O_p\left(\upsilon_{2n}\right)}{nh^d}\sum_{i=1}^{n}\left|K_{ix,\mathbf{j}}\tilde{G}_i q_{1,i}\left(\boldsymbol{\beta}^0\right)\frac{f'\left(\bar{\varepsilon}_i\right)}{f\left(\bar{\varepsilon}_i\right)}\right|.
\end{aligned}$$

Since the last two terms are of smaller order, it suffices to show the first term (denoted as $|\bar{T}_{1n\mathbf{j}}|$) is $O_p\left(h^{\epsilon}\right)$. By Lemma A.1, the definition of $\tilde{G}_i$, and Assumption A7,

$$\begin{aligned}
&\left|\frac{\tilde{f}_i'\left(\overrightarrow{\varepsilon}_i\right)}{\tilde{f}_i\left(\overrightarrow{\varepsilon}_i\right)}-\frac{f'\left(\bar{\varepsilon}_i\right)}{f\left(\bar{\varepsilon}_i\right)}\right|\tilde{G}_i=\left|\frac{\tilde{f}_i'\left(\overrightarrow{\varepsilon}_i\right)-f'\left(\bar{\varepsilon}_i\right)}{\tilde{f}_i\left(\overrightarrow{\varepsilon}_i\right)}+\frac{f'\left(\bar{\varepsilon}_i\right)\left[f\left(\bar{\varepsilon}_i\right)-\tilde{f}_i\left(\overrightarrow{\varepsilon}_i\right)\right]}{f\left(\bar{\varepsilon}_i\right)\tilde{f}_i\left(\overrightarrow{\varepsilon}_i\right)}\right|\tilde{G}_i \\
\leq\;& O_p\left(b^{-1}\nu_{3n,1}\right)+\left(f'\left(\bar{\varepsilon}_i\right)/f\left(\bar{\varepsilon}_i\right)\right)O_p\left(b^{-1}\nu_{3n,0}\right)=O_p\left(h^{\epsilon}\right)\left\{1+\left|f'\left(\bar{\varepsilon}_i\right)/f\left(\bar{\varepsilon}_i\right)\right|\right\}. \qquad \text{(C.25)}
\end{aligned}$$

Therefore $\left|\bar{T}_{1n\mathbf{j}}\right|=\frac{O_p(h^{\epsilon})}{nh^d}\sum_{i=1}^{n}\left|K_{ix,\mathbf{j}}q_{1,i}\left(\boldsymbol{\beta}^0\right)\left(1+\frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\right)\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1/2}\right|=O_p\left(h^{\epsilon}\right)$ by Markov inequality and the fact that

$$\begin{aligned}
\frac{1}{h^d}E\left|K_{ix,\mathbf{j}}q_{1,i}\left(\boldsymbol{\beta}^0\right)\left(1+\frac{f'\left(\bar{\varepsilon}_i\right)}{f\left(\bar{\varepsilon}_i\right)}\right)\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1/2}\right| &= \frac{1}{h^d}E\left|K_{ix,\mathbf{j}}\frac{f'\left(\bar{\varepsilon}_i\right)}{f\left(\bar{\varepsilon}_i\right)}\left(1+\frac{f'\left(\bar{\varepsilon}_i\right)}{f\left(\bar{\varepsilon}_i\right)}\right)\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1}\right| \\
&= \frac{1}{h^d}E\left|K_{ix,\mathbf{j}}\sigma^{-2}\left(X_i\right)\frac{f'\left(\varepsilon_i\right)}{f\left(\varepsilon_i\right)}\left(1+\frac{f'\left(\varepsilon_i\right)}{f\left(\varepsilon_i\right)}\right)\right|\{1+o\left(1\right)\} \\
&\leq \frac{f_X\left(x\right)}{\sigma^2\left(x\right)}\int\left|K\left(u\right)u^{\mathbf{j}}\right|du\left\{I^{1/2}\left(f\right)+I\left(f\right)\right\}=O\left(1\right),
\end{aligned}$$

where $I\left(f\right)\equiv E\left[\psi^2\left(\varepsilon_i\right)\right]$ and we use the fact that $\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))$ is the p-th order Taylor expansion of $\sigma^2\left(X_i\right)$ around x. This completes the proof of the lemma. ■

Proof of Lemma B.2. We only prove the case $(r,s)=(1,1)$ as the other cases are similar. For notational simplicity, write $T_{2n\mathbf{j}}=T_{2n\mathbf{j}}\left(1,1\right)$. That is, we will show

$$T_{2n\mathbf{j}}=\frac{1}{nh^d}\sum_{i=1}^{n}K_{ix,\mathbf{j}}\tilde{G}_i\left\{\tilde{q}_{1,i}(\tilde{\boldsymbol{\beta}})-q_1\left(Y_i;P_i\left(\boldsymbol{\beta}_1^0\right),P_i\left(\boldsymbol{\beta}_2^0\right)\right)\right\}^2=O_p\left(h^{\epsilon}\right).$$

By (C.24) and (C.25) in the proof of Lemma B.1, we can write

$$\begin{aligned}
&\left|\tilde{q}_{1,i}(\tilde{\boldsymbol{\beta}})-q_1\left(Y_i;P_i\left(\boldsymbol{\beta}_1^0\right),P_i\left(\boldsymbol{\beta}_2^0\right)\right)\right|^2\tilde{G}_i \\
=\;&\left[O_p\left(h^{\epsilon}\right)\left(1+\left|\frac{f'\left(\bar{\varepsilon}_i\right)}{f\left(\bar{\varepsilon}_i\right)}\right|\right)^2\left[\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1/2}+O_p(\upsilon_{2n})\right]^2+\left(\frac{f'\left(\bar{\varepsilon}_i\right)}{f\left(\bar{\varepsilon}_i\right)}\right)^2O_p(\upsilon_{2n}^2)\right]\tilde{G}_i \\
\leq\;&\left(1+\left|\frac{f'\left(\bar{\varepsilon}_i\right)}{f\left(\bar{\varepsilon}_i\right)}\right|\right)^2\left[\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1}+1\right]\tilde{G}_iO_p\left(h^{\epsilon}\right).
\end{aligned}$$

Thus

$$T_{2n\mathbf{j}}\leq\frac{O_p\left(h^{\epsilon}\right)}{nh^d}\sum_{i=1}^{n}\left|K_{ix,\mathbf{j}}\right|\left(1+\left|\frac{f'\left(\bar{\varepsilon}_i\right)}{f\left(\bar{\varepsilon}_i\right)}\right|\right)^2\left[\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1}+1\right]=O_p\left(h^{\epsilon}\right)$$

by Markov inequality and the fact that

$$
\begin{aligned}
& \frac{1}{h^d} E\left|K_{ix,\mathbf{j}}\left(1+\left|\frac{f'\left(\bar{\varepsilon}_i\right)}{f\left(\bar{\varepsilon}_i\right)}\right|\right)^2\left[\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1}+1\right]\right| \\
= & \frac{1}{h^d} E\left|K_{ix,\mathbf{j}}\left(1+\left|\frac{f'\left(\bar{\varepsilon}_i\right)}{f\left(\bar{\varepsilon}_i\right)}\right|\right)^2\left[\sigma^{-2}\left(X_i\right)+1+o(1)\right]\right|\{1+o(1)\} \\
\leq & f_X(x)\left[\sigma^{-2}(x)+1\right] \int\left|K(u) u^{\mathbf{j}}\right| du\left[1+I(f)+2 I^{1/2}(f)\right]\{1+o(1)\}=O(1) .
\end{aligned}
$$

This completes the proof of the lemma. ■

Proof of Lemma B.3. We only prove the case $(r, s)=(1,1)$ as the other cases are similar. For notational simplicity, write $T_{3n\mathbf{j}}=T_{3n\mathbf{j}}(1,1)$. That is, we will show

$$
T_{3n\mathbf{j}}=\frac{1}{nh^d} \sum_{i=1}^n K_{ix,\mathbf{j}}\left(1-\tilde{G}_i\right) q_{1,i}\left(\boldsymbol{\beta}^0\right)^2=O_p\left(h^\epsilon\right).
$$

We decompose $T_{3n\mathbf{j}}$ as follows

$$
\begin{aligned}
T_{3n\mathbf{j}} = & \frac{1}{nh^d} \sum_{i=1}^n K_{ix,\mathbf{j}}\left[1-G_b\left(f\left(\bar{\varepsilon}_i\right)\right)\right] \psi^2\left(\bar{\varepsilon}_i\right) \\
& +\frac{1}{nh^d} \sum_{i=1}^n K_{ix,\mathbf{j}}\left[G_b\left(f\left(\bar{\varepsilon}_i\right)\right)-G_b(\tilde{f}_i\left(\overrightarrow{\varepsilon}_i\right))\right] \psi^2\left(\bar{\varepsilon}_i\right) \\
\equiv & T_{3n\mathbf{j},1}+T_{3n\mathbf{j},2}, \text{ say.}
\end{aligned}
$$

By Lemma A.1,

$$
\begin{aligned}
\max_{1 \leq i \leq n}|\tilde{G}_i-G_i| & =\max_{1 \leq i \leq n}|G_b(\tilde{f}_i\left(\overrightarrow{\varepsilon}_i\right))-G_b\left(f\left(\bar{\varepsilon}_i\right)\right)| \\
& \leq \frac{C}{b} \max_{1 \leq i \leq n}|\tilde{f}_i\left(\overrightarrow{\varepsilon}_i\right)-f\left(\bar{\varepsilon}_i\right)|=b^{-1} O_p\left(\upsilon_{3n,0}\right)=O_p\left(h^\epsilon\right) . \qquad \text{(C.26)}
\end{aligned}
$$

With this, we can readily obtain $|T_{3n\mathbf{j},2}| \leq O_p\left(h^\epsilon\right) \frac{1}{nh^d} \sum_{i=1}^n|K_{ix,\mathbf{j}}| \psi^2\left(\bar{\varepsilon}_i\right)=O_p\left(h^\epsilon\right)$ by Markov inequality. For $T_{3n\mathbf{j},1}$, we have

$$
\begin{aligned}
E|T_{3n\mathbf{j},1}| & \leq E\left[\frac{1}{h^d}|K_{ix,\mathbf{j}}|\left[1-G_b\left(f\left(\bar{\varepsilon}_i\right)\right)\right] \psi^2\left(\bar{\varepsilon}_i\right)\right] \\
& =\frac{1}{h^d} E\left[|K_{ix,\mathbf{j}}|\right] E\left\{\left[1-G_b\left(f\left(\varepsilon_i\right)\right)\right] \psi^2\left(\varepsilon_i\right)\right\}\{1+o(1)\} .
\end{aligned}
$$

By the Hölder inequality,

$$
\begin{aligned}
E\left\{\left[1-G_b\left(f\left(\varepsilon_i\right)\right)\right] \psi^2\left(\varepsilon_i\right)\right\} & \leq E\left[\psi^2\left(\varepsilon_i\right) \mathbf{1}\{f(\varepsilon) \leq 2b\}\right] \\
& \leq\left\{E\left[\psi^{2\gamma}\left(\varepsilon_i\right)\right]\right\}^{1/\gamma}\left[P\left(f\left(\varepsilon_i\right) \leq 2b\right)\right]^{(\gamma-1)/\gamma} \\
& \leq C\left[P\left(f\left(\varepsilon_i\right) \leq 2b\right)\right]^{(\gamma-1)/\gamma}=O\left(b^{(\gamma-1)/(2\gamma)}\right)=O\left(h^\epsilon\right),
\end{aligned}
$$

where the last line follows from Lemma 6 of Robinson (1988) and the Markov inequality because by taking $\overline{B}=b^{-1/2}$, we have $P\left(f\left(\varepsilon_i\right) \leq 2b\right) \leq 2b \int_{|\varepsilon_i| \leq \overline{B}} dz+P\left(|\varepsilon_i|>\overline{B}\right) \leq 2b2b^{-1/2}+E|\varepsilon_i| b^{1/2}=O\left(b^{1/2}\right)=O\left(h^\epsilon\right)$. This, in conjunction with the fact that $\frac{1}{h^d} E\left[|K_{ix,\mathbf{j}}|\right]=O(1)$, implies that $T_{3n\mathbf{j},1}=O_p\left(h^\epsilon\right)$ by Markov inequality. Consequently, we have shown that $T_{3n\mathbf{j}}=O_p\left(h^\epsilon\right)$. ■

Proof of Lemma B.4. Let $\tilde{f}_i = \tilde{f}_i(\overrightarrow{\varepsilon}_i)$ and $f_i = f(\bar{\varepsilon}_i)$.Note that $\tilde{f}_i^{-1} = f_i^{-1} - (\tilde{f}_i - f_i)/f_i^2 + R_{2i}$, where $R_{2i} \equiv (\tilde{f}_i - f_i)^2/\{(f_i^2 \tilde{f}_i)$. First, we expand the trimming function to the second order:

$$G_b(\tilde{f}_i) - G_b(f_i) = g_b(f_i)\left(\tilde{f}_i - f_i\right) + \frac{1}{2} g_b'(f_i^*)\left(\tilde{f}_i - f_i\right)^2, \tag{C.27}$$

where f_i^* is an intermediate value between $\tilde{f}_i$ and f_i. Let $\rho_i(\boldsymbol{\beta}) \equiv \psi(\varepsilon_i(\boldsymbol{\beta}))\varepsilon_i(\boldsymbol{\beta}) + 1$, $\bar{\rho}_i \equiv \rho_i(\boldsymbol{\beta}^0)$, and $\rho_i \equiv \psi(\varepsilon_i)\varepsilon_i + 1$. Let $\varphi_i \equiv \varphi'(P_i(\boldsymbol{\beta}_2^0))/\varphi(P_i(\boldsymbol{\beta}_2^0))$. Then we have

$$\begin{aligned}
-\mathcal{S}_{1n\mathbf{j}} &= \frac{1}{2\sqrt{nh^d}} \sum_{i=1}^n K_{ix,\mathbf{j}} \varphi_i \left\{ \bar{\rho}_i \left[G_b(\tilde{f}_i) - 1 \right] + \log\left(f(\bar{\varepsilon}_i) \varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2} \right) g_b(f(\bar{\varepsilon}_i)) f'(\bar{\varepsilon}_i) \bar{\varepsilon}_i \right\} \\
&= \frac{1}{2\sqrt{nh^d}} \sum_{i=1}^n K_{ix,\mathbf{j}} \varphi_i \left\{ \bar{\rho}_i \left[G_b(f_i) - 1 \right] + \log\left(f(\bar{\varepsilon}_i) \varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2} \right) g_b(f(\bar{\varepsilon}_i)) f'(\bar{\varepsilon}_i) \bar{\varepsilon}_i \right\} \\
&\quad + \frac{1}{2\sqrt{nh^d}} \sum_{i=1}^n K_{ix,\mathbf{j}} \varphi_i \bar{\rho}_i g_b(f_i) \left(\tilde{f}_i - f_i \right) \\
&\quad + \frac{1}{4\sqrt{nh^d}} \sum_{i=1}^n K_{ix,\mathbf{j}} \varphi_i \bar{\rho}_i g_b'(f_i^*) \left(\tilde{f}_i - f_i \right)^2 \\
&\equiv \mathcal{S}_{1n\mathbf{j},1} + \mathcal{S}_{1n\mathbf{j},2} + \mathcal{S}_{1n\mathbf{j},3}, \text{ say.}
\end{aligned}$$

Using a crude bound on the last term, we have $|\mathcal{S}_{1n\mathbf{j},3}| = O_p\left(\upsilon_{3n,0}^2 b^{-2} n^{1/2} h^{d/2}\right) = o_p(1)$ by Lemma A.1, the fact that $\sup_s |g_b'(s)| = O(b^{-2})$, and Assumption A7.

To show the first term is $o_p(1)$, write

$$\begin{aligned}
\mathcal{S}_{1n\mathbf{j},1} &= \frac{-1}{2\sqrt{nh^d}} \sum_{i=1}^n K_{ix,\mathbf{j}} \varphi_i \bar{\rho}_i \\
&\quad + \frac{1}{2\sqrt{nh^d}} \sum_{i=1}^n K_{ix,\mathbf{j}} \varphi_i \left[\bar{\rho}_i G_b(f_i) + \log\left(f(\bar{\varepsilon}_i) \varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2} \right) g_b(f(\bar{\varepsilon}_i)) f'(\bar{\varepsilon}_i) \bar{\varepsilon}_i \right] \\
&= \frac{-1}{\sqrt{nh^d}} \sum_{i=1}^n K_{ix,\mathbf{j}} \xi_{1i} + \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K_{ix,\mathbf{j}} \xi_{2i} \\
&\equiv -\mathcal{S}_{1n\mathbf{j},11} + \mathcal{S}_{1n\mathbf{j},12}, \text{ say,}
\end{aligned}$$

where $\xi_{1i} = \frac{1}{2} \bar{\rho}_i \varphi_i$ and $\xi_{2i} = \frac{1}{2} \left\{ \bar{\rho}_i G_b(f_i) + \log(f(\bar{\varepsilon}_i) \varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}) g_b(f(\bar{\varepsilon}_i)) f'(\bar{\varepsilon}_i) \bar{\varepsilon}_i \right\} \varphi_i$.

Let $Q_{1n,i}(\boldsymbol{\beta}) \equiv \log\{f(\varepsilon_i(\boldsymbol{\beta})) \varphi(P_i(\boldsymbol{\beta}_2))^{-1/2}\} K_h(x - X_i)$, $Q_{2n,i}(\boldsymbol{\beta}) \equiv \log\{f(\varepsilon_i(\boldsymbol{\beta})) \varphi(P_i(\boldsymbol{\beta}_2))^{-1/2}\}$ $\times G_b(f(\varepsilon_i(\boldsymbol{\beta})) K_h(x - X_i)$, and $\varsigma_n(\boldsymbol{\beta}) \equiv E[Q_{2n,i}(\boldsymbol{\beta})] - E[Q_{1n,i}(\boldsymbol{\beta})]$. Then it is easy to show that (i) $\varsigma_n(\boldsymbol{\beta}^0) \to 0$, (ii) $\varsigma_n(\boldsymbol{\beta})$ is differentiable in a small ϵ_0-neighborhood $N_{\epsilon_0}(\boldsymbol{\beta}^0)$ of $\boldsymbol{\beta}^0$ with $N_{\epsilon_0}(\boldsymbol{\beta}^0) \equiv \{\boldsymbol{\beta} : \left\| \boldsymbol{\beta} - \boldsymbol{\beta}^0 \right\| \leq \epsilon_0\}$, (iii) $\varsigma_n'(\boldsymbol{\beta})$ converges uniformly on $N_{\epsilon_0}(\boldsymbol{\beta}^0)$. Then by Theorem 7.17 of Rudin (1976) and the fact that $h^{-|\mathbf{j}|} \partial Q_{1n,i}(\boldsymbol{\beta}^0) / \partial \boldsymbol{\beta}_{2\mathbf{j}} = -h^{-d} \xi_{1i} K_{ix,\mathbf{j}}$ and $h^{-|\mathbf{j}|} \partial Q_{2n,i}(\boldsymbol{\beta}^0) / \partial \boldsymbol{\beta}_{2\mathbf{j}} = -h^{-d} \xi_{2i} K_{ix,\mathbf{j}}$, we have

$$\begin{aligned}
E(\mathcal{S}_{1n\mathbf{j},12}) &= -\sqrt{n} h^{d/2} h^{-|\mathbf{j}|} E\left[\frac{\partial Q_{2n,i}(\boldsymbol{\beta}^0)}{\partial \boldsymbol{\beta}_{2\mathbf{j}}} \right] \\
&= -\sqrt{n} h^{d/2} h^{-|\mathbf{j}|} E\left[\frac{\partial Q_{1n,i}(\boldsymbol{\beta}^0)}{\partial \boldsymbol{\beta}_{2\mathbf{j}}} \right] \{1 + o(1)\} = E(\mathcal{S}_{1n\mathbf{j},11}) \{1 + o(1)\}.
\end{aligned}$$

Consequently, $E(\mathcal{S}_{1n\mathbf{j},1}) = o(1) E(\mathcal{S}_{1n\mathbf{j},11}) = o(1)$ as $\mathcal{S}_{1n\mathbf{j},11} = n^{1/2} h^{-d/2} E[K_{ix,\mathbf{j}} \xi_{1i}] = O\left(n^{1/2} h^{d/2} h^{p+1}\right)$ $= O(1)$. By straightforward calculations and the IID assumption, we can readily show that $\mathrm{Var}(\mathcal{S}_{1n\mathbf{j},1}) = o(1)$. Therefore, $\mathcal{S}_{1n\mathbf{j},1} = o_p(1)$ by the Chebyshev inequality.

Now, we show that $\mathcal{S}_{1n\mathbf{j},2}=o_p(1)$. Decompose $\mathcal{S}_{1n\mathbf{j},2}=\mathcal{S}_{1n\mathbf{j},21}+\mathcal{S}_{1n\mathbf{j},22}$, where $\mathcal{S}_{1n\mathbf{j},21}\equiv\frac{1}{2\sqrt{nh^d}}\sum_{i=1}^n K_{ix,\mathbf{j}}\varphi_i q_{2,i}\left(\boldsymbol{\beta}^0\right)g_b(f_i)\left(\tilde{f}_i(\overrightarrow{\varepsilon}_i)-\overline{f}(\bar{\varepsilon}_i)\right)$, and $\mathcal{S}_{1n\mathbf{j},22}\equiv\frac{1}{2\sqrt{nh^d}}\sum_{i=1}^n K_{ix,\mathbf{j}}\varphi_i q_{2,i}\left(\boldsymbol{\beta}^0\right)g_b(f_i)\left(\overline{f}(\bar{\varepsilon}_i)-f(\bar{\varepsilon}_i)\right)$. It suffices to show that $\mathcal{S}_{1n\mathbf{j},2s}=o_p(1)$, $s=1,2$. For $\mathcal{S}_{1n\mathbf{j},21}$, by a Taylor expansion and (C.9)-(C.10), we have

$$\begin{aligned}
\mathcal{S}_{1n\mathbf{j},21} &= \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^n K_{ix,\mathbf{j}}\varphi_i\bar{\rho}_i g_b(f_i)\frac{1}{nh_0}\sum_{j\neq i}\left[k_0\left(\frac{\overrightarrow{\varepsilon}_i-\tilde{\varepsilon}_j}{h_0}\right)-k_0\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\right] \\
&= \frac{1}{2n^{3/2}h^{d/2}h_0^2}\sum_{i=1}^n\sum_{j\neq i}K_{ix,\mathbf{j}}\varphi_i\bar{\rho}_i g_b(f_i)k_0'\left(\frac{\varepsilon_i-\varepsilon_j}{h_0}\right)(\overrightarrow{\varepsilon}_i-\tilde{\varepsilon}_j-\bar{\varepsilon}_i+\varepsilon_j)+o_p(1) \\
&= -\frac{1}{2n^{3/2}h^{d/2}h_0^2}\sum_{i=1}^n\sum_{j\neq i}K_{ix,\mathbf{j}}\varphi_i\bar{\rho}_i g_b(f_i)k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\frac{v_{1i}(x)}{\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2}} \\
&\quad+\frac{1}{2n^{3/2}h^{d/2}h_0^2}\sum_{i=1}^n\sum_{j\neq i}K_{ix,\mathbf{j}}\varphi_i\bar{\rho}_i g_b(f_i)k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\frac{\tilde{m}(X_j)-m(X_j)}{\sigma(X_j)} \\
&\quad+\frac{1}{2n^{3/2}h^{d/2}h_0^2}\sum_{i=1}^n\sum_{j\neq i}K_{ix,\mathbf{j}}\varphi_i\bar{\rho}_i g_b(f_i)k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\varepsilon_j\frac{\tilde{\sigma}(X_j)-\sigma(X_j)}{\tilde{\sigma}(X_j)} \\
&\quad+\frac{1}{2n^{3/2}h^{d/2}h_0^2}\sum_{i=1}^n\sum_{j\neq i}K_{ix,\mathbf{j}}\varphi_i\bar{\rho}_i g_b(f_i)k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\frac{u_i+\delta_i}{\varphi(P_i(\boldsymbol{\beta}_2^0))^{1/2}}\frac{\varphi(P_i(\boldsymbol{\beta}_2^0))^{1/2}-\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2}}{\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2}} \\
&\quad+o_p(1) \\
&\equiv -\mathcal{S}_{1n\mathbf{j},211}+\mathcal{S}_{1n\mathbf{j},212}+\mathcal{S}_{1n\mathbf{j},213}+\mathcal{S}_{1n\mathbf{j},214}+o_p(1).
\end{aligned}$$

For the first term, by Lemma A.2 and the fact that $v_{1i}(x)=O_p(\upsilon_{2n})$, $\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))=\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))+O_p(\upsilon_{2n})$ uniformly on the set $\{K_{ix}>0\}$, we have

$$\begin{aligned}
|\mathcal{S}_{1n\mathbf{j},211}| &= \left|\frac{1}{2\sqrt{nh^d}}\sum_{i=1}^n K_{ix,\mathbf{j}}\varphi_i\bar{\rho}_i g_b(f_i)\bar{f}_i'(\varepsilon_i)\frac{v_{1i}(x)}{\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2}}\right| \\
&= \left|\frac{1}{2\sqrt{nh^d}}\sum_{i=1}^n K_{ix,\mathbf{j}}\varphi_i\bar{\rho}_i g_b(f_i)f'(\varepsilon_i)\frac{v_{1i}(x)}{\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{1/2}}\right|+o_p(1) \\
&\leq \max_{\{K_{ix}>0\}}|v_{1i}(x)|\frac{1}{2\sqrt{nh^d}}\sum_{i=1}^n\left|K_{ix,\mathbf{j}}\varphi_i\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1/2}\bar{\rho}_i g_b(f_i)f'(\varepsilon_i)\right|+o_p(1).
\end{aligned}$$

The first term in the last expression is $o_p(1)$ if $n^{1/2}h^{-d/2}E\left|K_{ix,\mathbf{j}}\varphi_i\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1/2}\bar{\rho}_i g_b(f_i)f'(\varepsilon_i)\right|=o\left(\upsilon_{2n}^{-1}\right)$ by Markov inequality. Note that

$$\bar{\varepsilon}_i-\varepsilon_i=\{\varepsilon_i[\sigma(X_i)-\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{1/2}]+\delta_i\}/\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{1/2}=\varepsilon_i d_i+\overline{\delta}_i. \tag{C.28}$$

where $d_i\equiv\sigma(X_i)\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1/2}-1=O_p\left(h^{p+1}\right)$ and $\overline{\delta}_i\equiv\delta_i\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1/2}=O_p\left(h^{p+1}\right)$ uniformly on the set $\{K_{ix}>0\}$. Then by the triangle inequality,

$$\begin{aligned}
& h^{-d}E\left|K_{ix,\mathbf{j}}\varphi_i\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1/2}\bar{\rho}_i g_b\left(f_i\right)f'\left(\varepsilon_i\right)\right| \\
= & h^{-d}E\left|K_{ix,\mathbf{j}}\varphi_i\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1/2}\left[\bar{\rho}_i g_b\left(f_i\right)f'\left(\varepsilon_i\right)|X_i\right]\right| \\
= & h^{-d}E\left|\frac{K_{ix,\mathbf{j}}\varphi_i\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1/2}}{d_i+1}\int_{b\le f(\varepsilon)\le 2b}\left[\psi\left(\varepsilon\right)\varepsilon+1\right]f'\left(\varepsilon\right)g_b\left(f\left(\varepsilon\right)\right)f\left(\frac{\varepsilon-\overline{\delta}_i}{d_i+1}\right)d\varepsilon\right| \\
\le & h^{-d}E\left|\frac{K_{ix,\mathbf{j}}\varphi_i\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1/2}}{d_i+1}\int_{b\le f(\varepsilon)\le 2b}\rho\left(\varepsilon\right)f'\left(\varepsilon\right)g_b\left(f\left(\varepsilon\right)\right)f\left(\varepsilon\right)d\varepsilon\right| \\
& +h^{-d}E\left|\frac{K_{ix,\mathbf{j}}\varphi_i\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1/2}}{d_i+1}\int_{b\le f(\varepsilon)\le 2b}\rho\left(\varepsilon\right)f'\left(\varepsilon\right)g_b\left(f\left(\varepsilon\right)\right)\left[f\left(\frac{\varepsilon-\overline{\delta}_i}{d_i+1}\right)-f\left(\varepsilon\right)\right]d\varepsilon\right| \\
\equiv & S_{n1}+S_{n2},\ \text{say.}
\end{aligned}$$

For S_{n1}, we have

$$\begin{aligned}
S_{n1} &= h^{-d}E\left|K_{ix,\mathbf{j}}\varphi_i\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1/2}\left(d_i+1\right)^{-1}\int_{b\le f(\varepsilon)\le 2b}\rho\left(\varepsilon\right)f'\left(\varepsilon\right)g_b\left(f\left(\varepsilon\right)\right)f\left(\varepsilon\right)d\varepsilon\right| \\
&\le \sup_{b\le f(\varepsilon)\le 2b}\left[f\left(\varepsilon\right)g_b\left(f\left(\varepsilon\right)\right)\right]h^{-d}E\left|K_{ix,\mathbf{j}}\varphi_i\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1/2}\left(d_i+1\right)^{-1}\int_{b\le f(\varepsilon)\le 2b}\rho\left(\varepsilon\right)f'\left(\varepsilon\right)d\varepsilon\right| \\
&\le Ch^{-d}E\left|K_{ix,\mathbf{j}}\varphi_i\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1/2}\left(d_i+1\right)^{-1}\right|\left|\int_{b\le f(\varepsilon)\le 2b}\rho\left(\varepsilon\right)f'\left(\varepsilon\right)d\varepsilon\right| \\
&\le Ch^{-d}E\left|K_{ix,\mathbf{j}}\varphi_i\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1/2}\left(d_i+1\right)^{-1}\right|\left\{\int_{b\le f(\varepsilon)\le 2b}\rho\left(\varepsilon\right)^2f\left(\varepsilon\right)d\varepsilon\int_{b\le f(\varepsilon)\le 2b}\psi\left(\varepsilon\right)f\left(\varepsilon\right)d\varepsilon\right\}^{1/2} \\
&= O\left(h^{\epsilon}\right)
\end{aligned}$$

where the third inequality follows from the Hölder inequality and the independence between X_i and ε_i. By a Taylor expansion, $f\left(\frac{\varepsilon-\overline{\delta}_i}{1+d_i}\right)-f\left(\varepsilon\right)\simeq -f'\left(\varepsilon\right)\left(\overline{\delta}_i+d_i\varepsilon\right)$. With this, we can readily show that $S_{n2}=O\left(h^{\epsilon}\right)$. Consequently, $|\mathcal{S}_{1n\mathbf{j},211}|=O_p(\upsilon_{2n}\sqrt{nh^d}h^{\epsilon})=o_p\left(1\right)$.

For $\mathcal{S}_{1n\mathbf{j},212}$, using (C.2) we can write

$$\begin{aligned}
\mathcal{S}_{1n\mathbf{j},212} &= \frac{1}{2n^{3/2}h^{d/2}h_0^2}\sum_{i=1}^n\sum_{j\neq i}K_{ix,\mathbf{j}}\varphi_i\bar{\rho}_i g_b\left(f_i\right)k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\frac{\tilde{m}\left(X_j\right)-m\left(X_j\right)}{\sigma\left(X_j\right)} \\
&= \frac{1}{2n^{3/2}h^{d/2}h_0^2}\sum_{i=1}^n\sum_{j\neq i}\frac{K_{ix,\mathbf{j}}\varphi_i}{\sigma\left(X_j\right)}\bar{\rho}_i g_b\left(f_i\right)k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)e_1^{\top}M_n^{-1}\left(X_j\right)U_{1,n}\left(X_j\right) \\
&\quad -\frac{1}{2n^{3/2}h^{d/2}h_0^2}\sum_{i=1}^n\sum_{j\neq i}\frac{K_{ix,\mathbf{j}}\varphi_i}{\sigma\left(X_j\right)}\bar{\rho}_i g_b\left(f_i\right)k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)e_1^{\top}M_n^{-1}\left(X_j\right)B_{1,n}\left(X_j\right) \\
&\equiv \mathcal{S}_{1n\mathbf{j},212a}+\mathcal{S}_{1n\mathbf{j},212b}. \qquad\qquad \text{(C.29)}
\end{aligned}$$

Recall $\tilde{\mathbf{Z}}_i$ is defined analogously to $\mathbf{Z}_i$ with h_1 in place of h. So $\mathcal{S}_{1n\mathbf{j},212a}$ can be written as

$$\mathcal{S}_{1n\mathbf{j},212a}=\sum_{i=1}^n\sum_{j\neq i}\varsigma_{2n}\left(\varepsilon_i,\varepsilon_j\right)+\sum_{i=1}^n\sum_{j\neq i}\sum_{l\neq i,l\neq j}\varsigma_{3n}\left(\varepsilon_i,\varepsilon_j,\varepsilon_l\right),$$

where $\varsigma_{2n}\left(\varepsilon_i,\varepsilon_j\right)=\frac{1}{2n^{5/2}h^{d/2}h_1^dh_0^2}\frac{K_{ix,\mathbf{j}}\varphi_i}{\sigma(X_j)}\bar{\rho}_i g_b\left(f_i\right)k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)e_1^{\top}M_n^{-1}\left(X_j\right)\tilde{\mathbf{Z}}_iK\left(\frac{X_j-X_i}{h_1}\right)u_i$ and $\varsigma_{3n}(\varepsilon_i,\varepsilon_j,\varepsilon_l)$ $=\frac{1}{2n^{5/2}h^{d/2}h_1^dh_0^2}\frac{K_{ix,\mathbf{j}}\varphi_i}{2\sigma(X_j)}\bar{\rho}_i g_b\left(f_i\right)k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)e_1^{\top}M_n^{-1}\left(X_j\right)\tilde{\mathbf{Z}}_iK\left(\frac{X_j-X_l}{h_1}\right)u_l$. Let $\mathbb{X}\equiv\{X_1,...,X_n\}$. Then

$E[\mathcal{S}_{1n\mathbf{j},212a}|\mathbb{X}] = \sum_{i=1}^{n}\sum_{j\neq i} E\left[\varsigma_{2n}\left(z_i,z_j\right)|\mathbb{X}\right] = O_p\left(n^{-1/2}h^{d/2}b^{-1}\right) = o_p\left(1\right).$ For the variance of $\mathcal{S}_{1n\mathbf{j},212a}$, it is easy to show that

$$\begin{aligned}\text{Var}\left[\sum_{i=1}^{n}\sum_{j\neq i}\varsigma_{2n}\left(\varepsilon_i,\varepsilon_j\right)|\mathbb{X}\right] &= O\left(n^2\right)E\left[\varsigma_{2n}\left(\varepsilon_i,\varepsilon_j\right)^2+\varsigma_{2n}\left(\varepsilon_i,\varepsilon_j\right)\varsigma_{2n}\left(\varepsilon_j,\varepsilon_i\right)|\mathbb{X}\right]\\ &\quad +O\left(n^3\right)E\left[\varsigma_{2n}\left(\varepsilon_i,\varepsilon_j\right)\varsigma_{2n}\left(\varepsilon_l,\varepsilon_j\right)+\varsigma_{2n}\left(\varepsilon_i,\varepsilon_j\right)\varsigma_{2n}\left(\varepsilon_i,\varepsilon_i\right)|\mathbb{X}\right]\\ &= O_p\left(n^{-3}h^{-d-4}b^{-2}\right)+O_p\left(n^{-2}b^{-2}\right)=o_p\left(1\right).\end{aligned}$$

Similarly, one can show that $E\left(\varsigma_{3n}\left(z_i,z_j,z_l\right)|\mathbb{X}\right)=0$ and $\text{Var}\left[\sum_{i=1}^{n}\sum_{j\neq i}\sum_{l\neq i,l\neq j}\varsigma_{3n}\left(z_i,z_j,z_l\right)|\mathbb{X}\right] = o_p\left(1\right).$ Consequently, $\mathcal{S}_{1n\mathbf{j},212a}=o_p\left(1\right)$ by the conditional Chebyshev inequality. For $\mathcal{S}_{1n\mathbf{j},212b}$, we have $\mathcal{S}_{1n\mathbf{j},212b} = O_p(n^{1/2}h^{d/2}h_1^{p+1}) = o_p\left(1\right).$ Thus we have shown that $\mathcal{S}_{1n\mathbf{j},212} = o_p\left(1\right).$ By analogous arguments, Lemma A.1, and (C.8), we can show that $\mathcal{S}_{1n\mathbf{j},21s} = o_p\left(1\right)$ for $s=3,4$. It follows that $\mathcal{S}_{1n\mathbf{j},21}=o_p\left(1\right).$

For $\mathcal{S}_{1n\mathbf{j},22}$, we make the following decomposition:

$$\mathcal{S}_{1n\mathbf{j},22}=\frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n}K_{ix,\mathbf{j}}\varphi_i q_{2,i}\left(\boldsymbol{\beta}^0\right)g_b\left(f_i\right)\left\{\mathcal{V}\left(\bar{\varepsilon}_i\right)+\mathcal{B}\left(\bar{\varepsilon}_i\right)\right\}\equiv\mathcal{S}_{1n\mathbf{j},221}+\mathcal{S}_{1n\mathbf{j},222},$$

where

$$\mathcal{V}\left(\bar{\varepsilon}_i\right) = \frac{1}{nh_0}\sum_{j\neq i}\left\{k_0\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)-E_j\left[k_0\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\right]\right\}, \tag{C.30}$$

$$\mathcal{B}\left(\bar{\varepsilon}_i\right) = \frac{1}{nh_0}\sum_{j\neq i}E_j\left[k_0\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\right]-f\left(\bar{\varepsilon}_i\right), \tag{C.31}$$

and E_j indicates expectation with respect to the variable indexed by j. Writing $\mathcal{S}_{1n\mathbf{j},221}$ as a second order degenerate statistic we verify that $E\left[\mathcal{S}_{1n\mathbf{j},221}\right]^2=o\left(1\right)$ and thus $\mathcal{S}_{1n\mathbf{j},221}=o_p\left(1\right).$ For $\mathcal{S}_{1n\mathbf{j},222}$, we verify that $\mathcal{S}_{1n\mathbf{j},222}=O_p(n^{1/2}h^{d/2}h_0^{p+1})=o_p\left(1\right).$ Consequently, $\mathcal{S}_{1n\mathbf{j},22}=o_p\left(1\right).$ This concludes the proof of the lemma. ■

Proof of Lemma B.5. By a geometric expansion: $\tilde{f}_i = f^{-1}-(\tilde{f}_i-f)/f^2+(\tilde{f}_i-f)^2/(f^2\tilde{f}_i)$ where $\tilde{f}_i=\tilde{f}_i\left(\bar{\varepsilon}_i\right)$, we have

$$\begin{aligned}\mathcal{S}_{2n\mathbf{j}} &= \frac{1}{\sqrt{nh^d}}\sum_{i=1}^{n}K_{ix,\mathbf{j}}\left\{\tilde{G}_i\left[\tilde{q}_{2,i}\left(\boldsymbol{\beta}^0\right)-q_{2,i}\left(\boldsymbol{\beta}^0\right)\right]\right\}\\ &= -\frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n}K_{ix,\mathbf{j}}\varphi_i\frac{\tilde{f}_i'\left(\bar{\varepsilon}_i\right)-f'\left(\bar{\varepsilon}_i\right)}{f\left(\bar{\varepsilon}_i\right)}\bar{\varepsilon}_i\tilde{G}_i\\ &\quad +\frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n}K_{ix,\mathbf{j}}\varphi_i\frac{\tilde{f}_i'\left(\bar{\varepsilon}_i\right)\left[\tilde{f}_i\left(\bar{\varepsilon}_i\right)-f\left(\bar{\varepsilon}_i\right)\right]}{f\left(\bar{\varepsilon}_i\right)}\bar{\varepsilon}_i\tilde{G}_i\\ &\quad -\frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n}K_{ix,\mathbf{j}}\varphi_i\frac{\tilde{f}_i'\left(\bar{\varepsilon}_i\right)\left[\tilde{f}_i\left(\bar{\varepsilon}_i\right)-f\left(\bar{\varepsilon}_i\right)\right]^2}{f^2\left(\bar{\varepsilon}_i\right)\tilde{f}_i\left(\bar{\varepsilon}_i\right)}\bar{\varepsilon}_i\tilde{G}_i\\ &\equiv -\mathcal{S}_{2n\mathbf{j},1}+\mathcal{S}_{2n\mathbf{j},2}-\mathcal{S}_{2n\mathbf{j},3}.\end{aligned}$$

where recall $\varphi_i\equiv\varphi'(P_i\left(\boldsymbol{\beta}_2^0\right))/\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))$. It suffices to show that each of these three terms is $o_p\left(1\right).$ For $\mathcal{S}_{2n\mathbf{j},1}$, noticing that $G_b(\tilde{f}_i)-G_b\left(f_i\right)=g_b\left(f_i\right)(\tilde{f}_i-f_i)+\frac{1}{2}g_b'\left(f_i^*\right)(\tilde{f}_i-f_i)^2$, we can apply Lemma A.2

and show that

$$\begin{aligned}
\mathcal{S}_{2n\mathbf{j},1} &= \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n} K_{ix,\mathbf{j}}\varphi_i \frac{\tilde{f}_i'(\bar{\varepsilon}_i)-\overline{f}'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\bar{\varepsilon}_i G_b(f_i) \\
&\quad + \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n} K_{ix,\mathbf{j}}\varphi_i \frac{\overline{f}'(\bar{\varepsilon}_i)-f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\bar{\varepsilon}_i G_b(f_i) + o_p(1) \\
&\equiv \mathcal{S}_{2n\mathbf{j},11}+\mathcal{S}_{2n\mathbf{j},12}+o_p(1).
\end{aligned}$$

For the first term, we have

$$\begin{aligned}
\mathcal{S}_{2n\mathbf{j},11} &= \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n} \frac{K_{ix,\mathbf{j}}\varphi_i}{f(\bar{\varepsilon}_i)}\frac{1}{nh_0^2}\sum_{j\neq i}\left[k_0'\left(\frac{\bar{\varepsilon}_i-\tilde{\varepsilon}_j}{h_0}\right)-k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\right]\bar{\varepsilon}_i G_b(f_i) \\
&= \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n} \frac{K_{ix,\mathbf{j}}\varphi_i}{f(\bar{\varepsilon}_i)}\frac{1}{nh_0^3}\sum_{j\neq i}k_0''\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)(\tilde{\varepsilon}_j-\varepsilon_j)\bar{\varepsilon}_i G_b(f_i)+o_p(1) \\
&= \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n} \frac{K_{ix,\mathbf{j}}\varphi_i}{f(\bar{\varepsilon}_i)}\frac{1}{nh_0^3}\sum_{j\neq i}k_0''\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\frac{m(X_j)-\tilde{m}(X_j)}{\sigma(X_j)}\bar{\varepsilon}_i G_b(f_i) \\
&\quad + \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n} \frac{K_{ix,\mathbf{j}}\varphi_i}{f(\bar{\varepsilon}_i)}\frac{1}{nh_0^3}\sum_{j\neq i}k_0''\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\varepsilon_j\frac{\sigma(X_j)-\tilde{\sigma}(X_j)}{\sigma(X_j)}\bar{\varepsilon}_i G_b(f_i)+o_p(1) \\
&\equiv \mathcal{S}_{2n\mathbf{j},111}+\mathcal{S}_{2n\mathbf{j},112}+o_p(1), \text{ say.}
\end{aligned}$$

Write

$$\begin{aligned}
\mathcal{S}_{2n\mathbf{j},111} &= \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n} \frac{K_{ix,\mathbf{j}}\varphi_i}{f(\bar{\varepsilon}_i)}\frac{1}{nh_0^3}\sum_{j\neq i}k_0''\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\frac{e_1^\top M_n^{-1}(X_j)U_{1,n}(X_j)}{\sigma(X_j)}\bar{\varepsilon}_i G_b(f_i) \\
&\quad + \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n} \frac{K_{ix,\mathbf{j}}\varphi_i}{f(\bar{\varepsilon}_i)}\frac{1}{nh_0^3}\sum_{j\neq i}k_0''\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\frac{e_1^\top M_n^{-1}(X_j)B_{1,n}(X_j)}{\sigma(X_j)}\bar{\varepsilon}_i G_b(f_i) \\
&\equiv \mathcal{S}_{2n\mathbf{j},111a}+\mathcal{S}_{2n\mathbf{j},111b}.
\end{aligned}$$

Writing $\mathcal{S}_{2n\mathbf{j},111a}$ as a third order U-statistic, we can show that $\mathcal{S}_{2n\mathbf{j},111a}=O_p(h^{d/2})=o_p(1)$ by conditional moment calculations and conditional Chebyshev inequality. For $\mathcal{S}_{2n\mathbf{j},111b}$, we have $\mathcal{S}_{2n\mathbf{j},111b}=O_p(\sqrt{nh^d}h_1^{p+1})=o_p(1)$. Similarly, we can verify that $\mathcal{S}_{2n\mathbf{j},112}=o_p(1)$. Consequently $\mathcal{S}_{2n\mathbf{j},11}=o_p(1)$. For $\mathcal{S}_{2n\mathbf{j},12}$, we have

$$\begin{aligned}
\mathcal{S}_{2n\mathbf{j},12} &= \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n} \frac{K_{ix,\mathbf{j}}\varphi_i}{f(\bar{\varepsilon}_i)}\frac{\overline{f}'(\bar{\varepsilon}_i)-f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\bar{\varepsilon}_i G_b(f_i) \\
&= \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n} \frac{K_{ix,\mathbf{j}}\varphi_i}{f(\bar{\varepsilon}_i)}\left\{\frac{1}{nh_0^2}\sum_{j\neq i}\left\{k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)-E_j\left[k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\right]\right\}\right\}\bar{\varepsilon}_i G_b(f_i) \\
&\quad + \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n} \frac{K_{ix,\mathbf{j}}\varphi_i}{f(\bar{\varepsilon}_i)}\left\{\frac{1}{nh_0^2}\sum_{j\neq i}E_j\left[k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\right]-f'(\bar{\varepsilon}_i)\right\}\bar{\varepsilon}_i G_b(f_i) \\
&= \mathcal{S}_{2n\mathbf{j},121}+\mathcal{S}_{2n\mathbf{j},122},
\end{aligned}$$

where E_j indicates expectation with respect to the variable indexed by j. Noting $\mathcal{S}_{2n\mathbf{j},121}$ is a second order statistic, it is easy to verify that $E[\mathcal{S}_{2n\mathbf{j},121}]^2=O(h^d)=o(1)$, implying that $\mathcal{S}_{2n\mathbf{j},121}=o_p(1)$. For

$\mathcal{S}_{2n\mathbf{j},122}$, noticing that

$$\frac{1}{nh_0^2}\sum_{j\neq i}E_j\left[k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\right]-f'\left(\bar{\varepsilon}_i\right)=h_0^{p+1}f^{(p+2)}\left(\bar{\varepsilon}_i\right)\int k_0\left(u\right)u^{p+1}du,$$

we can show that $\mathcal{S}_{2n\mathbf{j},122}=O_p(\sqrt{nh^d}h_0^{p+1})=o_p\left(1\right).$ Consequently, $\mathcal{S}_{2n\mathbf{j},12}=o_p\left(1\right)$ and $\mathcal{S}_{2n\mathbf{j},1}=o_p\left(1\right).$

For $\mathcal{S}_{2n\mathbf{j},2}$, we can easily show that

$$\mathcal{S}_{2n\mathbf{j},2}=\frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n}\frac{K_{ix,\mathbf{j}}\varphi_i f'\left(\bar{\varepsilon}_i\right)}{f\left(\bar{\varepsilon}_i\right)}\left[\tilde{f}_i\left(\bar{\varepsilon}_i\right)-f\left(\bar{\varepsilon}_i\right)\right]\bar{\varepsilon}_i G_b\left(f_i\right)+o_p\left(1\right).$$

The rest of the proof is similar to that of $\mathcal{S}_{2n\mathbf{j},1}$ and thus omitted. For $\mathcal{S}_{2n\mathbf{j},3}$, by Lemma A.2, $\mathcal{S}_{2n\mathbf{j},3}=O_p(\sqrt{nh^d}b^{-2}\nu_{3n}^2)=o_p\left(1\right).$ This concludes the proof of the lemma. ■

Proof of Lemma B.6. Write $\mathcal{S}_{3n\mathbf{j}}=\frac{1}{2}\{\mathcal{S}_{3n\mathbf{j},1}-\mathcal{S}_{3n\mathbf{j},2}+\mathcal{S}_{3n\mathbf{j},3}-\mathcal{S}_{3n\mathbf{j},4}\}$, where

$$\begin{aligned}
\mathcal{S}_{3n\mathbf{j},1} &\equiv \frac{1}{\sqrt{nh^d}}\sum_{i=1}^{n}K_{ix,\mathbf{j}}\left\{\log\left(\tilde{f}_i\left(\overrightarrow{\varepsilon}_i\right)\right)g_b\left(\tilde{f}_i\left(\overrightarrow{\varepsilon}_i\right)\right)\tilde{f}_i'\left(\overrightarrow{\varepsilon}_i\right)\overrightarrow{\varepsilon}_i\varphi_i(\tilde{\boldsymbol{\beta}}_2)\right.\\
&\quad\left.-\log\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right)g_b\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right)\tilde{f}_i'\left(\bar{\varepsilon}_i\right)\bar{\varepsilon}_i\varphi_i\right\},\\
\mathcal{S}_{3n\mathbf{j},2} &\equiv \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n}K_{ix,\mathbf{j}}\left\{\log\varphi\left(P_i(\tilde{\boldsymbol{\beta}}_2)\right)g_b\left(\tilde{f}_i\left(\overrightarrow{\varepsilon}_i\right)\right)\tilde{f}_i'\left(\overrightarrow{\varepsilon}_i\right)\overrightarrow{\varepsilon}_i\varphi_i(\tilde{\boldsymbol{\beta}}_2)\right.\\
&\quad\left.-\log\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)g_b\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right)\tilde{f}_i'\left(\bar{\varepsilon}_i\right)\bar{\varepsilon}_i\varphi_i\right\},\\
\mathcal{S}_{3n\mathbf{j},3} &\equiv \frac{1}{\sqrt{nh^d}}\sum_{i=1}^{n}K_{ix,\mathbf{j}}\left\{\log\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right)g_b\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right)\tilde{f}_i'\left(\bar{\varepsilon}_i\right)-\log\left(f\left(\bar{\varepsilon}_i\right)\right)g_b\left(f\left(\bar{\varepsilon}_i\right)\right)f'\left(\bar{\varepsilon}_i\right)\right\}\bar{\varepsilon}_i\varphi_i,\\
\mathcal{S}_{3n\mathbf{j},4} &\equiv \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n}K_{ix,\mathbf{j}}\log\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)\left\{g_b\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right)\tilde{f}_i'\left(\bar{\varepsilon}_i\right)-g_b\left(f\left(\bar{\varepsilon}_i\right)\right)f'\left(\bar{\varepsilon}_i\right)\right\}\bar{\varepsilon}_i\varphi_i.
\end{aligned}$$

where $\varphi_i\left(\boldsymbol{\beta}_2\right)\equiv\varphi'(P_i\left(\boldsymbol{\beta}_2\right))/\varphi(P_i\left(\boldsymbol{\beta}_2\right))$ and $\varphi_i=\varphi_i\left(\boldsymbol{\beta}_2^0\right).$ We will only show that $\mathcal{S}_{3n\mathbf{j},1}=o_p\left(1\right)$ since the proofs of $\mathcal{S}_{3n\mathbf{j},s}=o_p\left(1\right)$ for $s=2,3,4$ are similar.

For $\mathcal{S}_{3n\mathbf{j},1}$, noticing that $\tilde{v}_{2i}\left(x\right)=(\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{1/2}-\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2})/\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2}$ and $\tilde{v}_{1i}\left(x\right)=v_{1i}\left(x\right)/\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2}$ are both $O_p\left(\upsilon_{2n}\right)$ uniformly in i on the set $\{K_{ix}>0\}$, and $\overrightarrow{\varepsilon}_i-\bar{\varepsilon}_i=\bar{\varepsilon}_i\tilde{v}_{2i}\left(x\right)-\tilde{v}_{1i}\left(x\right),$ we can show that

$$\begin{aligned}
\mathcal{S}_{3n\mathbf{j},1} &= \frac{1}{\sqrt{nh^d}}\sum_{i=1}^{n}K_{ix,\mathbf{j}}\varphi_i\left\{\tilde{\psi}_i\left(\bar{\varepsilon}_i\right)g_b\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right)\tilde{f}_i'\left(\bar{\varepsilon}_i\right)\bar{\varepsilon}_i\left\{\bar{\varepsilon}_i\tilde{v}_{2i}\left(x\right)-\tilde{v}_{1i}\left(x\right)\right\}\right.\\
&\quad+\frac{1}{\sqrt{nh^d}}\sum_{i=1}^{n}K_{ix,\mathbf{j}}\varphi_i\log\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right)g_b'\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right)\tilde{f}_i'\left(\bar{\varepsilon}_i\right)\bar{\varepsilon}_i\left\{\bar{\varepsilon}_i\tilde{v}_{2i}\left(x\right)-\tilde{v}_{1i}\left(x\right)\right\}\\
&\quad+\frac{1}{\sqrt{nh^d}}\sum_{i=1}^{n}K_{ix,\mathbf{j}}\varphi_i\log\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right)g_b\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right)\tilde{f}_i''\left(\bar{\varepsilon}_i\right)\bar{\varepsilon}_i\left\{\bar{\varepsilon}_i\tilde{v}_{2i}\left(x\right)-\tilde{v}_{1i}\left(x\right)\right\}+o_p\left(1\right)\\
&\equiv \mathcal{S}_{3n\mathbf{j},11}+\mathcal{S}_{3n\mathbf{j},12}+\mathcal{S}_{3n\mathbf{j},13}+o_p\left(1\right).
\end{aligned}$$

By Lemma A.1, we can show

$$\mathcal{S}_{3n\mathbf{j},11} \simeq \frac{1}{\sqrt{nh^d}}\sum_{i=1}^{n} K_{ix,\mathbf{j}}\varphi_i\psi\left(\bar{\varepsilon}_i\right) g_b\left(f\left(\bar{\varepsilon}_i\right)\right) f'\left(\bar{\varepsilon}_i\right)\bar{\varepsilon}_i\left\{\bar{\varepsilon}_i\tilde{v}_{2i}\left(x\right)-\tilde{v}_{1i}\left(x\right)\right\}, \tag{C.32}$$

$$\mathcal{S}_{3n\mathbf{j},12} \simeq \frac{1}{\sqrt{nh^d}}\sum_{i=1}^{n} K_{ix,\mathbf{j}}\varphi_i\log\left(f\left(\bar{\varepsilon}_i\right)\right) g'_b\left(f\left(\bar{\varepsilon}_i\right)\right) f'\left(\bar{\varepsilon}_i\right)\bar{\varepsilon}_i\left\{\bar{\varepsilon}_i\tilde{v}_{2i}\left(x\right)-\tilde{v}_{1i}\left(x\right)\right\}, \tag{C.33}$$

$$\mathcal{S}_{3n\mathbf{j},13} \simeq \frac{1}{\sqrt{nh^d}}\sum_{i=1}^{n} K_{ix,\mathbf{j}}\varphi_i\log\left(f\left(\bar{\varepsilon}_i\right)\right) g_b\left(f\left(\bar{\varepsilon}_i\right)\right) f''\left(\bar{\varepsilon}_i\right)\bar{\varepsilon}_i\left\{\bar{\varepsilon}_i\tilde{v}_{2i}\left(x\right)-\tilde{v}_{1i}\left(x\right)\right\}. \tag{C.34}$$

The rest of the proof relies on the repeated applications of the dominated convergence arguments. For example, the right hand side of (C.32) is smaller than

$$\frac{1}{\sqrt{nh^d}}\max_{\{K_{ix}>0\}}\left|\tilde{v}_{2i}\left(x\right)\right|\sum_{i=1}^{n}\left|K_{ix,\mathbf{j}}\varphi_i\psi\left(\bar{\varepsilon}_i\right) g_b\left(f\left(\bar{\varepsilon}_i\right)\right) f'\left(\bar{\varepsilon}_i\right)\bar{\varepsilon}_i^2\right|$$
$$+\frac{1}{\sqrt{nh^d}}\max_{\{K_{ix}>0\}}\left|\tilde{v}_{1i}\left(x\right)\right|\sum_{i=1}^{n}\left|K_{ix,\mathbf{j}}\varphi_i\psi\left(\bar{\varepsilon}_i\right) g_b\left(f\left(\bar{\varepsilon}_i\right)\right) f'\left(\bar{\varepsilon}_i\right)\bar{\varepsilon}_i\right|.$$

Noting that

$$\begin{aligned}
E\left|K_{ix,\mathbf{j}}\varphi_i\psi\left(\bar{\varepsilon}_i\right) g_b\left(f\left(\bar{\varepsilon}_i\right)\right) f'\left(\bar{\varepsilon}_i\right)\bar{\varepsilon}_i^r\right| &= E\left[\left|\frac{K_{ix,\mathbf{j}}\varphi_i}{d_i+1}\int\psi\left(\varepsilon\right) g_b\left(f\left(\varepsilon\right)\right) f'\left(\varepsilon\right)\varepsilon^r\right| f\left(\frac{\varepsilon-\overline{\delta}_i}{d_i+1}\right)d\varepsilon\right] \\
&\leq \sup_{\varepsilon}\left[g_b\left(f\left(\varepsilon\right)\right) f\left(\varepsilon\right)\right] E\left|\frac{K_{ix,\mathbf{j}}\varphi_i}{d_i+1}\right|\int_{b\leq f(\varepsilon)\leq 2b}\frac{f'\left(\varepsilon\right)^2}{f\left(\varepsilon\right)}\left|\varepsilon^r\right|d\varepsilon+O\left(h^{\epsilon}\right) \\
&\leq Cf\left(\varepsilon\right)\int_{b\leq f(\varepsilon)\leq 2b}\left|\frac{f'\left(\varepsilon\right)^2}{f\left(\varepsilon\right)}\varepsilon^r\right|d\varepsilon+O\left(h^{\epsilon}\right) \\
&\leq C\int_{b\leq f(\varepsilon)\leq 2b}\left|\frac{f'\left(\varepsilon\right)^2}{f\left(\varepsilon\right)}\varepsilon^r\right|d\varepsilon+O\left(h^{p+1}\right)=O\left(b^{(\gamma-1)/(2\gamma)}+h^{\epsilon}\right),
\end{aligned}$$

where the last equality follows from similar argument to the proof of Lemma B.3, we have $\mathcal{S}_{3n\mathbf{j},11} = O_p(\upsilon_{2n}\sqrt{nh^d}(b^{(\gamma-1)/(2\gamma)}+h^{\epsilon})) = o_p\left(1\right)$. Similarly, we can show that $\mathcal{S}_{3n\mathbf{j},1s} = o_p\left(1\right)$, $s=2,3$. ■

Proof of Lemma B.7. Observe that

$$\begin{aligned}
\mathcal{R}_{1n} &= \frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right)\bar{H}^{-1}[\tilde{G}_i\tilde{s}_i\left(\boldsymbol{\beta}^0\right)\tilde{s}_i\left(\boldsymbol{\beta}^0\right)^{\top}-G_is_i\left(\boldsymbol{\beta}^0\right)s_i\left(\boldsymbol{\beta}^0\right)^{\top}]\otimes(\tilde{\mathbf{X}}_i\tilde{\mathbf{X}}_i^{\top})\bar{H}^{-1} \\
&= \frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right)\bar{H}^{-1}G_i[\tilde{s}_i\left(\boldsymbol{\beta}^0\right)\tilde{s}_i\left(\boldsymbol{\beta}^0\right)^{\top}-s_i\left(\boldsymbol{\beta}^0\right)s_i\left(\boldsymbol{\beta}^0\right)^{\top}]\otimes(\tilde{\mathbf{X}}_i\tilde{\mathbf{X}}_i^{\top})\bar{H}^{-1} \\
&\quad+\frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right)\bar{H}^{-1}\left(\tilde{G}_i-G_i\right)s_i\left(\boldsymbol{\beta}^0\right)s_i\left(\boldsymbol{\beta}^0\right)^{\top}\otimes(\tilde{\mathbf{X}}_i\tilde{\mathbf{X}}_i^{\top})\bar{H}^{-1} \\
&\quad+\frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right)\bar{H}^{-1}\left(\tilde{G}_i-G_i\right)[\tilde{s}_i\left(\boldsymbol{\beta}^0\right)\tilde{s}_i\left(\boldsymbol{\beta}^0\right)^{\top}-s_i\left(\boldsymbol{\beta}^0\right)s_i\left(\boldsymbol{\beta}^0\right)^{\top}]\otimes(\tilde{\mathbf{X}}_i\tilde{\mathbf{X}}_i^{\top})\bar{H}^{-1} \\
&\equiv \mathcal{R}_{1n,1}+\mathcal{R}_{1n,2}+\mathcal{R}_{1n,3},\ \text{say.}
\end{aligned}$$

It suffices to prove the lemma by showing that $\mathcal{R}_{1n,r} = o_p\left(1\right)$ for $r=1,2,3$. We only prove $\mathcal{R}_{1n,1} = o_p\left(1\right)$ and $\mathcal{R}_{1n,2} = o_p\left(1\right)$ as $\mathcal{R}_{1n,3}$ is a smaller order term and can be studied analogously.

First, we show that $\mathcal{R}_{1n,1} = o_p(1)$. Note that

$$\tilde{s}_i\left(\boldsymbol{\beta}^0\right)\tilde{s}_i\left(\boldsymbol{\beta}^0\right)^\top = \begin{bmatrix} \frac{\tilde{\psi}_i^2(\bar{\varepsilon}_i)}{\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)} & \frac{\varphi'(P_i(\boldsymbol{\beta}_2))\tilde{\psi}_i(\bar{\varepsilon}_i)\left[\tilde{\psi}_i(\bar{\varepsilon}_i)\bar{\varepsilon}_i+1\right]}{2\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{3/2}} \\ \frac{\varphi'(P_i(\boldsymbol{\beta}_2))\tilde{\psi}_i(\bar{\varepsilon}_i)\left[\tilde{\psi}_i(\bar{\varepsilon}_i)\bar{\varepsilon}_i+1\right]}{2\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{3/2}} & \frac{\left[\varphi'(P_i(\boldsymbol{\beta}_2))\right]^2\left[\tilde{\psi}_i(\bar{\varepsilon}_i)\bar{\varepsilon}_i+1\right]^2}{4\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^2} \end{bmatrix},$$

and $s_i\left(\boldsymbol{\beta}^0\right)s_i\left(\boldsymbol{\beta}^0\right)^\top$ has a similar expression with $\psi(\bar{\varepsilon}_i)$ in the place of $\tilde{\psi}_i(\bar{\varepsilon}_i)$. It follows that

$$\begin{aligned}
\mathcal{R}_{1n,1} &= \frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)G_i \\
&\quad\times\begin{bmatrix} \frac{\tilde{\psi}_i^2(\bar{\varepsilon}_i)-\psi^2(\bar{\varepsilon}_i)}{\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)} & \frac{\varphi_i\left\{\tilde{\psi}_i(\bar{\varepsilon}_i)\left[\tilde{\psi}_i(\bar{\varepsilon}_i)\bar{\varepsilon}_i+1\right]-\psi(\bar{\varepsilon}_i)\left[\psi(\bar{\varepsilon}_i)\bar{\varepsilon}_i+1\right]\right\}}{2\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{1/2}} \\ \frac{\varphi_i\left\{\tilde{\psi}_i(\bar{\varepsilon}_i)\left[\tilde{\psi}_i(\bar{\varepsilon}_i)\bar{\varepsilon}_i+1\right]-\psi(\bar{\varepsilon}_i)\left[\psi(\bar{\varepsilon}_i)\bar{\varepsilon}_i+1\right]\right\}}{2\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{1/2}} & \frac{\varphi_i^2\left[\tilde{\psi}_i(\bar{\varepsilon}_i)\bar{\varepsilon}_i+1\right]^2-\left[\psi(\bar{\varepsilon}_i)\bar{\varepsilon}_i+1\right]^2}{4} \end{bmatrix}\otimes\left(\mathbf{Z}_i\mathbf{Z}_i^\top\right) \\
&\equiv \begin{bmatrix} \mathcal{R}_{1n,1,11} & \mathcal{R}_{1n,1,12} \\ \mathcal{R}_{1n,1,21} & \mathcal{R}_{1n,1,22} \end{bmatrix}, \text{ say,}
\end{aligned}$$

where recall $\varphi_i = \varphi'\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)/\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)$, $\mathcal{R}_{1n,1,21} = \mathcal{R}_{1n,1,12}^\top$, and $\mathcal{R}_{1n,1,rs}$, $r,s = 1,2$, are all $N \times N$ matrices. We need to show that $\mathcal{R}_{1n,1,11}$, $\mathcal{R}_{1n,1,12}$ and $\mathcal{R}_{1n,1,22}$ are all $o_p(1)$. Noting that

$$\begin{aligned}
\tilde{\psi}_i^2(\bar{\varepsilon}_i) - \psi^2(\bar{\varepsilon}_i) &= \frac{\tilde{f}_i'(\bar{\varepsilon}_i)^2 f(\bar{\varepsilon}_i)^2 - \tilde{f}_i(\bar{\varepsilon}_i)^2 f'(\bar{\varepsilon}_i)^2}{\tilde{f}_i(\bar{\varepsilon}_i)^2 f(\bar{\varepsilon}_i)^2} \\
&= \frac{\left[\tilde{f}_i'(\bar{\varepsilon}_i)^2 - f_i'(\bar{\varepsilon}_i)^2\right] f(\bar{\varepsilon}_i)^2 + \left[f(\bar{\varepsilon}_i)^2 - \tilde{f}_i(\bar{\varepsilon}_i)^2\right] f_i'(\bar{\varepsilon}_i)^2}{\tilde{f}_i(\bar{\varepsilon}_i)^2 f(\bar{\varepsilon}_i)^2},
\end{aligned}$$

we have

$$\begin{aligned}
\mathcal{R}_{1n,1,11} &= \frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)G_i\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{-1}\left[\tilde{\psi}_i^2(\bar{\varepsilon}_i) - \psi^2(\bar{\varepsilon}_i)\right]\mathbf{Z}_i\mathbf{Z}_i^\top \\
&= \frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)G_i\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{-1}\tilde{f}_i(\bar{\varepsilon}_i)^{-2}\left[\tilde{f}_i'(\bar{\varepsilon}_i)^2 - f_i'(\bar{\varepsilon}_i)^2\right]\mathbf{Z}_i\mathbf{Z}_i^\top \\
&\quad + \frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)G_i\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{-1}\tilde{f}_i(\bar{\varepsilon}_i)^{-2}\psi(\bar{\varepsilon}_i)^2\left[\tilde{f}_i(\bar{\varepsilon}_i)^2 - f_i(\bar{\varepsilon}_i)^2\right]\mathbf{Z}_i\mathbf{Z}_i^\top \\
&\equiv \mathcal{R}_{1n,1,11,a} + \mathcal{R}_{1n,1,11,b} \text{ say.}
\end{aligned}$$

Noting that $G_i\tilde{f}_i(\bar{\varepsilon}_i)^{-2} = O\left(b^{-2}\right)$, by Lemma A.2, we have

$$\begin{aligned}
\left\|\mathcal{R}_{1n,1,11,a}\right\| &\le O_p\left(\upsilon_{3n,1}b^{-2}\right)\frac{1}{nh^d}\sum_{i=1}^n\left\|K\left(\frac{x-X_i}{h}\right)G_i\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{-1}\mathbf{Z}_i\mathbf{Z}_i^\top\right\| \\
&= O_p\left(\upsilon_{3n,1}b^{-2}\right)O_p(1) = O_p\left(\upsilon_{3n,1}b^{-2}\right) = o_p(1).
\end{aligned}$$

By the same token, $|\mathcal{R}_{1n,1,11,b}| = o_p(1)$. Thus $\mathcal{R}_{1n,1,11} = o_p(1)$. Analogously, we can show $\mathcal{R}_{1n,1,12} = o_p(1)$ and $\mathcal{R}_{1n,1,22} = o_p(1)$. Hence we have shown that $\mathcal{R}_{1n,1} = o_p(1)$.

Now, we show that $\mathcal{R}_{1n,2} = o_p(1)$. By (C.26) and Markov inequality, we have

$$\begin{aligned}
|\mathcal{R}_{1n,2}| &\le O_p\left(h^\epsilon\right)\frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)\left\|s_i\left(\boldsymbol{\beta}^0\right)s_i\left(\boldsymbol{\beta}^0\right)^\top\otimes\left(\mathbf{Z}_i\mathbf{Z}_i^\top\right)\right\| \\
&= O_p\left(h^\epsilon\right)O_p(1) = O_p(1).
\end{aligned}$$

This completes the proof of the lemma. ■

Proof of Lemma B.8. Observe that

$$
\begin{aligned}
\mathcal{R}_{2n} &= \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\bar{H}^{-1}\left[\tilde{G}_i\frac{\partial \tilde{s}_i\left(\boldsymbol{\beta}^0\right)}{\partial \boldsymbol{\beta}^{\top}} - G_i\frac{\partial s_i\left(\boldsymbol{\beta}^0\right)}{\partial \boldsymbol{\beta}^{\top}}\right]\otimes \tilde{\mathbf{X}}_i\bar{H}^{-1} \\
&= \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\bar{H}^{-1}G_i\left\{\left[\frac{\partial \tilde{s}_i\left(\boldsymbol{\beta}^0\right)}{\partial \boldsymbol{\beta}^{\top}} - \frac{\partial s_i\left(\boldsymbol{\beta}^0\right)}{\partial \boldsymbol{\beta}^{\top}}\right]\otimes \tilde{\mathbf{X}}_i\right\}\bar{H}^{-1} \\
&\quad + \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\bar{H}^{-1}\left(\tilde{G}_i - G_i\right)\left\{\frac{\partial s_i\left(\boldsymbol{\beta}^0\right)}{\partial \boldsymbol{\beta}^{\top}}\otimes \tilde{\mathbf{X}}_i\right\}\bar{H}^{-1} \\
&\quad + \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\bar{H}^{-1}\left(\tilde{G}_i - G_i\right)\left\{\left[\frac{\partial \tilde{s}_i\left(\boldsymbol{\beta}^0\right)}{\partial \boldsymbol{\beta}^{\top}} - \frac{\partial s_i\left(\boldsymbol{\beta}^0\right)}{\partial \boldsymbol{\beta}^{\top}}\right]\otimes \tilde{\mathbf{X}}_i\right\}\bar{H}^{-1} \\
&\equiv \mathcal{R}_{2n,1} + \mathcal{R}_{2n,2} + \mathcal{R}_{2n,3}, \text{ say.}
\end{aligned}
$$

We prove the lemma by showing that $\mathcal{R}_{2n,s} = o_P(1)$ for $s = 1, 2, 3$. We will only show that $\mathcal{R}_{2n,1} = o_P(1)$ as the other two cases can be proved analogously. Recall $c_{i\varphi} = \varphi'\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^2 - \varphi''\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)$ and $\varphi_i \equiv \varphi'\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)/\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)$. Noting that

$$
\frac{\partial \tilde{s}_i\left(\boldsymbol{\beta}^0\right)}{\partial \boldsymbol{\beta}^{\top}} = \begin{pmatrix} \frac{\tilde{\psi}_i'(\bar{\varepsilon}_i)}{\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)} & \frac{\varphi_i\left[\tilde{\psi}_i'(\bar{\varepsilon}_i)\bar{\varepsilon}_i + \tilde{\psi}_i(\bar{\varepsilon}_i)\right]}{2\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{1/2}} \\ \frac{\varphi_i\left[\tilde{\psi}_i'(\bar{\varepsilon}_i)\bar{\varepsilon}_i + \tilde{\psi}_i(\bar{\varepsilon}_i)\right]}{2\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{1/2}} & \frac{2c_{i\varphi}\left[\tilde{\psi}_i(\bar{\varepsilon}_i)\bar{\varepsilon}_i + 1\right] + \varphi'\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^2\bar{\varepsilon}_i\left[\tilde{\psi}_i'(\bar{\varepsilon}_i)\bar{\varepsilon}_i + \tilde{\psi}_i(\bar{\varepsilon}_i)\right]}{4\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^2} \end{pmatrix} \otimes \tilde{\mathbf{X}}_i^{\top},
$$

and $\partial s_i\left(\boldsymbol{\beta}^0\right)/\partial \boldsymbol{\beta}^{\top}$ has similar expression with $\psi_i(\bar{\varepsilon}_i)$ in the place of $\tilde{\psi}_i(\bar{\varepsilon}_i)$, we have

$$
\begin{aligned}
\mathcal{R}_{2n,1} &= \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)G_i \\
&\quad \times \begin{pmatrix} \frac{\tilde{\psi}_i'(\bar{\varepsilon}_i) - \psi'(\bar{\varepsilon}_i)}{\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)} & \frac{\varphi_i\left\{\left[\tilde{\psi}_i'(\bar{\varepsilon}_i) - \psi'(\bar{\varepsilon}_i)\right]\bar{\varepsilon}_i + \left[\tilde{\psi}_i(\bar{\varepsilon}_i) - \psi(\bar{\varepsilon}_i)\right]\right\}}{2\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{1/2}} \\ \frac{\varphi_i\left\{\left[\tilde{\psi}_i'(\bar{\varepsilon}_i) - \psi'(\bar{\varepsilon}_i)\right]\bar{\varepsilon}_i + \left[\tilde{\psi}_i(\bar{\varepsilon}_i) - \psi(\bar{\varepsilon}_i)\right]\right\}}{2\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{1/2}} & \frac{2c_{i\varphi}\left[\tilde{\psi}_i(\bar{\varepsilon}_i) - \psi(\bar{\varepsilon}_i)\right]\bar{\varepsilon}_i}{4\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^2} + \frac{\tilde{d}_i}{4} \end{pmatrix} \otimes \left(\mathbf{Z}_i\mathbf{Z}_i^{\top}\right) \\
&\equiv \begin{bmatrix} \mathcal{R}_{2n,1,11} & \mathcal{R}_{2n,1,12} \\ \mathcal{R}_{2n,1,12}^{\top} & \mathcal{R}_{2n,1,22} \end{bmatrix}, \text{ say,}
\end{aligned}
$$

where $\tilde{d}_i \equiv \varphi_i^2\bar{\varepsilon}_i[\tilde{\psi}_i'(\bar{\varepsilon}_i) - \psi'(\bar{\varepsilon}_i)]\bar{\varepsilon}_i + [\tilde{\psi}_i(\bar{\varepsilon}_i) - \psi(\bar{\varepsilon}_i)]$. As in the analysis of $\mathcal{R}_{1n,1}$, using Lemma A.2, we can readily demonstrate that $\mathcal{R}_{2n,1,11} = o_p(1)$, $\mathcal{R}_{2n,1,12} = o_p(1)$ and $\mathcal{R}_{2n,1,22} = o_p(1)$. It follows that $\mathcal{R}_{2n,1} = o_p(1)$. Similarly, we can show that $\mathcal{R}_{2n,s} = o_P(1)$ for $s = 2, 3$. This completes the proof of the lemma. ■

D Derivative Matrices in the Proof of Proposition 2.1

In this appendix, we give explicit expressions for the elements of some derivative matrices of the log-likelihood function defined in the proof of Proposition 2.1. The elements of the Hessian matrix are

$$\begin{aligned}
q_{11}(y;\beta_1,\beta_2) &= \frac{\partial^2 \log f(\varepsilon(\beta))}{\partial\varepsilon^2}\frac{1}{\varphi(\beta_2)},\\
q_{12}(y;\beta_1,\beta_2) &= \left\{\frac{\partial^2 \log f(\varepsilon(\beta))}{\partial\varepsilon^2}\varepsilon(\beta)+\frac{\partial \log f(\varepsilon(\beta))}{\partial\varepsilon}\right\}\frac{\varphi'(\beta_2)}{2\varphi(\beta_2)^{3/2}},\\
q_{22}(y;\beta_1,\beta_2) &= \frac{-\varphi''(\beta_2)}{2\varphi(\beta_2)}\left[\frac{\partial \log f(\varepsilon(\beta))}{\partial\varepsilon}\varepsilon(\beta)+1\right]\\
&\quad+\frac{\varphi'(\beta_2)^2}{4\varphi(\beta_2)^2}\left\{\frac{\partial^2 \log f(\varepsilon(\beta))}{\partial\varepsilon^2}\varepsilon(\beta)^2+3\frac{\partial \log f(\varepsilon(\beta))}{\partial\varepsilon}\varepsilon(\beta)+2\right\},
\end{aligned}$$

and $q_{21}(y;\beta_1,\beta_2) = q_{12}(y;\beta_1,\beta_2)$ by Young's theorem, where, e.g., $\frac{\partial^2 \log f(\varepsilon)}{\partial\varepsilon^2} = \frac{f''(\varepsilon)f(\varepsilon)-f'(\varepsilon)^2}{f^2(\varepsilon)}$ and $\frac{\partial^2 \log f(\varepsilon(\beta))}{\partial\varepsilon^2} \equiv \left.\frac{\partial^2 \log f(\varepsilon)}{\partial\varepsilon^2}\right|_{\varepsilon=\varepsilon(\beta)}$. Note that when we restrict our attention to the case $\varphi(u) = u$ or $\exp(u)$, the above formulae can be greatly simplified.

In addition, in the proof of Proposition 2.1, we also need that $q_{rst}(y;\beta_1,\beta_2) \equiv \frac{\partial^3}{\partial\beta_r\partial\beta_s\partial\beta_t}\log(f(y;\beta_1,\beta_2))$, $r,s,t = 1,2$, should be well behaved. Using the expressions

$$\frac{\partial\varepsilon(\beta)}{\partial\beta} = \begin{pmatrix}\frac{\partial\varepsilon(\beta)}{\partial\beta_1}\\ \frac{\partial\varepsilon(\beta)}{\partial\beta_2}\end{pmatrix} = \begin{pmatrix}-\frac{1}{\varphi(\beta_2)^{1/2}}\\ -\frac{\varphi'(\beta_2)}{2\varphi(\beta_2)}\varepsilon(\beta)\end{pmatrix} \text{ and } \frac{\partial^2 \log f(\varepsilon)}{\partial\varepsilon^2} = \frac{f''(\varepsilon)f(\varepsilon)-f'(\varepsilon)^2}{f^2(\varepsilon)}$$

and by straightforward calculations, we have

$$\begin{aligned}
q_{111}(y;\beta_1,\beta_2) &= \frac{\partial^3 \log f(\varepsilon(\beta))}{\partial\varepsilon^3}\frac{1}{\varphi(\beta_2)},\\
q_{112}(y;\beta_1,\beta_2) &= \frac{\partial^3 \log f(\varepsilon(\beta))}{\partial\varepsilon^3}\frac{\partial\varepsilon(\beta)}{\partial\beta_2}\frac{1}{\varphi(\beta_2)}-\frac{\partial^2 \log f(\varepsilon(\beta))}{\partial\varepsilon^2}\frac{\varphi'(\beta_2)}{\varphi(\beta_2)^2},\\
q_{121}(y;\beta_1,\beta_2) &= \left\{\frac{\partial^3 \log f(\varepsilon(\beta))}{\partial\varepsilon^3}\frac{\partial\varepsilon(\beta)}{\partial\beta_1}\varepsilon(\beta)+2\frac{\partial^2 \log f(\varepsilon(\beta))}{\partial\varepsilon^2}\frac{\partial\varepsilon(\beta)}{\partial\beta_1}\right\}\frac{\varphi'(\beta_2)}{2\varphi(\beta_2)^{3/2}} = q_{112}(y;\beta_1,\beta_2),\\
q_{122}(y;\beta_1,\beta_2) &= \left\{\frac{\partial^3 \log f(\varepsilon(\beta))}{\partial\varepsilon^3}\frac{\partial\varepsilon(\beta)}{\partial\beta_2}\varepsilon(\beta)+2\frac{\partial^2 \log f(\varepsilon(\beta))}{\partial\varepsilon^2}\frac{\partial\varepsilon(\beta)}{\partial\beta_2}\right\}\frac{\varphi'(\beta_2)}{2\varphi(\beta_2)^{3/2}},\\
&\quad+\left\{\frac{\partial^2 \log f(\varepsilon(\beta))}{\partial\varepsilon^2}\varepsilon(\beta)+\frac{\partial \log f(\varepsilon(\beta))}{\partial\varepsilon}\right\}\frac{\varphi''(\beta_2)\varphi(\beta_2)^{3/2}-\frac{3}{2}\varphi'(\beta_2)^2\varphi(\beta_2)^{1/2}}{2\varphi(\beta_2)^3},\\
q_{221}(y;\beta_1,\beta_2) &= \frac{-\varphi''(\beta_2)}{2\varphi(\beta_2)}\left[\frac{\partial^2 \log f(\varepsilon(\beta))}{\partial\varepsilon^2}\varepsilon(\beta)+\frac{\partial \log f(\varepsilon(\beta))}{\partial\varepsilon}\right]\frac{\partial\varepsilon(\beta)}{\partial\beta_1}+\frac{\varphi'(\beta_2)^2}{4\varphi(\beta_2)^2}\kappa(\beta)\frac{\partial\varepsilon(\beta)}{\partial\beta_1}\\
&= q_{122}(y;\beta_1,\beta_2)\\
q_{222}(y;\beta_1,\beta_2) &= \frac{-\varphi''(\beta_2)}{2\varphi(\beta_2)}\left[\frac{\partial^2 \log f(\varepsilon(\beta))}{\partial\varepsilon^2}\varepsilon(\beta)+\frac{\partial \log f(\varepsilon(\beta))}{\partial\varepsilon}\right]\frac{\partial\varepsilon(\beta)}{\partial\beta_2}\\
&\quad-\frac{\varphi'''(\beta_2)\varphi(\beta_2)-\varphi''(\beta_2)\varphi'(\beta_2)}{2\varphi(\beta_2)^2}\left[\frac{\partial \log f(\varepsilon(\beta))}{\partial\varepsilon}\varepsilon(\beta)+1\right]\\
&\quad+\frac{\varphi'(\beta_2)^2}{4\varphi(\beta_2)^2}\kappa(\beta)\frac{\partial\varepsilon(\beta)}{\partial\beta_2}\\
&\quad+\frac{\varphi'(\beta_2)\varphi''(\beta_2)-\varphi'(\beta_2)^3\varphi(\beta_2)}{2\varphi(\beta_2)^4}\left\{\frac{\partial^2 \log f(\varepsilon(\beta))}{\partial\varepsilon^2}\varepsilon(\beta)^2+3\frac{\partial \log f(\varepsilon(\beta))}{\partial\varepsilon}\varepsilon(\beta)+2\right\},
\end{aligned}$$

$q_{211} = q_{121} = q_{112}$, and $q_{212} = q_{122} = q_{221}$ by Young's Theorem, where $\kappa(\beta) \equiv \frac{\partial^3 \log f(\varepsilon(\beta))}{\partial \varepsilon^3} \varepsilon(\beta)^2 + 2\frac{\partial^2 \log f(\varepsilon(\beta))}{\partial \varepsilon^2} \varepsilon(\beta) + 3\frac{\partial^2 \log f(\varepsilon(\beta))}{\partial \varepsilon^2} \varepsilon(\beta) + 3\frac{\partial \log f(\varepsilon(\beta))}{\partial \varepsilon}$. Note that under our assumptions (X_i has compact support, the parameter space is compact, $\sigma^2(x)$ is bounded away from 0) the terms associated with $\varphi(\cdot)$ or its derivatives are all well behaved when $\varphi(\cdot)$ is evaluated in the neighborhood of $\beta_2^0(x)$.

REFERENCES

Bickel, P. J. (1975) One-step Huber estimates in linear models. *Journal of the American Statistical Association* 70, 428-433.

Gozalo, P. & O. Linton (2000) Local non-linear least squares: using parametric information in nonparametric regression. *Journal of Econometrics* 99, 63-106.

Hansen, B. E. (2008) Uniform convergence rates for kernel regression. *Econometric Theory*, 24, 726-748.

Lee, A. J. (1990) *U-statistics: Theory and Practice*. Marcel Dekker, New York and Basel.

Linton, O., and Z. Xiao (2007) A nonparametric regression estimator that adapts to error distribution of unknown form. *Econometric Theory* 23, 371-413.

Masry, E. (1996a) Multivariate regression estimation: local polynomial fitting for time series. *Stochastic Processes and Their Applications* 65, 81-101.

Masry, E. (1996b) Multivariate local polynomial regression for time series: uniform strong consistency rates. *Journal of Time Series Analysis* 17, 571-599.

Newey, W. K. (1991) Uniform convergence in probability and stochastic equicontinuity. *Econometrica* 59, 1161-1167.

Nolan, D. and D. Pollard (1987) U-processes: rates of convergence. *Annals of Statistics* 15, 780-799.

Pakes, A. and D. Pollard (1989) Simulation and the asymptotics of optimization estimators. *Econometrica* 57, 1027-1057.

Pollard, D. (1984) *Convergence of Stochastic Processes*. Springer, New York.

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

Rudin, W. (1976) *Principles of Mathematical Analysis*. 3rd ed., McGraw-Hill, London.