

Linton, Oliver Bruce; Xiao, Zhijie

## Working Paper

# A nonparametric regression estimator that adapts to error distribution of unknown form

SFB 373 Discussion Paper, No. 2001,33

### Provided in Cooperation with:

Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

*Suggested Citation:* Linton, Oliver Bruce; Xiao, Zhijie (2001) : A nonparametric regression estimator that adapts to error distribution of unknown form, SFB 373 Discussion Paper, No. 2001,33, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin, <https://nbn-resolving.de/urn:nbn:de:kobv:11-10049681>

This Version is available at:

<https://hdl.handle.net/10419/62744>

#### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

#### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# A NONPARAMETRIC REGRESSION ESTIMATOR THAT ADAPTS TO ERROR DISTRIBUTION OF UNKNOWN FORM\*

Oliver Linton<sup>†</sup>

The London School of Economics and Political Science

Zhijie Xiao<sup>‡</sup>

University of Illinois at Urbana-Champaign

June 5, 2001

## Abstract

We propose a new estimator for nonparametric regression based on local likelihood estimation using an estimated error score function obtained from the residuals of a preliminary nonparametric regression. We show that our estimator is asymptotically equivalent to the infeasible local maximum likelihood estimator [Staniswalis (1989)], and hence improves on standard kernel estimators when the error distribution is not normal. We investigate the finite sample performance of our procedure on simulated data.

*Keywords:* Adaptive Estimation; Asymptotic Expansions; Efficiency; Kernel; Local Likelihood Estimation; Nonparametric Regression.

*AMS 1991 subject classifications.* 62G07, 62A10

*JEL Codes:* C13 C14 C24

---

\*We would like to thank Roger Koenker, Joel Horowitz, Whitney Newey, Peter Phillips, Steve Portnoy, and Tom Rothenberg for helpful comments. Financial support from the National Science Foundation is gratefully acknowledged.

<sup>†</sup>Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: O.Linton@lse.ac.uk.

<sup>‡</sup>Department of Economics, University of Illinois at Urbana-Champaign, 1206 South Sixth Street, Champaign, IL 61820, USA. Phone: (217) 333-4520. Fax: (217) 244-6678. E-mail address: zxiao@uiuc.edu.

# 1 Introduction

It has been known for some time that in certain parametric regression models it is possible to ‘adapt’ to an unknown error distribution by maximizing an estimated likelihood function based on an estimate of the error distribution. The common result is that you do as well in terms of asymptotic variance as if you knew the true error distribution, hence the term adaptive. In estimation problems where a Gaussian assumption on the underlying distribution of the data is inappropriate, adaptive estimation provides an alternative way to the conventional Gaussian maximum likelihood estimator by replacing the Gaussian density function with a nonparametric estimate of the score function of the log-likelihood. It has been proven that an efficiency gain over the least square methods can be achieved by adaptive estimators in many models.

Adaptive estimation was first studied by Stein (1956) who considered the problem of estimating and testing hypotheses about a parameter in the presence of an infinite dimensional “nuisance” parameter. Beran (1974) and Stone (1975) considered adaptive estimation in the symmetric location model, while Bickel (1982) extended this to linear regression and other models. This latter work provided a starting point for much future work in this area, most of which has exploited the property of Local Asymptotic Normality (LAN) of the class of likelihoods involved. Manski (1984) studied adaptive estimation in non-linear models, Kreiss (1987) considered stationary and invertible autoregressive moving average (ARMA) models, Steigerwald (1992) studied linear regression with ARMA error, and Linton (1993) considered the case of linear regression with autoregressive conditional heteroscedasticity (ARCH), which was extended by Drost and Klaassen (1997) to the GARCH(1,1) case. See Drost, Klaassen, and Werker (1997) for an excellent review and recent development for time series. Jeganathan (1995) extended the theory to nonstationary models with independent and identically distributed (i.i.d.) error, which involves the generalization to Local Asymptotic Mixed Normal (LAMN) likelihoods. Hodgson (1997) further extended this case but with ARMA errors.

We propose a new estimator for nonparametric regression that adapts to the unknown shape of the error term. Consider the following nonparametric regression

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (1)$$

where  $\varepsilon_i$  is an i.i.d. error term with  $E(\varepsilon_i|X_i) = 0$ . The assumption that  $m$  is a smooth function implies that for  $X_i$  close to  $x$ ,  $Y_i$  contains information of  $m(x)$ . A popular estimator of  $m(x)$  is the

Nadaraya-Watson kernel estimator, which is a locally weighted average

$$\frac{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) Y_i}{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)}, \quad (2)$$

where  $K$  is a kernel function and  $h = h(n)$  is a bandwidth sequence. Such an estimator can also be obtained from the following weighted least squares criterion

$$\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) (Y_i - \theta)^2. \quad (3)$$

Minimizing the above weighted sum of squared residuals gives precisely the estimator in (2), see, e.g., Härdle (1990) for discussion. When  $\varepsilon_i$  are Gaussian, (3) corresponds to the weighted likelihood criterion. In the absence of Gaussianity, asymptotic results of the above estimator generally still hold but this estimator is less efficient than estimators that exploit the distributional information. In this case, if the error density  $f$  were known, we may replace  $(Y_i - \theta)^2$  by the log likelihood  $\log f(Y_i - \theta)$  and obtain a likelihood-based estimator. Tibshirani (1984) introduced the local likelihood estimator in the context of nonparametric generalized linear models. This study was extended in Hastie and Tibshirani (1987). Staniswalis (1989) applied this idea to the estimation of a location parameter  $\theta$ ; she also derived the asymptotic properties of her procedure. There has been much recent work mostly focusing on density and hazard estimation. In particular see: Copas (1994), Hjort (1993), Hjort and Jones (1996) and Loader (1996). Work in other areas includes Robinson (1989) for a time series regression problem, and the recent paper by Gozalo and Linton (2000) for nonlinear regression models. These procedures are discussed in Härdle and Linton (1994, see p17 especially).

In practice,  $f$  is generally unknown and the local likelihood estimator is infeasible. We propose an adaptive procedure for the estimation of (1) based on estimating the score function of the errors, i.e., by replacing  $f$  by a nonparametric estimator thereof. We establish the pointwise asymptotic distribution of our estimator of  $m(x)$ , for interior  $x$ , and show that it adapts in the sense that it has the same variance as the infeasible procedure based on knowing the error distribution. This is true regardless of the dimensionality of the regressors. A more complicated implementation of our procedure involves undersmoothing at the first step and will result in an estimator with exactly the same bias as the infeasible local likelihood estimator. In this case, our adaptive estimator achieves the same mean squared error in large samples as the infeasible local likelihood estimator. In this sense, our estimator is efficient. This is not to say that our method has any special properties with

regard to minimax risk; as is well known it is not possible to achieve the lower bound here, see Fan (1993). But the pairwise comparison has been used elsewhere, see Linton (1997). With regard to regularity conditions we make rather strong assumptions about the smoothness of the regression function but impose very few requirements on the error density: in one set of conditions we allow it to have unbounded support, while in another we look at the case of bounded support. In the latter case, we must also require that the error density approaches zero at the boundary rather fast because otherwise the estimation problem is non-regular, i.e., there exist estimators that converge to the true value more rapidly.

Why is this important in practice? In many data sets, the error distribution is likely to be non-normally distributed and perhaps quite far from the normal distribution to such an extent that the local likelihood estimator has much lower variance than the Gaussian-based estimators [the relative efficiency is unbounded]. As a modelling issue it is hard to believe that we have better information about the error density than about the shape of the main regression effect and so it is quite natural to treat the error density as an unknown parameter just like the regression function. Thus our results are comforting in that they say that we can still use information from the error distribution to improve the performance of our location estimates; this is all the more important in nonparametric regression because the rate of convergence can be so slow.

The principle involved extends to other location models that depend on several functional parameters, like additive regression models. In more general nonparametric models, we may or may not find adaptivity just as in the parametric case, see Bickel, Klaassen, Ritov, and Wellner (1993).

The paper is organized as follows. The model and estimators are given in the next section. Asymptotic results of the estimator are given in Section 3. In section 4 we provide a small Monte Carlo experiment which evaluates the effectiveness of the adaptive regression estimator. We give the proofs of our main result in the appendix.

For notation, we use  $f^{(j)}$  to denote the  $j^{\text{th}}$  derivative of a function  $f$ . We also let  $\|A\|$  denote the Euclidean norm of the array  $A = (a_{i_1, \dots, i_s})$  defined as  $\|A\| = (\sum a_{i_1, \dots, i_s}^2)^{1/2}$ . For functions  $m$  and vectors  $\mathbf{k} = (k_1, \dots, k_d)$  and  $x = (x_1, \dots, x_d)$ , we use the following notations

$$\mathbf{k}! = k_1! \times \dots \times k_d!, \quad |\mathbf{k}| = \sum_{i=1}^d k_i, \quad x^{\mathbf{k}} = x_1^{k_1} \times \dots \times x_d^{k_d} \quad (4)$$

$$\sum_{0 \leq |\mathbf{k}| \leq p} = \sum_{j=0}^p \sum_{\substack{k_1=0 \\ \vdots \\ k_d=0 \\ k_1+\dots+k_d=j}}^j \cdots \sum_{k_d=0}^j, \quad (D^{\mathbf{k}}m)(x) = \frac{\partial^{|\mathbf{k}|}m(x)}{\partial x_1^{k_1} \cdots \partial x_d^{k_d}}.$$

## 2 The Model and Estimator

### 2.1 The Model and an Infeasible Estimator

Suppose that  $\{Y_i, X_i\}_{i=1}^n$  are i.i.d., where  $X_i \in \mathbb{R}^d$  and  $Y_i \in \mathbb{R}$ . We consider the following regression model

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

where  $\varepsilon_i$  is independent of  $X_i$  with  $E(\varepsilon_i) = 0$ . We are interested in estimation of  $\theta_0(x) = m(x)$ , where  $x$  is an interior point in the support of  $X_i$ , and the function  $m(\cdot)$  is assumed to be of unknown form, but smooth.

If  $\varepsilon_i$  have a known common density  $f$ , we may estimate  $\theta_0(x) = m(x)$  by maximizing the locally smoothed likelihood function

$$L_n(\theta) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \log f(Y_i - \theta),$$

with respect to the local parameter  $\theta \in \Theta \subset \mathbb{R}$ , where

$$K\left(\frac{x - X_i}{h}\right) = \prod_{j=1}^d k\left(\frac{x_j - X_{ij}}{h}\right)$$

and  $k$  is a kernel function, while  $h$  is a bandwidth parameter. The smoothed maximum likelihood estimator (SMLE) (or local maximum likelihood estimator) of  $\theta_0(x)$ , denoted  $\bar{\theta}(x)$ , can be obtained by setting the following smoothed score function

$$S_n(\theta; f) = -\frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'}{f}(Y_i - \theta) \quad (5)$$

equal to zero. If  $\varepsilon_i$  is assumed to be normally distributed, then  $\bar{\theta}(x)$  is the standard Nadaraya-Watson kernel smoother, while if  $\varepsilon_i$  is assumed to be Laplace distributed, then  $\bar{\theta}(x)$  is the standard local median smoother, see Chaudhuri (1991). In general, the estimator  $\bar{\theta}(x)$  is only implicitly defined and is a nonlinear function of  $Y_1, \dots, Y_n$ .

We next state the properties of  $\bar{\theta}(x)$ . Define

$$\mathcal{I}_\theta(x) = f_X(x)I(f), \quad (6)$$

where  $f_X(x)$  is the covariate density, and

$$I(f) = \int \left[ \frac{f'(\varepsilon)}{f(\varepsilon)} \right]^2 f(\varepsilon) d\varepsilon \quad (7)$$

is the Fisher information for the error density  $f$ .

PROPOSITION 1. *Under our assumptions A1-6 given below*

$$\sqrt{nh^d} [\bar{\theta}(x) - \theta_0(x)] \Rightarrow N(0, \|K\|^2 \mathcal{I}_\theta^{-1}(x)),$$

where  $\|K\|^2 = \int |K(u)|^2 du$ .

REMARK 1. This result is essentially the same as Staniswalis (1989, p277) except that her theorem allows a more general log likelihood function  $\log f(Y_i; \theta)$  rather than our location-based likelihood  $\log f(Y_i - \theta)$  [see also Gonzalez-Manteiga (1990)]. We allow for a more general multivariate random design for the covariates.

REMARK 2. The limiting variance of  $\bar{\theta}(x)$  is smaller than the variance of the Nadaraya-Watson estimator, which is  $\sigma^2 \|K\|^2 / f_X(x)$ , where  $\sigma^2$  is the error variance. It is also smaller than the variance of any other kernel  $M$ -estimator based on finding zeros of  $\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \psi(Y_i - \theta)$  for any function  $\psi$  with  $E\psi(\varepsilon_i) = 0$ . This follows from the classical Cramér-Rao inequality.

REMARK 3. The bias of  $\bar{\theta}(x)$  is, for second order kernels  $K$ , approximately  $h^2 [\int u^2 K(u) du] [m''(x) + 2m'(x)f'_X(x)/f_X(x)]/2$  regardless of the shape of  $f$  [this is assuming that  $E\ell'''(\varepsilon) = 0$  (where  $\ell(\varepsilon) = \log f(\varepsilon)$ ), which is certainly true for symmetric  $\varepsilon$ ]. In particular, it is the same as the bias of the Nadaraya-Watson estimator. A more general class of methods would involve local polynomials inside  $f(Y_i - \theta)$  in place of the constant  $\theta$ ; this would result in a simpler bias expression, see below for further discussion on this.

An alternative estimation method is given by taking one Newton-Raphson step [as proposed in Bickel (1975) for parametric estimation] from a preliminary consistent estimator  $\tilde{\theta}$  [say, the kernel estimator (2)], that is, let

$$\bar{\theta}_{NR} = \tilde{\theta} + \mathcal{I}_n(\tilde{\theta}; f)^{-1} S_n(\tilde{\theta}; f), \quad (8)$$

where  $\mathcal{I}_n(\tilde{\theta}; f)$  is a consistent estimate of the information  $\mathcal{I}_\theta(x)$  :

$$\mathcal{I}_n(\tilde{\theta}; f) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \left[ \frac{f'}{f}(Y_i - \tilde{\theta}) \right]^2.$$

This method is asymptotically equivalent to the SMLE  $\bar{\theta}$  in the sense that  $\sqrt{nh^d}(\bar{\theta}_{NR} - \bar{\theta}) = o_p(1)$  - see Fan and Chen (1999) for the proof of a similar result.

In practice,  $f$  is generally unknown and so neither  $\bar{\theta}$  nor  $\bar{\theta}_{NR}$  are feasible. However, the infeasible procedure defines an efficiency standard against which we should measure our feasible estimator.

## 2.2 Our Estimator

To obtain a feasible estimator for  $\theta(x)$ , we replace  $f$  by a nonparametric estimate, say  $\tilde{f}$ . Specifically, we let  $\tilde{f}(e_i)$  and  $\tilde{f}'(e_i)$  be the leave-one-out kernel estimates:

$$\tilde{f}(e_i) = \frac{1}{nh} \sum_{j \neq i} k \left( \frac{e_i - \tilde{\varepsilon}_j}{h} \right), \quad \tilde{f}'(e_i) = \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{e_i - \tilde{\varepsilon}_j}{h} \right), \quad (9)$$

where  $\tilde{\varepsilon}_j$  is a preliminary estimate of  $\varepsilon_j$  defined by  $\tilde{\varepsilon}_j = Y_j - \tilde{m}(X_j)$ , where  $\tilde{m}(X_j)$  is a preliminary nonparametric estimator of  $m(X_j)$ . We will discuss the construction of  $\tilde{m}(X_j)$  later in Section 2.3.

As in some other applications of kernel regression estimators, the estimator  $\tilde{f}(e_i)$  can be small and may cause technical difficulty since it enters into the estimated score function in the denominator. For this reason, we trim out small  $\tilde{f}(e_i)$  as do Bickel (1982) and Manski (1984). The simplest and probably most common trimming is the indicator trimming function:  $G_b(x) = I(|x| \geq b)$ , where  $b$  is the trimming parameter that goes to zero as  $n \rightarrow \infty$ . Instead, we consider the following smoothed trimming, which has been used recently by Andrews (1995) and Ai (1997). Let  $g(\cdot)$  be a density function that has support  $[0, 1]$ ,  $g(0) = g(1) = 0$ , and let

$$g_b(x) = \frac{1}{b} g \left( \frac{x}{b} - 1 \right),$$

where  $b$  is the trimming parameter; then  $g_b(x)$  has support on  $[b, 2b]$ . Letting

$$G_b(x) = \int_{-\infty}^x g_b(z) dz,$$

we have

$$G_b(x) = \begin{cases} 0, & x < b \\ \int_{-\infty}^x g_b(z) dz, & b \leq x \leq 2b \\ 1, & x > 2b. \end{cases}$$

For example, consider the following Beta density

$$g(z) = B(a+1)^{-1} z^a (1-z)^a, \quad 0 \leq z \leq 1,$$



for some positive integer  $a$ , where  $B(a)$  is the beta function defined by  $B(a) = \Gamma(a)^2/\Gamma(2a)$ , and  $\Gamma(a)$  is the Euler gamma function. Then, it can be verified that the function  $G_b(x)$  is  $a + 1$ -times continuously differentiable on  $[0, 1]$ . This property allows us to use standard Taylor series arguments, whereas indicator function trimming would preclude this. We will suppose that  $a \geq 3$ .

We now define the trimmed local score function as

$$\tilde{S}_n(\theta; \tilde{f}) = -\frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{\tilde{f}'}{\tilde{f}}(Y_i - \theta) G_b(\tilde{f}(Y_i - \theta)) \quad (10)$$

and the trimmed information by

$$\tilde{\mathcal{I}}_n(\theta; \tilde{f}) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \left[ \frac{\tilde{f}'}{\tilde{f}}(Y_i - \theta) \right]^2 G_b(\tilde{f}(Y_i - \theta)). \quad (11)$$

We may consider a profile smoothed maximum likelihood estimator  $\hat{\theta}(x)$  that sets  $\tilde{S}_n(\theta; \tilde{f})$  equal to zero. Similar to the case with known density  $f$ , a one-step Newton-Raphson estimator of  $\theta$  can also be obtained from a preliminary consistent estimator  $\tilde{\theta}(x)$ , which we take to be the Nadaraya-Watson estimator, i.e.,

$$\tilde{\theta}(x) = \frac{\sum_{j=1}^n K\left(\frac{X_j - x}{h}\right) Y_j}{\sum_{j=1}^n K\left(\frac{X_j - x}{h}\right)}. \quad (12)$$

We calculate the following one-step Newton-Raphson estimator

$$\hat{\theta}(x) = \tilde{\theta}(x) + \tilde{\mathcal{I}}_n(\tilde{\theta}(x); \tilde{f})^{-1} \tilde{S}_n(\tilde{\theta}(x); \tilde{f}). \quad (13)$$

In the next section we will give the asymptotic distribution of the one-step estimator (13).

The preliminary estimator  $\tilde{\theta}(x)$  in (12), the density estimators  $\tilde{f}$  and  $\tilde{f}'$  in (9), and the estimators of the score  $\tilde{S}_n$  in (10) and the information  $\tilde{\mathcal{I}}_n$  in (11) all involve choices of kernels and bandwidth values. In principle, we may consider more general devices that use different bandwidth parameters and kernels in different cases. However, the additional smoothing parameters bring complication to the analysis and comparisons. In this paper, we consider the simple case where the same bandwidth  $h$  and kernel  $k$  are used in  $\tilde{\theta}(x)$ ,  $\tilde{f}$ ,  $\tilde{f}'$ ,  $\tilde{S}_n$ , and  $\tilde{\mathcal{I}}_n$ . In the next subsection we describe our local polynomial estimator used in the construction of the residuals  $\tilde{\varepsilon}_j$ .

In the event that the error density is known to be symmetric, some improvement can be expected by symmetrizing the error density and derivative, thus  $\tilde{f}(e) \mapsto (\tilde{f}(e) + \tilde{f}(-e))/2$  and  $\tilde{f}'(e) \mapsto (\tilde{f}'(e) - \tilde{f}'(-e))/2$ , although this only affects the higher order terms in our case. Bickel (1982) proposes an

estimator that incorporates these restrictions. In his proofs he exploits the symmetry properties of the estimator. We do not impose symmetry on our estimator and our proof technique is quite different from Bickel's.

One can also expect better performance in practice by iterating (13) a few times or until convergence.

### 2.3 Estimation of the Residuals

An important input to the density estimate is the estimated residual  $\tilde{\varepsilon}_j = Y_j - \tilde{m}(X_j)$ , which requires an estimate of  $m(X_j)$ . For the choice of  $\tilde{m}(X_j)$ , natural candidates include the conventional Nadaraya-Watson estimator and the widely used local polynomial estimator or sieve estimators. When the ordinary kernel estimator is used, additional trimming is usually needed to remove the boundary bias because if we use all observations in estimating the error density, we are pushed into the boundary. To avoid introducing another trimming on  $\tilde{m}(X_j)$ , we use local polynomials instead of ordinary kernel estimators in the construction of residuals  $\tilde{\varepsilon}_j$ . See Fan (1992), and Fan and Gijbels (1996) for discussion on the attractive properties of local polynomials. Given observations  $\{Y_i, X_i\}_{i=1}^n$ , the preliminary estimate of the regression function  $m(x)$  can be obtained using the multivariate weighted least squares criterion

$$\sum_{i=1}^n \left[ Y_i - \sum_{0 \leq |k| \leq p} b_{\mathbf{k}} \cdot (X_i - x)^{\mathbf{k}} \right]^2 \mathcal{K}((X_i - x)/h), \tag{14}$$

where  $\mathcal{K}(u)$  is a nonnegative weight function on  $\mathbb{R}^d$  and  $h$  is a bandwidth parameter, while  $p$  is an integer with  $p \geq 2$ . Let  $\tilde{m}(x) = \hat{b}_0$ , where  $\hat{b}_0$  is the minimizing intercept in (14). We compute this estimator for each sample point and use it to construct the residuals  $\tilde{\varepsilon}_j = Y_j - \tilde{m}(X_j)$ , which are the key input to the density estimate. Again, for convenience of comparison, we choose  $p = q - 1$  and use the same bandwidth  $h$ , so that the bias and variance of the preliminary estimator are of the same orders of magnitude as the adaptive estimator. We give more discussion about the technical details of the local polynomial estimator in the appendix.

### 3 Main Results

#### 3.1 Assumptions

To facilitate the asymptotic analysis, we make the following assumptions on the distribution, the kernel function  $k(\cdot)$ , the bandwidth parameter  $h$ , and the trimming parameter  $b$  in  $\tilde{\theta}(x)$ ,  $\tilde{f}$ ,  $\tilde{f}'$ ,  $\tilde{S}_n$ , and  $\tilde{\mathcal{I}}_n$ .

A1.  $\varepsilon_i$  and  $X_i$  are independent and identically distributed (i.i.d.) random variables (vectors), mutually independent, and  $E(\varepsilon_i) = 0$  and  $E(\varepsilon_i^2) = \sigma^2 < \infty$ .

A2.  $\varepsilon_i$  has support  $\mathbb{R}$  and has symmetric Lebesgue density  $f(\varepsilon)$  which has uniformly bounded continuous partial derivatives up to the order  $r$ , and  $f^{(r)}(\varepsilon)$  is Lipschitz continuous, i.e., there exists a constant  $c$  such that for all  $\varepsilon, \varepsilon^*$  on its support, we have

$$|f^{(r)}(\varepsilon) - f^{(r)}(\varepsilon^*)| \leq c|\varepsilon - \varepsilon^*|.$$

A3. Let  $\ell(\varepsilon) = \log f(\varepsilon)$ , and suppose that  $E[(\ell'(\varepsilon))^2] < \infty$ ,  $E[|\ell''(\varepsilon)|] < \infty$  and  $E[|\ell'''(\varepsilon)|] < \infty$ .

A4. The kernel  $k$  has support  $[-1, 1]$  and is symmetric about zero and satisfies  $\int k(u)du = 1$ . There exists an even positive integer  $q$  with  $2 < q \leq r - 1$  such that

$$\int u^j k(u)du = 0, \quad j = 1, \dots, q - 1, \quad \text{and} \quad \int u^q k(u)du \neq 0.$$

Furthermore, it is four times differentiable on its support, while  $k'(0) = 0$ .

A5.  $h \rightarrow 0$ ,  $nh^{2q+d} \rightarrow 0$ , and  $nh^{d+9} \rightarrow \infty$ ,  $b = h^\tau$ , and  $0 < \tau < 1/2$ .

A6.  $X_i$  has Lebesgue density  $f_X(x)$  which is bounded away from zero on its support  $\mathcal{X}$ , a compact subset of  $\mathbb{R}^d$ .  $(D^{\mathbf{k}}f_X)(x)$  and  $(D^{\mathbf{k}}m)(x)$  are bounded and uniformly continuous on  $\mathbb{R}^d$ , and there exists finite  $C_2$  and  $C_3$  such that

$$|(D^{\mathbf{k}}f_X)(u) - (D^{\mathbf{k}}f_X)(v)| \leq C_2\|u - v\|, \quad |(D^{\mathbf{k}}m)(u) - (D^{\mathbf{k}}m)(v)| \leq C_3\|u - v\|,$$

with  $|\mathbf{k}| = r$ .

We assume the existence of a variance  $\sigma^2$  just for the purpose of verifying the properties of our pilot estimator. If for example the local median smoother were used, then it may be possible to make weaker assumptions about the error moments. Assumptions A2 and A3 ensure the adaptive property and Taylor expansions of the density function to appropriate orders. By dominated convergence, Assumption A3 also ensures that

$$\lim_{b \rightarrow 0} \int_{f(\varepsilon) < b} [\ell'(\varepsilon)]^2 f(\varepsilon) d\varepsilon = 0,$$

and

$$\lim_{b \rightarrow 0} \int_{f(\varepsilon) < b} \ell''(\varepsilon) f(\varepsilon) d\varepsilon = 0,$$

which guarantee that the trimming effect will be asymptotically ignorable. Further sufficient conditions may be found for this property. For example, we may replace assumptions A2 and A3 by the following (sufficient conditions) assumptions (A2' and A3' below), which assume that  $\varepsilon$  has bounded support.

A2'.  $\varepsilon_i$  has symmetric Lebesgue density  $f(\varepsilon)$  which has support  $\text{supp}(f) = [\underline{a}, \bar{a}]$ , where  $\underline{a}$  and  $\bar{a}$  are unknown boundary parameters that satisfy  $-\infty < \underline{a} < \bar{a} < \infty$ , and  $f(\varepsilon) > 0$  on  $(\underline{a}, \bar{a})$ . In addition, the density has uniformly bounded continuous partial derivatives up to the order  $r$ , and  $f^{(r)}(\varepsilon)$  is Lipschitz continuous on  $(\underline{a}, \bar{a})$ , i.e., there exists a constant  $c$  such that for all  $\varepsilon, \varepsilon^* \in (\underline{a}, \bar{a})$ , we have

$$|f^{(r)}(\varepsilon) - f^{(r)}(\varepsilon^*)| \leq c|\varepsilon - \varepsilon^*|.$$

In A2' we assume that  $f(\varepsilon)$  has bounded support. When  $f$  is strictly positive on  $[\underline{a}, \bar{a}]$ , the situation is non-regular. In some cases, this can lead to inconsistency of solutions of the likelihood score equations, but perhaps the potential for improved rates of convergence for other estimators. Therefore, we shall make an additional assumption

A3'.  $f(\varepsilon)$  and its first  $\varrho - 1$  derivatives vanish at  $\underline{a}$  and  $\bar{a}$ , while  $f^{(\varrho)}(\underline{a}) \neq 0$ , and  $f^{(\varrho)}(\bar{a}) \neq 0$  for some integer  $\varrho$  with  $2 \leq \varrho \leq r$ .

Assumption A3' guarantees that the density  $f$  vanishes at the boundary at a sufficiently fast rate so that the properties of regular estimation holds. In this case, one can not estimate a parameter  $\alpha$  of the density  $f$  at a rate better than root-n. See Akahira and Takeuchi (1995) for a discussion of

this issue. This assumption also implies that the Fisher information  $I(f)$  defined in (7) exists as do various other integrals used below.

These assumptions are similar to those used in the existing literature. Note that because the trimming parameter  $b$  is of larger magnitude than  $h$ , our estimator will not suffer from boundary bias from the estimation of the density  $f$ . Assumption A5 is quite strong and implies that we must have  $2q > 9$ . This assumption is stronger than necessary and arises partly because we have chosen the same bandwidth throughout and partly because of our proof technique. With a more complicated analysis, it is possible to show that  $2q > 6$  is enough. We believe that the result is true even for positive kernels. Assumption A6 is introduced to avoid the use of additional trimming parameters. Eliminating this assumption would substantially increase the mathematical complexity of our proofs without providing any further insight into our results.

We construct the residuals  $\tilde{\varepsilon}_j = Y_j - \tilde{m}(X_j)$  using  $p$ -th order local polynomials. This is only used in obtaining the residuals to avoid additional trimming or boundary modification when we use all  $n$  observations in estimating the error density. We use the same bandwidth  $h$  in the kernel estimations and the local polynomial estimation. For each  $j$  with  $0 \leq |j| \leq 2p + 1$  define the function

$$H_j(u) = u^j \mathcal{K}(u).$$

We make the following assumption on the kernel function  $\mathcal{K}(\cdot)$  in the local polynomial regression. This assumption and the assumptions on  $f_X$  and  $m$  given in the beginning of this section ensure uniform convergence results on  $\tilde{m}(X_j)$  (see Masry (1996)).

B1. *The kernel  $\mathcal{K}$  is symmetric about zero, bounded, and has compact connected support ( $\mathcal{K}(u) = 0$  for  $\|u\| > A_0$  some  $A_0$ ). For all  $j$  with  $0 \leq |j| \leq 2p + 1$ , there exists finite  $C_1$  such that*

$$|H_j(u) - H_j(v)| \leq C \|u - v\|.$$

REMARK 1. The regularity assumptions facilitate our asymptotic analysis. In practice, even when some of these conditions do not hold, if the error distribution is distant from normal, efficiency gain over the conventional kernel estimator may still be found in the adaptive estimator. Also see Monte Carlo results in Section 4.

### 3.2 Asymptotic Distribution

For notational simplicity, we write  $\tilde{\mathcal{I}}_n(\theta; \tilde{f})$  as  $\tilde{\mathcal{I}}_n(\theta)$ , and  $\tilde{S}_n(\theta; \tilde{f})$  as  $\tilde{S}_n(\theta)$ . The estimator  $\hat{\theta}(x)$  can be written, after standardization by  $\sqrt{nh^d}$ , as

$$\sqrt{nh^d}(\hat{\theta}(x) - \theta_0(x)) = \sqrt{nh^d}(\tilde{\theta}(x) - \theta_0(x)) + \sqrt{nh^d}\tilde{\mathcal{I}}_n(\tilde{\theta})^{-1}\tilde{S}_n(\tilde{\theta}), \quad (15)$$

where we suppress dependence of  $\theta$  on  $x$  wherever possible. Expanding  $\tilde{S}_n(\tilde{\theta})$  around the true value of  $\theta$ ,  $\theta_0$ , to the second term, we show that

$$\tilde{S}_n(\tilde{\theta}) = \tilde{S}_n(\theta_0) - \tilde{\mathcal{I}}_n(\theta_0)(\tilde{\theta} - \theta_0) + o_p(\tilde{\theta} - \theta_0). \quad (16)$$

Assuming that  $\tilde{\mathcal{I}}_n(\tilde{\theta})^{-1} = O_p(1)$  and substituting (16) into (15) we have

$$\sqrt{nh^d}(\hat{\theta} - \theta_0) = \left[ \sqrt{nh^d}(\tilde{\theta} - \theta_0) - \tilde{\mathcal{I}}_n(\tilde{\theta})^{-1}\tilde{\mathcal{I}}_n(\theta_0)\sqrt{nh^d}(\tilde{\theta} - \theta_0) \right] + \sqrt{nh^d}\tilde{\mathcal{I}}_n(\tilde{\theta})^{-1}\tilde{S}_n(\theta_0) + o_p(1),$$

since  $\tilde{\theta}$  is a  $\sqrt{nh^d}$ -consistent estimator of  $\theta_0$  under our assumptions. Furthermore, it can be shown that  $\tilde{\mathcal{I}}_n(\tilde{\theta}) - \tilde{\mathcal{I}}_n(\theta_0) \xrightarrow{p} 0$  and  $\tilde{\mathcal{I}}_n(\theta_0) \xrightarrow{p} \mathcal{I}_\theta(x)$ . Therefore,

$$\sqrt{nh^d}(\tilde{\theta} - \theta_0) - \tilde{\mathcal{I}}_n(\tilde{\theta})^{-1}\tilde{\mathcal{I}}_n(\theta_0)\sqrt{nh^d}(\tilde{\theta} - \theta_0) \xrightarrow{p} 0 \quad (17)$$

and in fact

$$\sqrt{nh^d}(\hat{\theta} - \theta_0) = \mathcal{I}_\theta(x)^{-1}\sqrt{nh^d}\tilde{S}_n(\theta_0) + o_p(1).$$

Finally, we show that

$$\sqrt{nh^d}\tilde{S}_n(\theta_0) \Rightarrow N(0, \mathcal{I}_\theta(x) \|K\|^2). \quad (18)$$

Thus, we obtain the following result for our adaptive nonparametric regression estimator.

**THEOREM 1.** *Suppose that Assumptions A1 to A6 and B1 hold. Then, as  $n \rightarrow \infty$*

$$\sqrt{nh^d} \left[ \hat{\theta}(x) - \theta_0(x) \right] \Rightarrow N \left( 0, \|K\|^2 \mathcal{I}_\theta^{-1}(x) \right).$$

**REMARK 1.** It may appear obvious that the effect of estimating  $f$  should not effect the distribution of the resulting estimator of  $m$ , at least when  $\varepsilon$  is of lower dimension than  $X$  i.e.,  $d > 1$ . However, this is a bit misleading because to first estimate  $f$  we need to estimate  $m$ ; our estimator of  $f$  has convergence rate determined by the preliminary estimation of  $m$ . Therefore, our result is quite surprising, especially when compared with other results for nonparametrically generated data [see

for example Ahn (1995)], where the dominant term is that due to the higher dimensional estimation problem. On the other hand, under symmetry the location score is orthogonal to the score for the error density in parametric models, which is why one achieves adaptivity in that case. The same is true here. In addition, the estimator  $\widehat{\theta}(x)$  depends on an estimate of a functional of the error density rather than the error density at a single point.

REMARK 2. The notion of efficiency that is employed here is similar to that used in Linton (1997) in the context of additive nonparametric regression. We are aiming to do as well as the corresponding estimator one would compute if one knew the error density  $f$ . Of course any specific estimator can be bettered for some combination of  $(f, m, f_X)$ , see Fan (1993).

REMARK 3. In the proofs, we decompose the estimation errors into different types of effects and separately deal with these effects. In particular, to prove the asymptotic results, we have to deal with at least four types of terms in the estimation of the score and Hessian. The first type of errors, in the form  $m(X_i) - m(x)$ , come from the local deviation from  $X_i$  to  $x$  due to smoothing; The second type of errors,  $\widetilde{m}(X_i) - m(X_i)$  and  $\widetilde{m}(x) - m(x)$ , are from the preliminary estimation; these effects are largely determined by the smoothness property of  $m$  and, of course, the kernel smoothing. The third type of errors are due to conventional nonparametric kernel smoothing in density estimation. The orders of magnitude of the bias effects and variance effects in nonparametric kernel estimation are determined by the properties of the density function and the kernel function and have been extensively studied in the literature. The fourth type is the trimming effect whose order of magnitude is generally determined by the tail behavior of the density function. In the proof we need a bit more than the mean squared consistency of the score function estimator, as required in Bickel (1982) for the parametric problem. We require uniform rates of convergence on the density and derivative estimates.

REMARK 4. From the proof of the results, we can see that our estimator has a bias effect of order  $h^q$ . Under Assumption A5, our presentation of the asymptotics has eliminated the bias term.

REMARK 5. The above result is also useful for deriving nonparametric confidence intervals for the regression function. Under the bandwidth assumptions, the bias term can be ignored and a confidence interval at significance level  $\alpha$  can be constructed as follows

$$\left[ \widehat{\theta} - \frac{\widehat{s}}{\sqrt{nh^d}} z_{\alpha/2}, \widehat{\theta} + \frac{\widehat{s}}{\sqrt{nh^d}} z_{\alpha/2} \right], \quad (19)$$

where  $\Phi(z_{\alpha/2}) = 1 - \alpha/2$  with  $\Phi(\cdot)$  the standard normal distribution, and  $\widehat{s}^2$  is a consistent estimate of the asymptotic variance. Suitable estimators include

$$\widehat{s}_1^2 = \frac{\|K\|^2}{\widehat{f}_X(x)\widehat{I}(f)}, \text{ with } \widehat{I}(f) = \frac{1}{n} \sum_{i=1}^n \frac{\widetilde{f}'(\widetilde{\varepsilon}_i)}{\widetilde{f}(\widetilde{\varepsilon}_i)},$$

or

$$\widehat{s}_2^2 = \|K\|^2 \left\{ \frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \left[ \frac{\widetilde{f}'}{\widetilde{f}}(Y_i - \widetilde{\theta}) \right]^2 \right\}^{-1}.$$

REMARK 6. Assumption A5 gives general bandwidth conditions that are sufficient for the asymptotic analysis. Cross-validation may be used for bandwidth selection. In particular, for each  $j$ , we estimate  $m(X_j)$  by removing the  $j$ -th observation and denote the corresponding estimator as  $\widehat{\theta}_{h,j}(X_j)$ , then  $h$  is chosen to maximize

$$\frac{1}{n} \sum_{j=1}^n (Y_j - \widehat{\theta}_{h,j}(X_j))^2 \text{ or } \frac{1}{n} \sum_{j=1}^n \log \widetilde{f}(Y_j - \widehat{\theta}_{h,j}(X_j)).$$

Alternatively, a (complicated) higher order analysis of the nonparametric estimator may be conducted and a bandwidth choice could be determined based on optimizing the second order effects with respect to the bandwidth. See Fan, Farmen, and Gijbels (2000) for the development of a bandwidth selection method in a local likelihood context.

REMARK 7. Our analysis in the current paper is conducted based on the Nadaraya-Watson kernel estimator. In principle, the same idea may be applied to other types of estimators, like local polynomials. This leads to a difference in the bias expression [which has been omitted by undersmoothing anyway] but the same variance for comparable implementations.

## 4 Simulations

We conducted a small Monte Carlo simulation to evaluate the finite sample performance of the proposed estimation procedure. The data were generated from (1) with  $X_i$  being i.i.d. standard normal truncated at  $\pm 5$ , and  $m(x) = x^2$ . We compare the nonparametric adaptive estimator with the Nadaraya-Watson kernel estimator for different error distributions. Several different specifications of  $\varepsilon_i$  were considered. In particular, we considered cases where  $\varepsilon_i$  are i.i.d.  $t$ -distributions with different degrees of freedom, and the case where  $\varepsilon_i$  are i.i.d. standard normal variates. Different sample sizes



were tried,  $n = 100, 200, 300, 1000$ . The number of replications is 200 in each case. The following kernel function was used in the nonparametric estimation  $K(u) = 15(7u^4 - 10u^2 + 3)^2 1(|u| \leq 1)/32$ , which has  $q = 4$ . For the estimation of the residuals, we use third order ( $p = 3$ ) local polynomial with kernel  $\mathcal{K}(u) = 0.75(1 - u^2)1(|u| \leq 1)$ .

The sampling performance of both the Nadaraya-Watson kernel estimator and the nonparametric adaptive estimators were examined for each case. In particular, we compared the biases and mean squared errors of these estimators given different choices of innovation processes and bandwidth values.

We first estimated  $m(x)$  at a fixed point  $x = 0$  for various data generation and bandwidth choices. We compared the Bias and Mean Squared Errors of these estimates in Tables 1 to 3. Results based on three sample sizes,  $n = 100, 200, 1000$ , are reported. In each case, five different bandwidth values were considered. Considering the relationship between the bandwidth and sample sizes, slightly different bandwidth choices were considered for different sample sizes. In particular, smaller bandwidth were used in larger sample sizes. Table 1 reports the results when  $\varepsilon_i$  are i.i.d.  $t$ -distributions with 2 degrees of freedom. Notice that in this case the regularity conditions do not hold because  $\varepsilon_i$  has infinite variance. However, the error distribution in this case is distant from normal and, as shown in Table 1, the adaptive estimator still brings efficiency gain. Table 2 reports the estimation results when  $\varepsilon_i$  are i.i.d.  $t$ -distributions with 3 degrees of freedom. From both Table 1 and Table 2 we can see that the efficiency gain from adaptive estimation is more apparent when sample size is large, corroborating the asymptotic theory. Table 3 gives the result when the error terms are iid standardized normal variates. As anticipated, the conventional kernel estimator has slightly better performance since, in this case, it corresponds to the likelihood estimator based on known density.

TABLE 1: Error Term =  $t(2)$ 

	Bandwidth	NW Estimator		Adaptive Estimator	
		Bias	MSE	Bias	MSE
n=100	0.08	-0.045	0.2752	-0.050	0.2626
	0.09	-0.078	0.2235	-0.052	0.2084
	0.10	-0.082	0.2202	-0.083	0.1793
	0.15	-0.042	0.1863	-0.046	0.1856
	0.20	0.063	0.1557	-0.056	0.1543
n=200	0.05	-0.032	0.1717	-0.031	0.1552
	0.08	-0.038	0.1487	-0.036	0.1467
	0.09	0.048	0.1194	0.031	0.1136
	0.10	-0.034	0.1003	-0.036	0.1006
	0.15	-0.037	0.0945	-0.028	0.0836
n=1000	0.02	-0.018	0.0858	-0.015	0.0826
	0.04	0.0085	0.0695	0.0043	0.0646
	0.06	-0.017	0.0491	-0.009	0.0310
	0.08	0.026	0.0338	0.018	0.0223
	0.10	-0.025	0.0262	0.019	0.0177

TABLE 2: Error Term =  $t(3)$ 

	Bandwidth	NW Estimator		Adaptive Estimator	
		Bias	MSE	Bias	MSE
n=100	0.08	-0.034	0.2538	-0.027	0.2626
	0.09	-0.045	0.1221	-0.020	0.1068
	0.10	0.049	0.1171	0.043	0.1122
	0.15	-0.036	0.0858	-0.051	0.0855
	0.2	0.061	0.0802	0.054	0.0735
n=200	0.05	-0.024	0.1281	-0.027	0.1273
	0.08	-0.026	0.0786	-0.019	0.0785
	0.09	0.032	0.0718	0.031	0.0700
	0.10	-0.043	0.0612	-0.042	0.0545
	0.15	-0.038	0.0568	0.041	0.0537
n=1000	0.02	-0.011	0.0489	-0.009	0.0414
	0.04	-0.015	0.0366	-0.007	0.0308
	0.06	0.015	0.0338	-0.011	0.0221
	0.08	-0.027	0.0302	-0.016	0.0241
	0.10	-0.018	0.0155	-0.014	0.0109

TABLE 3: Error Term = Normal

	Bandwidth	NW Estimator		Adaptive Estimator	
		Bias	MSE	Bias	MSE
n=100	0.08	0.0146	0.2113	0.0046	0.2176
	0.09	0.0018	0.1057	0.012	0.1171
	0.10	0.0026	0.1011	0.018	0.1086
	0.15	0.0113	0.0847	0.027	0.0908
	0.20	0.0252	0.0729	0.033	0.0786
n=200	0.05	-0.027	0.116	-0.022	0.123
	0.08	-0.021	0.0721	-0.018	0.0772
	0.09	0.025	0.0649	0.029	0.0671
	0.10	-0.038	0.0508	-0.036	0.0535
	0.15	-0.031	0.0489	0.039	0.0514
n=1000	0.02	-0.0083	0.0408	-0.0091	0.0411
	0.04	-0.0046	0.0303	-0.0045	0.0305
	0.06	0.0092	0.0256	-0.0088	0.0255
	0.08	-0.0161	0.0216	-0.0172	0.0222
	0.10	-0.0131	0.0105	-0.0240	0.0105

Figures 1 to 5 depict the estimated function  $\hat{\theta}(x)$  for the model (1) with  $m(x) = x^2 - 1$ ,  $n = 300$ , and, again,  $X_i$  is i.i.d. standard normal truncated at  $\pm 5$ .  $\varepsilon_i$  are i.i.d. student- $t$  variates with degrees of freedom equals 3. In particular, Figures 1 to 5 corresponds to bandwidth values  $h = 0.06, 0.08, 0.1, 0.15, 0.2$ , respectively. Our simulations indicate that the proposed estimator does a reasonable job for curve estimation.

Figures 1 - 5 Here

## 5 Appendix

We denote  $\phi(x, y, z, \dots)$  as a general function whose exact form may change from case to case. For simplicity, we denote  $G_b(\tilde{f}_i)$  and  $G_b(f_i)$  as  $\tilde{G}_i$  and  $G_i$ . For two random variables  $X_n, Y_n$ , we say that  $X_n \simeq Y_n$  whenever  $X_n = Y_n(1 + o_p(1))$  as  $n \rightarrow \infty$ . Let  $E_i$  denote expectation conditional on

$$z_i = (X_i, Y_i).$$

### 5.1 Preliminaries

The asymptotic properties of local polynomial estimator have been well developed and documented, see, e.g., Fan and Gijbels (1996) and Masry (1996) and the references therein. For convenience, we first give some general definitions for our local polynomial kernel nonparametric regression estimators. Let  $N_\ell = \binom{\ell + d - 1}{d - 1}$  be the number of distinct  $d$ -tuples  $j$  with  $|j| = \ell$ . Arrange these  $N_\ell$   $d$ -tuples as a sequence in a lexicographical order (with highest priority to last position so that  $(0, \dots, 0, \ell)$  is the first element in the sequence and  $(\ell, 0, \dots, 0)$  the last element) and let  $\phi_\ell^{-1}$  denote this one-to-one map. Arrange the distinct values of  $(\widetilde{D^{\mathbf{k}}})(m)$ ,  $0 \leq |\mathbf{k}| \leq p$ , as a column vector of dimension  $N \times 1$ , where  $N = \sum_{\ell=0}^p N_\ell \times 1$ , where the  $i^{\text{th}}$  element of that vector is obtained by the following relation  $i = \phi_{|j|}^{-1}(j) + \sum_{k=0}^{|j|-1} N_k$ . Similarly, arrange the vector  $(D^{\mathbf{k}})(m)$ . For each  $j$  with  $0 \leq |j| \leq 2p$ , let

$$\mu_j(\mathcal{K}) = \int_{\mathbb{R}^d} u^j \mathcal{K}(u) du, \quad \nu_j(\mathcal{K}) = \int_{\mathbb{R}^d} u^j \mathcal{K}^2(u) du,$$

and define the  $N \times N$  dimensional matrices  $M$  and  $\Gamma$  and  $N \times 1$  vector  $B$  by

$$M = \begin{bmatrix} M_{0,0} & M_{0,1} & \cdots & M_{0,p} \\ M_{1,0} & M_{1,1} & \cdots & M_{1,p} \\ \vdots & & & \vdots \\ M_{p,0} & M_{p,1} & \cdots & M_{p,p} \end{bmatrix}, \quad \Gamma = \begin{bmatrix} \Gamma_{0,0} & \Gamma_{0,1} & \cdots & \Gamma_{0,p} \\ \Gamma_{1,0} & \Gamma_{1,1} & \cdots & \Gamma_{1,p} \\ \vdots & & & \vdots \\ \Gamma_{p,0} & \Gamma_{p,1} & \cdots & \Gamma_{p,p} \end{bmatrix}, \quad B = \begin{bmatrix} M_{0,p+1} \\ M_{1,p+1} \\ \vdots \\ M_{p,p+1} \end{bmatrix}, \quad (20)$$

where  $M_{i,j}$  and  $\Gamma_{i,j}$  are  $N_i \times N_j$  dimensional matrices whose  $(\ell, m)$  element are, respectively,  $\mu_{\phi_i(\ell) + \phi_j(m)}$  and  $\nu_{\phi_i(\ell) + \phi_j(m)}$ . Note that the elements of the matrices  $M$  and  $\Gamma$  are simply multivariate moments of the kernel  $\mathcal{K}$  and  $\mathcal{K}^2$ , respectively. Define also we denote

$$M^{-1} = \begin{bmatrix} M^{0,0} & M^{0,1} & \cdots & M^{0,p} \\ M^{1,0} & M^{1,1} & \cdots & M^{1,p} \\ \vdots & & & \vdots \\ M^{p,0} & M^{p,1} & \cdots & M^{p,p} \end{bmatrix}.$$

Finally, arrange the  $N_{p+1}$  elements of the derivatives  $(1/j!)(D^j m)(x)$  for  $|j| = p + 1$  as a column vector  $\mathcal{D}_{p+1}(x; m)$  using the lexicographical order introduced earlier.

Minimizing (14) with respect to  $b_{\mathbf{k}}$  gives an estimate  $\hat{b}_{\mathbf{k}}(x)$  and  $\tilde{m}(x) = \hat{b}_0(x) = e_1' M_n^{-1} \Psi_n$ , where  $e_1 = (1, 0, \dots, 0)'$  is the vector with the one in the first position,  $M_n(x)$  and  $\Psi_n(x)$  are symmetric  $N \times N$  ( $N = \sum_{\ell=0}^p N_\ell \times 1$ ) matrix and  $N \times 1$  dimensional column vector respectively and are defined as

$$M_n(x) = \begin{bmatrix} M_{n,0,0}(x) & M_{n,0,1}(x) & \dots & M_{n,0,p}(x) \\ \vdots & M_{n,1,1}(x) & \dots & M_{n,1,p}(x) \\ \vdots & & \ddots & \vdots \\ M_{n,p,0}(x) & \dots & \dots & M_{n,p,p}(x) \end{bmatrix}, \quad \Psi_n(x) = \begin{bmatrix} \Psi_{n,0}(x) \\ \Psi_{n,1}(x) \\ \vdots \\ \Psi_{n,p}(x) \end{bmatrix},$$

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

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

and  $\Psi_{n,|j|}(x)$  is a  $N_{|j|}$  dimensional subvector whose  $r$ -th element is given by

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

The estimate of  $m(x)$  is given by  $\tilde{m}(x) = e_1 M_n^{-1} \Psi_n$  and its bias and variance effects can be written as  $\tilde{m}(x) - m(x) = e_1' M_n^{-1}(x) U_n(x) + e_1' M_n^{-1}(x) B_n(x)$ . The stochastic term  $U_n(x)$  and the bias term  $B_n(x)$  are  $N \times 1$  vectors

$$U_n(x) = \begin{bmatrix} U_{n,0}(x) \\ U_{n,1}(x) \\ \vdots \\ U_{n,p}(x) \end{bmatrix}, \quad B_n(x) = \begin{bmatrix} B_{n,0}(x) \\ B_{n,1}(x) \\ \vdots \\ B_{n,d}(x) \end{bmatrix},$$

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

$$[U_{n,|j|}]_r = \frac{1}{nh^d} \sum_{i=1}^n \left( \frac{x - X_i}{h} \right)^{\phi_{|j|}(r)} \mathcal{K} \left( \frac{x - X_i}{h} \right) \varepsilon_i.$$

and

$$[B_{n,|j|}]_r = \frac{1}{nh^d} \sum_{i=1}^n \left( \frac{x - X_i}{h} \right)^{\phi_{|j|}(r)} \mathcal{K} \left( \frac{x - X_i}{h} \right) \Delta_i(x),$$

where  $\Delta_i(x) = m(X_i) - \frac{1}{k!} \sum_{0 \leq |\mathbf{k}| \leq p} (D^{\mathbf{k}}m)(x)(X_i - x)^{\mathbf{k}}$ .

Under assumptions A1-A6 and B1, we have the following results:

$$\begin{aligned} \sup_{x \in \mathcal{X}} |M_n(x) - f(x)M| &= O_p(h + n^{-1/2}h^{-d/2} \log n) \\ \sup_{x \in \mathcal{X}} |\tilde{m}(x) - m(x)| &= O_p(h^{p+1} + n^{-1/2}h^{-d/2} \log n), \end{aligned} \quad (21)$$

which follow from the results of Masry (1996). For notational convenience, now define

$$\bar{\varepsilon}_i = Y_i - \theta_0 = \varepsilon_i + \delta_i,$$

where  $\delta_i = m(X_i) - m(x)$ , and define

$$\check{\varepsilon}_i = Y_i - \tilde{\theta} = \varepsilon_i + [m(X_i) - m(x)] - [\tilde{m}(x) - m(x)] = \varepsilon_i + \delta_i - \nu = \bar{\varepsilon}_i - \nu,$$

where  $\nu = \tilde{m}(x) - m(x)$ . Thus  $\tilde{f}(Y_i - \tilde{\theta})$  can be written as  $\tilde{f}(\check{\varepsilon}_i)$ . To facilitate asymptotic analysis, we also define the kernel density and derivative estimator based on the unobserved errors:

$$\bar{f}(e_i) = \frac{1}{nh} \sum_{j \neq i} k\left(\frac{e_i - \varepsilon_j}{h}\right), \quad \bar{f}'(e_i) = \frac{1}{nh^2} \sum_{j \neq i} k'\left(\frac{e_i - \varepsilon_j}{h}\right).$$

LEMMA A. *Under our conditions*

$$\tilde{f}(\check{\varepsilon}_i) - f(\bar{\varepsilon}_i) = O_p(h^q + n^{-1/2}h^{-d/2}) \text{ for each } i, \quad (22)$$

$$\max_{1 \leq i \leq n} \left| \tilde{f}(\check{\varepsilon}_i) - f(\bar{\varepsilon}_i) \right| = O_p(h^{q-1} + n^{-1/2}h^{-d/2-1} \log(n)). \quad (23)$$

PROOF. First (22). Notice that

$$\begin{aligned} \tilde{f}(\check{\varepsilon}_i) - f(\bar{\varepsilon}_i) &= [\bar{f}(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)] + [\tilde{f}(\check{\varepsilon}_i) - \bar{f}(\bar{\varepsilon}_i)] \\ &= \left[ \frac{1}{nh} \sum_{j \neq i} k\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) - f(\bar{\varepsilon}_i) \right] + \frac{1}{nh} \sum_{j \neq i} \left[ k\left(\frac{\check{\varepsilon}_i - \tilde{\varepsilon}_j}{h}\right) - k\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \right]. \end{aligned} \quad (24)$$

The first term,  $\bar{f}(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)$ , is just the conventional density estimator error and satisfies

$$\bar{f}(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i) = O_p(h^q + n^{-1/2}h^{-1/2}). \quad (25)$$

The second term,  $\tilde{f}(\check{\varepsilon}_i) - \bar{f}(\bar{\varepsilon}_i)$ , contains the effect coming from the preliminary estimation; we show that it is of order  $O_p(h^q + n^{-1/2}h^{-d/2})$ . By second order Taylor expansion, we have

$$\tilde{f}(\check{\varepsilon}_i) - \bar{f}(\bar{\varepsilon}_i) = \frac{1}{nh} \sum_{j \neq i} \left[ k\left(\frac{\check{\varepsilon}_i - \tilde{\varepsilon}_j}{h}\right) - k\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \right]$$

$$\begin{aligned}
&= -\frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) (\check{\varepsilon}_i - \bar{\varepsilon}_i + \tilde{\varepsilon}_j - \varepsilon_j) + \frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\varepsilon_i^* - \varepsilon_j^*}{h} \right) (\check{\varepsilon}_i - \bar{\varepsilon}_i + \tilde{\varepsilon}_j - \varepsilon_j)^2 \\
&= \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) [\tilde{m}(x) - m(x)] - \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) [\tilde{m}(X_j) - m(X_j)] \\
&\quad + \frac{1}{2nh^3} \sum_{j \neq i} k'' \left( \frac{\varepsilon_i^* - \varepsilon_j^*}{h} \right) [\tilde{m}(X_j) - m(X_j) - (\tilde{m}(x) - m(x))]^2 \\
&= A_I + A_{II} + A_{III},
\end{aligned}$$

where  $\varepsilon_i^*$  and  $\varepsilon_j^*$  are intermediate values.

The first term,

$$A_I = [\tilde{m}(x) - m(x)] \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) = O_p(h^q + n^{-1/2}h^{-d/2}),$$

since  $\tilde{m}(x) - m(x) = O_p(h^q + n^{-1/2}h^{-d/2})$  and

$$\frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \simeq f'(\bar{\varepsilon}_i)$$

by standard kernel theory.

Regarding  $A_{III}$ , notice that  $\max_{1 \leq j \leq n} |\tilde{m}(X_j) - m(X_j)| = O_p(h^{p+1} + n^{-1/2}h^{-d/2} \log n)$ , so that

$$\begin{aligned}
&\frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\varepsilon_i^* - \varepsilon_j^*}{h} \right) [\tilde{m}(X_j) - m(X_j) - (\tilde{m}(x) - m(x))]^2 \\
&\leq \left( \max_{1 \leq j \leq n} |\tilde{m}(X_j) - m(X_j)| + \sup_{x \in \mathcal{X}} |\tilde{m}(x) - m(x)| \right)^2 \frac{1}{nh^3} \sum_{j \neq i} \left| k'' \left( \frac{\varepsilon_i^* - \varepsilon_j^*}{h} \right) \right| \\
&\leq \left( \max_{1 \leq j \leq n} |\tilde{m}(X_j) - m(X_j)| + \sup_{x \in \mathcal{X}} |\tilde{m}(x) - m(x)| \right)^2 \frac{1}{h^3} \sup_u |k''(u)| \\
&\leq O_p(h^{2p+2} + n^{-1}h^{-d} \log n) h^{-3} = O_p(h^{2p-1} + n^{-1}h^{-(d+3)} \log n).
\end{aligned}$$

Under the bandwidth condition A5, the remainder term is of order  $O_p(h^{p+1} + n^{-1/2}h^{-d/2})$ .

We now turn to the proof of the magnitude of  $A_{II}$ . This calculation is quite long. The general strategy is to expand out the random denominator of  $\tilde{m}(X_j) - m(X_j)$  around its probability limit and then calculate the moments of the resulting degenerate U-statistics term by term. Notice that

$$\tilde{m}(X_j) - m(X_j) = e_1' M_n^{-1}(X_j) U_n(X_j) + e_1' M_n^{-1}(X_j) B_n(X_j),$$



so we have

$$\begin{aligned}
A_{II} &= \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) [\tilde{m}(X_j) - m(X_j)] \\
&= \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) e_1' M_n^{-1}(X_j) U_n(X_j) + \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) e_1' M_n^{-1}(X_j) B_n(X_j) \\
&= A_{IIA} + A_{IIB}.
\end{aligned}$$

We expand  $M_n^{-1}(X_j)$  around its limit  $(Mf(X_j))^{-1}$  and get

$$\begin{aligned}
A_{IIA} &= \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) e_1' [Mf_X(X_j)]^{-1} U_n(X_j) \tag{26} \\
&+ \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \sum_{\rho=1}^{p+1} e_1' \{ [Mf_X(X_j)]^{-1} [M_n(X_j) - Mf_X(X_j)] \}^\rho [Mf_X(X_j)]^{-1} U_n(X_j) \\
&+ \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) e_1' \{ [Mf_X(X_j)]^{-1} [M_n(X_j) - Mf_X(X_j)] \}^{p+2} M_n^{-1}(X_j) U_n(X_j).
\end{aligned}$$

Since  $M^{0,l}$  are  $1 \times N_l$  row vectors, we have

$$\begin{aligned}
&\frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) e_1' [Mf_X(X_j)]^{-1} U_n(X_j) \\
&= \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) f_X(X_j)^{-1} \sum_{l=0}^p M^{0,l} U_{n,l}(X_j) \\
&= \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) f_X(X_j)^{-1} \sum_{\kappa} \omega^{0,\kappa} \left( \frac{1}{nh_n^d} \sum_{l=1}^n \left( \frac{X_j - X_l}{h} \right)^\kappa \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \varepsilon_l \right),
\end{aligned}$$

where  $\omega^{0,\kappa}$  are elements in the first row of  $M^{-1}$  and the sum over  $\kappa$  is over a finite index set. Thus

$$\begin{aligned}
&\frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) e_1' [Mf_X(X_j)]^{-1} U_n(X_j) \tag{27} \\
&= \sum_{j \neq i} \sum_{\kappa} \omega^{0,\kappa} \frac{1}{n^2 h^{2+d}} f_X(X_j)^{-1} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \mathcal{K} \left( \frac{X_j - X_i}{h} \right) \left( \frac{X_j - X_i}{h} \right)^\kappa \varepsilon_i \\
&+ \sum_{j \neq i} \sum_{l \neq j, l \neq i} \sum_{\kappa} \omega^{0,\kappa} \frac{1}{n^2 h^{2+d}} f_X(X_j)^{-1} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \left( \frac{X_j - X_l}{h} \right)^\kappa \varepsilon_l.
\end{aligned}$$

By the i.i.d. assumption, we have

$$\begin{aligned}
&\frac{1}{n} \sum_{j \neq i} \sum_{\kappa} \omega^{0,\kappa} \frac{1}{nh^{2+d}} f_X(X_j)^{-1} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \mathcal{K} \left( \frac{X_j - X_i}{h} \right) \left( \frac{X_j - X_i}{h} \right)^\kappa \varepsilon_i \\
&\simeq \sum_{\kappa} \omega^{0,\kappa} \frac{1}{nh^{2+d}} E_i \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \varepsilon_i \right] E_i \left[ f_X(X_j)^{-1} \mathcal{K} \left( \frac{X_j - X_i}{h} \right) \left( \frac{X_j - X_i}{h} \right)^\kappa \right] = O_p(n^{-1}),
\end{aligned}$$

since [using integration by parts, a change of variables and dominated convergence]:

$$E_i k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) = h \int k \left( \frac{\bar{\varepsilon}_i - \varepsilon}{h} \right) f'(\varepsilon) d\varepsilon = h^2 \int k(u) f'(\bar{\varepsilon}_i - uh) du,$$

and

$$E_i \frac{1}{f_X(X_j)} \mathcal{K} \left( \frac{X_j - X_i}{h} \right) \left( \frac{X_j - X_i}{h} \right)^\kappa = \int \mathcal{K} \left( \frac{X_j - X_i}{h} \right) \left( \frac{X_j - X_i}{h} \right)^\kappa dX_j = h^d \int K(u) u^\kappa du.$$

For the second term in (27), if we denote

$$\phi_1(i, j, l) = \frac{1}{n^2 h^{2+d}} f_X(X_j)^{-1} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \left( \frac{X_j - X_l}{h} \right)^\kappa \varepsilon_l,$$

we only need to verify the magnitude of

$$\sum_{j \neq i} \sum_{l \neq j, l \neq i} \phi_1(i, j, l). \quad (28)$$

Notice that (28) is mean zero with variance

$$O(n^2) E[\phi_1(i, j, l)^2 + \phi_1(i, j, l) \phi_1(i, l, j)] + O(n^3) E[\phi_1(i, j, l) \phi_1(i, r, l)].$$

The orders of  $E[\phi_1(i, j, l)^2]$ ,  $E\phi_1(i, j, l) \phi_1(i, l, j)$ ,  $E[\phi_1(i, j, l) \phi_1(i, r, l)]$  can be verified. Under our assumptions, we have:

$$E_i k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right)^2 = h \int k'(u)^2 f(\bar{\varepsilon}_i - uh) du = O_p(h), \quad (29)$$

$$\begin{aligned} & E_i k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \varepsilon_j k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_l}{h} \right) \varepsilon_l \quad (30) \\ &= \int k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \varepsilon_j k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_l}{h} \right) \varepsilon_l f(\varepsilon_j) f(\varepsilon_l) d\varepsilon_j d\varepsilon_l \\ &= h^2 \int k'(u_j) (\bar{\varepsilon}_i - hu_j) f(\bar{\varepsilon}_i - hu_j) du_j \int k'(u_l) (\bar{\varepsilon}_i - hu_l) f(\bar{\varepsilon}_i - hu_l) du_l \\ &= h^4 \bar{\varepsilon}_i^2 \int k(u_j) f'(\bar{\varepsilon}_i - hu_j) du_j \int k(u_l) f'(\bar{\varepsilon}_i - hu_l) du_l \\ &\quad - h^4 \bar{\varepsilon}_i \int k'(u_j) u_j f(\bar{\varepsilon}_i - hu_j) du_j \int k(u_l) f'(\bar{\varepsilon}_i - hu_l) du_l \\ &\quad - h^4 \bar{\varepsilon}_i \int k(u_j) f'(\bar{\varepsilon}_i - hu_j) du_j \int k'(u_l) u_l f(\bar{\varepsilon}_i - hu_l) du_l \\ &\quad + h^4 \int k'(u_j) u_j f(\bar{\varepsilon}_i - hu_j) du_j \int k'(u_l) u_l f(\bar{\varepsilon}_i - hu_l) du_l \\ &= O_p(h^4), \end{aligned}$$

$$Ek' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_r}{h} \right) \varepsilon_l^2 = O_p(h^4), \quad (31)$$

$$E_i \left[ \frac{1}{f_X(X_j)} \mathcal{K} \left( \frac{X_j - X_i}{h} \right) \left( \frac{X_j - X_i}{h} \right)^\kappa \right]^2 = O_p(h^d), \quad (32)$$

$$Ef_X(X_j)^{-1} f_X(X_l)^{-1} \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \mathcal{K} \left( \frac{X_l - X_j}{h} \right) \left( \frac{X_j - X_l}{h} \right)^\kappa \left( \frac{X_l - X_j}{h} \right)^\kappa = O(h^d), \quad (33)$$

and

$$E_l \left[ f_X(X_j)^{-1} \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \left( \frac{X_j - X_l}{h} \right)^\kappa \right] E_l \left[ f_X(X_r)^{-1} \mathcal{K} \left( \frac{X_r - X_l}{h} \right) \left( \frac{X_r - X_l}{h} \right)^\kappa \right] = O(h^{2d}). \quad (34)$$

Therefore, it can be verified that:

$$\begin{aligned} E[\phi_1(i, j, l)^2] &= E \left[ \frac{1}{n^2 h^{2+d}} f_X(X_j)^{-1} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \left( \frac{X_j - X_l}{h} \right)^\kappa \varepsilon_l \right]^2 \\ &= E \left[ \frac{1}{n^4 h^{4+2d}} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right)^2 \varepsilon_l^2 f_X(X_j)^{-2} \mathcal{K} \left( \frac{X_j - X_l}{h} \right)^2 \left( \frac{X_j - X_l}{h} \right)^{2\kappa} \right] \\ &= O\left(\frac{1}{n^4 h^{3+d}}\right), \end{aligned}$$

$$\begin{aligned} E[\phi_1(i, j, l)\phi_1(i, l, j)] &= \frac{1}{n^4 h^{4+2d}} Ek' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \varepsilon_j k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_l}{h} \right) \varepsilon_l \\ &\quad Ef_X(X_j)^{-1} f_X(X_l)^{-1} \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \mathcal{K} \left( \frac{X_l - X_j}{h} \right) \left( \frac{X_j - X_l}{h} \right)^\kappa \left( \frac{X_l - X_j}{h} \right)^\kappa \\ &= O\left(\frac{1}{n^4 h^d}\right), \end{aligned}$$

and

$$\begin{aligned} E[\phi_1(i, j, l)\phi_1(i, r, l)] &= \frac{1}{n^4 h^{4+2d}} Ek' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_r}{h} \right) \varepsilon_l^2 \\ &\quad f_X(X_j)^{-1} \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \left( \frac{X_j - X_l}{h} \right)^\kappa f_X(X_r)^{-1} \mathcal{K} \left( \frac{X_r - X_l}{h} \right) \left( \frac{X_r - X_l}{h} \right)^\kappa \\ &= O(n^{-4}). \end{aligned}$$

Thus,

$$\frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) e_1' [M f_X(X_j)]^{-1} U_n(X_j) = O_p(n^{-1/2}).$$

Similarly, we can verify that the higher order terms in  $A_{IIA}$

$$\frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) e'_1 \{ [Mf_X(X_j)]^{-1} [M_n(X_j) - Mf_X(X_j)] \}^p [Mf_X(X_j)]^{-1} U_n(X_j)$$

are of small order of magnitude. For example, for

$$\frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) e'_1 [Mf_X(X_j)]^{-1} [M_n(X_j) - Mf_X(X_j)] [Mf_X(X_j)]^{-1} U_n(X_j),$$

notice that

$$\begin{aligned} & e'_1 [Mf_X(X_j)]^{-1} [M_n(X_j) - Mf_X(X_j)] [Mf_X(X_j)]^{-1} U_n(X_j) \\ &= f_X(X_j)^{-2} \sum_{l=0}^p \sum_{r=0}^p \sum_{s=0}^p M^{0,l} [M_{n,l,r} - f(X_j) M_{l,r}] M^{r,s} U_{ns}(X_j) \\ &= f_X(X_j)^{-2} \sum_{\gamma} \sum_{\beta} \sum_{\kappa} \omega^{0,\gamma} \mu_{\gamma,\beta}(X_j) \omega^{\beta,\kappa} u_{\kappa}(X_j), \end{aligned}$$

where  $\omega^{\bullet,\bullet}$  are elements in  $M^{-1}$ ,

$$u_{\kappa}(X_j) = \frac{1}{nh_n^d} \sum_{l=1}^n \left( \frac{X_j - X_l}{h} \right)^{\kappa} \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \varepsilon_l$$

and

$$\begin{aligned} \mu_{\gamma,\beta}(X_j) &= \frac{1}{nh_n^d} \sum_{l \neq j} \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \left( \frac{X_j - X_l}{h} \right)^{\gamma+\beta} - f(X_j) \omega_{\gamma,\beta} \\ &= \frac{1}{nh_n^d} \sum_{l \neq j} \left[ \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \left( \frac{X_j - X_l}{h} \right)^{\gamma+\beta} - h^d f(X_j) \omega_{\gamma,\beta} \right], \end{aligned}$$

where  $\omega_{\bullet,\bullet}$  are elements in  $M$ . Thus

$$\begin{aligned} & \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) e'_1 [Mf_X(X_j)]^{-1} [M_n(X_j) - Mf_X(X_j)] [Mf_X(X_j)]^{-1} U_n(X_j) \\ &= \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) f_X(X_j)^{-2} \sum_{\gamma} \sum_{\beta} \sum_{\kappa} \omega^{0,\gamma} \omega^{\beta,\kappa} \mu_{\gamma,\beta}(X_j) u_{\kappa}(X_j) \\ &= \frac{1}{n^3 h^{2+2d}} \sum_{\gamma} \sum_{\beta} \sum_{\kappa} \omega^{0,\gamma} \omega^{\beta,\kappa} \sum_{j \neq i} \sum_{s \neq j} \sum_{l \neq j} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \varepsilon_s \\ & \quad \times \left[ \mathcal{K} \left( \frac{X_j - X_s}{h} \right) \left( \frac{X_j - X_s}{h} \right)^{\kappa} \right] \left[ \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \left( \frac{X_j - X_l}{h} \right)^{\gamma+\beta} - h^d f(X_j) \omega_{\gamma,\beta} \right] \\ &= \frac{1}{n^3 h^{2+2d}} \sum_{\gamma} \sum_{\beta} \sum_{\kappa} \omega^{0,\gamma} \omega^{\beta,\kappa} \sum_{j \neq i} \sum_{l \neq j} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \varepsilon_i \end{aligned}$$

$$\begin{aligned}
& \left[ \mathcal{K} \left( \frac{X_j - X_i}{h} \right) \left( \frac{X_j - X_i}{h} \right)^\kappa \right] \left[ \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \left( \frac{X_j - X_l}{h} \right)^{\gamma+\beta} - h^d f(X_j) \omega_{\gamma,\beta} \right] \\
& + \frac{1}{n^3 h^{2+2d}} \sum_{\gamma} \sum_{\beta} \sum_{\kappa} \omega^{0,\gamma} \omega^{\beta,\kappa} \sum_{\substack{j \neq i \\ s \neq j \\ l \neq j \\ s \neq i}} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \varepsilon_s \\
& \left[ \mathcal{K} \left( \frac{X_j - X_s}{h} \right) \left( \frac{X_j - X_s}{h} \right)^\kappa \right] \left[ \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \left( \frac{X_j - X_l}{h} \right)^{\gamma+\beta} - h^d f(X_j) \omega_{\gamma,\beta} \right].
\end{aligned}$$

Thus,

$$\begin{aligned}
& \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) e_1' [Mf_X(X_j)]^{-1} [M_n(X_j) - Mf_X(X_j)] [Mf_X(X_j)]^{-1} U_n(X_j) \\
& = \sum_{\gamma} \sum_{\beta} \sum_{\kappa} \omega^{0,\gamma} \omega^{\beta,\kappa} \sum_{j \neq i} \sum_{l \neq j} \phi_2(j, l) + \sum_{\gamma} \sum_{\beta} \sum_{\kappa} \omega^{0,\gamma} \omega^{\beta,\kappa} \sum_{\substack{j \neq i \\ s \neq j \\ l \neq j \\ s \neq i}} \phi_3(j, s, l) \\
& = A_{IIA1} + A_{IIA2},
\end{aligned}$$

where

$$\begin{aligned}
\phi_2(j, l) & = \frac{1}{n^3 h^{2+2d}} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \varepsilon_i \left[ \mathcal{K} \left( \frac{X_j - X_i}{h} \right) \left( \frac{X_j - X_i}{h} \right)^\kappa \right] \\
& \quad \times \left[ \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \left( \frac{X_j - X_l}{h} \right)^{\gamma+\beta} - h^d f(X_j) \omega_{\gamma,\beta} \right] \\
\phi_3(j, s, l) & = \frac{1}{n^3 h^{2+2d}} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \varepsilon_s \left[ \mathcal{K} \left( \frac{X_j - X_s}{h} \right) \left( \frac{X_j - X_s}{h} \right)^\kappa \right] \\
& \quad \times \left[ \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \left( \frac{X_j - X_l}{h} \right)^{\gamma+\beta} - h^d f(X_j) \omega_{\gamma,\beta} \right].
\end{aligned}$$

It can be easily verified that

$$E \left[ \sum_{j \neq i} \sum_{l \neq j} \phi_2(j, l) \right] = O(n^2 h^{3+2d} \times \frac{1}{n^3 h^{2+2d}}) = O(n^{-1} h)$$

and

$$E \left[ \sum_{j \neq i} \sum_{l \neq j} \phi_2(j, l) \right]^2 = O(n^2 h^{1+2d} \times \frac{1}{n^6 h^{4+4d}} + n^4 h^{4d+6} \times \frac{1}{n^6 h^{4+4d}}) = O(n^{-4} h^{-3-2d} + n^{-2} h^2),$$

thus  $A_{IIA1} = O_p(n^{-1} h)$ . For  $A_{IIA2}$ , notice it is mean zero and

$$\text{var} \left[ \sum_{\substack{j \neq i \\ s \neq j \\ l \neq j \\ s \neq i}} \phi_3(j, s, l) \right] = O(n^3)I + O(n^4)II + O(n^5)III,$$

where  $I$  includes terms like  $E\phi_3(j, s, l)^2 + E\phi_3(j, s, l)\phi_3(s, j, l)$ ,  $II$  includes terms with four different indices like  $E\phi_3(j, s, l)\phi_3(j, s, \beta) + E\phi_3(j, s, l)\phi_3(\alpha, s, l)$ , and  $III$  includes terms with five different indices like  $E\phi_3(j, s, l)\phi_3(\alpha, s, \beta)$ . By a verification of moments, we can see that

$$\text{var} \left[ \sum_{\substack{j \neq i \\ s \neq j \\ l \neq j \\ s \neq i}} \phi_3(j, s, l) \right] = O(n^3)I + O(n^4)II + O(n^5)III = O(n^{-1}h^2).$$

Thus,  $A_{IIA2} = O_p(n^{-1/2}h)$ . For

$$\begin{aligned} & \left| \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) e'_1 \{ [Mf_X(X_j)]^{-1} [M_n(X_j) - Mf_X(X_j)] \}^{p+2} M_n^{-1}(X_j) U_n(X_j) \right| \\ & \leq \max_{1 \leq j \leq n} |e'_1 [Mf_X(X_j)]^{-1} [M_n(X_j) - Mf_X(X_j)]^{p+2} [Mf_X(X_j)]^{-1} [M_n(X_j) - Mf_X(X_j)] M_n^{-1}(X_j) U_n(X_j)| \\ & \quad \times \frac{1}{nh^2} \sum_{j \neq i} \left| k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right| \\ & = O_p(h^{p+2} \times h \times \frac{1}{nh^2} \times n) = O_p(h^{p+1}). \end{aligned}$$

For the bias term  $A_{IIB}$ ,

$$\begin{aligned} A_{IIB} &= \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) e'_1 [Mf_X(X_j)]^{-1} B_n(X_j) \\ & \quad + \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) e'_1 [Mf_X(X_j)]^{-1} [M_n(X_j) - Mf_X(X_j)] M_n^{-1}(X_j) B_n(X_j). \end{aligned}$$

Similarly, by a verification on the moments of U-statistics, we have

$$\begin{aligned} & \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) e'_1 [Mf_X(X_j)]^{-1} B_n(X_j) \\ & \simeq \sum_{j \neq i} \sum_{l \neq j} \sum_{\kappa} \omega^{1, \kappa} \frac{1}{n^2 h^{2+d}} f_X(X_j)^{-1} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \left( \frac{X_j - X_l}{h} \right)^{p+\kappa} h^{p+1} m_{p+1}(X_j) \\ & = O_p(h^{p+1}) \end{aligned}$$

and

$$\begin{aligned} & \left| \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) e'_1 [Mf_X(X_j)]^{-1} [M_n(X_j) - Mf_X(X_j)] M_n^{-1}(X_j) B_n(X_j) \right| \\ & \leq \max_{1 \leq j \leq n} |e'_1 [Mf_X(X_j)]^{-1} [M_n(X_j) - Mf_X(X_j)] M_n^{-1}(X_j) B_n(X_j)| \frac{1}{nh^2} \sum_{j \neq i} \left| k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right| \\ & = O_p(h^{p+2}). \end{aligned}$$

In conclusion,  $A_{II} = o_p(n^{-1/2}h^{-d/2} + h^q)$ .

Now to the proof of the uniform convergence result (23). As above, we decompose  $\tilde{f}(\tilde{\varepsilon}_i) - f(\tilde{\varepsilon}_i)$  as in (24). Notice that the first term,  $\bar{f}(\tilde{\varepsilon}_i) - f(\tilde{\varepsilon}_i)$ , which is just the conventional density estimator error, satisfies

$$\max_{1 \leq i \leq n} |\bar{f}(\tilde{\varepsilon}_i) - f(\tilde{\varepsilon}_i)| = O_p(h^q + n^{-1/2}h^{-1/2} \log(n)).$$

The second term in (24) can be further decomposed as

$$\begin{aligned} \tilde{f}(\tilde{\varepsilon}_i) - \bar{f}(\tilde{\varepsilon}_i) &= \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\tilde{\varepsilon}_i - \varepsilon_j}{h} \right) [\tilde{m}(x) - m(x)] - \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\tilde{\varepsilon}_i - \varepsilon_j}{h} \right) [\tilde{m}(X_j) - m(X_j)] \\ &\quad + \frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\varepsilon_i^* - \varepsilon_j^*}{h} \right) [\tilde{m}(X_j) - m(X_j) - (\tilde{m}(x) - m(x))]^2, \end{aligned} \quad (35)$$

where  $\varepsilon_i^*$  and  $\varepsilon_j^*$  are intermediate values. Notice that  $\frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\tilde{\varepsilon}_i - \varepsilon_j}{h} \right)$  is the conventional kernel estimator of  $f'(\tilde{\varepsilon}_i)$ , we have denoted as  $\bar{f}'(\tilde{\varepsilon}_i)$ , and write the first term in (35) as

$$[\tilde{m}(x) - m(x)] \bar{f}'(\tilde{\varepsilon}_i) = [\tilde{m}(x) - m(x)] f'(\tilde{\varepsilon}_i) + [\tilde{m}(x) - m(x)] [\bar{f}'(\tilde{\varepsilon}_i) - f'(\tilde{\varepsilon}_i)].$$

By Assumptions A2 that  $f'(\cdot)$  is uniformly bounded and by standard results that

$$\begin{aligned} \sup_{x \in \mathcal{X}} |\tilde{m}(x) - m(x)| &= O_p(h^q + n^{-1/2}h^{-d/2} \log(n)), \\ \max_{1 \leq i \leq n} |\bar{f}'(\tilde{\varepsilon}_i) - f'(\tilde{\varepsilon}_i)| &= O_p(h^q + n^{-1/2}h^{-3/2} \log(n)) = o_p(1), \end{aligned}$$

we have

$$\max_{1 \leq i \leq n} \left| \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\tilde{\varepsilon}_i - \varepsilon_j}{h} \right) [\tilde{m}(x) - m(x)] \right| = O_p(h^q + n^{-1/2}h^{-d/2} \log(n)).$$

For the second term in (35),

$$\begin{aligned} &\left| \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\tilde{\varepsilon}_i - \varepsilon_j}{h} \right) [\tilde{m}(X_j) - m(X_j)] \right| \\ &\leq \max_{1 \leq j \leq n} |\tilde{m}(X_j) - m(X_j)| \frac{1}{nh^2} \sum_{j \neq i} \left| k' \left( \frac{\tilde{\varepsilon}_i - \varepsilon_j}{h} \right) \right|. \end{aligned}$$

By i.i.d. property and conditional on  $\varepsilon_i$  and  $X_i$ ,

$$\frac{1}{nh^2} \sum_{j \neq i} \left| k' \left( \frac{\tilde{\varepsilon}_i - \varepsilon_j}{h} \right) \right| \simeq E_i \frac{1}{h^2} \left| k' \left( \frac{\tilde{\varepsilon}_i - \varepsilon_j}{h} \right) \right|,$$

where

$$E_i \frac{1}{h^2} \left| k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right| = \int \frac{1}{h^2} \left| k' \left( \frac{\bar{\varepsilon}_i - \varepsilon}{h} \right) \right| f(\varepsilon) d\varepsilon = \frac{1}{h} \int |k'(u)| f(\bar{\varepsilon}_i - hu) du.$$

Again, notice that  $f(\bar{\varepsilon}_i - hu)$  is uniformly bounded, thus

$$\max_{1 \leq i \leq n} \left| \frac{1}{h} \int |k'(u)| f(\bar{\varepsilon}_i - hu) du \right| = O_p\left(\frac{1}{h}\right).$$

In addition, notice that

$$\max_{1 \leq j \leq n} |\tilde{m}(X_j) - m(X_j)| = O_p(h^q + n^{-1/2} h^{-d/2} \log(n)),$$

so that

$$\max_{1 \leq i \leq n} \left| \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) [\tilde{m}(X_j) - m(X_j)] \right| = O_p(h^{q-1} + n^{-1/2} h^{-d/2-1} \log(n)).$$

For the third term in (35)

$$\begin{aligned} & \frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\varepsilon_i^* - \varepsilon_j^*}{h} \right) [\tilde{m}(X_j) - m(X_j) - (\tilde{m}(x) - m(x))]^2 \\ & \leq \left( \max_{1 \leq j \leq n} |\tilde{m}(X_j) - m(X_j)| + \sup_{x \in \mathcal{X}} |\tilde{m}(x) - m(x)| \right)^2 \frac{1}{h^3} \sup_u |k''(u)| \\ & = o_p(h^{q-1} + n^{-1/2} h^{-d/2-1} \log(n)). \end{aligned}$$

Thus,

$$\begin{aligned} \max_{1 \leq i \leq n} \left| \tilde{f}(\check{\varepsilon}_i) - f(\bar{\varepsilon}_i) \right| & \leq \max_{1 \leq i \leq n} \left| \tilde{f}(\check{\varepsilon}_i) - \bar{f}(\bar{\varepsilon}_i) \right| + \max_{1 \leq i \leq n} \left| \bar{f}(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i) \right| \\ & = O_p(h^{q-1} + n^{-1/2} h^{-d/2-1} \log(n)) \end{aligned}$$

as required. ■

REMARK 1. The uniform rate given by (23) is not the best result, but suffices our purpose of proofs in this paper. In fact, following a similar analysis as Masry (1996), a better rate ( $O_p(h^q + n^{-1/2} h^{-d/2} \log(n))$ ) could be obtained with substantially more complicated analysis.

REMARK 2. The above results can be extended to estimates of derivatives based on similar analysis. In particular, if

$$\tilde{f}^{(j)}(\check{\varepsilon}_i) = \frac{1}{nh^{j+1}} \sum_{j \neq i} k^{(j)} \left( \frac{\check{\varepsilon}_i - \check{\varepsilon}_j}{h} \right), \quad j = 1, 2, 3,$$



under our regularity conditions

$$\max_{1 \leq i \leq n} \left| \tilde{f}^{(j)}(\check{\varepsilon}_i) - f^{(j)}(\bar{\varepsilon}_i) \right| = O_p(h^{q-j-1} + n^{-1/2}h^{-d/2-j-1} \log(n)). \quad (36)$$

The above results also hold for any  $\check{\varepsilon}_i$  in a small neighborhood of  $\bar{\varepsilon}_i$ , say,  $\check{\varepsilon}_i - \bar{\varepsilon}_i = O_p(n^{-1/2}h^{-d/2})$ .

## 5.2 Proof of Theorem 1

We expand  $\tilde{S}_n(\tilde{\theta})$  around the true value of  $\theta$ ,  $\theta_0$ , and obtain

$$\tilde{S}_n(\tilde{\theta}) = \tilde{S}_n(\theta_0) - \tilde{\mathcal{I}}_n(\theta_0)(\tilde{\theta} - \theta_0) + \frac{1}{2} \frac{\partial^2 \tilde{S}_n}{\partial \theta^2}(\theta^*)(\tilde{\theta} - \theta_0)^2, \quad (37)$$

where  $\theta^*$  is an intermediate value between  $\theta_0$  and  $\tilde{\theta}$ . This is the complete expansion corresponding to (16). Under the given assumptions, the preliminary estimator  $\tilde{\theta}(x)$  given by (12) is consistent and indeed satisfies  $\tilde{\theta}(x) - \theta_0(x) = O_p(n^{-1/2}h^{-d/2} + h^q) = O_p(n^{-1/2}h^{-d/2})$ . In the following three subsections, we derive the asymptotic results for the Hessian  $\tilde{\mathcal{I}}_n(\tilde{\theta})$ , the score  $\tilde{S}_n(\theta_0)$ , and the remainder term  $R_n(\theta^*) = (\partial^2 \tilde{S}_n(\theta^*)/\partial \theta^2)(\tilde{\theta} - \theta_0)^2/2$ .

### 5.2.1 The Hessian

We want to show that

$$\tilde{\mathcal{I}}_n(\tilde{\theta}; \tilde{f}) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \left[ \frac{\tilde{f}'}{\tilde{f}}(Y_i - \tilde{\theta}) \right]^2 G_b(\tilde{f}_i) \xrightarrow{p} f_X(x) I(f). \quad (38)$$

We write  $\tilde{f}(Y_i - \tilde{\theta})$  and  $\tilde{f}'(Y_i - \tilde{\theta})$  as  $\tilde{f}(\check{\varepsilon}_i)$  and  $\tilde{f}'(\check{\varepsilon}_i)$  respectively. Thus

$$\tilde{\mathcal{I}}_n(\tilde{\theta}) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \left[ \frac{\tilde{f}'(\check{\varepsilon}_i)}{\tilde{f}(\check{\varepsilon}_i)} \right]^2 G_i.$$

We decompose  $\tilde{f}(\check{\varepsilon}_i)$  as  $f(\bar{\varepsilon}_i) + [\tilde{f}(\check{\varepsilon}_i) - f(\bar{\varepsilon}_i)]$ , and  $\tilde{f}'(\check{\varepsilon}_i)$  as  $f'(\bar{\varepsilon}_i) + [\tilde{f}'(\check{\varepsilon}_i) - f'(\bar{\varepsilon}_i)]$ . By a geometric expansion  $\tilde{f}(\check{\varepsilon}_i)^{-2}$  can be written as

$$f(\bar{\varepsilon}_i)^{-2} - f(\bar{\varepsilon}_i)^{-4}(\tilde{f}(\check{\varepsilon}_i) + f(\bar{\varepsilon}_i))(\tilde{f}(\check{\varepsilon}_i) - f(\bar{\varepsilon}_i)) + f(\bar{\varepsilon}_i)^{-4}\tilde{f}(\check{\varepsilon}_i)^{-2}(\tilde{f}(\check{\varepsilon}_i) + f(\bar{\varepsilon}_i))^2(\tilde{f}(\check{\varepsilon}_i) - f(\bar{\varepsilon}_i))^2.$$

Thus

$$\begin{aligned} \tilde{\mathcal{I}}_n(\tilde{\theta}) &= \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) G_b(\tilde{f}_i) \left( f'(\bar{\varepsilon}_i) + [\tilde{f}'(\check{\varepsilon}_i) - f'(\bar{\varepsilon}_i)] \right)^2 \times \\ &\quad \left( \frac{1}{f(\bar{\varepsilon}_i)^2} - \frac{[\tilde{f}(\check{\varepsilon}_i) + f(\bar{\varepsilon}_i)][\tilde{f}(\check{\varepsilon}_i) - f(\bar{\varepsilon}_i)]}{f(\bar{\varepsilon}_i)^4} + \frac{[\tilde{f}(\check{\varepsilon}_i) + f(\bar{\varepsilon}_i)]^2[\tilde{f}(\check{\varepsilon}_i) - f(\bar{\varepsilon}_i)]^2}{f(\bar{\varepsilon}_i)^4 \tilde{f}(\check{\varepsilon}_i)^2} \right) \\ &= \mathcal{J}_1 + \mathcal{J}_2 + \mathcal{J}_3 - \mathcal{J}_4 + \mathcal{J}_5 + \mathcal{J}_6 - \mathcal{J}_7 - \mathcal{J}_8 + \mathcal{J}_9 + \mathcal{J}_{10}, \end{aligned}$$

where:

$$\begin{aligned}
\mathcal{J}_1 &= \frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \left[ \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \right]^2 \\
\mathcal{J}_2 &= \frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \left[ \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \right]^2 [1 - G_b(\tilde{f}_i)] \\
\mathcal{J}_3 &= \frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) [\tilde{f}'(\check{\varepsilon}_i) - f'(\bar{\varepsilon}_i)]}{f(\bar{\varepsilon}_i)^2} G_b(\tilde{f}_i) \\
\mathcal{J}_4 &= \frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)^2}{f(\bar{\varepsilon}_i)^4} (\tilde{f}(\check{\varepsilon}_i) + f(\bar{\varepsilon}_i)) (\tilde{f}(\check{\varepsilon}_i) - f(\bar{\varepsilon}_i)) G_b(\tilde{f}_i) \\
\mathcal{J}_5 &= \frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) [\tilde{f}'(\check{\varepsilon}_i) - f'(\bar{\varepsilon}_i)]^2}{f(\bar{\varepsilon}_i)^2} G_b(\tilde{f}_i) \\
\mathcal{J}_6 &= \frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)^2}{f(\bar{\varepsilon}_i)^4 \tilde{f}(\check{\varepsilon}_i)^2} (\tilde{f}(\check{\varepsilon}_i) + f(\bar{\varepsilon}_i))^2 (\tilde{f}(\check{\varepsilon}_i) - f(\bar{\varepsilon}_i))^2 G_b(\tilde{f}_i) \\
\mathcal{J}_7 &= \frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) [\tilde{f}'(\check{\varepsilon}_i) - f'(\bar{\varepsilon}_i)]}{f(\bar{\varepsilon}_i)^4} (\tilde{f}(\check{\varepsilon}_i) + f(\bar{\varepsilon}_i)) (\tilde{f}(\check{\varepsilon}_i) - f(\bar{\varepsilon}_i)) G_b(\tilde{f}_i) \\
\mathcal{J}_8 &= \frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{[\tilde{f}'(\check{\varepsilon}_i) - f'(\bar{\varepsilon}_i)]^2}{f(\bar{\varepsilon}_i)^4} (\tilde{f}(\check{\varepsilon}_i) + f(\bar{\varepsilon}_i)) (\tilde{f}(\check{\varepsilon}_i) - f(\bar{\varepsilon}_i)) G_b(\tilde{f}_i) \\
\mathcal{J}_9 &= \frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) [\tilde{f}'(\check{\varepsilon}_i) - f'(\bar{\varepsilon}_i)]}{f(\bar{\varepsilon}_i)^4 \tilde{f}(\check{\varepsilon}_i)^2} (\tilde{f}(\check{\varepsilon}_i) + f(\bar{\varepsilon}_i))^2 (\tilde{f}(\check{\varepsilon}_i) - f(\bar{\varepsilon}_i))^2 G_b(\tilde{f}_i) \\
\mathcal{J}_{10} &= \frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{[\tilde{f}'(\check{\varepsilon}_i) - f'(\bar{\varepsilon}_i)]^2}{f(\bar{\varepsilon}_i)^4 \tilde{f}(\check{\varepsilon}_i)^2} (\tilde{f}(\check{\varepsilon}_i) + f(\bar{\varepsilon}_i))^2 (\tilde{f}(\check{\varepsilon}_i) - f(\bar{\varepsilon}_i))^2 G_b(\tilde{f}_i).
\end{aligned}$$

We show that

$$\mathcal{J}_1 \xrightarrow{p} f_X(x)I(f),$$

and

$$\mathcal{J}_j = o_p(1), \quad j = 2, \dots, 10.$$

This is carried out in a series of lemmas given below.

LEMMA H1. *Under our conditions*

$$\mathcal{J}_1 \xrightarrow{p} f_X(x)I(f).$$

PROOF.  $\mathcal{J}_1$  contains the errors coming from the local deviation from  $X_i$  to  $x$  in terms of  $\delta_i = m(X_i) - m(x) = \bar{\varepsilon}_i - \varepsilon_i$ , which is largely determined by the smoothness property of  $m(\cdot)$ . For simplicity of exposition, we denote  $f'/f$  as  $\psi$ . Thus,

$$\begin{aligned}\mathcal{J}_1 &= \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \psi(\bar{\varepsilon}_i)^2 \\ &= \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \psi(\varepsilon_i)^2 + \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) [\psi(\bar{\varepsilon}_i)^2 - \psi(\varepsilon_i)^2] \\ &= \mathcal{J}_{11} + \mathcal{J}_{12}.\end{aligned}$$

Since  $\{\varepsilon_i\}$  and  $\{X_i\}$  are i.i.d. and are mutually independent,

$$E\left[K\left(\frac{x - X_i}{h}\right) \psi(\varepsilon_i)^2\right] = I(f)h^d \int K(U) f_X(x - hU) dU \simeq h^d f_X(x) I(f)$$

by a change of variables and dominated convergence. By a law of large numbers for independent random variables we have

$$\mathcal{J}_{11} = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \left[\frac{f'(\varepsilon_i)}{f(\varepsilon_i)}\right]^2 \xrightarrow{p} \frac{1}{h^d} E\left\{K\left(\frac{x - X_i}{h}\right) \left[\frac{f'(\varepsilon_i)}{f(\varepsilon_i)}\right]^2\right\} \simeq f_X(x) I(f).$$

Now we examine  $\mathcal{J}_{12}$ . Define

$$\rho(\delta) = \int [\psi(\varepsilon + \delta)^2 - \psi(\varepsilon)^2] f(\varepsilon) d\varepsilon$$

for any  $\delta$ . This quantity is finite, differentiable, and satisfies  $\rho(0) = 0$ . By independence of  $\varepsilon_i$  and  $X_i$  we have

$$\mathcal{J}_{12} \simeq E\left[\frac{1}{h^d} K\left(\frac{x - X_i}{h}\right) \rho(\delta_i)\right].$$

By dominated convergence, this expectation is zero because as  $h \rightarrow 0$

$$\max_{i: |X_i - x| \leq h} |m(X_i) - m(x)| = \max_{i: |X_i - x| \leq h} |\delta_i| \rightarrow 0$$

by the differentiability of  $m$  at  $x$ . ■

LEMMA H2. *Under our conditions*

$$\mathcal{J}_2 = o_p(1).$$

PROOF. Under our conditions  $G_b(\tilde{f}_i) = G_b(f_i) + o_p(1)$ . In particular,

$$\max_{1 \leq i \leq n} \left|G_b(\tilde{f}_i) - G_b(f_i)\right| = \max_{1 \leq i \leq n} \left|g_b(f_i^*)(\tilde{f}_i - f_i)\right| \leq b^{-1} \max_{1 \leq i \leq n} \left|\tilde{f}_i - f_i\right|,$$

where  $f_i^*$  is an intermediate point between  $\tilde{f}_i$  and  $f_i$ . Notice that under bandwidth assumption A5,

$$\max_{1 \leq i \leq n} |\tilde{f}_i - f_i| = o_p(b), \quad (39)$$

and thus  $\max_{1 \leq i \leq n} |G_b(\tilde{f}_i) - G_b(f_i)| = o_p(1)$ . Let  $G_i = G_b(f_i)$ . Then,

$$\begin{aligned} & \left| \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \left[ \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \right]^2 [G_i - G_b(\tilde{f}_i)] \right| \\ & \leq \frac{1}{nh^d} \sum_{i=1}^n \left| K\left(\frac{x - X_i}{h}\right) \right| \psi(\bar{\varepsilon}_i)^2 \cdot \max_{1 \leq i \leq n} |G_b(\tilde{f}_i) - G_b(f_i)| \\ & = O_p(1) \cdot o_p(1) = o_p(1). \end{aligned}$$

Therefore, we can ignore the estimation errors in trimming effect. Furthermore, making an expansion of  $\psi(\bar{\varepsilon}_i)$  we obtain

$$\begin{aligned} \mathcal{J}_2 &= \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \psi(\varepsilon_i)^2 [1 - G_i] + \frac{2}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \psi(\varepsilon_i) \psi'(\varepsilon_i) \delta_i [1 - G_i] \\ & \quad + \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \psi'(\varepsilon_i)^2 \delta_i^2 [1 - G_i] + o_p(1). \end{aligned}$$

We verify the orders of magnitude for these terms. It can be shown that the leading trimming term is determined by

$$\frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \psi(\varepsilon_i)^2 [1 - G_b(f(\varepsilon_i))].$$

Notice that

$$\begin{aligned} & E [\psi(\varepsilon_i)^2 \{1 - G_b(f(\varepsilon_i))\}] \\ &= \int_{0 < f(\varepsilon) < b} \psi(\varepsilon_i)^2 f(\varepsilon) d\varepsilon + \int_{b < f(\varepsilon) < 2b} \psi(\varepsilon_i)^2 \left( \int_{f(\varepsilon)}^{\infty} g_b(z) dz \right) f(\varepsilon) d\varepsilon, \end{aligned}$$

under Assumption A3 (or A3'), we can show that

$$\int_{0 < f(\varepsilon) < b} \psi(\varepsilon_i)^2 f(\varepsilon) d\varepsilon + \int_{b < f(\varepsilon) < 2b} \psi(\varepsilon_i)^2 \left( \int_{f(\varepsilon)}^{\infty} g_b(z) dz \right) f(\varepsilon) d\varepsilon = o(1).$$

The precise order of magnitude of the trimming effect will depend on the tail behavior of  $f$ . For example, under Assumptions A2' and A3', when  $\varepsilon$  approaches the boundary  $\underline{a}$ ,

$$\begin{aligned} f(\varepsilon) &\simeq \frac{1}{\varrho!} f^{(\varrho)}(\underline{a})(\varepsilon - \underline{a})^\varrho, \\ f'(\varepsilon) &\simeq \frac{1}{(\varrho - 1)!} f^{(\varrho)}(\underline{a})(\varepsilon - \underline{a})^{\varrho-1}, \end{aligned}$$

for  $\underline{a} \leq \varepsilon \leq \underline{a} + \delta$ , and  $\delta$  is small. Similar results hold in small neighborhoods around the upper bound  $\bar{a}$ . It is easy to show that

$$\int_{0 < f(\varepsilon) < b} \psi(\varepsilon_i)^2 f(\varepsilon) d\varepsilon + \int_{b < f(\varepsilon) < 2b} \psi(\varepsilon_i)^2 \left( \int_{f(\varepsilon)}^{\infty} g_b(z) dz \right) f(\varepsilon) d\varepsilon = O(b^{(e-1)/e}) = o(1).$$

In addition, by a straightforward calculation, we have

$$E \left[ K \left( \frac{x - X_i}{h} \right)^2 \right] = h^d f_X(x) \mu_2(K), \quad (40)$$

where  $\mu_2(K)$  is a constant depending only on the kernel  $K$ . By i.i.d. assumption, we have

$$\frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \psi(\varepsilon_i)^2 [1 - G_b(f(\varepsilon_i))] = o_p(1).$$

Other terms can be analyzed similarly. ■

For  $j = 3, 4, \dots, 10$ , terms  $\mathcal{J}_j$  are functions of  $\tilde{f}'(\check{\varepsilon}_i) - f'(\bar{\varepsilon}_i)$  and/or  $\tilde{f}(\check{\varepsilon}_i) - f(\bar{\varepsilon}_i)$ . To facilitate asymptotic analysis, we decompose  $\tilde{f}'(\check{\varepsilon}_i) - f'(\bar{\varepsilon}_i)$  into the sum of  $\tilde{f}'(\check{\varepsilon}_i) - \bar{f}'(\bar{\varepsilon}_i)$  and  $\bar{f}'(\bar{\varepsilon}_i) - f'(\bar{\varepsilon}_i)$  (and  $\tilde{f}(\check{\varepsilon}_i) - f(\bar{\varepsilon}_i)$  into the sum of  $\tilde{f}(\check{\varepsilon}_i) - \bar{f}(\bar{\varepsilon}_i)$  and  $\bar{f}(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)$ ). The first term,  $\tilde{f}'(\check{\varepsilon}_i) - \bar{f}'(\bar{\varepsilon}_i)$ , contains errors coming from preliminary estimation  $\tilde{m}(X_i) - m(X_i)$  and  $\tilde{m}(x) - m(x)$ , and the second term,  $\bar{f}'(\bar{\varepsilon}_i) - f'(\bar{\varepsilon}_i)$ , contains the errors coming from nonparametric kernel smoothing.

LEMMA H3. *Under our conditions*

$$\mathcal{J}_3 = o_p(1).$$

PROOF. We decompose  $\tilde{f}'(\check{\varepsilon}_i) - f'(\bar{\varepsilon}_i)$  and obtain:

$$\begin{aligned} \mathcal{J}_3 &= \frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) \left[ \tilde{f}'(\check{\varepsilon}_i) - \bar{f}'(\bar{\varepsilon}_i) \right]}{f(\bar{\varepsilon}_i)^2} G_i \\ &\quad + \frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) \left[ \bar{f}'(\bar{\varepsilon}_i) - f'(\bar{\varepsilon}_i) \right]}{f(\bar{\varepsilon}_i)^2} G_i \\ &= \mathcal{J}_{31} + \mathcal{J}_{32} \end{aligned}$$

$$\mathcal{J}_{31} = \frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) \left[ \tilde{f}'(\check{\varepsilon}_i) - \bar{f}'(\bar{\varepsilon}_i) \right]}{f(\bar{\varepsilon}_i)^2} G_i$$

$$\begin{aligned}
&= \frac{2}{nh^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{\frac{1}{nh^2} \sum_{j \neq i} \left[ k' \left( \frac{\check{\varepsilon}_i - \check{\varepsilon}_j}{h} \right) - k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right]}{f(\bar{\varepsilon}_i)} G_i \\
&\simeq \frac{2}{nh^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{\frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) (\check{\varepsilon}_i - \bar{\varepsilon}_i + \check{\varepsilon}_j - \varepsilon_j)}{f(\bar{\varepsilon}_i)} G_i \\
&= [\tilde{m}(x) - m(x)] \frac{2}{nh^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{\frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right)}{f(\bar{\varepsilon}_i)} G_i \\
&\quad + \frac{2}{nh^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{\frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) [\tilde{m}(X_j) - m(X_j)]}{f(\bar{\varepsilon}_i)} G_i \\
&= \mathcal{J}_{31A} + \mathcal{J}_{31B}.
\end{aligned}$$

Since  $|\tilde{m}(x) - m(x)| = O_p(h_n^q + n^{-1/2}h_n^{-d/2})$ , to show that  $\mathcal{J}_{31A} = o_p(1)$ , we only need to show that

$$\begin{aligned}
&\frac{2}{nh^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{\frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right)}{f(\bar{\varepsilon}_i)} G_i \\
&= \sum_{i=1}^n \sum_{j \neq i} \frac{2}{n^2 h^{d+3}} K\left(\frac{x-X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{G_i}{f(\bar{\varepsilon}_i)} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \\
&= \sum_{i=1}^n \sum_{j \neq i} \varphi_{An}(z_i, z_j) = o_p(1/(h_n^q + n^{-1/2}h_n^{-d/2})),
\end{aligned}$$

where

$$\varphi_{An}(z_i, z_j) = \frac{2}{n^2 h^{d+3}} K\left(\frac{x-X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{G_i}{f(\bar{\varepsilon}_i)} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right).$$

Notice that conditional on  $z_i$ ,

$$E_i k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) = h^3 \int k(u) f''(\bar{\varepsilon}_i - uh) du,$$

and it can be verified that:

$$\begin{aligned}
E \varphi_{An}(z_i, z_j) &\leq E \frac{2}{n^2 h^{d+3}} K\left(\frac{x-X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{1}{b} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \\
&= O\left(\frac{n^2}{n^2 h^{d+3}} \times h^d \times h^3 \times b^{-1}\right)
\end{aligned}$$

$$\begin{aligned}
E \varphi_{An}(z_i, z_j)^2 &= \frac{4}{n^4 h^{2d+6}} E K\left(\frac{x-X_i}{h}\right)^2 \left[ \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \right]^2 \left[ \frac{G_i}{f(\bar{\varepsilon}_i)} \right]^2 k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right)^2 \\
&= O\left(\frac{1}{n^4 h^{2d+6}} \times h^d \times h \times b^{-2}\right) = O\left(\frac{4}{n^4 h^{d+5} b^2}\right)
\end{aligned}$$

$$\begin{aligned}
E\varphi_{An}(z_i, z_j)\varphi_{An}(z_i, z_l) &= \frac{2}{n^4 h^{2d+6}} EK \left( \frac{x - X_i}{h} \right)^2 \left[ \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{G_i}{f(\bar{\varepsilon}_i)} \right]^2 k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_l}{h} \right) \\
&= O\left(\frac{1}{n^4 h^{2d+6}} \times h^d \times h^6 \times b^{-2}\right) = O\left(\frac{1}{n^4 h^d b^2}\right).
\end{aligned}$$

Thus the first two moments of  $\sum_{i=1}^n \sum_{j \neq i} \varphi_{An}(z_i, z_j)$  are  $o\left(1/(h_n^q + \frac{1}{\sqrt{nh_n^d}})\right)$ .

For  $\mathcal{J}_{31B}$ , since  $\tilde{m}(X_j) - m(X_j) = e'_1 M_n^{-1}(X_j) U_n(X_j) + e'_1 M_n^{-1}(X_j) B_n(X_j)$ , where  $e_1 = (1, 0, \dots, 0, 0)'$ , we expand  $M_n^{-1}(X_j)$  around its limit  $(Mf(X_j))^{-1}$  and get  $M_n^{-1}(X_j) = (Mf(X_j))^{-1} - (Mf(X_j))^{-1} \times [M_n(X_j) - Mf_X(X_j)] M_n^{-1}(X_j)$ . Thus

$$\begin{aligned}
\mathcal{J}_{31B} &= \frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{\frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) [\tilde{m}(X_j) - m(X_j)]}{f(\bar{\varepsilon}_i)} G_i \\
&= \frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{\frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) e'_1 M^{-1} U_n(X_j)}{f(\bar{\varepsilon}_i)} G_i + \\
&\quad \frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{\frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) e'_1 M^{-1} B_n(X_j)}{f(\bar{\varepsilon}_i)} G_i - \\
&\quad \frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{\frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right)}{f(\bar{\varepsilon}_i)} \times \\
&\quad \frac{e'_1 M^{-1} [M_n(X_j) - Mf_X(X_j)] M_n^{-1}(X_j) U_n(X_j)}{f_X(X_j)} G_i - \\
&\quad \frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{\frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right)}{f(\bar{\varepsilon}_i)} \times \\
&\quad \frac{e'_1 M^{-1} [M_n(X_j) - Mf_X(X_j)] M_n^{-1}(X_j) B_n(X_j)}{f_X(X_j)} G_i \\
&= \mathcal{J}_{311} + \mathcal{J}_{312} + \mathcal{J}_{313} + \mathcal{J}_{314}.
\end{aligned}$$

We verify that each of these terms are  $o_p(1)$ . Recalling the definition of  $M^{-1}$  we have

$$\begin{aligned}
\mathcal{J}_{311} &= \frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{\frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) e'_1 M^{-1} U_n(X_j)}{f(\bar{\varepsilon}_i)} G_i \\
&= \frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{\frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \sum_{l=0}^p M^{0,l} U_{nl}(X_j)}{f(\bar{\varepsilon}_i)} G_i.
\end{aligned}$$

Denoting elements in the first row of  $M^{-1}$  as  $\omega^{0,\bullet}$ , we have

$$\begin{aligned}
\mathcal{J}_{311} &= \frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{\frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right)}{f(\bar{\varepsilon}_i)} \\
&\quad \frac{\sum_{\kappa} \omega^{0,\kappa} \frac{1}{nh^d} \sum_{l \neq j} \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \left( \frac{X_j - X_l}{h} \right)^\kappa \varepsilon_l}{f_X(X_j)} G_i
\end{aligned}$$

$$\begin{aligned}
&= 2 \sum_{i=1}^n \sum_{j \neq i} \sum_{l \neq j} \sum_{\kappa} \omega^{0,\kappa} \frac{1}{n^3 h^{2d+3}} \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{G_i}{f(\bar{\varepsilon}_i) f_X(X_j)} \\
&\quad K \left( \frac{x - X_i}{h} \right) k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \left( \frac{X_j - X_l}{h} \right)^\kappa \varepsilon_l \\
&= 2 \sum_{i=1}^n \sum_{j=1}^n \sum_{\substack{\kappa \\ j \neq i}} \omega^{0,\kappa} \frac{1}{n^3 h^{2d+3}} \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{G_i}{f(\bar{\varepsilon}_i) f_X(X_j)} \\
&\quad K \left( \frac{x - X_i}{h} \right) k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \mathcal{K} \left( \frac{X_j - X_i}{h} \right) \left( \frac{X_j - X_i}{h} \right)^\kappa \varepsilon_i \\
&\quad + 2 \sum_{i=1}^n \sum_{j \neq i} \sum_{\substack{l \neq j \\ l \neq i}} \sum_{\kappa} \omega^{0,\kappa} \frac{1}{n^3 h^{2d+3}} \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{G_i}{f(\bar{\varepsilon}_i) f_X(X_j)} \\
&\quad K \left( \frac{x - X_i}{h} \right) k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \left( \frac{X_j - X_l}{h} \right)^\kappa \varepsilon_l \\
&= 2 \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \varphi_{1n}(z_i, z_j) + 2 \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i, j \neq l, l \neq i}}^n \sum_{l=1}^n \varphi_{2n}(z_i, z_j, z_l),
\end{aligned}$$

where

$$\varphi_{1n}(z_i, z_j) = \frac{1}{n^3 h^{2d+3}} \sum_{\kappa} \omega^{0,\kappa} \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{G_i}{f(\bar{\varepsilon}_i) f_X(X_j)} K \left( \frac{x - X_i}{h} \right) k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \mathcal{K} \left( \frac{X_j - X_i}{h} \right) \left( \frac{X_j - X_i}{h} \right)^\kappa \varepsilon_i$$

is a second order U-statistic, while

$$\begin{aligned}
\varphi_{2n}(z_i, z_j, z_l) &= \sum_{\kappa} \omega^{0,\kappa} \frac{1}{n^3 h^{2d+3}} \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{G_i}{f(\bar{\varepsilon}_i) f_X(X_j)} \\
&\quad \times K \left( \frac{x - X_i}{h} \right) k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \left( \frac{X_j - X_l}{h} \right)^\kappa \varepsilon_l
\end{aligned}$$

is a third order U-statistics with degeneracy, i.e.,  $E_{i,j} \varphi_{2n}(z_i, z_j, z_l) = 0$ . Therefore

$$\begin{aligned}
E\mathcal{J}_{311} &= 2 \sum_{i=1}^n \sum_{j=1}^n E \varphi_{1n}(z_i, z_j) \\
&= \frac{2}{n^3 h^{2d+3}} \sum_{i=1}^n \sum_{j=1}^n \sum_{\kappa} \omega^{0,\kappa} E \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{\varepsilon_i G_i}{f(\bar{\varepsilon}_i)} K \left( \frac{x - X_i}{h} \right) E_i k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \\
&\quad \times E_i \frac{1}{f_X(X_j)} \mathcal{K} \left( \frac{X_j - X_i}{h} \right) \left( \frac{X_j - X_i}{h} \right)^\kappa.
\end{aligned}$$

Notice that for large  $n$

$$E_i \frac{1}{f_X(X_j)} \mathcal{K} \left( \frac{X_j - X_i}{h} \right) \left( \frac{X_j - X_i}{h} \right)^\kappa = \int \mathcal{K} \left( \frac{X_j - X_i}{h} \right) \left( \frac{X_j - X_i}{h} \right)^\kappa dX_j = h^d \int K(u) u^\kappa du,$$



$$EK \left( \frac{x - X_i}{h} \right) = O_p(h^d) \text{ and } E_i k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) = O_p(h^3),$$

$$E\mathcal{J}_{311} = O(n^{-1}b^{-1}).$$

For the variances,  $\text{var}[\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \varphi_{1n}(z_i, z_j)]$  is determined by

$$O(n^2)E[\varphi_{1n}(z_i, z_j)^2 + \varphi_{1n}(z_i, z_j)\varphi_{1n}(z_j, z_i)] \quad (41)$$

$$+ O(n^3)E[\varphi_{1n}(z_i, z_j)\varphi_{1n}(z_l, z_j) + \varphi_{1n}(z_i, z_l)\varphi_{1n}(z_i, z_j)]$$

A calculation of moments shows that (41) is of order  $O\left(\frac{1}{n^4 h^{2d+5} b^2}\right) + O\left(\frac{1}{n^3 h^d b^2}\right) = o(1)$ , since, for each  $\kappa$ ,

$$E_i \left[ \frac{1}{f_X(X_j)} \mathcal{K} \left( \frac{X_j - X_i}{h} \right) \left( \frac{X_j - X_i}{h} \right)^\kappa \right]^2 = O(h^d)$$

$$E_i \left[ k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right]^2 = O_p(h)$$

$$E \left[ \frac{K \left( \frac{x - X_j}{h} \right) K \left( \frac{x - X_i}{h} \right) \mathcal{K} \left( \frac{X_j - X_i}{h} \right) \mathcal{K} \left( \frac{X_i - X_j}{h} \right) \left( \frac{X_j - X_i}{h} \right)^\kappa \left( \frac{X_i - X_j}{h} \right)^\kappa}{f_X(X_i) f_X(X_j)} \right] = O(h^d)$$

thus

$$E[\varphi_{1n}(z_i, z_j)^2 + \varphi_{1n}(z_i, z_j)\varphi_{1n}(z_j, z_i)] = O\left(\frac{1}{n^6 h^{2d+5} b^2}\right),$$

and

$$E[\varphi_{1n}(z_i, z_j)\varphi_{1n}(z_l, z_j) + \varphi_{1n}(z_i, z_l)\varphi_{1n}(z_i, z_j)] = O\left(\frac{1}{n^6 h^d b^2}\right).$$

For the third order U-statistics,

$$\text{var} \left[ \sum_{\substack{i=1 \\ j \neq i}}^n \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{\substack{l=1 \\ l \neq i, j}}^n \varphi_{2n}(z_i, z_j, z_l) \right] = O(n^3)E[\Phi_{21}] + O(n^4)E[\Phi_{22}] + O(n^5)E[\Phi_{23}],$$

where  $\Phi_{21}$  includes terms with three different indices, say  $\varphi_{2n}(z_i, z_j, z_l)^2$ ,  $\Phi_{22}$  includes terms like  $\varphi_{2n}(z_i, z_j, z_l)\varphi_{2n}(z_i, z_r, z_l)$ , and  $\Phi_{23}$  contains terms of type  $\varphi_{2n}(z_i, z_j, z_l)\varphi_{2n}(z_r, z_t, z_l)$ . We can verify that: for terms with three different indices

$$E\varphi_{2n}(z_i, z_j, z_l)^2 = O\left(\frac{1}{n^6 h^{4d+6} b^2} \times h^{2d+1}\right),$$

for terms with four different indices

$$E\varphi_{2n}(z_i, z_j, z_l)\varphi_{2n}(z_i, z_r, z_l) = O\left(\frac{1}{n^6 h^{4d+6} b^2} \times h^{3d+6}\right),$$

for terms with five different indices

$$E\varphi_{2n}(z_i, z_j, z_l)\varphi_{2n}(z_r, z_t, z_l) = O\left(\frac{1}{n^6 h^{4d+6} b^2} \times h^{4d+6}\right).$$

Thus

$$\text{var} \left[ \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \varphi_{2n}(z_i, z_j, z_l) \right]_{\substack{j \neq i, j \neq l, l \neq i}} = O(n^{-1}b^{-2} + n^{-2}h^{-d}b^{-2} + n^{-3}h^{-2d-5}b^{-2}).$$

Consequently,  $\mathcal{J}_{311} = o_p(1)$ . For  $\mathcal{J}_{312}$ ,

$$\begin{aligned} & \frac{2}{nh^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{\frac{1}{nh^3} \sum_{j \neq i} k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right)}{f(\bar{\varepsilon}_i)} \frac{e'_1 M^{-1} B_n(X_j)}{f_X(X_j)} G_i \\ &= \frac{2}{nh^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{\frac{1}{nh^3} \sum_{j \neq i} k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right)}{f(\bar{\varepsilon}_i)} \frac{\sum_l M^{0,l} B_{nl}(X_j)}{f_X(X_j)} G_i \\ &\simeq \frac{2}{nh^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{\frac{1}{nh^3} \sum_{j \neq i} k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right)}{f(\bar{\varepsilon}_i)} \\ &\quad \times \frac{\sum_{\kappa} \omega^{0,\kappa} \left( \frac{1}{nh^d} \sum_{l \neq j} K\left(\frac{X_j - X_l}{h}\right) \left(\frac{X_j - X_l}{h}\right)^{p+\kappa} h^{p+1} m_{p+1}(X_j) \right)}{f_X(X_j)} G_i \\ &= \frac{2}{n^3 h^{2d+2-p}} \sum_{i=1}^n \sum_{j \neq i} \sum_{l \neq j} \sum_{\kappa} \omega^{0,\kappa} \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{G_i}{f(\bar{\varepsilon}_i)} \frac{1}{f_X(X_j)} \\ &\quad \times K\left(\frac{x-X_i}{h}\right) k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) K\left(\frac{X_j - X_l}{h}\right) \left(\frac{X_j - X_l}{h}\right)^{p+\kappa} m_{p+1}(X_j) \\ &= O_p(h^{p+1}), \end{aligned}$$

$$\begin{aligned} \mathcal{J}_{313} &= \left| \frac{2}{nh^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{\frac{1}{nh^3} \sum_{j \neq i} k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right)}{f(\bar{\varepsilon}_i)} \times \right. \\ &\quad \left. \frac{e'_1 M^{-1} [M_n(X_j) - M f_X(X_j)] M_n^{-1}(X_j) U_n(X_j)}{f_X(X_j)} G_i \right| \\ &\leq \max_{1 \leq j \leq n} \left| \frac{e'_1 M^{-1} [M_n(X_j) - M f_X(X_j)] M_n^{-1}(X_j) U_n(X_j)}{f_X(X_j)} \right| \\ &\quad \times \frac{2}{n^2 h^{d+3}} \sum_{i=1}^n \sum_{j \neq i} \left| K\left(\frac{x-X_i}{h}\right) k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{G_i}{f(\bar{\varepsilon}_i)} \right| \\ &\leq \max_{1 \leq j \leq n} \|e'_1 M^{-1} [M_n(X_j) - M f_X(X_j)]\| \times \left\| \frac{M_n^{-1}(X_j) U_n(X_j)}{f_X(X_j)} \right\| \\ &\quad \times \frac{2}{n^2 h^{d+3}} \sum_{i=1}^n \sum_{j \neq i} \left| K\left(\frac{x-X_i}{h}\right) k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{G_i}{f(\bar{\varepsilon}_i)} \right| \end{aligned}$$

$$\begin{aligned}
&= O_p(h + n^{-1/2}h^{-d/2} \log n)b^{-1} \\
&= o_p(1).
\end{aligned}$$

Similarly

$$\begin{aligned}
\mathcal{J}_{314} &= \frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{\frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right)}{f(\bar{\varepsilon}_i)} \times \\
&\quad \frac{e'_1 M^{-1} [M_n(X_j) - M f_X(X_j)] M_n^{-1}(X_j) B_n(X_j)}{f_X(X_j)} G_i \\
&= O_p(h + n^{-1/2}h^{-d/2} \log n)h^{p+1}b^{-1} \\
&= o_p(1).
\end{aligned}$$

Thus,  $\mathcal{J}_{31B} = o_p(1)$  and  $\mathcal{J}_{31} = o_p(1)$ .

In order to show  $\mathcal{J}_{32} = o_p(1)$ , we decompose  $\bar{f}'(\bar{\varepsilon}_i) - f'(\bar{\varepsilon}_i)$  into a bias effect  $B'(\bar{\varepsilon}_i)$  and a variance effect  $V'(\bar{\varepsilon}_i)$  as

$$\begin{aligned}
B'(\bar{\varepsilon}_i) &= \frac{1}{nh^2} \sum_{j \neq i} E_i \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right] - f'(\bar{\varepsilon}_i) \\
V'(\bar{\varepsilon}_i) &= \frac{1}{nh^2} \sum_{j \neq i} \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) - E_i \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right] \right],
\end{aligned}$$

thus

$$\begin{aligned}
\mathcal{J}_{32} &= \frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) [\bar{f}'(\bar{\varepsilon}_i) - f'(\bar{\varepsilon}_i)]}{f(\bar{\varepsilon}_i)^2} G_i \\
&= \frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) V'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)^2} G_i + \frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) B'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)^2} G_i \\
&= \mathcal{J}_{32A} + \mathcal{J}_{32B}.
\end{aligned}$$

The variance term is

$$\begin{aligned}
\mathcal{J}_{32A} &= \frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) V'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)^2} G_i \\
&= \frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)^2} \left\{ \frac{1}{nh^2} \sum_{j \neq i} \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) - E_i \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right] \right] G_i \right\} \\
&= \sum_{i=1}^n \sum_{j \neq i} \frac{2}{n^2 h^{d+2}} K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)^2} \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) - E_i \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right] \right] G_i.
\end{aligned}$$

Denote

$$\varphi_{3n}(z_i, z_j) = \frac{2}{n^2 h^{d+2}} K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)^2} \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) - E_i \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right] \right] G_i,$$

again, this is a second order U-statistic with degeneracy. We should verify the orders of magnitude of  $E[\varphi_{3n}(z_i, z_j)^2]$ ,  $E[\varphi_{3n}(z_i, z_j)\varphi_{3n}(z_j, z_i)]$ , and  $E[\varphi_{3n}(z_i, z_j)\varphi_{3n}(z_l, z_j)]$ . By a similar calculation as we did before (also see similar treatment later in the analysis of score), we can show that

$$\frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)V'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)^2} = O_p(b^{-1}n^{-1/2} + b^{-1}n^{-1}h^{-3/2-d/2}).$$

The leading bias term is

$$\frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)B'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)^2} = O_p(h^q) = o_p(1).$$

Consequently  $\mathcal{J}_{32} = o_p(1)$  and thus  $\mathcal{J}_3 = o_p(1)$ . ■

Similar analysis can be applied to other terms that involve  $\tilde{f}'(\check{\varepsilon}_i) - f'(\bar{\varepsilon}_i)$  and  $\tilde{f}(\check{\varepsilon}_i) - f(\bar{\varepsilon}_i)$ .

LEMMA H4. *Under our conditions*

$$\mathcal{J}_4 = o_p(1).$$

PROOF. For  $\mathcal{J}_4$ , we decompose  $\tilde{f}(\check{\varepsilon}_i) - f(\bar{\varepsilon}_i)$  into the sum of  $\tilde{f}(\check{\varepsilon}_i) - \bar{f}(\bar{\varepsilon}_i)$  and  $\bar{f}(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)$ , and the analysis for

$$\frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)^2}{f(\bar{\varepsilon}_i)^4} (\tilde{f}(\check{\varepsilon}_i) + f(\bar{\varepsilon}_i)) (\tilde{f}(\check{\varepsilon}_i) - \bar{f}(\bar{\varepsilon}_i)) G_i$$

and

$$\frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)^2}{f(\bar{\varepsilon}_i)^4} (\tilde{f}(\check{\varepsilon}_i) + f(\bar{\varepsilon}_i)) (\bar{f}(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)) G_i$$

is similar to that of  $\mathcal{J}_{31}$  and  $\mathcal{J}_{32}$ . ■

LEMMA H5. *Under our conditions*

$$\mathcal{J}_5 = o_p(1).$$

PROOF. Write

$$\mathcal{J}_5 = \frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) \left[ \tilde{f}'(\check{\varepsilon}_i) - f'(\bar{\varepsilon}_i) \right]^2}{f(\bar{\varepsilon}_i)^2} G_i$$

$$\begin{aligned}
&= \frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) \left[ \tilde{f}'(\check{\varepsilon}_i) - \bar{f}'(\bar{\varepsilon}_i) + \bar{f}'(\bar{\varepsilon}_i) - f'(\bar{\varepsilon}_i) \right]^2}{f(\bar{\varepsilon}_i)^2} G_i \\
&= \frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) \left[ \tilde{f}'(\check{\varepsilon}_i) - \bar{f}'(\bar{\varepsilon}_i) \right]^2}{f(\bar{\varepsilon}_i)^2} G_i \\
&\quad + \frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) \left[ \bar{f}'(\bar{\varepsilon}_i) - f'(\bar{\varepsilon}_i) \right]^2}{f(\bar{\varepsilon}_i)^2} G_i \\
&\quad + \frac{2}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) \left[ \tilde{f}'(\check{\varepsilon}_i) - \bar{f}'(\bar{\varepsilon}_i) \right] \left[ \bar{f}'(\bar{\varepsilon}_i) - f'(\bar{\varepsilon}_i) \right]}{f(\bar{\varepsilon}_i)^2} G_i \\
&= \mathcal{J}_{51} + \mathcal{J}_{52} + \mathcal{J}_{53}
\end{aligned}$$

$$\begin{aligned}
\mathcal{J}_{51} &= \frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) \left[ \tilde{f}'(\check{\varepsilon}_i) - \bar{f}'(\bar{\varepsilon}_i) \right]^2}{f(\bar{\varepsilon}_i)^2} G_i \\
&\simeq \frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) \left[ \frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) (\check{\varepsilon}_i - \bar{\varepsilon}_i + \tilde{\varepsilon}_j - \varepsilon_j) \right]^2}{f(\bar{\varepsilon}_i)^2} G_i \\
&= \frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) \left[ (\check{\varepsilon}_i - \bar{\varepsilon}_i) \frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right]^2}{f(\bar{\varepsilon}_i)^2} G_i \\
&\quad + \frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) \left[ \frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) (\tilde{\varepsilon}_j - \varepsilon_j) \right]^2}{f(\bar{\varepsilon}_i)^2} G_i \\
&\quad + \frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) \left[ \frac{1}{n^2 h^6} (\check{\varepsilon}_i - \bar{\varepsilon}_i) \sum_{j \neq i} \sum_{l \neq i} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_l}{h} \right) (\tilde{\varepsilon}_j - \varepsilon_j) \right]}{f(\bar{\varepsilon}_i)^2} G_i \\
&\quad + \frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) \left[ \frac{1}{n^2 h^6} (\check{\varepsilon}_i - \bar{\varepsilon}_i) \sum_{j \neq i} \sum_{l \neq i} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_l}{h} \right) (\tilde{\varepsilon}_l - \varepsilon_l) \right]}{f(\bar{\varepsilon}_i)^2} G_i,
\end{aligned}$$

where

$$\begin{aligned}
&\frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) \left[ (\check{\varepsilon}_i - \bar{\varepsilon}_i) \frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right]^2}{f(\bar{\varepsilon}_i)^2} G_i \\
&= |\tilde{m}(x) - m(x)|^2 \frac{1}{n^3 h^{d+6}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) G_i}{f(\bar{\varepsilon}_i)^2} \sum_{j \neq i} \sum_{l \neq i} k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) k'' \left( \frac{\bar{\varepsilon}_i - \varepsilon_l}{h} \right) \\
&= o_p(1)
\end{aligned}$$

$$\begin{aligned}
& \left| \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i) \left[ \frac{1}{nh^3} \sum_{j \neq i} k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) (\tilde{\varepsilon}_j - \varepsilon_j) \right]^2}{f(\bar{\varepsilon}_i)^2} G_i \right| \\
&= \left| \frac{1}{n^3 h^{d+6}} \sum_{i=1}^n \sum_{j \neq i} \sum_{l \neq i} K\left(\frac{x-X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i) G_i k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_l}{h}\right)}{f(\bar{\varepsilon}_i)^2} [\tilde{m}(X_j) - m(X_j)] [\tilde{m}(X_l) - m(X_l)] \right| \\
&\leq \left\{ \max_{1 \leq j \leq n} [\tilde{m}(X_j) - m(X_j)] \right\}^2 \\
&\quad \times \frac{1}{n^3 h^{d+6}} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \left| \frac{f'(\bar{\varepsilon}_i) G_i}{f(\bar{\varepsilon}_i)^2} \right| \left| \sum_{j \neq i} \left| k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \right| \right| \left| \sum_{l \neq i} \left| k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_l}{h}\right) \right| \right|.
\end{aligned}$$

Conditional on  $i$ , by the i.i.d. assumption,

$$\begin{aligned}
\frac{1}{n} \sum_{j \neq i} \left| k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \right| &\simeq E_i \left| k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \right| = h \int |k''(u_j)| f(\bar{\varepsilon}_i - hu_j) du_j \simeq hf(\bar{\varepsilon}_i) \int |k''(u_j)| du_j, \\
\frac{1}{n} \sum_{l \neq i} \left| k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_l}{h}\right) \right| &\simeq E_i \left| k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_l}{h}\right) \right| = h \int |k''(u_l)| f(\bar{\varepsilon}_i - hu_l) du_l \simeq hf(\bar{\varepsilon}_i) \int |k''(u_l)| du_l,
\end{aligned}$$

and notice that

$$\max_{1 \leq j \leq n} [\tilde{m}(X_j) - m(X_j)] = O_p(h^{p+1} + n^{-1/2} h^{-d/2} \log n),$$

we have

$$\frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i) \left[ \frac{1}{nh^3} \sum_{j \neq i} k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) (\tilde{\varepsilon}_j - \varepsilon_j) \right]^2}{f(\bar{\varepsilon}_i)^2} G_i = o_p(1).$$

Similarly, we can show

$$\frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i) \left[ \frac{1}{n^2 h^6} (\check{\varepsilon}_i - \bar{\varepsilon}_i) \sum_{j \neq i} \sum_{l \neq i} k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_l}{h}\right) (\tilde{\varepsilon}_j - \varepsilon_j) \right]}{f(\bar{\varepsilon}_i)^2} G_i = o_p(1),$$

and

$$\mathcal{J}_{52} = o_p(1), \quad \mathcal{J}_{53} = o_p(1).$$

■

The analysis for  $\mathcal{J}_7$  and  $\mathcal{J}_8$  is similar. The terms  $\mathcal{J}_6$ ,  $\mathcal{J}_9$ , and  $\mathcal{J}_{10}$  contain  $\tilde{f}(\check{\varepsilon}_i)$  in the denominator.

Under our assumptions,

$$\begin{aligned}
|\mathcal{J}_6| &= \left| \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)^2}{f(\bar{\varepsilon}_i)^4 \tilde{f}(\check{\varepsilon}_i)^2} (\tilde{f}(\check{\varepsilon}_i) + f(\bar{\varepsilon}_i))^2 (\tilde{f}(\check{\varepsilon}_i) - f(\bar{\varepsilon}_i))^2 G_b(\tilde{f}_i) \right| \\
&\leq \max_{1 \leq i \leq n} |(\tilde{f}(\check{\varepsilon}_i) - f(\bar{\varepsilon}_i))^2| \left| \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)^2}{f(\bar{\varepsilon}_i)^4 \tilde{f}(\check{\varepsilon}_i)^2} (\tilde{f}(\check{\varepsilon}_i) + f(\bar{\varepsilon}_i))^2 G_b(\tilde{f}_i) \right| \\
&= O_p(h^{2q-2} + n^{-1} h^{-d-2}) b^{-2} = o_p(1).
\end{aligned}$$

We have used Lemma A here. Similar analysis can be conducted for  $\mathcal{J}_9$  and  $\mathcal{J}_{10}$ . ■

### 5.2.2 The Score Function

We want to show that

$$\sqrt{nh^d}\tilde{S}_n(\theta_0) \Rightarrow N(0, f_X(x)I(f) \|K\|^2).$$

By definition

$$\sqrt{nh^d}\tilde{S}_n(\theta_0) = -\frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{\tilde{f}'}{\tilde{f}}(Y_i - \theta_0) G_b(\tilde{f}_i).$$

For the denominator

$$\frac{1}{\tilde{f}(Y_i - \theta_0)} = \frac{1}{f(Y_i - \theta_0)} - \frac{\tilde{f}(Y_i - \theta_0) - f(Y_i - \theta_0)}{f(Y_i - \theta_0)^2} + R_2,$$

where

$$R_2 = \frac{\{\tilde{f}(Y_i - \theta_0) - f(Y_i - \theta_0)\}^2}{f^2(Y_i - \theta_0)\tilde{f}(Y_i - \theta_0)}.$$

Therefore,

$$\begin{aligned} \sqrt{nh^d}\tilde{S}_n(\theta_0) &= -\frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{\tilde{f}'}{\tilde{f}}(Y_i - \theta_0) G_b(\tilde{f}_i) \\ &= -\frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) G_b(\tilde{f}_i) \\ &\quad \left\{ \frac{\tilde{f}'(Y_i - \theta_0)}{f(Y_i - \theta_0)} - \frac{\tilde{f}'(Y_i - \theta_0) [\tilde{f}(Y_i - \theta_0) - f(Y_i - \theta_0)]}{f(Y_i - \theta_0)^2} + \tilde{f}'(Y_i - \theta_0) R_2 \right\} \\ &= -\frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{\tilde{f}'(Y_i - \theta_0)}{f(Y_i - \theta_0)} G_b(\tilde{f}_i) \\ &\quad + \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{\tilde{f}'(Y_i - \theta_0) [\tilde{f}(Y_i - \theta_0) - f(Y_i - \theta_0)]}{f(Y_i - \theta_0)^2} G_b(\tilde{f}_i) \\ &\quad - \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) G_b(\tilde{f}_i) \tilde{f}'(Y_i - \theta_0) R_2. \end{aligned}$$

Thus

$$\sqrt{nh^d}\tilde{S}_n(\theta_0) = -\frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{f'(Y_i - \theta_0)}{f(Y_i - \theta_0)}$$

$$\begin{aligned}
& + \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(Y_i - \theta_0)}{f(Y_i - \theta_0)} [1 - G_b(\tilde{f}_i)] \\
& - \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{\tilde{f}'(Y_i - \theta_0) - f'(Y_i - \theta_0)}{f(Y_i - \theta_0)} G_b(\tilde{f}_i) \\
& + \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{\tilde{f}'(Y_i - \theta_0) \left[ \tilde{f}(Y_i - \theta_0) - f(Y_i - \theta_0) \right]}{f(Y_i - \theta_0)^2} G_b(\tilde{f}_i) \\
& + \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) G_b(\tilde{f}_i) \tilde{f}'(Y_i - \theta_0) R_2 \\
& = T_1 + T_2 - T_3 + T_4 + T_5.
\end{aligned}$$

We show that

$$T_1 \Rightarrow N(0, f_X(x)I(f) \|K\|^2),$$

and

$$T_j \xrightarrow{p} 0, \quad j = 2, \dots, 5.$$

LEMMA S1. *Under our conditions*

$$T_1 \Rightarrow N(0, f_X(x)I(f) \|K\|^2).$$

PROOF. Notice that  $Y_i - \theta_0 = \varepsilon_i + m(X_i) - m(x) = \varepsilon_i + \delta_i$  and

$$\begin{aligned}
T_1 &= -\frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(Y_i - \theta_0)}{f(Y_i - \theta_0)} \\
&= -\frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(\varepsilon_i + \delta_i)}{f(\varepsilon_i + \delta_i)} \\
&= -\frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) [f'(\varepsilon_i) + f'(\varepsilon_i + \delta_i) - f'(\varepsilon_i)] \left[ \frac{1}{f(\varepsilon_i)} + \frac{[f(\varepsilon_i + \delta_i) - f(\varepsilon_i)]}{f(\varepsilon_i)f(\varepsilon_i + \delta_i)} \right] \\
&= -\frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(\varepsilon_i)}{f(\varepsilon_i)} \\
&\quad - \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(\varepsilon_i + \delta_i) - f'(\varepsilon_i)}{f(\varepsilon_i)} \\
&\quad - \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(\varepsilon_i)[f(\varepsilon_i + \delta_i) - f(\varepsilon_i)]}{f(\varepsilon_i)f(\varepsilon_i + \delta_i)} \\
&\quad - \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{[f'(\varepsilon_i + \delta_i) - f'(\varepsilon_i)][f(\varepsilon_i + \delta_i) - f(\varepsilon_i)]}{f(\varepsilon_i)f(\varepsilon_i + \delta_i)} \\
&= T_{11} + T_{12} + T_{13} + T_{14}.
\end{aligned}$$



We verify each of these terms. First, by a central limit theorem for i.i.d. random variables,

$$T_{11} = -\frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \psi(\varepsilon_i) \Rightarrow N(0, f_X(x)I(f) \|K\|^2).$$

By calculation of moments, it is easy to verify that, under our conditions,  $T_{12}$ ,  $T_{13}$ , and  $T_{14}$  are  $o_p(1)$ . For example,

$$T_{12} \simeq -\frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f''(\varepsilon_i)\delta_i}{f(\varepsilon_i)}$$

by a law of large numbers for independent random variables,

$$\begin{aligned} \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f''(\varepsilon_i)\delta_i}{f(\varepsilon_i)} &\simeq \frac{n}{\sqrt{nh^d}} EK\left(\frac{x - X_i}{h}\right) \frac{f''(\varepsilon_i)\delta_i}{f(\varepsilon_i)} \\ &= \frac{n}{\sqrt{nh^d}} \int f''(\varepsilon_i)d\varepsilon_i \int K\left(\frac{x - X_i}{h}\right) [m(X_i) - m(x)] f_X(X_i)dX_i \\ &= O\left(\frac{nh^d h^q}{\sqrt{nh^d}}\right) = O\left(\sqrt{nh^d} h^q\right). \end{aligned}$$

Similarly,

$$T_{13} = -\frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(\varepsilon_i)[f(\varepsilon_i + \delta_i) - f(\varepsilon_i)]}{f(\varepsilon_i)f(\varepsilon_i + \delta_i)} = O_p\left(\sqrt{nh^d} h^q\right).$$

■

LEMMA S2. *Under our conditions*

$$T_2 \xrightarrow{p} 0.$$

PROOF. First, we expand the trimming function to the second order,

$$G_b(\tilde{f}_i) - G_b(f_i) = g_b(f_i)(\tilde{f}_i - f_i) + \frac{1}{2}g'_b(f_i^*)(\tilde{f}_i - f_i)^2,$$

where  $f_i^*$  is an intermediate point between  $\tilde{f}_i$  and  $f_i$ . Then

$$\begin{aligned} &\frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(Y_i - \theta_0)}{f(Y_i - \theta_0)} [G_i - G_b(\tilde{f}_i)] \\ &= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(Y_i - \theta_0)}{f(Y_i - \theta_0)} g_b(f_i)(\tilde{f}_i - f_i) \\ &\quad + \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(Y_i - \theta_0)}{f(Y_i - \theta_0)} \frac{1}{2} g'_b(f_i^*)(\tilde{f}_i - f_i)^2. \end{aligned}$$

We use a crude bound on the last term so that

$$\begin{aligned}
& \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(Y_i - \theta_0)}{f(Y_i - \theta_0)} \frac{1}{2} g'_b(f_i^*) (\tilde{f}_i - f_i)^2 \\
& \leq \frac{1}{nh^d} \sum_{i=1}^n \left| K \left( \frac{x - X_i}{h} \right) \frac{f'(Y_i - \theta_0)}{f(Y_i - \theta_0)} \right| \times \sup_t |g'(t)| \\
& \quad \times \frac{1}{2} b^{-2} \sqrt{nh^d} \left( \max_{1 \leq i \leq n} |\tilde{f}_i - f_i| \right)^2,
\end{aligned}$$

where the first term can be shown to be  $O_p(1)$ , and the second term is  $o_p(1)$  by result (23). Therefore, we must analyze the term

$$\frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(Y_i - \theta_0)}{f(Y_i - \theta_0)} g_b(f_i) (\tilde{f}_i - f_i). \quad (42)$$

This again involves some further U-statistic calculation. Similar to the previous analysis, we decompose (42) into  $T_{21} + T_{22}$ , where

$$\begin{aligned}
T_{21} &= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(Y_i - \theta_0)}{f(Y_i - \theta_0)} g_b(f_i) (\tilde{f}(\check{\varepsilon}_i) - \bar{f}(\bar{\varepsilon}_i)), \\
T_{22} &= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(Y_i - \theta_0)}{f(Y_i - \theta_0)} g_b(f_i) (\bar{f}(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)).
\end{aligned}$$

$$\begin{aligned}
T_{21} &= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \frac{1}{nh} \sum_{j \neq i} \left[ k \left( \frac{\check{\varepsilon}_i - \tilde{\varepsilon}_j}{h} \right) - k \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right] \\
&= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) (\check{\varepsilon}_i - \bar{\varepsilon}_i + \tilde{\varepsilon}_j - \varepsilon_j) \\
&\quad + \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\bar{\varepsilon}_i^* - \varepsilon_j^*}{h} \right) (\check{\varepsilon}_i - \bar{\varepsilon}_i + \tilde{\varepsilon}_j - \varepsilon_j)^2,
\end{aligned}$$

where the second term is  $o_p(1)$  under our assumptions:

$$\begin{aligned}
& \left| \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \frac{1}{nh^3} \sum_{j \neq i} k'' \left( \frac{\bar{\varepsilon}_i^* - \varepsilon_j^*}{h} \right) (\check{\varepsilon}_i - \bar{\varepsilon}_i + \tilde{\varepsilon}_j - \varepsilon_j)^2 \right| \\
& \leq O_p(h^{2q} + n^{-1}h^{-d} \log^2(n)) \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n \left| K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \right| \frac{1}{nh^3} \sum_{j \neq i} \left| k'' \left( \frac{\bar{\varepsilon}_i^* - \varepsilon_j^*}{h} \right) \right| \\
& = O_p(h^{2q} + n^{-1}h^{-d} \log^2(n)) \times \frac{n}{\sqrt{nh^d}} \times h^d \times h^{-3} \\
& = O_p(h^{2q} + n^{-1}h^{-d} \log^2(n)) \times \sqrt{nh^d} \times h^{-3} = o_p(1).
\end{aligned}$$

Thus,

$$\begin{aligned}
T_{21} &= [\tilde{m}(x) - m(x)] \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \frac{1}{nh^2} \sum_{j \neq i} k'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \\
&\quad + \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \frac{1}{nh^2} \sum_{j \neq i} k'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) [\tilde{m}(X_j) - m(X_j)] + o_p(1) \\
&\simeq T_{21A} + T_{21B}.
\end{aligned}$$

For the leading term  $T_{21A}$ , since  $\tilde{m}(x) - m(x) = O_p\left(h_n^q + \frac{1}{\sqrt{nh_n^d}}\right)$ , we need to show that

$$\sum_{i=1}^n \frac{1}{\sqrt{nh^d}} K\left(\frac{x - X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \left[ \frac{1}{nh^2} \sum_{j \neq i} k'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \right] = o_p(1/(h_n^q + n^{-1/2}h_n^{-d/2})). \quad (43)$$

Notice that

$$\begin{aligned}
&\frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \left[ \frac{1}{nh^2} \sum_{j \neq i} k'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \right] \\
&= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \hat{f}'(\bar{\varepsilon}_i) \\
&\simeq \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) f'(\bar{\varepsilon}_i) \\
&\simeq \frac{n}{\sqrt{nh^d}} E \left[ K\left(\frac{x - X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) f'(\bar{\varepsilon}_i) \right].
\end{aligned}$$

We need to show that

$$E \left[ K\left(\frac{x - X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) f'(\bar{\varepsilon}_i) \right] = o(h^d).$$

Conditional on  $X_i$ ,

$$\begin{aligned}
&E \left[ K\left(\frac{x - X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) f'(\bar{\varepsilon}_i) | X_i \right] \\
&= K\left(\frac{x - X_i}{h}\right) \int_{b \leq f(\bar{\varepsilon}_i) \leq 2b} \frac{f'(\bar{\varepsilon}_i)^2}{f(\bar{\varepsilon}_i)} \frac{1}{b} g\left(\frac{f(\bar{\varepsilon}_i)}{b} - 1\right) f(\varepsilon_i) d\varepsilon_i \\
&= K\left(\frac{x - X_i}{h}\right) \int_{b \leq f(v) \leq 2b} \frac{f'(v)^2}{f(v)} \frac{1}{b} g\left(\frac{f(v)}{b} - 1\right) f(v - \delta_i) dv
\end{aligned}$$

Notice that  $g(\cdot) \geq 0$  is bounded, say,  $g(\cdot) \leq C$ , then

$$\int_{b \leq f(v) \leq 2b} \frac{f'(v)^2}{f(v)} \frac{1}{b} g\left(\frac{f(v)}{b} - 1\right) f(v - \delta_i) dv \leq \frac{C}{b} \int_{b \leq f(v) \leq 2b} \frac{f'(v)^2}{f(v)} f(v - \delta_i) dv.$$

Thus

$$\begin{aligned}
& E \left[ K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) f'(\bar{\varepsilon}_i) \right] \\
& \leq \frac{C}{b} \int \int_{b \leq f(v) \leq 2b} K \left( \frac{x - X}{h} \right) \frac{f'(v)^2}{f(v)} f(v - m(x) + m(X)) f_X(X) dv dX \\
& = \frac{C}{b} \times h^d \int \int_{b \leq f(v) \leq 2b} K(U) \frac{f'(v)^2}{f(v)} f(v + m(x - hU) - m(x)) f_X(x - hU) dv dU.
\end{aligned}$$

By Assumption A3, we have

$$\begin{aligned}
\int_{b \leq f(v) \leq 2b} f'(v)^2 dv & = \int_{b \leq f(v) \leq 2b} \frac{f'(v)^2}{f(v)} f(v) dv \\
& \leq \sup_{b \leq f(v) \leq 2b} f(v) \int_{b \leq f(v) \leq 2b} \frac{f'(v)^2}{f(v)} dv \\
& = 2b \cdot \int_{b \leq f(v) \leq 2b} \frac{f'(v)^2}{f(v)} dv = o(b)
\end{aligned}$$

by dominated convergence, since  $\int_{b \leq f(v) \leq 2b} \frac{f'(v)^2}{f(v)} dv \leq I(f) < \infty$ . Therefore,

$$E \left[ K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) f'(\bar{\varepsilon}_i) \right] \leq \frac{C}{b} \times h^d \times o(b) = o(h^d).$$

Consequently, under our bandwidth conditions,

$$\sum_{i=1}^n \frac{1}{\sqrt{nh^d}} K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \left[ \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right] = o_p(1/(h_n^q + n^{-1/2} h_n^{-d/2})).$$

Thus

$$T_{21A} = o_p(1).$$

For  $T_{21B}$ , following the analysis for  $\mathcal{J}_{31B}$ , we have

$$\begin{aligned}
& \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) [\tilde{m}(X_j) - m(X_j)] \\
& = \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \frac{e'_1 M^{-1} U_n(X_j)}{f_X(X_j)} \\
& \quad + \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \frac{e'_1 M^{-1} B_n(X_j)}{f_X(X_j)} \\
& \quad - \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right)
\end{aligned}$$

$$\begin{aligned}
& \times \frac{e_1' M^{-1} [M_n(X_j) - M f_X(X_j)] M_n^{-1}(X_j) U_n(X_j)}{f_X(X_j)} \\
& - \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \frac{1}{nh^2} \sum_{j \neq i} k'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \\
& \times \frac{e_1' M^{-1} [M_n(X_j) - M f_X(X_j)] M_n^{-1}(X_j) B_n(X_j)}{f_X(X_j)} \\
& = T_{211} + T_{212} + T_{213} + T_{214}. \tag{44}
\end{aligned}$$

Following similar analysis as those in the proof of Lemma H3, we can verify that each of these terms is  $o_p(1)$ . In particular,

$$\begin{aligned}
T_{211} &= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \frac{1}{nh^2} \sum_{j \neq i} k'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \frac{e_1' M^{-1} U_n(X_j)}{f_X(X_j)} \\
&= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \frac{1}{nh^2} \sum_{j \neq i} k'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \\
&\quad \times \frac{\sum_{\kappa} \omega^{0,\kappa} \frac{1}{nh^d} \sum_{l \neq j} \mathcal{K}\left(\frac{X_j - X_l}{h}\right) \left(\frac{X_j - X_l}{h}\right)^{\kappa} \varepsilon_l}{f_X(X_j)} \\
&= \frac{1}{n^{5/2} h^{2+3d/2}} \sum_{i=1}^n \sum_{j \neq i} \sum_{l \neq j} \sum_{\kappa} \omega^{0,\kappa} \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) k'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) K\left(\frac{x - X_i}{h}\right) \frac{\mathcal{K}\left(\frac{X_j - X_l}{h}\right) \left(\frac{X_j - X_l}{h}\right)^{\kappa} \varepsilon_l}{f_X(X_j)} \\
&= \sum_{i=1}^n \sum_{j \neq i} \sum_{\kappa} \frac{1}{n^{5/2} h^{2+3d/2}} \omega^{0,\kappa} \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) k'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) K\left(\frac{x - X_i}{h}\right) \frac{\mathcal{K}\left(\frac{X_j - X_i}{h}\right) \left(\frac{X_j - X_i}{h}\right)^{\kappa} \varepsilon_i}{f_X(X_j)} \\
&\quad + \sum_{i=1}^n \sum_{j \neq i} \sum_{\substack{l \neq j \\ l \neq i}} \sum_{\kappa} \frac{1}{n^{5/2} h^{2+3d/2}} \omega^{0,\kappa} \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) k'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) K\left(\frac{x - X_i}{h}\right) \frac{\mathcal{K}\left(\frac{X_j - X_l}{h}\right) \left(\frac{X_j - X_l}{h}\right)^{\kappa} \varepsilon_l}{f_X(X_j)} \\
&= \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \gamma_{1n}(z_i, z_j) + \sum_{i=1}^n \sum_{j=1}^n \sum_{\substack{l=1 \\ j \neq i, j \neq l, l \neq i}}^n \gamma_{2n}(z_i, z_j, z_l),
\end{aligned}$$

where

$$\gamma_{1n}(z_i, z_j) = \frac{1}{n^{5/2} h^{2+3d/2}} \sum_{\kappa} \omega^{0,\kappa} \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) k'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) K\left(\frac{x - X_i}{h}\right) \frac{\mathcal{K}\left(\frac{X_j - X_i}{h}\right) \left(\frac{X_j - X_i}{h}\right)^{\kappa} \varepsilon_i}{f_X(X_j)}$$

is a second order U-statistic, while

$$\gamma_{2n}(z_i, z_j, z_l) = \frac{1}{n^{5/2} h^{2+3d/2}} \sum_{\kappa} \omega^{0,\kappa} \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) k'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) K\left(\frac{x - X_i}{h}\right) \frac{\mathcal{K}\left(\frac{X_j - X_l}{h}\right) \left(\frac{X_j - X_l}{h}\right)^{\kappa} \varepsilon_l}{f_X(X_j)}$$

is a third order U-statistics with degeneracy, i.e.,  $E_{i,j}\gamma_{2n}(z_i, z_j, z_l) = 0$ . Therefore,

$$\begin{aligned}
& ET_{211} \\
&= \sum_{i=1}^n \sum_{j=1}^n E\gamma_{1n}(z_i, z_j) \\
&= \sum_{i=1}^n \sum_{j=1}^n \frac{1}{n^{5/2}h^{2+3d/2}} \sum_{\kappa} \omega^{0,\kappa} E \frac{f'(\bar{\varepsilon}_i)\varepsilon_i}{f(\bar{\varepsilon}_i)} g_b(f_i) K\left(\frac{x-X_i}{h}\right) E_i k' \left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) E_i \frac{\mathcal{K}\left(\frac{X_j - X_i}{h}\right) \left(\frac{X_j - X_i}{h}\right)^{\kappa}}{f_X(X_j)} \\
&= \sum_{i=1}^n \sum_{j=1}^n \frac{h^2 \times h^d}{n^{5/2}h^{2+3d/2}} \sum_{\kappa} \omega^{0,\kappa} E \frac{f'(\bar{\varepsilon}_i)\varepsilon_i}{f(\bar{\varepsilon}_i)} g_b(f_i) K\left(\frac{x-X_i}{h}\right) f'(\bar{\varepsilon}_i) \int K(u)u^{\kappa} du \\
&= \sum_{i=1}^n \sum_{j=1}^n \sum_{\kappa} \int K(u)u^{\kappa} du \frac{1}{n^{5/2}h^{d/2}} \omega^{0,\kappa} E \left\{ \frac{f'(\bar{\varepsilon}_i)^2 \varepsilon_i}{f(\bar{\varepsilon}_i)} g_b(f_i) K\left(\frac{x-X_i}{h}\right) \right\} \\
&= o(1).
\end{aligned}$$

For the variance of  $\sum_{i=1}^n \sum_{j=1}^n \gamma_{1n}(z_i, z_j)$ , the order of magnitude is determined by

$$O(n^2)E[\gamma_{1n}(z_i, z_j)^2 + \gamma_{1n}(z_i, z_j)\gamma_{1n}(z_j, z_i)] + O(n^3)E[\gamma_{1n}(z_i, z_j)\gamma_{1n}(z_l, z_j) + \gamma_{1n}(z_i, z_l)\gamma_{1n}(z_i, z_j)].$$

thus we need to verify that

$$E[\gamma_{1n}(z_i, z_j)]^2 = o(n^{-2}) \quad (45)$$

$$E[\gamma_{1n}(z_i, z_j)\gamma_{1n}(z_j, z_i)] = o(n^{-2}) \quad (46)$$

$$E[\gamma_{1n}(z_i, z_j)\gamma_{1n}(z_l, z_j)] = o(n^3) \quad (47)$$

$$E[\gamma_{1n}(z_i, z_l)\gamma_{1n}(z_i, z_j)] = o(n^{-3}). \quad (48)$$

To verify (45), we need to calculate the expectation of

$$\frac{1}{n^5 h^{4+3d}} \left(\frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\right)^2 \left[k' \left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right)\right]^2 \left[K\left(\frac{x-X_i}{h}\right)\right]^2 \frac{g_b(f_i)^2}{f_X(X_j)^2} \left[\mathcal{K}\left(\frac{X_j - X_i}{h}\right)\right]^2 \left(\frac{X_j - X_i}{h}\right)^{2\kappa} \varepsilon_i^2.$$

Conditional on  $\varepsilon_i$  and  $X_i$ ,

$$E_i \left\{ \frac{1}{f_X(X_j)^2} \left[\mathcal{K}\left(\frac{X_j - X_i}{h}\right)\right]^2 \left(\frac{X_j - X_i}{h}\right)^{2\kappa} \right\} = O(h^d),$$

and then

$$E \left\{ \left(\frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\right)^2 \varepsilon_i^2 g_b(f_i)^2 \left[k' \left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right)\right]^2 \left[K\left(\frac{x-X_i}{h}\right)\right]^2 \right\}$$

$$\begin{aligned}
&= E \left\{ \left( \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \right)^2 \varepsilon_i^2 g_b(f_i)^2 \left[ K \left( \frac{x - X_i}{h} \right) \right]^2 E \left\{ \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right]^2 | X \right\} \right\} \\
&\simeq E \left\{ \left( \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \right)^2 \varepsilon_i^2 g_b(f_i)^2 \left[ K \left( \frac{x - X_i}{h} \right) \right]^2 h f(\bar{\varepsilon}_i) \int k'(u)^2 du \right\} \\
&= h \left( \int k'(u)^2 du \right) E \left\{ \left[ K \left( \frac{x - X_i}{h} \right) \right]^2 E \left\{ \left( \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \right)^2 g_b(f_i)^2 f(\bar{\varepsilon}_i) \varepsilon_i^2 | X \right\} \right\} \\
&= h \left( \int k'(u)^2 du \right) E \left\{ \left[ K \left( \frac{x - X_i}{h} \right) \right]^2 E \left\{ \frac{f'(\bar{\varepsilon}_i)^2}{f(\bar{\varepsilon}_i)} g_b(f_i)^2 \varepsilon_i^2 | X \right\} \right\} \\
&= o(h^{1+db-2}).
\end{aligned}$$

For (46), we calculate the expectation of

$$\begin{aligned}
&\frac{1}{n^5 h^{4+3d}} \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \frac{f'(\bar{\varepsilon}_j)}{f(\bar{\varepsilon}_j)} g_b(f_j) k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) k' \left( \frac{\bar{\varepsilon}_j - \varepsilon_i}{h} \right) \varepsilon_i \varepsilon_j \\
&\frac{1}{f_X(X_i) f_X(X_j)} K \left( \frac{x - X_i}{h} \right) K \left( \frac{x - X_j}{h} \right) \mathcal{K} \left( \frac{X_i - X_j}{h} \right) \left( \frac{X_i - X_j}{h} \right)^\kappa \mathcal{K} \left( \frac{X_j - X_i}{h} \right) \left( \frac{X_j - X_i}{h} \right)^\kappa
\end{aligned}$$

We first condition on  $X_1, \dots, X_n$  and take expectation with respect to  $\varepsilon_j$ :

$$\begin{aligned}
&E \left\{ \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \frac{f'(\bar{\varepsilon}_j)}{f(\bar{\varepsilon}_j)} g_b(f_j) k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) k' \left( \frac{\bar{\varepsilon}_j - \varepsilon_i}{h} \right) \varepsilon_i \varepsilon_j | X_1, \dots, X_n \right\} \\
&= E \left\{ \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \varepsilon_i \int_{b \leq f(\bar{\varepsilon}_j) \leq 2b} \frac{f'(\bar{\varepsilon}_j)}{f(\bar{\varepsilon}_j)} g_b(f_j) k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) k' \left( \frac{\varepsilon_j + \delta_j - \varepsilon_i}{h} \right) \varepsilon_j f(\varepsilon_j) d\bar{\varepsilon}_j \right\} \\
&= \frac{1}{b} E \left\{ \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \varepsilon_i \int_{b \leq f(v) \leq 2b} \frac{f'(v)}{f(v)} g \left( \frac{f(v)}{b} - 1 \right) k' \left( \frac{\bar{\varepsilon}_i - v - \delta_j}{h} \right) k' \left( \frac{v - \varepsilon_i}{h} \right) (v - \delta_j) f(v - \delta_j) dv \right\} \\
&= \frac{1}{b} \int \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \varepsilon_i \int_{b \leq f(v) \leq 2b} \frac{f'(v)}{f(v)} g \left( \frac{f(v)}{b} - 1 \right) k' \left( \frac{\bar{\varepsilon}_i - v - \delta_j}{h} \right) k' \left( \frac{v - \varepsilon_i}{h} \right) \\
&\quad \times (v - \delta_j) f(v - \delta_j) dv f(\varepsilon_i) d\varepsilon_i \\
&= O \left( \frac{h}{b^2} \right),
\end{aligned}$$

while

$$\begin{aligned}
&E \left[ \frac{K \left( \frac{x - X_i}{h} \right) K \left( \frac{x - X_j}{h} \right) \mathcal{K} \left( \frac{X_i - X_j}{h} \right) \left( \frac{X_i - X_j}{h} \right)^\kappa \mathcal{K} \left( \frac{X_j - X_i}{h} \right) \left( \frac{X_j - X_i}{h} \right)^\kappa}{f_X(X_i) f_X(X_j)} \right] \\
&= h^{2d} \int \int K(U) K(U + V) \mathcal{K}(V) \mathcal{K}(V) (-V)^\kappa V^\kappa dV dU = O(h^{2d}).
\end{aligned}$$

Therefore,

$$E[\gamma_{1n}(z_i, z_j) \gamma_{1n}(z_j, z_i)] = \frac{1}{n^5 h^{3+db^2}} = o(n^{-2}),$$

Similarly

$$E[\gamma_{1n}(z_i, z_j)\gamma_{1n}(z_l, z_j) + \gamma_{1n}(z_i, z_l)\gamma_{1n}(z_i, z_j)] = o(n^{-3}).$$

For the third order U-statistics,

$$\text{var} \left[ \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i, j \neq l, l \neq i}}^n \sum_{l=1}^n \gamma_{2n}(z_i, z_j, z_l) \right] = O(n^3)E[\Gamma_{21}] + O(n^4)E[\Gamma_{22}] + O(n^5)E[\Gamma_{23}],$$

where  $\Gamma_{21}$  includes terms with three different indices, say  $\gamma_{2n}(z_i, z_j, z_l)^2$ ,  $\Gamma_{22}$  includes terms like  $\gamma_{2n}(z_i, z_j, z_l)\gamma_{2n}(z_i, z_r, z_l)$ , and  $\Gamma_{23}$  contains terms of type  $\gamma_{2n}(z_i, z_j, z_l)\gamma_{2n}(z_r, z_t, z_l)$ . We can verify that: for terms with three different indices

$$E[\Gamma_{21}] = O\left(\frac{1}{n^5 h^{4+3d}} \times h \times h^{2d}\right) = o(n^{-3}),$$

for terms with four different indices

$$E[\Gamma_{22}] = O\left(\frac{1}{n^5 h^{4+3d}} h^{d+4} h^{2d}\right) = o(n^{-4}).$$

The leading term in the variance of the third order U-statistics is the term with five different indices.

To calculate  $E\gamma_{2n}(z_i, z_j, z_l)\gamma_{2n}(z_r, z_t, z_l)$ , we consider terms

$$\begin{aligned} & E \frac{1}{n^5 h^{4+3d}} \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_r)}{f(\bar{\varepsilon}_r)} g_b(f_r) k' \left( \frac{\bar{\varepsilon}_r - \varepsilon_t}{h} \right) K \left( \frac{x - X_r}{h} \right) \\ & \frac{1}{f_X(X_j)} \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \left( \frac{X_j - X_l}{h} \right)^{\kappa_1} \frac{1}{f_X(X_t)} \mathcal{K} \left( \frac{X_t - X_l}{h} \right) \left( \frac{X_t - X_l}{h} \right)^{\kappa_2} \varepsilon_l^2. \end{aligned}$$

Conditional on  $\varepsilon_i$  and  $X_i, X_r, X_l$ , we take expectation with respect to  $X_t$  and  $X_l$ , get

$$E \left\{ \frac{1}{f_X(X_j)} \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \left( \frac{X_j - X_l}{h} \right)^{\kappa_1} \right\} E \left\{ \frac{1}{f_X(X_t)} \mathcal{K} \left( \frac{X_t - X_l}{h} \right) \left( \frac{X_t - X_l}{h} \right)^{\kappa_2} \right\} = O(h^{2d}).$$

Then, conditional on  $\bar{\varepsilon}_i, \bar{\varepsilon}_r$ , and  $X$ , we take expectation with respect to  $\varepsilon_t$  and  $\varepsilon_j$ , get

$$E k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) k' \left( \frac{\bar{\varepsilon}_r - \varepsilon_t}{h} \right) \simeq f'(\bar{\varepsilon}_i) f'(\bar{\varepsilon}_r) h^4.$$

Finally we take expectations on

$$\begin{aligned} & \frac{1}{n^5 h^{4+3d}} \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_r)}{f(\bar{\varepsilon}_r)} g_b(f_r) k' \left( \frac{\bar{\varepsilon}_r - \varepsilon_t}{h} \right) K \left( \frac{x - X_r}{h} \right) \\ & \frac{1}{f_X(X_j)} \mathcal{K} \left( \frac{X_j - X_l}{h} \right) \left( \frac{X_j - X_l}{h} \right)^{\kappa_1} \frac{1}{f_X(X_t)} \mathcal{K} \left( \frac{X_t - X_l}{h} \right) \left( \frac{X_t - X_l}{h} \right)^{\kappa_2} \varepsilon_l^2 \end{aligned}$$



and get

$$\begin{aligned} & \frac{1}{n^5 h^{4+3d}} \times h^4 h^{2d} E(\varepsilon_l^2) \int \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) f'(\bar{\varepsilon}_i) K\left(\frac{x - X_i}{h}\right) \int \frac{f'(\bar{\varepsilon}_r)}{f(\bar{\varepsilon}_r)} g_b(f_r) f'(\bar{\varepsilon}_r) K\left(\frac{x - X_r}{h}\right) \\ &= \frac{1}{n^5 h^{4+3d}} \times h^4 h^{2d} \times o(h^{2d}) = o\left(\frac{h^d}{n^5}\right). \end{aligned}$$

Thus

$$\text{var} \left[ \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{\substack{l=1 \\ l \neq i}}^n \varphi_{2n}(z_i, z_j, z_l) \right] = o(1),$$

and thus  $T_{211} = o_p(1)$ . Similarly,  $T_{213} = o_p(1)$ . Now we consider the leading bias effect  $T_{212}$ , it is easy to check that

$$\begin{aligned} & \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \frac{1}{nh^2} \sum_{j \neq i} k' \left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \frac{e_1' M^{-1} B_n(X_j)}{f_X(X_j)} \\ & \simeq \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \frac{1}{nh^2} \sum_{j \neq i} k' \left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \frac{1}{f_X(X_j)} \\ & \quad \times \left[ \frac{1}{nh^d} \sum_{\kappa} \omega^{1,\kappa} \sum_l \mathcal{K} \left(\frac{X_j - X_l}{h}\right) \left(\frac{X_j - X_l}{h}\right)^{p+\kappa} h^{p+1} m_{p+1}(X_j) \right] \\ &= \sum_{\kappa} \omega^{1,\kappa} \frac{h^{p+1}}{n^{5/2} h^{2+3d/2}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \sum_{j \neq i} k' \left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \frac{1}{f_X(X_j)} \\ & \quad \sum_l \mathcal{K} \left(\frac{X_j - X_l}{h}\right) \left(\frac{X_j - X_l}{h}\right)^{p+\kappa} m_{p+1}(X_j) \\ &= o_p\left(\frac{h^{p+1}}{n^{5/2} h^{2+3d/2}} \times h^2 \times h^{2d} \times n^3\right) = o_p(h^{p+1+d/2} \times n^{1/2}) = o_p(1), \end{aligned}$$

and  $T_{212} = o_p(1)$ . Similarly  $T_{214} = o_p(1)$ .

The second component of errors comes from standard kernel estimation and can be analyzed similarly.

$$T_{22} = \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(Y_i - \theta_0)}{f(Y_i - \theta_0)} g_b(f_i) (\bar{f}(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)).$$

We decompose  $\bar{f}(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)$  into a bias effect  $B(\bar{\varepsilon}_i)$  and a variance effect  $V(\bar{\varepsilon}_i)$  as

$$\begin{aligned} B(\bar{\varepsilon}_i) &= \frac{1}{nh} \sum_{j \neq i} E_i \left[ k \left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \right] - f(\bar{\varepsilon}_i) \\ V(\bar{\varepsilon}_i) &= \frac{1}{nh} \sum_{j \neq i} \left[ k \left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) - E_i \left[ k \left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \right] \right], \end{aligned}$$

thus

$$\begin{aligned}
T_{22} &= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(Y_i - \theta_0)}{f(Y_i - \theta_0)} g_b(f_i) V(\bar{\varepsilon}_i) \\
&\quad + \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(Y_i - \theta_0)}{f(Y_i - \theta_0)} g_b(f_i) B(\bar{\varepsilon}_i) \\
&= T_{22A} + T_{22B}.
\end{aligned}$$

The variance term is

$$\begin{aligned}
T_{22A} &= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \left\{ \frac{1}{nh} \sum_{j \neq i} \left[ k\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) - E_i \left[ k\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \right] \right] \right\} \\
&= \sum_{i=1}^n \sum_{j \neq i} \frac{1}{nh\sqrt{nh^d}} K\left(\frac{x - X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \left\{ \left[ k\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) - E_i \left[ k\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \right] \right] \right\}.
\end{aligned}$$

Denote

$$\gamma_{3n}(z_i, z_j) = \frac{1}{nh\sqrt{nh^d}} K\left(\frac{x - X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \left\{ \left[ k\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) - E_i \left[ k\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \right] \right] \right\},$$

again, this is a second order U-statistic with degeneracy and

$$E[T_{22A}]^2 = O(n^2) [E\gamma_{3n}(z_i, z_j)^2 + E[\gamma_{3n}(z_i, z_j)\gamma_{3n}(z_j, z_i)]] + O(n^3)E[\gamma_{3n}(z_i, z_j)\gamma_{3n}(z_l, z_j)]$$

We should verify the orders of magnitude of  $E[\gamma_{3n}(z_i, z_j)^2]$ ,  $E[\gamma_{3n}(z_i, z_j)\gamma_{3n}(z_j, z_i)]$ , and  $E[\gamma_{3n}(z_i, z_j)\gamma_{3n}(z_l, z_j)]$ . By a similar calculation as we did for  $\mathcal{J}_{32}$ , we can show that

$$\begin{aligned}
E\gamma_{3n}(z_i, z_j)^2 + E[\gamma_{3n}(z_i, z_j)\gamma_{3n}(z_j, z_i)] &= o(n^{-2}) \\
E[\gamma_{3n}(z_i, z_j)\gamma_{3n}(z_l, z_j)] &= o(n^{-3}).
\end{aligned}$$

In particular, the leading term in  $E[T_{22A}]^2$  is  $O(n^3)E[\gamma_{3n}(z_i, z_j)\gamma_{3n}(z_l, z_j)]$ .

$$\begin{aligned}
&E[\gamma_{3n}(z_i, z_j)\gamma_{3n}(z_l, z_j)] \\
&= \frac{1}{n^3 h^{2+d}} EK\left(\frac{x - X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \left[ k\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) - E_i \left[ k\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \right] \right] \\
&\quad K\left(\frac{x - X_l}{h}\right) \frac{f'(\bar{\varepsilon}_l)}{f(\bar{\varepsilon}_l)} g_b(f_l) \left[ k\left(\frac{\bar{\varepsilon}_l - \varepsilon_j}{h}\right) - E_l \left[ k\left(\frac{\bar{\varepsilon}_l - \varepsilon_j}{h}\right) \right] \right] \\
&= \frac{1}{n^3 h^{2+d}} EK\left(\frac{x - X_i}{h}\right) K\left(\frac{x - X_l}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \frac{f'(\bar{\varepsilon}_l)}{f(\bar{\varepsilon}_l)} g_b(f_l) k\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) k\left(\frac{\bar{\varepsilon}_l - \varepsilon_j}{h}\right) \\
&\quad + \frac{1}{n^3 h^{2+d}} EK\left(\frac{x - X_i}{h}\right) K\left(\frac{x - X_l}{h}\right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \frac{f'(\bar{\varepsilon}_l)}{f(\bar{\varepsilon}_l)} g_b(f_l) E_i \left[ k\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \right] E_l \left[ k\left(\frac{\bar{\varepsilon}_l - \varepsilon_j}{h}\right) \right]
\end{aligned}$$

$$\begin{aligned}
& -\frac{1}{n^3 h^{2+d}} EK \left( \frac{x - X_i}{h} \right) K \left( \frac{x - X_l}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \frac{f'(\bar{\varepsilon}_l)}{f(\bar{\varepsilon}_l)} g_b(f_l) k \left( \frac{\bar{\varepsilon}_l - \varepsilon_j}{h} \right) E_i \left[ k \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right] \\
& -\frac{1}{n^3 h^{2+d}} EK \left( \frac{x - X_i}{h} \right) K \left( \frac{x - X_l}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \frac{f'(\bar{\varepsilon}_l)}{f(\bar{\varepsilon}_l)} g_b(f_l) k \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) E_l \left[ k \left( \frac{\bar{\varepsilon}_l - \varepsilon_j}{h} \right) \right].
\end{aligned}$$

By direct calculations of moments, we can verify that

$$\begin{aligned}
& \frac{1}{n^3 h^{2+d}} EK \left( \frac{x - X_i}{h} \right) K \left( \frac{x - X_l}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} g_b(f_i) \frac{f'(\bar{\varepsilon}_l)}{f(\bar{\varepsilon}_l)} g_b(f_l) k \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) k \left( \frac{\bar{\varepsilon}_l - \varepsilon_j}{h} \right) \\
& = \frac{1}{n^3 h^{2+d}} \times o(h^{2d} \times h^2) \\
& = o \left( \frac{h^d}{n^3} \right) = o(n^{-3}),
\end{aligned}$$

and similarly we can verify that other terms are also  $o(n^{-3})$ . Thus  $T_{22A} = o_p(1)$ . The leading bias term is

$$\begin{aligned}
T_{22B} & = \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(Y_i - \theta_0)}{f(Y_i - \theta_0)} g_b(f_i) B(\bar{\varepsilon}_i) \\
& = o_p(\sqrt{nh^d} h^q) = o_p(1).
\end{aligned}$$

Thus  $T_{22} = o_p(1)$ . Consequently the term (42) =  $o_p(1)$ .

Finally we turn to the leading term of  $T_2$ ,

$$\frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(Y_i - \theta_0)}{f(Y_i - \theta_0)} [1 - G_i],$$

again, we denote  $f'/f$  as  $\psi$  and under given assumptions we obtain a Taylor expansion that  $\psi(Y_i - \theta_0) \simeq \psi(\varepsilon_i) + \psi'(\varepsilon_i)\delta_i$ . Thus

$$\begin{aligned}
& \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(Y_i - \theta_0)}{f(Y_i - \theta_0)} [1 - G_i] \\
& \simeq \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \psi(\varepsilon_i) [1 - G_i] \tag{49}
\end{aligned}$$

$$+ \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \psi'(\varepsilon_i) \delta_i [1 - G_i]. \tag{50}$$

To verify the order of magnitude of (49), notice that by Assumptions A2 and A3,  $\varepsilon_i$  is symmetrically distributed with zero mean, it is easy to verify that  $\psi(\varepsilon_i)[1 - G_i]$  is i.i.d. with zero mean, thus we just need to verify the second moment of (49). Notice that

$$E \left[ K \left( \frac{x - X_i}{h} \right)^2 \right] = h^d f_X(x) \mu_2(K), \tag{51}$$

and

$$E[\psi(\varepsilon_i)^2[1 - G_i]] = \int \left[ \frac{f'}{f}(\varepsilon_i) \right]^2 f(\varepsilon_i)[1 - G_i] d\varepsilon_i = o(1) \quad (52)$$

under assumption A3 (or A3'). Thus, combining (51) and (52), we can show that the order of (49) is  $o_p(1)$ . If we assume that Assumption A3' holds, it can be shown that

$$E[\psi(\varepsilon_i)^2(1 - G_i)] = O(b^{(e-1)/e}). \quad (53)$$

and the order of magnitude of (49) is  $O_p(b^{(e-1)/(2e)}) = o_p(1)$ .

Now we look at the term (50), notice that

$$\delta_i = m(X_i) - m(x) \simeq \frac{1}{\mathbf{k}!} \sum_{0 \leq |\mathbf{k}| \leq q} (D^{\mathbf{k}}m)(x)(X_i - x)^{\mathbf{k}}.$$

thus, by Assumptions A5,

$$\begin{aligned} E \left[ K \left( \frac{x - X_i}{h} \right) \delta_i \right] &\simeq E \left[ K \left( \frac{x - X_i}{h} \right) \frac{1}{\mathbf{k}!} \sum_{0 \leq |\mathbf{k}| \leq p} (D^{\mathbf{k}}m)(x)(X_i - x)^{\mathbf{k}} \right] \\ &= h^{d+q} \phi(K, m, f_X), \end{aligned}$$

where  $\phi(K, m, f_X)$  is a function of the kernel and derivatives of  $m$  and  $f_X$  evaluated at  $x$ . And

$$E[\psi'(\varepsilon_i)(1 - G_i)] = o(1).$$

Consequently, by calculation of moments we have, under bandwidth assumption A5, (50) =  $o_p(1)$ . ■

LEMMA S3. *Under our conditions*

$$T_3 \xrightarrow{p} 0.$$

PROOF. The analysis of  $T_3$  is similar to that of  $\mathcal{J}_3$ .

$$\begin{aligned} T_3 &= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{\tilde{f}'(\bar{\varepsilon}_i) - f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} G_i \\ &= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{[\tilde{f}'(\bar{\varepsilon}_i) - \bar{f}'(\bar{\varepsilon}_i)] + [\bar{f}'(\bar{\varepsilon}_i) - f'(\bar{\varepsilon}_i)]}{f(\bar{\varepsilon}_i)} G_i \\ &= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{[\tilde{f}'(\bar{\varepsilon}_i) - \bar{f}'(\bar{\varepsilon}_i)]}{f(\bar{\varepsilon}_i)} G_i \\ &\quad + \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{[\bar{f}'(\bar{\varepsilon}_i) - f'(\bar{\varepsilon}_i)]}{f(\bar{\varepsilon}_i)} G_i \\ &= T_{31} + T_{32} \end{aligned}$$

$$\begin{aligned}
T_{31} &= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{[\tilde{f}'(\bar{\varepsilon}_i) - \bar{f}'(\bar{\varepsilon}_i)]}{f(\bar{\varepsilon}_i)} G_i \\
&= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{\frac{1}{nh^2} \sum_{j \neq i} \left[ k'\left(\frac{\bar{\varepsilon}_i - \bar{\varepsilon}_j}{h}\right) - k'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \right]}{f(\bar{\varepsilon}_i)} G_i \\
&\simeq \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{\frac{1}{nh^3} \sum_{j \neq i} k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) [\tilde{m}(X_j) - m(X_j)]}{f(\bar{\varepsilon}_i)} G_i \\
&= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{1}{f(\bar{\varepsilon}_i)} \frac{1}{nh^3} \sum_{j \neq i} k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) [e'_1 M_n^{-1}(X_j) U_n(X_j)] G_i \\
&\quad + \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{1}{f(\bar{\varepsilon}_i)} \frac{1}{nh^3} \sum_{j \neq i} k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) [e'_1 M_n^{-1}(X_j) B_n(X_j)] G_i,
\end{aligned}$$

analysis of these terms will then be the similar to the previous analysis. In particular, the leading variance term is

$$\begin{aligned}
&\frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{1}{f(\bar{\varepsilon}_i)} \frac{1}{nh^3} \sum_{j \neq i} k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) G_i \frac{[e'_1 M^{-1} U_n(X_j)]}{f_X(X_j)} \\
&= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{1}{f(\bar{\varepsilon}_i)} \frac{1}{nh^3} \sum_{j \neq i} k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) G_i \frac{\sum_{\kappa} \omega^{0,\kappa} \frac{1}{nh^d} \sum_{l \neq j} \mathcal{K}\left(\frac{X_j - X_l}{h}\right) \left(\frac{X_j - X_l}{h}\right)^{\kappa} \varepsilon_l}{f_X(X_j)} \\
&= \sum_{i=1}^n \sum_{j \neq i} \sum_{l \neq j} \sum_{\kappa} \omega^{0,\kappa} \frac{1}{n^{5/2} h^{3d/2+3}} \frac{G_i}{f(\bar{\varepsilon}_i)} \frac{1}{f_X(X_j)} K\left(\frac{x-X_i}{h}\right) k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \\
&\quad \times \mathcal{K}\left(\frac{X_j - X_l}{h}\right) \left(\frac{X_j - X_l}{h}\right)^{\kappa} \varepsilon_l \\
&= \sum_{i=1}^n \sum_{j \neq i} \sum_{\kappa} \omega^{0,\kappa} \frac{1}{n^{5/2} h^{3d/2+3}} \frac{G_i}{f(\bar{\varepsilon}_i)} \frac{1}{f_X(X_j)} K\left(\frac{x-X_i}{h}\right) k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \\
&\quad \times \mathcal{K}\left(\frac{X_j - X_i}{h}\right) \left(\frac{X_j - X_i}{h}\right)^{\kappa} \varepsilon_i \\
&\quad + \sum_{i=1}^n \sum_{j \neq i} \sum_{l \neq j} \sum_{\kappa} \omega^{0,\kappa} \frac{1}{n^{5/2} h^{3d/2+3}} \frac{G_i}{f(\bar{\varepsilon}_i)} \frac{1}{f_X(X_j)} K\left(\frac{x-X_i}{h}\right) k''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \\
&\quad \times \mathcal{K}\left(\frac{X_j - X_l}{h}\right) \left(\frac{X_j - X_l}{h}\right)^{\kappa} \varepsilon_l \\
&= \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \gamma_{3n}(z_i, z_j) + \sum_{i=1}^n \sum_{j=1}^n \sum_{\substack{l=1 \\ j \neq i, j \neq l, l \neq i}}^n \gamma_{4n}(z_i, z_j, z_l),
\end{aligned}$$

$$\begin{aligned}
E \left[ \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \gamma_{3n}(z_i, z_j) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{\substack{l=1 \\ l \neq i, l \neq j}}^n \gamma_{4n}(z_i, z_j, z_l) \right] &= E \left[ \sum_{i=1}^n \sum_{j=1}^n \gamma_{3n}(z_i, z_j) \right] \\
&= O(n^{-1/2} h^{d/2} b^{-1}) = o(1)
\end{aligned}$$

and the variances are verified similar to  $\mathcal{J}_{31B}$ .

$$\begin{aligned}
T_{32} &= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{[\bar{f}'(\bar{\varepsilon}_i) - f'(\bar{\varepsilon}_i)]}{f(\bar{\varepsilon}_i)} G_i \\
&= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{\frac{1}{nh^2} \sum_{j \neq i} k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) - f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} G_i \\
&= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{\frac{1}{nh^2} \sum_{j \neq i} \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) - E_i \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right] \right]}{f(\bar{\varepsilon}_i)} G_i \\
&\quad + \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{\frac{1}{nh^2} \sum_{j \neq i} E_i \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right] - f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} G_i \\
&= T_{32A} + T_{32B}
\end{aligned}$$

$$\begin{aligned}
T_{32A} &= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{\frac{1}{nh^2} \sum_{j \neq i} \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) - E_i \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right] \right]}{f(\bar{\varepsilon}_i)} G_i \\
&= \frac{1}{nh^2} \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n \sum_{j \neq i} K \left( \frac{x - X_i}{h} \right) \frac{k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) - E_i \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right]}{f(\bar{\varepsilon}_i)} G_i \\
&= \sum_{i=1}^n \sum_{j \neq i} \gamma_{5n}(z_i, z_j),
\end{aligned}$$

where

$$\gamma_{5n}(z_i, z_j) = \frac{1}{n^{3/2} h^{2+d/2}} K \left( \frac{x - X_i}{h} \right) \frac{k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) - E_i \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right]}{f(\bar{\varepsilon}_i)} G_i.$$

Notice that  $E_i[\gamma_{5n}(z_i, z_j)] = 0$ ,  $T_{32A}$  is a second order U-statistic with first order degeneracy.  $ET_{32A} = 0$  and

$$\text{var}(T_{32A}) = O(n^2) E [\gamma_{5n}(z_i, z_j)^2 + \gamma_{5n}(z_i, z_j) \gamma_{5n}(z_j, z_i)] + O(n^3) E [\gamma_{5n}(z_i, z_j) \gamma_{5n}(z_l, z_j)]$$

in which  $i, j, l$  are distinct indices. Conditional on  $X_1, \dots, X_n$ , using integration by parts and change of variables,

$$E_i \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right] = \int k' \left( \frac{\bar{\varepsilon}_i - \varepsilon}{h} \right) f(\varepsilon) d\varepsilon = h \int k \left( \frac{\bar{\varepsilon}_i - \varepsilon}{h} \right) f'(\varepsilon) d\varepsilon = h^2 \int k(u) f'(\bar{\varepsilon}_i - uh) du$$

and

$$\begin{aligned} E_j \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right] &= \int k' \left( \frac{\varepsilon + \delta_i - \varepsilon_j}{h} \right) f(\varepsilon) d\varepsilon \\ &= h \int k \left( \frac{\varepsilon + \delta_i - \varepsilon_j}{h} \right) f'(\varepsilon) d\varepsilon = h^2 \int k(u) f'(\delta_i - \varepsilon_j - uh) du. \end{aligned}$$

Then, conditional on  $\varepsilon_j$ ,

$$E \left[ \frac{\left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) k' \left( \frac{\bar{\varepsilon}_l - \varepsilon_j}{h} \right) \right]}{f(\bar{\varepsilon}_i) f(\bar{\varepsilon}_l)} G_i G_l \right] = E \left[ \frac{E_j k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) E_j \left[ k' \left( \frac{\bar{\varepsilon}_l - \varepsilon_j}{h} \right) \right]}{f(\bar{\varepsilon}_i) f(\bar{\varepsilon}_l)} G_i G_l \right] = b^{-2} h^4 \phi_1(\delta_i, \delta_l)$$

similarly

$$\begin{aligned} E \left[ \frac{\left[ E_i k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) k' \left( \frac{\bar{\varepsilon}_l - \varepsilon_j}{h} \right) \right]}{f(\bar{\varepsilon}_i) f(\bar{\varepsilon}_l)} G_i G_l \right] &= b^{-2} h^4 \phi_2(\delta_i, \delta_l), \\ E \left[ \frac{\left[ E_i k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) E_l k' \left( \frac{\bar{\varepsilon}_l - \varepsilon_j}{h} \right) \right]}{f(\bar{\varepsilon}_i) f(\bar{\varepsilon}_l)} G_i G_l \right] &= b^{-2} h^4 \phi_3(\delta_i, \delta_l), \end{aligned}$$

where  $\phi_i, i = 1, 2, 3$  are uniformly bounded functions. Finally, notice that when we integrate over the distribution of  $X$ , we pick up an additional factor of order  $h^{2d}$ , we have

$$\begin{aligned} &EK \left( \frac{x - X_i}{h} \right) K \left( \frac{x - X_l}{h} \right) \left[ \frac{k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) - E_i \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right]}{f(\bar{\varepsilon}_i)} \frac{k' \left( \frac{\bar{\varepsilon}_l - \varepsilon_j}{h} \right) - E_l \left[ k' \left( \frac{\bar{\varepsilon}_l - \varepsilon_j}{h} \right) \right]}{f(\bar{\varepsilon}_l)} G_i G_l \right] \\ &= O(b^{-2} h^{2d+4}). \end{aligned}$$

And for

$$\begin{aligned} &E \left[ K \left( \frac{x - X_i}{h} \right) \right]^2 \left[ \frac{k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) - E_i \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right]}{f(\bar{\varepsilon}_i)} \right]^2 \\ &= E \left[ K \left( \frac{x - X_i}{h} \right) \right]^2 \left[ \frac{k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right)^2 - 2k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) E_i \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right] + \left\{ E_i \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right] \right\}^2}{f(\bar{\varepsilon}_i)^2} \right], \end{aligned}$$

notice the following facts:

$$E_i \left[ \frac{k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right)^2}{f(\bar{\varepsilon}_i)^2} \right] = \frac{1}{f(\bar{\varepsilon}_i)^2} \int k' \left( \frac{\bar{\varepsilon}_i - \varepsilon}{h} \right)^2 f(\varepsilon) d\varepsilon = h \frac{1}{f(\bar{\varepsilon}_i)^2} \int k'(u)^2 f(\bar{\varepsilon}_i - uh) du$$

and

$$E \left[ K \left( \frac{x - X_i}{h} \right)^2 \right] = h^d f_X(x) \mu_2(K)$$

it can be verified that

$$E \left[ K \left( \frac{x - X_i}{h} \right) \right]^2 \left[ \frac{k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) - E_i \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right]}{f(\bar{\varepsilon}_i)} \right]^2 = O(b^{-2} h^{d+1}).$$

Similarly,

$$\begin{aligned} & EK \left( \frac{x - X_i}{h} \right) K \left( \frac{x - X_j}{h} \right) \left[ \frac{k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) - E_i \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right]}{f(\bar{\varepsilon}_i)} \frac{k' \left( \frac{\bar{\varepsilon}_j - \varepsilon_i}{h} \right) - E_j \left[ k' \left( \frac{\bar{\varepsilon}_j - \varepsilon_i}{h} \right) \right]}{f(\bar{\varepsilon}_j)} \right] \\ &= O(b^{-2} h^{2d+1}). \end{aligned}$$

Thus we have

$$\begin{aligned} T_{32A} &= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{\frac{1}{nh^2} \sum_{j \neq i} \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) - E_i \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right] \right]}{f(\bar{\varepsilon}_i)} \\ &= O_p(n^{-1/2} h^{-3/2} b^{-1} + h^{d/2} b^{-1}). \end{aligned}$$

For  $T_{32B}$ , notice that, conditional on  $X_1, \dots, X_n$ ,

$$E_i \left[ \frac{1}{nh^2} \sum_{j \neq i} E_i \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right] - f'(\bar{\varepsilon}_i) \right] = h^q f^{(q+1)}(\bar{\varepsilon}_i) \int k(u) u^q du$$

it is easy to show that

$$T_{32B} = O_p(\sqrt{nh^d} h^q).$$

■

LEMMA S4. *Under our conditions*

$$T_4 \xrightarrow{p} 0.$$

PROOF. The proof is similar to that of Lemma S3. Again, we decompose  $\tilde{f}(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)$  into  $[\tilde{f}(\bar{\varepsilon}_i) - \bar{f}(\bar{\varepsilon}_i)] + [\bar{f}(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)]$ , thus

$$T_4 = T_{41} + T_{42},$$



with

$$T_{41} = \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{\tilde{f}'(\bar{\varepsilon}_i) [\tilde{f}(\bar{\varepsilon}_i) - \bar{f}(\bar{\varepsilon}_i)]}{f(\bar{\varepsilon}_i)^2} G_i,$$

$$T_{42} = \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i) [\bar{f}(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)]}{f(\bar{\varepsilon}_i)^2} G_i.$$

The orders of magnitude of  $T_{41}$  and  $T_{42}$  can be verified similarly as  $T_{31}$  and  $T_{32}$ . For example, we can write  $T_{42}$  as the variance effect  $T_{42A}$  and the bias effect  $T_{42B}$ ,

$$T_{42A} = \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i) \frac{1}{nh} \sum_{j \neq i} \left[ k\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) - E_i \left[ k\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \right] \right]}{f(\bar{\varepsilon}_i)^2} G_i$$

$$= \sum_{i=1}^n \sum_{j \neq i} \frac{1}{n^{3/2} h^{1+d/2}} K\left(\frac{x - X_i}{h}\right) \frac{f'(\bar{\varepsilon}_i) \left[ k\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) - E_i \left[ k\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \right] \right]}{f(\bar{\varepsilon}_i)^2} G_i$$

again, notice that it is a second order U-statistic with first order degeneracy,  $ET_{42A} = 0$ . As the proof for  $T_{32A}$ , we show

$$T_{42A} = O_p(n^{-1/2} h^{-1/2} b^{-1} + h^{d/2} b^{-1}).$$

In particular, conditional on  $X_1, \dots, X_n$ , using integration by parts and change of variables,

$$E_i \left[ k\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \right] = \int k\left(\frac{\bar{\varepsilon}_i - \varepsilon}{h}\right) f(\varepsilon) d\varepsilon = h \int k(u) f(\bar{\varepsilon}_i - uh) du,$$

and

$$E_j \left[ k\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) \right] = \int k\left(\frac{\varepsilon + \delta_i - \varepsilon_j}{h}\right) f(\varepsilon) d\varepsilon = h \int k(u) f(\delta_i - \varepsilon_j - uh) du.$$

Conditional on  $\varepsilon_j$ ,

$$E \left[ \frac{f'(\bar{\varepsilon}_i) f'(\bar{\varepsilon}_l) \left[ k\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) k\left(\frac{\bar{\varepsilon}_l - \varepsilon_j}{h}\right) \right]}{f(\bar{\varepsilon}_i)^2 f(\bar{\varepsilon}_l)^2} G_i G_l \right]$$

$$= E \left[ \frac{f'(\bar{\varepsilon}_i) f'(\bar{\varepsilon}_l) E_j k\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) E_j \left[ k\left(\frac{\bar{\varepsilon}_l - \varepsilon_j}{h}\right) \right]}{f(\bar{\varepsilon}_i) f(\bar{\varepsilon}_l) f(\bar{\varepsilon}_i) f(\bar{\varepsilon}_l)} G_i G_l \right]$$

$$= b^{-2} h^2 \phi_4(\delta_i, \delta_l),$$

similarly,

$$E \left[ \frac{f'(\bar{\varepsilon}_i) f'(\bar{\varepsilon}_l) \left[ E_i k\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) k\left(\frac{\bar{\varepsilon}_l - \varepsilon_j}{h}\right) \right]}{f(\bar{\varepsilon}_i)^2 f(\bar{\varepsilon}_l)^2} G_i G_l \right] = b^{-2} h^2 \phi_5(\delta_i, \delta_l),$$

$$E \left[ \frac{f'(\bar{\varepsilon}_i) f'(\bar{\varepsilon}_l) \left[ E_i k'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h}\right) E_l k'\left(\frac{\bar{\varepsilon}_l - \varepsilon_j}{h}\right) \right]}{f(\bar{\varepsilon}_i)^2 f(\bar{\varepsilon}_l)^2} G_i G_l \right] = b^{-2} h^2 \phi_6(\delta_i, \delta_l),$$

where  $\phi_i, i = 4, 5, 6$  are uniformly bounded functions. Finally, notice that when we integrate over the distribution of  $X$ , we pick up an additional factor of order  $h^{2d}$ , we have

$$\begin{aligned} & EK \left( \frac{x - X_i}{h} \right) K \left( \frac{x - X_l}{h} \right) \times \\ & \left[ \frac{k \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) - E_i \left[ k \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right]}{f(\bar{\varepsilon}_i)^2} f'(\bar{\varepsilon}_i) \frac{k \left( \frac{\bar{\varepsilon}_l - \varepsilon_j}{h} \right) - E_i \left[ k \left( \frac{\bar{\varepsilon}_l - \varepsilon_j}{h} \right) \right]}{f(\bar{\varepsilon}_l)^2} G_i G_l \right] \\ & = O(b^{-2} h^{2d+2}). \end{aligned}$$

And

$$E \left[ \frac{f'(\bar{\varepsilon}_i)^2 k \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right)^2}{f(\bar{\varepsilon}_i)^4} \middle| \bar{\varepsilon}_i \right] = \frac{f'(\bar{\varepsilon}_i)^2}{f(\bar{\varepsilon}_i)^4} \int k \left( \frac{\bar{\varepsilon}_i - \varepsilon}{h} \right)^2 f(\varepsilon) d\varepsilon = h \frac{f'(\bar{\varepsilon}_i)^2}{f(\bar{\varepsilon}_i)^4} \int k(u)^2 f(\bar{\varepsilon}_i - uh) du$$

it can be verified that

$$E \left[ K \left( \frac{x - X_i}{h} \right) \right]^2 \left[ \frac{f'(\bar{\varepsilon}_i) \left( k \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) - E_i \left[ k \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right] \right)}{f(\bar{\varepsilon}_i)^2} G_i \right]^2 = O(b^{-2} h^{d+1}).$$

By the same approach as for  $T_{32B}$ , it is easy to show that

$$T_{42B} = \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i) \left[ \frac{1}{nh} \sum_{j \neq i} E_i \left[ k' \left( \frac{\bar{\varepsilon}_i - \varepsilon_j}{h} \right) \right] - f(\bar{\varepsilon}_i) \right]}{f(\bar{\varepsilon}_i)^2} = O_p(\sqrt{nh^d} h^q).$$

■

LEMMA S5. *Under our conditions*

$$T_5 \xrightarrow{p} 0.$$

PROOF. For the remainder terms  $T_5$ , we need to show that

$$\frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(Y_i - \theta_0) \{ \tilde{f}(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i) \}^2}{f^2(\bar{\varepsilon}_i) \tilde{f}(\bar{\varepsilon}_i)} G_i$$

goes to zero as  $n \rightarrow \infty$ . Notice that

$$\begin{aligned} & \left| \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(Y_i - \theta_0) \{ \tilde{f}(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i) \}^2}{f^2(\bar{\varepsilon}_i) \tilde{f}(\bar{\varepsilon}_i)} G_i \right| \\ & \leq \max_{1 \leq i \leq n} \left| \{ \tilde{f}(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i) \}^2 \right| \times \left| \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)} \frac{G_i}{f(\bar{\varepsilon}_i) \tilde{f}(\bar{\varepsilon}_i)} \right|. \end{aligned}$$

Since  $|\tilde{f}_i| > b$ , and using the result (23), the remainder term is of order

$$O_p(n^{1/2}h^{d/2})O_p(b^{-2})O_p(h^{2q-2} + n^{-1}h^{-d-2}\log^2(n)) = O_p(h^{2q+d/2-2}n^{1/2}b^{-2} + n^{-1/2}h^{-d/2-2}b^{-2}(\log^2(n))),$$

which is  $o_p(1)$  under our bandwidth conditions. ■

### 5.2.3 Remainder term

It suffices to show that

$$\sup_{|\theta - \theta_0| \leq cn^{-1/2}h^{-d/2}} \left| \frac{\partial^2 \tilde{S}_n(\theta)}{\partial \theta^2} \right| = o_p(\sqrt{nh^d}) \quad (54)$$

because  $nh^d(\tilde{\theta} - \theta_0)^2 = O_p(1)$ . The expression for  $R_n(\theta^*)$  is quite complicated but its analysis is similar to that of the Hessian except that: (a) we require only bounds in probability that are quite weak; (b) those bounds, however, have to be locally uniform in the argument  $\theta$ . By an analysis similar to that of the Hessian, it can be shown that  $\partial^2 \tilde{S}_n(\theta)/\partial \theta^2 = O_p(1)$  for any given  $\theta$ . The extension to local uniform over  $\{\theta : |\theta - \theta_0| \leq n^{-1/2}h^{-d/2}\}$  follows from the smoothness properties on the kernel that we have imposed. We just examine a single key term

$$R_{n1}(\theta) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{\tilde{f}'''}{\tilde{f}}(Y_i - \theta) G_b(\tilde{f}_i).$$

By construction

$$\frac{G_b(\tilde{f}_i)}{\tilde{f}_i} \leq \frac{1}{b},$$

while

$$\sup_{t \in \mathbb{R}} |\tilde{f}'''(t)| \leq h^{-4} \sup_u |k'''(u)|.$$

Therefore,

$$|R_{n1}^*(\theta)| = O_p(h^{-4}b^{-1}).$$

Therefore, (54) is satisfied because  $h^{-4}b^{-1}/\sqrt{nh^d} = 1/\sqrt{nh^{d+8}b^2}$ , which is assumed to go to zero under our condition A5. By expanding out we can obtain better results but with considerably more calculations. ■

## REFERENCES

- Ahn, H. (1995). Nonparametric two-stage estimation of conditional choice probabilities in a binary choice model under uncertainty, *Journal of Econometrics* 67, 337-378.
- Ai, C., (1997). A Semiparametric Maximum Likelihood Estimator. *Econometrica*, 65, 933-963.
- Akahira, M., and Takeuchi. K, (1995). *Non-Regular Statistical Estimation*. Springer, New York.
- Andrews, D.W.K., (1995). Nonparametric Kernel Estimation for Semiparametric Models. *Econometric Theory*, 11,560-596.
- Beran, R., (1974). Asymptotically efficient adaptive rank estimates in location models. *Annals of Statistics* 2, 248-266.
- Bickel, P.J., (1975). One-step Huber estimates in the linear model. *Journal of the American Statistical Association* 70, 428-434.
- Bickel, P. J., (1982). On Adaptive Estimation. *Annals of Statistics*, 10, 647-671.
- Bickel, P.J., C.A.J. Klaassen, Y. Ritov, and J.A. Wellner (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer: Berlin.
- Chaudhuri, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Annals of Statistics* 19, 760-777.
- Copas, J.B. (1994). Local likelihood based on kernel censoring. *Journal of the Royal Statistical Society, Series B*, 57, 221-235.
- Drost, F., and C.A.J. Klaassen (1997). Efficient estimation in semiparametric GARCH models. *Journal of Econometrics* 81, 193-221.
- Drost, F., C.A.J. Klaassen, and B.J.M. Werker (1997). Adaptive Estimation in Time Series Models. *Annals of Statistics* 25, 786-817.
- 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. *The Annals of Statistics*, 21, 196-216.

- Fan, J., and J. Chen (1999). One-step Local Quasi-Likelihood Estimation. *Journal of the Royal Statistical Society, Series B.* 61, 927-943.
- Fan, J., M. Farnen, and I. Gijbels (2000). Local Maximum Likelihood Estimation and Inference. Forthcoming in *Journal of the American Statistical Association*.
- Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and Its Applications*, Chapman & Hall.
- Gonzalez-Manteiga, W. (1990). Asymptotic normality of Generalized Functional Estimators dependent on Covariates. *Journal of Statistical Planning and Inference.* 24, 377-390.
- Gozalo, P. and O. Linton. (2000) Local Nonlinear Least Squares: Using Parametric Information in Nonparametric Regression. *Journal of Econometrics* 99, 63-106.
- Härdle, W. (1990). *Applied Nonparametric Regression*, Cambridge University Press.
- Härdle, W., and O. B. Linton (1994), *Applied Nonparametric Methods*, *The Handbook of Econometrics*, Vol. IV, 2295-2339, Eds. D. McFadden and R. Engle, North Holland.
- Hjort, N.L. (1993). Dynamic likelihood hazard estimation. *Biometrika*. To appear.
- Hjort, N.L., and M.C. Jones (1994). Local fitting of regression models by likelihood: what is important? Institute of Mathematics, University of Oslo, Statistical Research Report no. 10.
- Hjort, N.L., and M.C. Jones (1996). Locally parametric nonparametric density estimation. *Annals of Statistics* 24, 1619-1647.
- Hodgson, D., (1997). Adaptive Estimation of Cointegrating Regressions with ARMA Errors. manuscript, University of Rochester. Forthcoming in *Journal of Econometrics*.
- Jeganathan, P. (1995). Some Aspects of Asymptotic Theory with Applications to Time Series Models. *Econometric Theory*, 11, 818-887.
- Kreiss, J., (1987). On Adaptive Estimation in Stationary ARMA process. *Annals of Statistics*, 15, 112-133.
- Linton, O.B. (1993). Adaptive estimation in ARCH Models. *Econometric Theory* 9, 539-569.

- Linton, O.B. (1997). Efficient estimation of additive nonparametric regression models. *Biometrika*, 84, 469-474.
- Linton, O.B., and Z. Xiao (1998). Second order approximation for adaptive regression estimators. Forthcoming in *Econometric Theory*.
- Loader, C.R. (1996). Local likelihood density estimation. *Annals of Statistics* 24, 1602-1618.
- Manski, C.F., (1984). Adaptive Estimation of Nonlinear Regression Models. *Econometric Reviews*, 3, 187-208.
- Masry, E. (1996). Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *Journal of Time Series Analysis* 17, 571-599.
- Robinson, P.M. (1989). Time varying nonlinear regression. in *Statistical Analysis and Forecasting of Economic Structural Change*. Eds. P. Hackl and A. Westland. Springer-Verlag.
- Rothenberg, T., and C. T. Leenders (1964). Efficient Estimation of Simultaneous Equation Systems. *Econometrica* 32, 57-76.
- Staniswalis, J.G. (1989). The Kernel Estimate of a Regression Function in Likelihood Based Models. *Journal of the American Statistical Association*, 84, 276-283.
- Steigerwald, D., (1992). Adaptive Estimation in Time Series Regression Models. *Journal of Econometrics*, 54, 251-276.
- Stein, C. (1956). Efficient nonparametric testing and estimation. in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, ed. by J. Neyman. University of California Press.
- Stone, C., (1975). Adaptive maximum likelihood estimation of a location parameter. *Annals of Statistics*, 3, 267-284.
- Tibshirani, R. (1984). Local Likelihood estimation. PhD Thesis, Stanford University.
- Tibshirani, R. and T. Hastie (1987). Local likelihood Estimation. *Journal of the American Statistical Association*, 82, 559-567.

Wand, M.P., and M.C. Jones (1995). Kernel Smoothing. Chapman and Hall: London.