Oliver B. Linton and Zhijie Xiao

Second-order approximation for adaptive regression estimators

Originally published in Econometric theory, 17 (5). pp. 984-1024 © 2001 Cambridge University Press.

Econometric Theory, **17**, 2001, 984–1024. Printed in the United States of America.

SECOND-ORDER APPROXIMATION FOR ADAPTIVE REGRESSION ESTIMATORS

Oliver Linton
London School of Economics
and
Yale University

Zhijie Xiao
University of Illinois at Urbana-Champaign

We derive asymptotic expansions for semiparametric adaptive regression estimators. In particular, we derive the asymptotic distribution of the second-order effect of an adaptive estimator in a linear regression whose error density is of unknown functional form. We then show how the choice of smoothing parameters influences the estimator through higher order terms. A method of bandwidth selection is defined by minimizing the second-order mean squared error. We examine both independent and time series regressors; we also extend our results to a *t*-statistic. Monte Carlo simulations confirm the second order theory and the usefulness of the bandwidth selection method.

1. INTRODUCTION

In estimation problems where a Gaussian assumption on the underlying distribution of the data is inappropriate, the so-called adaptive estimator provides an alternative to the conventional Gaussian maximum likelihood estimator (MLE) 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 MLE can be achieved by adaptive estimators in many econometric 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, whereas Bickel (1982) extended this to linear regression and other models. This

We thank Doug Hodgson, Joel Horowitz, Peter Phillips, Roger Koenker, Tom Rothenberg and two referees for helpful comments. Financial support from the National Science Foundation is gratefully acknowledged. Gauss programs are available from the second author upon request. Address correspondence to: Oliver Linton, Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, UK; e-mail: O.Linton@lse.ac.uk; Zhijie Xiao, Department of Economics, University of Illinois at Urbana-Champaign, 1206 South Sixth Street, Champaign, IL 61820, USA; e-mail: zxiao@uiuc.edu.

  0266-4666/01 $9.50

latter work provided a starting point for much future work in this area. Manski (1984) studied adaptive estimation in nonlinear 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 heteroskedasticity (ARCH). Jeganathan (1995) extended the theory to nonstationary models with i.i.d. error, and Hodgson (1998) further studied this case but with ARMA errors.

Much of this literature has been devoted to first-order theoretical results and has used devices from mathematical statistics, such as sample splitting and discretization, that do not appeal to practitioners. As we argued elsewhere (Linton, 1995), the first-order asymptotics by no means always provide a good approximation to the sampling behavior of the semiparametric estimators; for confirmation of this see the simulation evidence in Hsieh and Manski (1987). Furthermore, computing the semiparametric estimates requires the selection of a smoothing parameter h, called the bandwidth, that determines the effective degree of parameterization taken by the nuisance function for given sample size n. Although the first-order approximation does not reflect the choice of $h(n)$, the finite sample performance of the estimators depends greatly on the choice of bandwidth.

We shall use higher order expansions as a means to solve some of the problems presented by the first-order theory. Higher order expansions have a long history of application in econometrics (see, among others, Sargan, 1976; Phillips, 1978; Rothenberg, 1984). Applications of higher order approximations to bandwidth choice in semiparametric models have been studied by Härdle, Hart, Marron, and Tsybakov (1992), Linton (1995, 1996, 1998), Linton and Xiao (1997), Nishiyama and Robinson (1997), Powell and Stoker (1996), and Xiao and Phillips (1996) among others. In this paper, we derive higher order expansions for an adaptive estimator in linear regression. We do not require the error to be symmetrically distributed. In fact, we show how choices of smoothing parameters influence the semiparametric adaptive estimator by deriving the asymptotic distribution of the second-order effect. This distribution reflects the bandwidth and kernel used and suggests a method of bandwidth choice. We develop rule-of-thumb plug-in bandwidth selection methods for the estimation problem that are convenient to implement and reasonably insensitive to the true underlying density. We also extend the analysis to the t-ratio and to the case of regressors that are not strictly exogenous. The adaptive estimator is quite promising relative to other semiparametric procedures because the nonparametric estimation only involves one dimensional smoothing and so does not suffer from the curse of dimensionality. In this case, the kernel procedures we employ can work well provided they are implemented appropriately. The main purpose of our asymptotic approximations is to show how the semiparametric adaptive estimator is affected by the smoothing parameters to a higher order and to provide the tools to effect good implementation. Throughout we allow the error density to be zero at the boundary, which is required to make the situation "regular." This necessitates the use of a trimming function. We use the smooth trimming adopted in Andrews (1995) and Ai (1997).

The paper is organized as follows. The model and estimators are described in the next section. Results of the expansion are given in Section 3, and the details of these expansions can be found in the Appendix. In Section 4 we give some extensions to dependent regressors and t-statistics. Bandwidth selection is discussed in Section 5. In Section 6 we provide a small Monte Carlo experiment that evaluates the effectiveness of the second-order approximation. Section 7 concludes.

For notation, we use $f^{(j)}$ to denote the jth derivative of a function f and for a function g of functions $a_1,\dots,a_d$, define the linear differential operator

$$\mathcal{D}_q g(a_1,\dots,a_d)(x) = \sum_{j=1}^{d} \frac{\partial g}{\partial a_j}(a_1,\dots,a_d)(x)\cdot a_j^{(q)}(x).$$

We also let $\|A\|$ denote the Euclidean norm of the array $A=(a_{i_1,\dots,i_s})$ defined as $\|A\| = (\sum a_{i_1,\dots,i_s}^2)^{1/2}$.

2. THE MODEL AND ESTIMATOR

We consider the problem of estimating $\beta \in \mathbb{R}^p$ in the following regression model:

$$y_i = \beta^T x_i + \varepsilon_i, \qquad i=1,\dots,n, \tag{1}$$

where x_i and ε_i satisfy the following assumptions.

A1. ε_i and x_i are independent and identically distributed (i.i.d.) random variables and are mutually independent. Furthermore, $E(x_i)=0$, $\Omega_x = E(x_i x_i^T)$ is positive definite, and for some $\eta>0$ we have $E[\|x\|^{4+\eta}]<\infty$.

A2. ε_i has 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 $(\underline{a},\bar{a})$.

A3. The density function $f(\cdot)$ 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

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

Because we do not impose any additional restrictions on the density function of ε, we cannot separately identify an intercept. Therefore, we shall absorb the intercept into the error density (which can have arbitrary mean) and assume for convenience that the regressors are mean zero in A1. Our other assumptions on the covariates are very weak.

In A2 we assume that $f(\varepsilon)$ has bounded support. Even though the second-order analysis on adaptive regression estimators can be extended to the case with unbounded support, our discussion in this paper is confined to the bounded support case, partly for simplification and partly for some technical reasons. We discuss this point further later (see Remark 5 in Section 3). When f is strictly positive on $[\underline{a}, \bar{a}]$, the situation is nonregular. In some cases, this can lead to inconsistency of solutions of the likelihood score equations but perhaps to the potential for improved rates of convergence for other estimators. Therefore, we shall make an additional assumption.

A4. $f(\varepsilon)$ and its first $\varrho - 1$ derivatives vanish at $\underline{a}$ and $\bar{a}$, whereas $f^{(\varrho)}(\underline{a}) \neq 0$ and $f^{(\varrho)}(\bar{a}) \neq 0$ for some integer ϱ with $2 \leq \varrho \leq r$.

Assumption A4 guarantees that the density f vanishes at the boundary at a sufficiently fast rate so that the properties of regular estimation hold. In this case, one cannot estimate β 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

$$I(f) = \int \ell'(\varepsilon)^2 f(\varepsilon)\, d\varepsilon,$$

where $\ell(\varepsilon) = \log f(\varepsilon)$, exists as do various other integrals used subsequently.

In the sequel we shall let β_0 be the true parameter value. If the density f were known, the MLE of β_0, denoted $\bar{\beta}$, could be obtained by setting the following average score function,

$$\begin{aligned} s(\beta) = s(\beta; f) &= \frac{1}{n}\sum_{i=1}^{n} s_i(\beta; f) \\ &= -\frac{1}{n}\sum_{i=1}^{n} \frac{f'(\varepsilon_i(\beta))}{f(\varepsilon_i(\beta))}\, x_i, \end{aligned} \tag{2}$$

equal to zero, assuming an interior solution of course. Here, for any parameter value β, $\varepsilon_i(\beta) = y_i - \beta^T x_i$. This method works well in regular situations but can lead to inconsistent estimates in some cases of interest to us (for a discussion of this issue, see Bickel, 1975). An alternative method is given by taking one Newton–Raphson step from a preliminary root-n consistent estimator $\tilde{\beta}$. That is, let

$$\bar{\beta}_{NR} = \tilde{\beta} + \tilde{\mathcal{I}}(\tilde{\beta}; f)^{-1} s(\tilde{\beta}; f),$$

where $\tilde{\mathcal{I}}$ is a consistent estimate of the information matrix $\mathcal{I} = \Omega_x I(f)$. For example,

$$\tilde{\mathcal{I}}(\tilde{\beta}; f) = -\frac{1}{n}\sum_{i=1}^{n} x_i x_i^T \left[\frac{f''(\varepsilon_i(\tilde{\beta}))}{f(\varepsilon_i(\tilde{\beta}))} - \frac{f'(\varepsilon_i(\tilde{\beta}))^2}{f(\varepsilon_i(\tilde{\beta}))^2} \right].$$

This method has been investigated in Rothenberg and Leenders (1964) and Bickel (1975). It is first-order equivalent to the MLE when the MLE is consistent, and it has the added advantage of working in certain nonregular cases where the MLE is inconsistent. In econometrics it is common to refer to the estimator as linearized maximum likelihood or two-step, whereas the statistical literature uses one-step. In the regression case we study, there are many preliminary root-n consistent estimators: e.g., the ordinary least squares estimator.

When f is unknown, we have to replace it by a nonparametric estimate $\tilde{f}$, say, and we thereby obtain the estimated average score function

$$\tilde{s}(\beta) = s(\beta;\tilde{f}) = \frac{1}{n}\sum_{i=1}^{n}\tilde{s}_i(\beta;\tilde{f}) = -\frac{1}{n}\sum_{i=1}^{n}\frac{\tilde{f}'(\varepsilon_i(\beta))}{\tilde{f}(\varepsilon_i(\beta))}x_i. \tag{3}$$

The semiparametric profile likelihood estimator $\hat{\beta}_{PL}$ sets $\tilde{s}(\beta)$ equal to zero. Similar to the case where f is known, a one-step Newton–Raphson estimator of β can be obtained from a preliminary root-n consistent estimator $\tilde{\beta}$,

$$\hat{\beta}_{NR} = \tilde{\beta} + \tilde{\mathcal{I}}(\tilde{\beta};\tilde{f})^{-1}s(\tilde{\beta};\tilde{f}), \tag{4}$$

where

$$\tilde{\mathcal{I}}(\tilde{\beta};\tilde{f}) = \frac{1}{n}\sum_{i=1}^{n}x_i x_i^T\left[\frac{\tilde{f}_i''(\varepsilon_i(\tilde{\beta}))}{\tilde{f}_i(\varepsilon_i(\tilde{\beta}))} - \frac{\tilde{f}_i'(\varepsilon_i(\tilde{\beta}))^2}{\tilde{f}_i(\varepsilon_i(\tilde{\beta}))^2}\right].$$

We shall work with this one-step estimator. An important ingredient in our estimator is the error density estimate $\tilde{f}$. We consider the following leave-one-out kernel estimates of $f(t)$ and $f'(t)$ at the point $t = \varepsilon_i(\beta)$ using the residuals $\varepsilon_j(\beta)$ as data:

$$\begin{aligned}\tilde{f}_i(\varepsilon_i(\beta)) &= \frac{1}{(n-1)h_n}\sum_{j\neq i}K\left(\frac{\varepsilon_i(\beta)-\varepsilon_j(\beta)}{h_n}\right)\\ &= \frac{1}{(n-1)}\sum_{j\neq i}K_{h_n}(\varepsilon_i(\beta)-\varepsilon_j(\beta)),\\ \tilde{f}_i'(\varepsilon_i(\beta)) &= \frac{1}{(n-1)h_n^2}\sum_{j\neq i}K'\left(\frac{\varepsilon_i(\beta)-\varepsilon_j(\beta)}{h_n}\right)\\ &= \frac{1}{(n-1)}\sum_{j\neq i}K_{h_n}'(\varepsilon_i(\beta)-\varepsilon_j(\beta)),\end{aligned}$$

where $K_{h_n}(t) = K(t/h_n)/h_n$ and $K_{h_n}'(t) = K'(t/h_n)/h_n^2$. Here, $K(\cdot)$ is the kernel function whose properties are given in Assumption A5, which follows, whereas h_n is the bandwidth parameter. In principle, we may consider more general devices that use different bandwidth parameters in the estimation of f and f'. However, the additional smoothing parameter brings substantial complication to the higher order analysis, and we consider the simple case where the same h_n is used in estimating f and f' (also see the subsequent discussion on trimming). Asymptotic results for the bias and variance in these nonparametric density estimates are given in the Appendix. The estimator can be computed using only matrix computations, which makes it very fast.

As in some other applications of kernel regression estimators, the random denominator $\tilde{f}_i$ can be small and may cause technical difficulty. For this reason, we trim out small $\tilde{f}_i$ as do Bickel (1982) and Manski (1984) (for a more recent discussion, also see Ai, 1997) . However, trimming brings an additional parameter into the estimation and complicates the higher order expansions.

We consider the following smoothed trimming (Andrews, 1995; 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

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

For example, if we consider the following Beta density

$$g(z) = B(k+1)^{-1} z^k (1-z)^k, \qquad 0 \le z \le 1,$$

for some integer k, where $B(k)$ is the beta function defined by

$$B(k) = \Gamma(k)^2/\Gamma(2k), \qquad \Gamma(k) \text{ is the Euler gamma function,}$$

then it can be verified that for $b \le x \le 2b$

$$G_b(x) = B(k+1)^{-1}\left\{\frac{(k!)^2}{(2k+1)!} - \sum_{l=0}^{k} \frac{(k!)^2}{(k-l)!(k+l+1)!}\left(\frac{x-b}{b}\right)^{k-l} \times \left(1 - \frac{x-b}{b}\right)^{k+l+1}\right\}, \tag{5}$$

which is a $(2k+1)$th order polynomial in $(x-b)/b$. The function $G_b(x)$ is continuously differentiable on $[0,1]$. This property is important because it allows us to use standard Taylor series arguments, whereas indicator function trimming would preclude this. We now estimate the average score function (3) by

$$\tilde{s}_n(\tilde{\beta}) = -\frac{1}{n}\sum_{i=1}^{n}\frac{\tilde{f}_i'}{\tilde{f}_i}x_i G_b(\tilde{f}_i), \tag{6}$$

and the information matrix by

$$\tilde{\mathcal{I}}_n(\tilde{\beta};\tilde{f}) = -\frac{1}{n}\sum_{i=1}^{n}x_i x_i^T\left[\frac{\tilde{f}_i''}{\tilde{f}_i} - \left(\frac{\tilde{f}_i'}{\tilde{f}_i}\right)^2\right]G_b(\tilde{f}_i),$$

where $\tilde{f}_i = \tilde{f}_i(\varepsilon_i(\tilde{\beta}))$ and $\tilde{f}_i' = \tilde{f}_i'(\varepsilon_i(\tilde{\beta}))$. Thus we estimate β by the following one-step Newton–Raphson estimator:

$$\hat{\beta} = \tilde{\beta} + \tilde{\mathcal{I}}_n(\tilde{\beta};\tilde{f})^{-1}\tilde{s}_n(\tilde{\beta};\tilde{f}). \tag{7}$$

We study the higher order property of the adaptive estimator $\hat{\beta}$ given in (7). We make the following assumptions on the kernel function $K(\cdot)$ and the trimming parameter b.

A5. The kernel K has support $[-1,1]$ and is symmetric about zero and satisfies $\int K(u)du = 1$. It is twice differentiable on its support and K'' is Lipschitz continuous, whereas $K'(0) = 0$. Furthermore, there exists an even positive integer q with $2 < q \leq r-3$ such that

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

A6. The trimming function $G_b(x)$ is $(L+1)$th order differentiable for some $L > 4$. In addition, $h \to 0$ and $nh^5/\log n \to \infty$, $b \to 0$, and $h/b \to 0$ as $n \to \infty$.

These assumptions are similar to those used in the existing literature. Note that because b is of larger magnitude than h, our estimator will not suffer from boundary bias (for a discussion of boundary issues, see Müller 1988, pp. 32–36).[1]

Define for any function K and integer q

$$\mu_q(K) = \frac{(-1)^q}{q!}\int u^q K(u)du; \qquad \|K\|_2 = \left\{\int |K(u)|^2 du\right\}^{1/2}.$$

These notations will be used in the following sections for higher order asymptotics.

3. THE EXPANSION

Making a Taylor series expansion of $\tilde{s}_n(\tilde{\beta})$ about $\tilde{s}_n(\beta_0)$ and collecting terms, we obtain

$$\sqrt{n}(\hat{\beta} - \beta_0) = \tilde{\mathcal{I}}_n(\beta_0)^{-1}\sqrt{n}\tilde{s}_n(\beta_0) + \{\tilde{\mathcal{I}}_n(\tilde{\beta})^{-1} - \tilde{\mathcal{I}}_n(\beta_0)^{-1}\}\sqrt{n}\tilde{s}_n(\beta_0) + \{I + \tilde{\mathcal{I}}_n(\tilde{\beta})^{-1}\tilde{s}_n'(\beta^*)\}\sqrt{n}(\tilde{\beta} - \beta_0), \tag{8}$$

where β^* is an intermediate point, $\tilde{s}_n(\beta) = \tilde{s}_n(\beta, \tilde{f})$, and $\tilde{\mathcal{I}}_n(\beta) = \tilde{\mathcal{I}}_n(\beta, \tilde{f})$, whereas $\tilde{s}_n'(\beta) = \partial\tilde{s}_n(\beta)/\partial\beta$. A first-order analysis shows that $\sqrt{n}\tilde{s}_n(\beta_0) = O_p(1)$ and $\tilde{\mathcal{I}}_n(\beta_0)^{-1} = O_p(1)$. Note that in the parametric case, both terms in (8) would be $O_p(n^{-1/2})$. In our case, this is true apart from some "trimming terms," which turn out to be of smaller order than our leading trimming terms (see the discussion that follows). Specifically, we obtain in the Appendix that

$$\sqrt{n}(\hat{\beta} - \beta_0) = \tilde{\mathcal{I}}_n(\beta_0)^{-1}\sqrt{n}\tilde{s}_n(\beta_0) + \mathcal{T}_I + O_p(n^{-1/2}), \tag{9}$$

where $\mathcal{T}_I$ is a small trimming term.

We next derive an approximation to $\tilde{\mathcal{I}}_n(\beta_0)^{-1}\sqrt{n}\tilde{s}_n(\beta_0)$. The random variables $\Delta_H = \tilde{\mathcal{I}}_n(\beta_0) - \mathcal{I}$ and $\Delta_s = \sqrt{n}\{\tilde{s}_n(\beta_0) - s(\beta_0)\}$, which are functions of nonparametric estimates of the residual densities and their derivatives, are now the two key elements in the expansion. Both quantities can be decomposed into the sum of different terms that are functions of the bias and variance effects in estimating the densities f and their derivatives. We show in the Appendix that

$$\Delta_s = \mathcal{T}_s + O_p(h^q) + O_p\left(\frac{1}{\sqrt{nh_n^3}}\right), \tag{10}$$

$$\Delta_H = \mathcal{T}_H + O_p(h^q) + O_p\left(\frac{1}{\sqrt{n^2h_n^5}}\right) + O_p(n^{-1/2}), \tag{11}$$

where $\mathcal{T}_s$ and $\mathcal{T}_H$ are $o_p(1)$ trimming effects (and note that $\mathcal{T}_I$ from (9) is of smaller order than $\mathcal{T}_s$ and $\mathcal{T}_H$). These terms depend on the parameter b and on the boundary behavior of the densities f, g. By a geometric series expansion of $\tilde{\mathcal{I}}_n(\beta_0)^{-1}$ about $\mathcal{I}^{-1}$ we obtain

$$\tilde{\mathcal{I}}_n(\beta_0)^{-1} = \mathcal{I}_n^{-1} - \mathcal{I}_n^{-1}\Delta_H\mathcal{I}_n^{-1} + o_p(\delta_n) = \mathcal{I}^{-1} - \mathcal{I}^{-1}\Delta_H\mathcal{I}^{-1} + o_p(\delta_n),$$

where $\delta_n = \max\{h_n^q, 1\big/\sqrt{nh_n^3}\}$ is larger than $n^{-1/2}$. Here, $\mathcal{I}_n = -s'(\beta_0)$, and by the central limit theorem for independent random variables we obtain that $\mathcal{I}_n = \mathcal{I} + O_p(n^{-1/2})$. This yields that

$$\sqrt{n}(\hat{\beta} - \beta_0) = \mathcal{I}^{-1}\sqrt{n}s(\beta_0) + \mathcal{I}^{-1}\Delta_H\mathcal{I}^{-1}\sqrt{n}s(\beta_0) + \mathcal{I}^{-1}\Delta_s + o_p(\delta_n). \tag{12}$$

We then obtain the following stochastic expansion of the standardized estimator:

$$\sqrt{n}(\hat{\beta} - \beta_0) = X_0 + \mathcal{T} + h_n^q\mathcal{B} + \frac{1}{\sqrt{nh_n^3}}\,\mathcal{V} + o_p(\delta_n), \tag{13}$$

where X_0, $\mathcal{B}$, and $\mathcal{V}$ are zero mean $O_p(1)$ quantities and $\mathcal{T}$ is an $o_p(1)$ quantity. These quantities are defined in the Appendix. Here, $X_0 = \mathcal{I}^{-1}\sqrt{n}s(\beta_0)$ is the leading term, $\mathcal{T}$ is the trimming effect, $\mathcal{B}$ reflects the bias effect in nonparametric density estimation, and $\mathcal{V}$ reflects the variance effect in the nonparametric estimation. The random variables X_0, $\mathcal{T}$, and $\mathcal{B}$ are all mean zero and sums of i.i.d. random variables, whereas $\mathcal{V}$ is a degenerate U-statistic. Note that $\sqrt{n}(\bar{\beta} - \beta_0) = X_0 + O_p(n^{-1/2})$, where $\bar{\beta}$ is the infeasible estimator of β, so that

$$\sqrt{n}(\hat{\beta} - \bar{\beta}) = \mathcal{T} + h_n^q \mathcal{B} + \frac{\mathcal{V}}{\sqrt{nh_n^3}} + o_p(\delta_n).$$

The trimming effect $\mathcal{T}$ is an $o_p(1)$ quantity whose magnitude is determined jointly by the trimming parameter b and the rate that the density f approaches zero on the boundary, but does not to first order depend on the bandwidth parameter h_n in the nonparametric density estimation. We are now ready to state the main result of the paper.

THEOREM 1. *Suppose that Assumptions A1–A6 hold and denote* $\tau = \mathcal{T} + h_n^q \mathcal{B} + \mathcal{V}/\sqrt{nh_n^3}$. *We have the following results.*

(1a) If $nh^{2q+3} \to \infty$, *then*

$$h^{-q}(\tau - \mathcal{T}) \Rightarrow N(0, \Sigma_1),$$

where

$$\Sigma_1 = \mu_q^2(K)[\mathcal{I}^{-1}\mathcal{M}_1\mathcal{I}^{-1} + \mathcal{I}^{-1}\mathcal{M}_3\mathcal{I}^{-1}\mathcal{M}_3\mathcal{I}^{-1} + 2\mathcal{I}^{-1}\mathcal{M}_3\mathcal{I}^{-1}\mathcal{M}_2\mathcal{I}^{-1}]$$

with $\mathcal{I} = \Omega_x I(f)$ *and* $\mathcal{M}_1 = \Omega_x \operatorname{var}[\mathcal{D}_q \ell^{(1)}(\varepsilon)]$, $\mathcal{M}_2 = \Omega_x \operatorname{cov}[\mathcal{D}_q \ell^{(1)}(\varepsilon), \ell^{(1)}(\varepsilon)]$, $\mathcal{M}_3 = -\Omega_x E[\mathcal{D}_q \ell^{(2)}(\varepsilon)]$.

(1b) If $nh^{2q+3} \to 0$, *then*

$$n^{1/2}h^{3/2}(\tau - \mathcal{T}) \Rightarrow N(0, \Sigma_2),$$

where $\Sigma_2 = \|K'\|_2^2 \mathcal{I}^{-1}\mathcal{S}_1\mathcal{I}^{-1}$ *with* $\mathcal{S}_1 = \Omega_x(\bar{a} - \underline{a})$.

(1c) If $h_n = \gamma n^{-1/(2q+3)}$ *for some* γ *with* $0 < \gamma < \infty$,

$$n^{q/(2q+3)}(\tau - \mathcal{T}) \Rightarrow N(0, \Sigma),$$

where $\Sigma = \gamma^{2q}\Sigma_1 + \gamma^{-3}\Sigma_2$.

(2) Finally,

$$b^{-(\varrho-1)/2\varrho}\mathcal{T} \Rightarrow N(0, \underline{\Sigma}), \tag{14}$$

where $\underline{\Sigma} = \varphi(\varrho, \underline{a}, \bar{a}, f)\Omega_x$ *and* $\varphi(\varrho, \underline{a}, \bar{a}, f)$ *depends on both the trimming function and the boundary behavior of the density. In particular, if we use the trimming function (5),*

$$\varphi(\varrho, \underline{a}, \bar{a}, f) = c(\varrho)[f^{(\varrho)}(\underline{a})^{1/\varrho} + f^{(\varrho)}(\bar{a})^{1/\varrho}],$$

where $c(\varrho)$ *is a coefficient depending on* ϱ *and* G.

Remarks.

1. To minimize the "smoothing effect," we should set $h_n = \gamma n^{-1/(2q+3)}$ so that the second-order bias and variance effects are balanced, i.e., the "in probability" magnitude of $\tau - \mathcal{T}$ is minimized. Note that

$$\mathcal{D}_q \ell^{(1)} = \frac{f^{(q+1)}}{f} - \frac{f^{(q)} f^{(1)}}{f^2},$$

$$\mathcal{D}_q \ell^{(2)} = \frac{f^{(q+2)}}{f} - \frac{f^{(q)} f^{(2)}}{f^2} - 2\frac{f^{(1)}}{f}\left[\frac{f^{(q+1)}}{f} - \frac{f^{(q)} f^{(1)}}{f^2}\right].$$

   The terms $\mathcal{M}_j$, $j = 1,2,3$, arise from the bias of the nonparametric estimates $\tilde{f}$ and $\tilde{f}'$, whereas the term $\mathcal{S}_1$ comes from the variance of $\tilde{f}'$. Both terms are positive, and the overall effect is to increase variance above the first-order limiting variance of $\hat{\beta}$.
2. The magnitude of the variance of the trimming effect is $O(b^{(\varrho-1)/\varrho})$, which increases with b and which is of larger order than h_n^{2q} under our assumptions. The limiting variance of the trimming effect is given by (14); this depends on both the trimming function and the boundary behavior of the density function in a complicated manner. Nevertheless, the limiting variance of the trimming effect can be consistently estimated without knowledge of the parameter ϱ; specifically, $b^{(\varrho-1)/\varrho}\varphi$ can be estimated by

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\left[\frac{\tilde{f}'(\tilde{\varepsilon}_i)}{\tilde{f}(\tilde{\varepsilon}_i)}\right]^2 [1 - G_b(\tilde{f}_i)]^2\right\}. \quad \textbf{(15)}$$

3. When f is strictly positive on $[\underline{a}, \bar{a}]$, the situation is nonregular, and there is the potential for improved rate of convergence by other estimators. In this case, the two-step estimator $\hat{\beta}$ may not necessarily have adaptive properties, at least when $\underline{a} > -\infty$ or $\bar{a} < \infty$, because it has too slow a rate of convergence; specifically, when f is known it is possible to obtain estimates with faster rate of convergence. However, $\hat{\beta}$ is consistent and asymptotically normal under our conditions in this case. Of course, trimming is no longer needed, and the untrimmed estimator then has the stochastic expansion $X_0 + h_n^q \mathcal{B} + \mathcal{V}/\sqrt{nh_n^3} + o_p(\delta_n)$.
4. In the regular case, the two-step estimator $\hat{\beta}$ has the exact same second-order effect as the profile likelihood estimator (for a similar result, see Linton, 1998).
5. Finally we consider what happens when the error support is unbounded. As indicated in the analysis in the Appendix, in this case, the second-order variance effect involves terms such as $n^{-1}\sum_{i=1}^{n} 1/f(\varepsilon_i)$ (which is related to the $\mathcal{S}_1$ term defined previously) that do not satisfy a law of large numbers if f has unbounded support like the Gaussian distribution. In fact, this random sequence grows to infinity in probability at a rate determined by the tails of the distribution. In the Gaussian case, the rate is logarithmic. Thus the order in probability of the second-order terms will be larger and will depend on the tails of the distribution. Also, whether a central limit theorem for these terms operates remains to be seen.

4. EXTENSIONS

4.1. *t*-Statistic

In this section we derive the second order expansions for t-ratio statistics. Consider the linear hypothesis $H_0: c^T\beta = c_0$, where c is a $p \times 1$ vector of constants and c_0 is a scalar. The corresponding t-statistic is

$$\hat{t} = \frac{c^T\hat{\beta} - c_0}{\widehat{\mathrm{se}}(c^T\hat{\beta})} = \frac{c^T\hat{\beta} - c_0}{\sqrt{n^{-1}c^T\tilde{\mathcal{I}}_n(\tilde{\beta})^{-1}c}}.$$

Under the null hypothesis that $c^T\beta = c_0$, $\hat{t}$ is asymptotically standard normal and first-order equivalent to the corresponding (infeasible) MLE based t-ratio

$$\bar{t} = \frac{c^T\sqrt{n}(\bar{\beta} - \beta)}{\sqrt{c^T\bar{\mathcal{I}}^{-1}c}},$$

where $\bar{\mathcal{I}} = -s'(\bar{\beta})$. Under our conditions,

$$[c^T\tilde{\mathcal{I}}_n(\tilde{\beta})^{-1}c]^{-1/2} = [c^T\mathcal{I}^{-1}c]^{-1/2} - \frac{1}{2}[c^T\mathcal{I}^{-1}c]^{-3/2}c^T\mathcal{I}^{-1}\Delta_H\mathcal{I}^{-1}c + \text{higher order terms}, \quad \mathbf{(16)}$$

where Δ_H is defined by (11). Denote the second-order effect as $\tau_t = \hat{t} - \bar{t}$ and the $o_p(1)$ trimming effect as t_r. We have the following result.

THEOREM 2. *Suppose that Assumptions A1–A6 hold and that* $h_n = \gamma n^{-1/(2q+3)}$ *for some* γ *with* $0 < \gamma < \infty$. *Then, under* H_0,

$$n^{q/(2q+3)}(\tau_t - t_r) \Rightarrow N(0, \sigma_t^2),$$

where

$$\sigma_t^2 = \frac{c^T\Sigma c}{c^T\mathcal{I}^{-1}c} - \gamma^{2q}\mu_q^2(K) \times \left\{\frac{c^T\mathcal{I}^{-1}\mathcal{M}_2\mathcal{I}^{-1}cc^T\mathcal{I}^{-1}\mathcal{M}_3\mathcal{I}^{-1}c}{(c^T\mathcal{I}^{-1}c)^2} + \frac{3}{4}\left(\frac{c^T\mathcal{I}^{-1}\mathcal{M}_3\mathcal{I}^{-1}c}{c^T\mathcal{I}^{-1}c}\right)^2\right\}.$$

The rate of convergence for $\tau_t - t_r$ is the same as for $\tau - \mathcal{T}$, but the asymptotic variance is slightly different, reflecting the estimation of the asymptotic variance of $\hat{\beta}$. The trimming terms are similar to those in Theorem 1.

4.2. Time Series Regressors

In this section, we extend our second-order analysis to more general models where the regressors contain lagged disturbances and thus are serially correlated. In particular, we consider the case where the regressor $x_i = (x_{i1}, \ldots, x_{ip})^T$ satisfies Assumption A1′.

A1′. The stochastic process $\{x_i\}$ satisfies

$$x_i = x_i^* + \sum_{k=1}^{\infty} \Psi_k \varepsilon_{i-k}, \quad \text{where } \Psi_k = (\psi_{k1}, \dots, \psi_{kp})^T, \tag{17}$$

where x_i^*, ε_i are i.i.d. random variables and are mutually independent. Furthermore, there exists a ρ with $0 < \rho < 1$ such that $|\psi_{kj}| < \rho^k$ for all j, k. We require also that $E(x_i) = 0$, that $\Omega_x = E(x_i x_i^T)$ is positive definite, and that for some $\eta > 0$, we have $E[\|x\|^{4+\eta}] < \infty$.

This setting is general enough to include leading cases in time series models such as, say, stationary ARMA time series regression models. For example, we consider the case of a first-order univariate autoregressive regression described as follows:

$$y_t = \beta y_{t-1} + \varepsilon_t, \qquad t = 0, 1, \dots, n, \tag{18}$$

where $|\beta| < 1$ and $\{\varepsilon_i\}$ are i.i.d. random variables with mean zero and finite variance σ_ε^2 and satisfy Assumptions A1–A3 in Section 2. Then regression (18) corresponds to the special case in models (1) and (17) with $x_i^* = 0$, and $\Psi_k = \beta^{k-1}$.

The semiparametric adaptive estimator $\hat{\beta}$ is still consistent and asymptotically normal for this specification of the covariate process. A similar expansion for $\hat{\beta}$ can be performed. The following theorem summarizes the higher order effects.

THEOREM 3. *Suppose that Assumptions A1′ and A2–A6 hold and that $h_n = \gamma n^{-1/(2q+3)}$ for some γ with $0 < \gamma < \infty$. Then,*

$$n^{q/(2q+3)}(\tau - \mathcal{T}) \Rightarrow N(0, \Sigma^*),$$

where

$$\begin{aligned}\Sigma^* &= \gamma^{2q} \mu_q^2(K) \\ &\quad \times [\mathcal{I}^{-1}(\mathcal{M}_1 + \Gamma)\mathcal{I}^{-1} + \mathcal{I}^{-1}\mathcal{M}_3\mathcal{I}^{-1}\mathcal{M}_3\mathcal{I}^{-1} + 2\mathcal{I}^{-1}\mathcal{M}_3\mathcal{I}^{-1}\mathcal{M}_2\mathcal{I}^{-1}] \\ &\quad + \gamma^{-3}\|K'\|_2^2 \mathcal{I}^{-1}\mathcal{S}_1\mathcal{I}^{-1},\end{aligned}$$

where $\mathcal{I}$, $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$ are defined as in Theorem 1 with $\Omega_x = E(x_i^* x_i^{*T}) + \sigma_\varepsilon^2(\sum_{k=1}^\infty \Psi_k \Psi_k^T)$ and

$$\Gamma = \sigma_\varepsilon^2 \left(\sum_{i=0}^{\infty} \sum_{j=1}^{\infty} \Psi_i \Psi_{i+j}^T \right) E^2[\mathcal{D}_q \ell^{(1)}].$$

Remarks.

1. The second-order effects are similar in Theorem 3 to those in Theorem 1, but in model (17) the serial correlation in the regressors brings additional terms into the second order effect; these additional terms are summarized in Γ. They arise from

autocorrelations within some of the "bias related" terms. When ε is symmetric about zero, $\Gamma = 0$, because $\mathcal{D}_q \ell^{(1)}$ is an odd function for q an even integer.

2. In the special case that $x_i^* = 0$ and $\Psi_k = \beta^{k-1}$, $|\beta| < 1$, we reduce to the case of autoregression of order one, i.e., $x_t = y_{t-1}$, or,

$$y_t = \beta y_{t-1} + \varepsilon_t, \qquad t = 0,1,\ldots,n,$$

and $\mathcal{I}$, $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$ are defined as in Theorem 1 with $\Omega_x = \sigma_\varepsilon^2/(1-\beta^2)$, whereas

$$\Gamma = \sigma_\varepsilon^2 \frac{2\beta}{1-\beta} E^2[\mathcal{D}_q \ell^{(1)}(\varepsilon)].$$

5. BANDWIDTH SELECTION

The results in the previous sections can be used to select bandwidth parameter h_n for the semiparametric estimator $\hat{\beta}$ and t-ratio. Here, we just consider the estimator in the i.i.d. setting, although similar comments apply to the test statistic and to the dependent data design. The higher order effects generally depend on the bandwidth parameters and the trimming procedure. However, although in principle joint optimization over the trimming and bandwidth parameters may be considered, the analysis would be substantially more complicated, not least because there is only a lower bound on b. In this paper, we confine our attention to the effect of bandwidth and keep the choice of trimming parameter fixed. Our analysis is not the best over all possibilities; however, it provides a second best choice, and our analysis shows how the estimator is affected by these parameters. We shall try to minimize the second-order term $\tau - \mathcal{T} = h_n^q \mathcal{B} + n^{-1/2} h_n^{-3/2} \mathcal{V}$, which is mean zero and has asymptotic variance

$$\begin{aligned}\Sigma(h_n) &= h_n^{2q} \mu_q^2(K) \\ &\quad \times [\mathcal{I}^{-1}\mathcal{M}_1\mathcal{I}^{-1} + \mathcal{I}^{-1}\mathcal{M}_3\mathcal{I}^{-1}\mathcal{M}_3\mathcal{I}^{-1} + 2\mathcal{I}^{-1}\mathcal{M}_3\mathcal{I}^{-1}\mathcal{M}_2\mathcal{I}^{-1}] \\ &\quad + \frac{1}{nh_n^3} \|K'\|_2^2 \mathcal{I}^{-1}\mathcal{S}_1\mathcal{I}^{-1} \\ &\equiv h_n^{2q}\mathcal{Q}_1 + \frac{1}{nh_n^3} \mathcal{Q}_2.\end{aligned}$$

Specifically, we define an optimal bandwidth as one that minimizes some scalar-valued convex loss function defined on the second-order mean square error matrix $\Sigma(h_n)$. If the loss function is denoted as $l(\Sigma)$, then, by Taylor expansion, we obtain the following optimal bandwidth formula:

$$h_n^{opt} = \left[\frac{3\underline{l}^T \text{vec}\{\mathcal{Q}_2\}}{2q\underline{l}^T \text{vec}\{\mathcal{Q}_1\}} \right]^{1/(2q+3)} n^{-1/(2q+3)}, \tag{19}$$

where $\underline{l} = \partial l(0)/\partial \mathrm{vec}\Sigma$. Replacing the unknown quantities $\underline{l}$, $\mathcal{Q}_1$, and $\mathcal{Q}_2$ in the bandwidth formula by their estimates $\hat{\underline{l}}$, $\hat{\mathcal{Q}}_1$, $\hat{\mathcal{Q}}_2$, we obtain a feasible optimal bandwidth choice.

One way of estimating the optimal bandwidth parameter is the plug-in method. We consider the following rule-of-thumb method for bandwidth selection as in Silverman (1986) and Andrews (1991). We specify a parametric model for the error structure $\{f_p(\cdot;\theta), \theta \in \Theta\}$, and estimators of these parameters, denoted $\hat{\theta}$, are used to obtain preliminary estimates of the density functions $f_p(\cdot;\hat{\theta})$ and their derivatives $f_p^{(j)}(\cdot;\hat{\theta})$. These preliminary estimates are then plugged into the formulae of $\mathcal{I}$, $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$, and $\mathcal{S}_1$ to get estimates of them. Let

$$\hat{\mathcal{I}} = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T \left[\frac{f_p'(\hat{\varepsilon}_i;\hat{\theta})}{f_p(\hat{\varepsilon}_i;\hat{\theta})}\right]^2,$$

$$\hat{\mathcal{M}}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T \left\{\left[\frac{f_p^{(q+1)}(\hat{\varepsilon}_i;\hat{\theta})}{f_p(\hat{\varepsilon}_i;\hat{\theta})}\right]^2 + \left[\frac{f_p'(\hat{\varepsilon}_i;\hat{\theta}) f_p^{(q)}(\hat{\varepsilon}_i;\hat{\theta})}{f_p(\hat{\varepsilon}_i;\hat{\theta})^2}\right]^2\right\}$$
$$- \frac{2}{n}\sum_{i=1}^{n} x_i x_i^T \frac{f_p'(\hat{\varepsilon}_i;\hat{\theta}) f_p^{(q)}(\hat{\varepsilon}_i;\hat{\theta}) f_p^{(q+1)}(\hat{\varepsilon}_i;\hat{\theta})}{f_p(\hat{\varepsilon}_i;\hat{\theta})^3},$$

$$\hat{\mathcal{M}}_2 = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T \left\{\left[\frac{f_p'(\hat{\varepsilon}_i;\hat{\theta}) f_p^{(q+1)}(\hat{\varepsilon}_i;\hat{\theta})}{f_p(\hat{\varepsilon}_i;\hat{\theta})^2}\right] - \left[\frac{f_p'(\hat{\varepsilon}_i;\hat{\theta})^2 f_p^{(q)}(\hat{\varepsilon}_i;\hat{\theta})}{f_p(\hat{\varepsilon}_i;\hat{\theta})^3}\right]\right\},$$

$$\hat{\mathcal{M}}_3 = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T \left\{\left[\frac{f_p^{(q+2)}(\hat{\varepsilon}_i;\hat{\theta})}{f_p(\hat{\varepsilon}_i;\hat{\theta})}\right] - \left[\frac{2 f_p'(\hat{\varepsilon}_i;\hat{\theta}) f_p^{(q+1)}(\hat{\varepsilon}_i;\hat{\theta})}{f_p(\hat{\varepsilon}_i;\hat{\theta})^2}\right]\right\}$$
$$+ \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T \left\{\left[\frac{2 f_p'(\hat{\varepsilon}_i;\hat{\theta})^2 f_p^{(q)}(\hat{\varepsilon}_i;\hat{\theta})^2}{f_p(\hat{\varepsilon}_i;\hat{\theta})^3}\right] - \left[\frac{f_p''(\hat{\varepsilon}_i;\hat{\theta}) f_p^{(q)}(\hat{\varepsilon}_i;\hat{\theta})}{f_p(\hat{\varepsilon}_i;\hat{\theta})^2}\right]\right\},$$

$$\hat{\mathcal{S}}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T \frac{1}{f_p(\hat{\varepsilon}_i;\hat{\theta})},$$

where $\hat{\varepsilon}_i = y_i - \hat{\beta}^T x_i$. Then let $\mathcal{Q}_1$ and $\mathcal{Q}_2$ be estimated by plugging $\hat{\mathcal{I}}$, $\hat{\mathcal{M}}_j$, and $\hat{\mathcal{S}}_1$ into the corresponding formulae. When the parametric specification is correct, these estimates are consistent. More generally, they will be not far from the truth. Plugging them into formula (19), we get an estimate of the optimal bandwidth. Under further conditions, this data-based method is second-order efficient in the sense that the corresponding effect $\tau_{\hat{h}} - \mathcal{T}$ has the same asymptotic distribution as $\tau_{h_{opt}} - \mathcal{T}$. See Linton (1998) for a similar result.

We now discuss further the choice of trimming parameter. Suppose we take $b = h_n^{1-\eta}$ for some $\eta > 0$. Then under Assumptions A1–A6 with $q > 2$, we obtain the following expansion:

$$\sqrt{n}(\hat{\beta} - \beta_0) = X_0 + h_n^{(1-\eta)(\varrho-1)/2\varrho}\mathcal{T}_0 + h_n^q \mathcal{B} + \frac{1}{\sqrt{nh_n^3}}\mathcal{V} + o_p(\delta_n),$$

where $\mathcal{T}_0$, $\mathcal{B}$, and $\mathcal{V}$ are all $O_p(1)$. The bias related term $h_n^q\mathcal{B}$ is of smaller order than the trimming term, and the optimal choice of h_n will now trade off $h_n^{(1-\eta)(\varrho-1)/2\varrho}\mathcal{T}_0$ against $\mathcal{V}/\sqrt{nh_n^3}$. The optimal choice of h_n would be

$$h^* = \gamma(\varrho, \underline{a}, \bar{a}, f)n^{-\varrho/(4\varrho-(\varrho-1)\eta-1)},$$

whose rate depends on η and ϱ, the first of which is arbitrary and the second of which is unknown. Note also that this bandwidth may not satisfy the restrictions in A6 for some values of ϱ. Therefore, this method is not appealing.

If the trimming term is of some concern, one can estimate it using (15) and correct standard errors accordingly.

6. MONTE CARLO RESULTS

We conducted a Monte Carlo experiment to evaluate the second-order theory of the semiparametric adaptive regression estimators. We show by simulation how the semiparametric estimators are affected by the choices of smoothing parameters in finite sample. We evaluate the effect of a bandwidth selection criterion that minimizes the second-order mean squared error and the sampling performance of estimators that use different bandwidth choices.

The model used for data generation was the following:

$$y_i = \beta x_i + \varepsilon_i, \tag{20}$$

with $\beta = 1$ and x_i i.i.d. standard normal variates. Two different specifications of ε_i were considered. In the first case, ε_i are i.i.d. t-distributions with degree of freedom 5 and truncated at ± 10. The second case considers the centered i.i.d. *Beta*(4,4) variates whose probability density vanishes on the boundary and is thus consistent with the requirements for regular estimation. See Devroye (1995) for a discussion on generating Beta random variables. These two specifications of the residuals are denoted DGP(1) and DGP(2) in our analysis. The second-order effects for these examples are also calculated; readers are referred to an early version of this paper (Linton and Xiao, 1998) for the formulae. Two sample sizes are tried, $n = 100$, 200. In our experiment, x_i and ε_i are independent of each other, and the number of replications is 500 in each case.

The sampling performances of both the ordinary least squares (OLS) estimator and the semiparametric adaptive estimators were examined for each case. For the adaptive estimator, the following kernel function was used in the semiparametric estimation: $K(u) = 15(1-u^2)^2 1(|u| \le 1)/16$. For purpose of comparison, we also considered the MLE, which uses knowledge of the density function. In particular, we calculated the two-step Newton–Raphson estimator from the OLS preliminary estimator.

Because we are especially interested in the effect of smoothing parameters on the finite sample performance of the adaptive estimators, different choices of bandwidth parameters were considered and compared. We examined the properties of the adaptive estimators with optimal bandwidth selection and several fixed bandwidth choices. Different trimming parameter values were used in the Monte Carlo experiment, and the effects of trimming parameter value on the sample performance of these estimators were also examined.

Denoting the jth replication of estimator b as $b(j)$, we calculated in each case the (average) bias ($R^{-1}\sum_{j=1}^{R} b(j) - 1$), the median bias (median of $b - 1$), the variance ($R^{-1}\sum_{j=1}^{R}(b(j) - \bar{b})^2$), the mean squared error ($R^{-1}\sum_{j=1}^{R}(b(j) - 1)^2$), and the interquartile range (IQR = the 75% quantile–the 25% quantile), where R is the number of replications.

Table 1 provides the simulation results for the nonregular case where ε_i are i.i.d. truncated t-distributions whose density is strictly positive on its bounded support. Both $n = 100$ and $n = 200$ are reported. We calculated the OLS estimator, the MLE, and the adaptive estimators using optimal bandwidth (19), which was close to 0.035, and fixed bandwidth values $h = 0.01, 0.03, 0.05, 0.1$, in each case without any trimming. For this case, we can see that the mean squared errors of the MLE, the adaptive estimator with optimal bandwidth, and the OLS estimator are close, although small difference does exist. Substantial difference can be found among adaptive estimators using different bandwidth values. From these results we can see the influence of bandwidth choice on the adaptive estimator.

Table 2 reports the results for DGP(2) where ε_i are i.i.d. $Beta(4,4)$ random variates and $n = 100$. The results for the $n = 200$ case are similar. Besides the

TABLE 1. Simulation results where ε_i are i.i.d. truncated t-distributions

Estimators	Bias	Median Bias	Variance	MSE	IQR
		$n = 100$			
OLS estimator	0.00176	−0.00144	0.0302	0.0303	0.0810
MLE: 2-step from OLS	0.00316	−0.00166	0.0294	0.0294	0.0798
ADAP1: optimal band	0.00320	−0.00178	0.0294	0.0294	0.0801
ADAP2: $h = 0.1$	0.04815	−0.00132	1.2063	1.2086	0.0958
ADAP3: $h = 0.05$	0.00406	−0.00267	0.0612	0.0612	0.1207
ADAP4: $h = 0.03$	−0.00238	−0.00209	0.0437	0.0437	0.0823
ADAP5: $h = 0.01$	−0.01958	−0.00239	0.2617	0.2621	0.0853
		$n = 200$			
OLS estimator	0.00251	−0.00163	0.0135	0.0139	0.0810
MLE: 2-step from OLS	0.00254	−0.00208	0.0131	0.0135	0.0812
ADAP1: optimal band	0.00246	−0.00131	0.0132	0.0136	0.0801
ADAP2: $h = 0.1$	0.00372	−0.00166	0.0297	0.0304	0.0799
ADAP3: $h = 0.05$	0.00255	−0.00239	0.0139	0.0145	0.0852
ADAP4: $h = 0.03$	−0.0027	−0.00178	0.0131	0.0136	0.0801
ADAP5: $h = 0.01$	−0.0021	−0.00207	0.0172	0.0178	0.0958

Table 2. Simulation results where ε_i are i.i.d. *Beta*(4,4) random variates

Estimators	Bias	Median Bias	Variance	MSE	IQR
		$b = 0.005$			
OLS estimator	−0.00251	−0.00214	0.000385	0.000391	0.0254
MLE: 2-step from OLS	−0.00213	−0.00132	0.000319	0.000323	0.0243
ADAP1: optimal band	−0.00250	−0.00154	0.000369	0.000375	0.0253
ADAP2: $h = 0.01$	−0.00253	−0.00174	0.000372	0.000379	0.0250
ADAP3: $h = 0.003$	−0.00236	−0.00084	0.000372	0.000378	0.0238
ADAP4: $h = 0.001$	−0.00253	−0.00186	0.000376	0.000382	0.0249
		$b = 0.05$			
OLS estimator	−0.00251	−0.00214	0.000385	0.000391	0.0254
MLE: 2-step from OLS	−0.00213	−0.00132	0.000327	0.000332	0.0243
ADAP1: optimal band	−0.00279	−0.00262	0.000382	0.000390	0.0242
ADAP2: $h = 0.01$	−0.00262	−0.00247	0.000390	0.000397	0.0257
ADAP3: $h = 0.003$	−0.00264	−0.00244	0.000386	0.000393	0.0258
ADAP4: $h = 0.001$	−0.00249	−0.00208	0.000408	0.000414	0.0263

OLS, MLE, and adaptive estimator with optimal bandwidth, which was close to 0.006, we also considered the adaptive estimators with fixed bandwidth $h =$ 0.001, 0.003, 0.01. We use the trimming function (5) and the following two values of trimming parameter b: 0.005 and 0.05, corresponding to the two parts of Table 2. Alternative choices of the trimming parameters were tried, and the results are quantitatively similar. These results confirm the previous finding from Table 1 that the finite sample performance of the adaptive estimator is affected by the choice of h. We also see that the efficiency gain from using the density information is relatively higher for DGP(2). A comparison within Table 2 indicates that the choices of trimming parameter values have important influence on the finite sample performance.

In summary, these Monte Carlo results illustrated the influence of choices of smoothing parameters on the finite sample performance of the semiparametric adaptive regression estimators, and confirms the effectiveness of the second order theory.

7. CONCLUSION

The results of this paper readily extend to the multivariate SUR case where ε_i is a vector, see Jeganathan (1995) and Hodgson (1998) for first-order theory. In this case the corresponding second-order rate is $n^{q/(2q+d+2)}$, which worsens with dimensions. Our results also extend to the nonlinear regression function case as in Manski (1984). Finally, when ε_i has higher order dependence on x_i, it may still be possible to justify our results provided that

$$E\left[x_i \frac{f'(\varepsilon_i)}{f(\varepsilon_i)}\right] = 0.$$

This would happen for example if the conditional distribution of $\varepsilon_i | x_i$ were symmetric about zero. See Hodgson (1999) for a first-order result in this direction. However, in cases where ARCH effects are strong, it may be preferable to work with the adaptive ARCH estimator of Linton (1993).

NOTE

1. In any case, the density and its derivatives are zero at the boundary, so that the bias would be $O(h^{\varrho})$ were we to be estimating there.

REFERENCES

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

Akahira, M. & K. Takeuchi (1995) *Non-Regular Statistical Estimation*. New York: Springer-Verlag.

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

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.

De Jong, P. (1987) A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields* 75, 261–277.

Devroye, Luc (1995) *Non-Uniform Random Variate Generation*. New York: Springer-Verlag.

Dieudonné, J. (1969) *Foundations of Modern Analysis*. New York: Academic Press.

Fan, J. & Q. Li (1996) Central Limit Theorem for Degenerate U-statistics of Absolutely Regular Processes with Applications to Model Specification Testing. Discussion paper 1996-8, University of Guelph.

Härdle, W., J. Hart, J.S. Marron, & A.B. Tsybakov (1992) Bandwidth choice for average derivative estimation. *Journal of the American Statistical Association* 87, 218–226.

Hodgson, D. (1998) Adaptive estimation of cointegrating regressions with ARMA errors. *Journal of Econometrics* 85, 231–268.

Hodgson, D. (1999) Partial Maximum Likelihood and Adaptive Estimation in the Presence of Conditional Heterogeneity of Unknown Form. Manuscript, University of Rochester.

Hsieh, D.A. & C.F. Manski (1987) Monte Carlo evidence on adaptive maximum likelihood estimation of a regression. *Annals of Statistics* 15, 541–551.

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. (1995) Second order approximation in the partially linear regression model. *Econometrica* 63, 1079–1112.

Linton, O.B. (1996) Second order approximation in a linear regression with heteroskedasticity of unknown form. *Econometric Reviews* 15, 1–32.
Linton, O.B. (1998) Edgeworth Approximations for Semiparametric Instrumental Variable Estimators and Test Statistics. Manuscript, Yale University.
Linton, O.B. & Z. Xiao (1997) Second Order Approximation in a Semiparametric Binary Choice Model. Manuscript, Yale University.
Linton, O.B. & Z. Xiao (1998) Second Order Approximation for Adaptive Regression Estimators. Manuscript, Yale University.
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.
Müller, H.G. (1988) *Nonparametric Regression Analysis of Longitudinal Data*. London: Springer-Verlag.
Nishiyama, Y. & P.M. Robinson (1997) Edgeworth expansions for semiparametric averaged derivatives. *Econometrica*, forthcoming.
Phillips, P.C.B. (1978) A general theorem in the theory of asymptotic expansion as approximations to the finite sample distributions of econometric estimators. *Econometrica* 45, 463–485.
Powell, J.L. & T.M. Stoker (1996) Optimal bandwidth choice for density-weighted averages. *Journal of Econometrics* 75, 291–316.
Rothenberg, T. (1984) Approximating the distributions of econometric estimators and test statistics. In Z. Griliches & M. Intriligator (eds.), *The Handbook of Econometrics*, vol. 2. Amsterdam: North-Holland.
Rothenberg, T. & C.T. Leenders (1964) Efficient estimation of simultaneous equation systems. *Econometrica* 32, 57–76.
Sargan, J.D. (1976) Econometric estimators and the Edgeworth approximation. *Econometrica* 44, 421–448.
Silverman, B. (1978) Weak and strong uniform consistency of the kernel estimate of a density function and its derivatives. *Annals of Statistics* 8, 1175–1176.
Silverman, B. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.
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 J. Neyman (ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press.
Stone, C. (1975) Adaptive maximum likelihood estimation of a location parameter. *Annals of Statistics* 3, 267–284.
Xiao, Z. & P.C.B. Phillips (1996) Higher order approximation for a frequency domain regression estimator. *Journal of Econometrics* 86, 297–336.

APPENDIX

We shall use the notation $E_i(\cdot)$ to denote $E_i(\cdot|\mathcal{F}_i)$, where $\mathcal{F}_i = \{\varepsilon_i; x_1, \dots, x_n\}$. Note that the Lebesgue density of $\varepsilon_i(\beta) = \varepsilon_i - (\beta - \beta_0)^T x_i$, denoted $f_\beta(\cdot;\beta)$, is the convolution of f with the density or probability mass function of x_i. We shall just treat explicitly the case where x_i has a Lebesgue density f_x, because the discrete case is similar. Note that if

x_i has unbounded support, then so does $\varepsilon_i(\beta)$. Let $K_{h_n}^{(j)}(u) = (1/h_n^{j+1})K^{(j)}(u/h_n)$, $j = 0,1,2$.

LEMMA 1. *For $j = 0,1,2$, we have as $n \to \infty$,*

$$E_i[K_{h_n}^{(j)}(\varepsilon_i(\beta) - \varepsilon_j(\beta))] = f_\beta^{(j)}(\varepsilon_i(\beta)) + \frac{h_n^q}{q!} f_\beta^{(q+j)}(\varepsilon_i(\beta)) \int u^q K(u)du + o(h_n^q), \tag{A.1}$$

$$E_i[K_{h_n}^{(j)}(\varepsilon_i(\beta) - \varepsilon_j(\beta))^2] = \frac{1}{h_n^{2j+1}} f_\beta(\varepsilon_i(\beta))\|K^{(j)}\|_2^2 + o(h_n^{-2j+1}) \tag{A.2}$$

uniformly in $S = \{i : f(\varepsilon_i) \geq b\} \cap \{\beta : \|\beta - \beta_0\| \leq c/\sqrt{n}\}$ with probability one. Here, $f_\beta^{(j)}(u) = \int f^{(j)}(u - (\beta - \beta_0)^T x) f_x(x)dx$.

LEMMA 2. *Suppose that Assumptions A1–A6 hold. Then, for $j = 0,1,2$,*

$$\sup_{\|\beta-\beta_0\|\leq c/\sqrt{n}} \max_{1\leq i\leq n} |\tilde{f}_\beta(\varepsilon_i(\beta)) - f_\beta(\varepsilon_i(\beta))| = O_p(h_n^\varrho) + O_p\left(\sqrt{\frac{\log n}{nh_n^{2j+1}}}\right), \tag{A.3}$$

$$\sup_{\|\beta-\beta_0\|\leq c/\sqrt{n}} \max_{1\leq i\leq n} |\tilde{f}_\beta^{(j)}(\varepsilon_i(\beta)) - f_\beta^{(j)}(\varepsilon_i(\beta))| G_b(|f_i|) = O_p(h_n^q) + O_p\left(\sqrt{\frac{\log n}{nh_n^{2j+1}}}\right). \tag{A.4}$$

In the remainder of this section we write $\tilde{f}_i$, $\tilde{f}_i'$, $\partial\tilde{f}_i/\partial\beta$, and $\partial\tilde{f}_i'/\partial\beta$ (at the true β_0) in terms of their probability limits and correction terms involving "bias terms" and "variance terms." We decompose $\tilde{f}(\varepsilon_i) - f(\varepsilon_i)$ in the following way:

$$\tilde{f}_i(\varepsilon_i) - f(\varepsilon_i) = [\tilde{f}_i(\varepsilon_i) - E_i\{\tilde{f}_i(\varepsilon_i)\}] + [E_i\{\tilde{f}_i(\varepsilon_i)\} - f(\varepsilon_i)] \equiv V_i + B_i, \tag{A.5}$$

where

$$E_i\{\tilde{f}_i(\varepsilon_i)\} = E_i\{K_{h_n}(\varepsilon_i - \varepsilon_j)\} = f(\varepsilon_i) + \frac{h_n^q}{q!} f^{(q)}(\varepsilon_i) \int u^q K(u)du + o(h_n^q)$$

by identity of distribution and Lemma 1. Likewise, the derivative estimate can be written as follows:

$$\tilde{f}_i' = \tilde{f}_i'(\varepsilon_i) = f'(\varepsilon_i) + [E_i \tilde{f}_i'(\varepsilon_i) - f'(\varepsilon_i)] + [\tilde{f}_i'(\varepsilon_i) - E_i \tilde{f}_i'(\varepsilon_i)] = f'(\varepsilon_i) + B_i' + V_i', \tag{A.6}$$

where

$$E_i\{\tilde{f}_i'(\varepsilon_i)\} = \frac{1}{n-1}\sum_{j\neq i} E_i\{K_{h_n}'(\varepsilon_i - \varepsilon_j)\} = f'(\varepsilon_i) + \frac{h_n^q}{q!} f^{(q+1)}(\varepsilon_i) \int u^q K(u)du + o(h_n^q),$$

using integration by parts.

For the derivatives with respect to β of the density estimates, we have

$$\frac{\partial \tilde{f}_i(\varepsilon_i(\beta))}{\partial \beta} = \frac{1}{n-1}\sum_{j\neq i} K'_{h_n}(\varepsilon_i - \varepsilon_j)(x_j - x_i),$$

$$\frac{\partial \tilde{f}_i'(\varepsilon_i(\beta))}{\partial \beta} = \frac{1}{n-1}\sum_{j\neq i} K''_{h_n}(\varepsilon_i - \varepsilon_j)(x_j - x_i),$$

and we make the following decomposition:

$$\begin{aligned}\frac{\partial \tilde{f}_i(\varepsilon_i(\beta))}{\partial \beta} &= \bar{f}_i' + \left[E_i\left\{\frac{\partial \tilde{f}(\varepsilon_i(\beta))}{\partial \beta}\right\} - \bar{f}_i'\right] + \left[\frac{\partial \tilde{f}(\varepsilon_i(\beta))}{\partial \beta} - E_i\left\{\frac{\partial \tilde{f}(\varepsilon_i(\beta))}{\partial \beta}\right\}\right] \\ &\equiv \bar{f}_i' + \bar{B}_i' + \bar{V}_i', \qquad \textbf{(A.7)}\end{aligned}$$

where

$$\bar{f}_i' = \frac{1}{n-1}\sum_{j\neq i}(x_j - x_i) f'(\varepsilon_i) = f'(\varepsilon_i)(\bar{x}_{-i} - x_i) = -x_i f'(\varepsilon_i) + O_p(n^{-1/2}),$$

$$\begin{aligned}\bar{B}_i' &= \frac{1}{n-1}\sum_{j\neq i}(x_j - x_i) E_i K'_{h_n}(\varepsilon_i - \varepsilon_j) - \bar{f}_i' \\ &= f'(\varepsilon_i)(\bar{x}_{-i} - x_i) + \frac{h_n^q}{q!} f^{(q+1)}(\varepsilon_i)\int u^q K(u)\,du(\bar{x}_{-i} - x_i) - \bar{f}_i' + o(h_n^q) \\ &= -x_i \frac{h_n^q}{q!} f^{(q+1)}(\varepsilon_i)\int u^q K(u)\,du + o(h_n^q) + O_p(n^{-1/2}),\end{aligned}$$

where $\bar{x}_{-i} = 1/(n-1)\sum_{j\neq i} x_j = O_p(n^{-1/2})$. Finally,

$$\begin{aligned}\frac{\partial \tilde{f}_i'(\varepsilon_i(\beta))}{\partial \beta} &= \bar{f}_i'' + \left[E_i\left\{\frac{\partial \tilde{f}_i'(\varepsilon_i(\beta))}{\partial \beta}\right\} - \bar{f}_i''\right] + \left[\frac{\partial \tilde{f}_i'(\varepsilon_i(\beta))}{\partial \beta} - E_i\left\{\frac{\partial \tilde{f}_i'(\varepsilon_i(\beta))}{\partial \beta}\right\}\right] \\ &= \bar{f}_i'' + \bar{B}_i'' + \bar{V}_i'', \qquad \textbf{(A.8)}\end{aligned}$$

where

$$\bar{f}_i'' = \frac{1}{n-1}\sum_{j\neq i}(x_j - x_i) f''(\varepsilon_i) = f''(\varepsilon_i)(\bar{x}_{-i} - x_i) = -x_i f''(\varepsilon_i) + O_p(n^{-1/2}),$$

$$\begin{aligned}\bar{B}_i'' &= \frac{1}{n-1}\sum_{j\neq i}(x_j - x_i) E_i K''_{h_n}(\varepsilon_i - \varepsilon_j) - \bar{f}_i'' \\ &= f''(\varepsilon_i)(\bar{x}_{-i} - x_i) + \frac{h_n^q}{q!} f^{(q+2)}(\varepsilon_i)\int u^q K(u)\,du(\bar{x}_{-i} - x_i) - \bar{f}_i'' + o(h_n^q) \\ &= -x_i \frac{h_n^q}{q!} f^{(q+2)}(\varepsilon_i)\int u^q K(u)\,du + o(h_n^q) + O_p(n^{-1/2}).\end{aligned}$$

As to the stochastic terms in (A.5)–(A.8), we have, for example,

$$E_i[\bar{V}_i' \bar{V}_i'^T] = \frac{1}{(n-1)^2} \sum_{j \neq i} (x_j - x_i)(x_j - x_i)^T \{E_i[K_{h_n}'(\varepsilon_i - \varepsilon_j)^2] - [E_i K_{h_n}'(\varepsilon_i - \varepsilon_j)]^2\}$$

$$= \frac{1}{n} E[(x_j - x_i)(x_j - x_i)^T] E_i[K_{h_n}'(\varepsilon_i - \varepsilon_j)^2] + O(n^{-1}) = O(n^{-1} h_n^{-3}).$$

In conclusion, B_i, B_i', $\bar{B}_i'$, $\bar{B}_i''$ are $O_p(h_n^q)$ uniformly in i. The stochastic terms V_i, V_i', $\bar{V}_i'$, and $\bar{V}_i''$ are mean zero sums of independent random variables with probability orders $O_p(n^{-1/2} h_n^{-1/2})$, $O_p(n^{-1/2} h_n^{-3/2})$, $O_p(n^{-1/2} h_n^{-3/2})$, and $O_p(n^{-1/2} h_n^{-5/2})$, respectively, which follows from Lemma 1.

Proof of Lemma 1. We first use the law of iterated expectations to write

$$E_i[K_{h_n}(\varepsilon_i(\beta) - \varepsilon_j(\beta))] = \int \left[\int K_{h_n}(\varepsilon_i(\beta) - \varepsilon_j(\beta)) f(\varepsilon_j) d\varepsilon_j \right] f_x(x_j) dx_j.$$

We work on the inner integral. Given the moment conditions in assumption A1 and the bandwidth conditions in A6, and when $\|\beta - \beta_0\| \leq c/\sqrt{n}$, we have (see the proof of Lemma 2) $\max_{1 \leq i \leq n} |(\beta - \beta_0)^T (x_i - x_j)| = o_p(b^{1/\rho})$. Notice that on S we have $f(\varepsilon_i) \geq b$; thus, for small values of $f(\varepsilon_i)$, $\varepsilon_i - \underline{a}$ or $\bar{a} - \varepsilon_i$ are of order $b^{1/\varrho}$ by A4. Under Assumption A6, they are larger in order of magnitude than h. As a result, by a change of variable $\varepsilon_j \mapsto u = (\varepsilon_i - \varepsilon_j - (\beta - \beta_0)^T (x_i - x_j))/h_n$, we have the following approximation:

$$\int K_{h_n}(\varepsilon_i(\beta) - \varepsilon_j(\beta)) f(\varepsilon_j) d\varepsilon_j \approx \int_{-1}^{1} f(\varepsilon_i - (\beta - \beta_0)^T (x_i - x_j) + u h_n) K(u) du$$

on S. The following Taylor expansion of f around $\varepsilon_{ij}^* = \varepsilon_i - (\beta - \beta_0)^T (x_i - x_j)$ is valid for $q \leq r$

$$f(\varepsilon_{ij}^* + u h_n) = f(\varepsilon_{ij}^*) + u h_n f'(\varepsilon_{ij}^*) + \cdots + \frac{(u h_n)^q}{q!} f^{(q)}(\varepsilon_{ij}^*)$$

$$+ \frac{(u h_n)^q}{q!} \int_0^1 q(1-t)^{q-1} \{ f^{(q)}(\varepsilon_{ij}^* + t u h_n) - f(\varepsilon_{ij}^*)\} dt$$

by Dieudonné (1969, Theorem 8.14.3). Therefore,

$$\int K_{h_n}(\varepsilon_i(\beta) - \varepsilon_j(\beta)) f(\varepsilon_j) d\varepsilon_j = f(\varepsilon_{ij}^*) + \frac{h_n^q}{q!} f^{(q)}(\varepsilon_{ij}^*) \int u^q K(u) du + R_{ij},$$

where, by Assumption A4, uniformly in i with probability one

$$|R_{ij}| = \frac{h_n^q}{q!} \left| \int u^q K(u) \int_0^1 q(1-t)^{q-1} \{ f^{(q)}(\varepsilon_{ij}^* + t u h_n) - f^{(q)}(\varepsilon_{ij}^*)\} dt du \right|$$

$$\leq c h_n^{q+1} \left| \int u^{q+1} K(u) du \right|$$

for some constant c, by Lipschitz continuity of $f^{(q)}$. The final step is to integrate with respect to f_x—the integrals are finite because of Assumptions A1–A3.

As for the derivatives, we use integration by parts (which is justified by Assumption A4). We have, e.g.,

$$\int \frac{1}{h_n^2} K'\left(\frac{\varepsilon_i(\beta)-\varepsilon_j(\beta)}{h_n}\right) f(\varepsilon_j)d\varepsilon_j = \frac{1}{h_n}\int f'(\varepsilon_j)K\left(\frac{\varepsilon_i(\beta)-\varepsilon_j(\beta)}{h_n}\right)d\varepsilon_j$$

$$= \int f'(\varepsilon_{ij}^* + uh_n)K(u)du.$$

We can then apply the same Taylor expansion as previously, this time of f' around ε_{ij}^*. ■

Proof of Lemma 2. The proof follows from an extension of Silverman (1978), as treated in Andrews (1995). In the second part of the theorem, note that $G_b(|f_i|)$ excludes observations too close to the boundary, so that we can apply the usual Taylor series expansion to treat the bias terms. In the first part of the theorem, the bias terms are small because the density (and its derivatives up to order $\varrho - 1$) is zero at the boundary. With regard to the stochastic part in both theorems, uniform convergence over i follows from Masry (1996, Theorem 2). We concentrate on the uniformity with respect to β. We first show that $\Pr[\mathcal{A}] = o(1)$ for any $c_n \to \infty$, where

$$\mathcal{A} = \left\{ \sup_{\|\beta-\beta_0\|\le c/\sqrt{n}} |\tilde{f}_\beta^{(j)}(\varepsilon_i(\beta)) - E_i\{\tilde{f}_\beta^{(j)}(\varepsilon_i(\beta))\}| > c_n\sqrt{\frac{\log n}{nh_n^{2j+1}}}\right\}.$$

The proof is made more complicated by the fact that we have not assumed that the support of x_i is finite so that the density function f_β could have unbounded support—or more relevantly, the range of the evaluation points $\{\varepsilon_i(\beta)\}_{i=1}^n$ increases as $n \to \infty$. Define the events

$$\mathcal{C}_n(d) = \left\{\max_{1\le i\le n} \|x_i\| \le dn^{1/2}b^{1/\varrho}\right\}.$$

Then, by Assumption A1, we have

$$\Pr[\mathcal{C}_n^c(d)] \le n\Pr[\|x_i\| > dn^{1/2}b^{1/\varrho}]$$

$$\le n\frac{E[\|x_i\|^{4+\eta}]}{d^{4+\eta}n^{2+\eta/2}b^{(4+\eta)/\varrho}} \to 0, \qquad \textbf{(A.9)}$$

i.e., $\max_{1\le i\le n}\|x_i\|/n^{1/2}b^{1/\varrho} = o_p(1)$. We shall restrict attention to $\mathcal{A} \cap \mathcal{C}_n(d)$, which can be justified by the argument that $\Pr[\mathcal{A}] \le \Pr[\mathcal{A} \cap \mathcal{C}] + \Pr[\mathcal{C}^c]$ for any event $\mathcal{A}$.

Write $\beta = \beta_0 + b/\sqrt{n}$ for any $\beta \in \mathcal{N}_n(c) = \{\beta : \|\beta - \beta_0\| \le c/\sqrt{n}\}$. Because $\mathcal{N}_n(c)$ is compact it can be covered by a finite number of cubes I_{nl} with centers $\beta_l = \beta_0 + b_l/\sqrt{n}$ having sides of length $\delta(L)$, for $l = 1,\dots,L$. Note that $\delta(L) = O(L^{-1/p})$. We take

$$L(n) = \left(\frac{n}{h^3\log n}\right)^{p/2}.$$

We now make a standard decomposition

$$
\begin{aligned}
&\sup_{\mathcal{N}_n(c)} |\tilde f_\beta^{(j)}(\varepsilon_i(\beta)) - E_i\{\tilde f_\beta^{(j)}(\varepsilon_i(\beta))\}| \\
&\quad \le \max_{1\le l\le L} \sup_{\mathcal{N}_n(c)\cap I_{nl}} |\tilde f_\beta^{(j)}(\varepsilon_i(\beta)) - \tilde f_\beta^{(j)}(\varepsilon_i(\beta_l))| \\
&\qquad + \max_{1\le l\le L} |\tilde f_\beta^{(j)}(\varepsilon_i(\beta_l)) - E_i\{\tilde f_\beta^{(j)}(\varepsilon_i(\beta_l))\}| \\
&\qquad + \max_{1\le l\le L} \sup_{\mathcal{N}_n(c)\cap I_{nl}} |E_i\{\tilde f_\beta^{(j)}(\varepsilon_i(\beta_l))\} - E_i\{\tilde f_\beta^{(j)}(\varepsilon_i(\beta))\}| \\
&\quad \equiv Q_1 + Q_2 + Q_3.
\end{aligned}
$$

By the Lipschitz condition on the kernel and the Cauchy–Schwarz inequality, we have for any $\beta \in \mathcal{N}_n(c) \cap I_{nl}$,

$$
\begin{aligned}
&\frac{1}{h_n^{j+1}} \left| K^{(j)}\left(\frac{\varepsilon_i(\beta_l) - \varepsilon_j(\beta_l)}{h_n}\right) - K^{(j)}\left(\frac{\varepsilon_i(\beta) - \varepsilon_j(\beta)}{h_n}\right)\right| \\
&\quad \le \frac{c}{h_n^{j+2}} \max_{1\le i\le n} |\varepsilon_i(\beta_l) - \varepsilon_i(\beta)| \\
&\quad \le \frac{c}{h_n^{j+2}} \|\beta_l - \beta\| \max_{1\le i\le n} \|x_i\| \\
&\quad \le \frac{c\delta(L)}{h_n^{j+2}} = \frac{c}{h_n^{j+2}}\left(\frac{h^3 \log n}{n}\right)^{1/2} = c\left(\frac{\log n}{nh_n^{2j+1}}\right)^{1/2}
\end{aligned}
$$

for some constant c (which can be different from expression to expression) with probability tending to one by (A.9). This gives a bound on Q_1 similar to (3.21) in Masry (1996). The Bonferroni inequality and an exponential bound are used to treat Q_2. Specifically, for any $\lambda > 0$, we have

$$
\begin{aligned}
&\Pr\left[\max_{1\le l\le L} |\tilde f_\beta^{(j)}(\varepsilon_i(\beta_l)) - E_i\{\tilde f_\beta^{(j)}(\varepsilon_i(\beta_l))\}| > \lambda\right] \\
&\quad \le 2L(n)\exp\left[-\frac{\lambda^2}{2\left(\sum_{i=1}^n \sigma_{ni}^2 + m\lambda\right)}\right], \qquad \textbf{(A.10)}
\end{aligned}
$$

where $m = c/nh_n^{j+1}$ is a bound on the random variable

$$
Z_{ni} = \frac{1}{nh_n^{j+1}}\left[K^{(j)}\left(\frac{\varepsilon_i(\beta_l) - \varepsilon_j(\beta_l)}{h_n}\right) - E_i K^{(j)}\left(\frac{\varepsilon_i(\beta_l) - \varepsilon_j(\beta_l)}{h_n}\right)\right]
$$

and $\sigma_{ni}^2 = c/n^2 h_n^{2j+1} = \operatorname{var}[Z_{ni}]$. We take $\lambda = c_n\sqrt{\log n/nh_n^{2j+1}}$. Note that the $\sum_{i=1}^n \sigma_{ni}^2$ term in the denominator of (A.10) dominates when $nh_n/\log n \to \infty$. It now follows that the right hand side of (A.10) is $o(1)$, provided the constants are taken large enough. Finally, the term Q_3 is bounded in the same way as Q_1. ■

Proof of Theorem 1. By a Taylor series expansion of $\tilde{s}_n(\tilde{\beta})$ about $\tilde{s}_n(\beta_0)$, we have

$$\tilde{s}_n(\tilde{\beta}) = \tilde{s}_n(\beta_0) + \tilde{s}_n'(\beta^*)(\tilde{\beta} - \beta_0),$$

where $\tilde{s}_n'(\beta) = \partial \tilde{s}_n(\beta)/\partial\beta$ and β^* is an intermediate point. Thus

$$\begin{aligned}\sqrt{n}(\hat{\beta} - \beta_0) &= \tilde{\mathcal{I}}_n(\tilde{\beta})^{-1}[\sqrt{n}\tilde{s}_n(\beta_0) + \tilde{s}_n'(\beta^*)\sqrt{n}(\tilde{\beta} - \beta_0)] + \sqrt{n}(\tilde{\beta} - \beta_0)\\ &= \tilde{\mathcal{I}}_n(\beta_0)^{-1}\sqrt{n}\tilde{s}_n(\beta_0)\\ &\quad + \{\tilde{\mathcal{I}}_n(\tilde{\beta})^{-1} - \tilde{\mathcal{I}}_n(\beta_0)^{-1}\}\sqrt{n}\tilde{s}_n(\beta_0) + \{I + \tilde{\mathcal{I}}_n(\tilde{\beta})^{-1}\tilde{s}_n'(\beta^*)\}\sqrt{n}(\tilde{\beta} - \beta_0).\end{aligned}$$

Noticing that $\tilde{s}_n'(\beta^*) = \tilde{s}_n'(\beta_0) + \tilde{s}_n''(\beta^{**})(\beta^* - \beta_0)$, a first-order analysis shows that $\tilde{s}_n''(\beta^{**}) = O_p(1)$, with limit $\Omega_x E\{f'''/f - 3f''f'/(f)^2 + 2(f'/f)^3\}$. Also observing the fact that $\tilde{\beta} - \beta_0 = O_p(n^{-1/2})$, and β^* is an intermediate point between β_0 and $\tilde{\beta}$, we have $\tilde{s}_n'(\beta^*) = \tilde{s}_n'(\beta_0) + O_p(n^{-1/2})$. In addition,

$$\tilde{s}_n'(\beta_0) = -\tilde{\mathcal{I}}_n(\beta_0) - \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T \frac{\tilde{f}_i'(\varepsilon_i(\beta_0))}{\tilde{f}_i(\varepsilon_i(\beta_0))} g_b(\tilde{f}_i(\varepsilon_i(\beta_0)))\tilde{f}_i(\varepsilon_i(\beta_0)).$$

The second term, $(1/n)\sum_{i=1}^n x_i x_i^T g_b(\tilde{f}_i)(\tilde{f}_i')^2/\tilde{f}_i$, is a higher order term that depends on the trimming parameter and on the boundary behavior of the densities. By a similar argument used elsewhere in this paper, we can show that it is $o_p(b^{(\varrho-1)/2\varrho})$. In particular, notice that for small enough b, $|g_b(f)| \le b^{-1}$, the leading term is asymptotically

$$\begin{aligned}&E\left[x_i x_i^T \frac{(f_i')^2}{f_i} g_b(f_i)\right]\\ &\quad = \Omega_x \int \left[\frac{f'(\varepsilon)^2}{f(\varepsilon)} g_b(f(\varepsilon))\right] 1(b \le f(\varepsilon) \le 2b) f(\varepsilon)\, d\varepsilon\\ &\quad \approx \Omega_x \int_{\underline{a}+\underline{\delta}_1}^{\underline{a}+\underline{\delta}_2} c(\underline{a})(\varepsilon - \underline{a})^{2(\varrho-1)} g_b(f(\varepsilon))\, d\varepsilon\\ &\qquad + \Omega_x \int_{\bar{a}-\bar{\delta}_2}^{\bar{a}-\bar{\delta}_1} c(\bar{a})(\varepsilon - \bar{a})^{2(\varrho-1)} g_b(f(\varepsilon))\, d\varepsilon\\ &\quad \le \Omega_x b^{-1}\left[c(\underline{a})\int_{\underline{a}+\underline{\delta}_1}^{\underline{a}+\underline{\delta}_2} (\varepsilon - \underline{a})^{2(\varrho-1)} d\varepsilon + c(\bar{a})\int_{\bar{a}-\bar{\delta}_2}^{\bar{a}-\bar{\delta}_1} (\varepsilon - \bar{a})^{2(\varrho-1)} d\varepsilon\right]\\ &\quad = \Omega_x b^{-1}[c(\underline{a}) + c(\bar{a})] b^{(2\varrho-1)/\varrho}\\ &\quad = o(b^{(\varrho-1)/2\varrho}),\end{aligned}$$

where $\underline{\delta}_1$, $\underline{\delta}_2$, $\bar{\delta}_1$, $\bar{\delta}_2$ are defined as in (A.15) and (A.16), which follow. Thus, we have

$$\{I + \tilde{\mathcal{I}}_n(\tilde{\beta})^{-1}\tilde{s}_n'(\bar{\beta})\}\sqrt{n}(\tilde{\beta} - \beta_0) = \mathcal{T}_I + O_p(n^{-1/2}),$$

where $\mathcal{T}_I$ is a trimming effect of order $o_p(b^{(\varrho-1)/2\varrho})$.

For $\{\tilde{\mathcal{I}}_n(\tilde{\beta})^{-1} - \tilde{\mathcal{I}}_n(\beta_0)^{-1}\}\sqrt{n}\tilde{s}_n(\beta_0)$, first, it can be shown that $\tilde{\mathcal{I}}_n(\tilde{\beta}) - \tilde{\mathcal{I}}_n(\beta_0) = O_p(n^{-1/2})$. Using the results given previously in this Appendix, we have

$$\begin{aligned}
&\tilde{f}_i(\varepsilon_i(\tilde{\beta})) - \tilde{f}_i(\varepsilon_i(\beta_0))\\
&\qquad = \frac{1}{n-1}\sum_{j\neq i}\{K_{h_n}[(\varepsilon_i - \varepsilon_j) + (\tilde{\beta} - \beta)^T(x_j - x_i)] - K_{h_n}[\varepsilon_i - \varepsilon_j]\}\\
&\qquad \approx \frac{1}{n-1}\sum_{j\neq i}K'_{h_n}(\varepsilon_i - \varepsilon_j)(x_j - x_i)^T(\tilde{\beta} - \beta)\\
&\qquad \approx -x_i f'(\varepsilon_i)(\tilde{\beta} - \beta)\\
&\qquad = O_p(n^{-1/2}),
\end{aligned}$$

and similarly

$$\tilde{f}_i'(\varepsilon_i(\tilde{\beta})) - \tilde{f}_i'(\varepsilon_i(\beta_0)) \approx -x_i f''(\varepsilon_i)(\tilde{\beta} - \beta) = O_p(n^{-1/2}).$$

Thus, using the definition of $\tilde{\mathcal{I}}_n(\beta, \tilde{f})$, we can show that $\tilde{\mathcal{I}}_n(\tilde{\beta}) - \tilde{\mathcal{I}}_n(\beta_0) = O_p(n^{-1/2})$. In addition, $\sqrt{n}s_n(\beta_0) = O_p(1)$. Thus, by a geometric expansion we have

$$\{\tilde{\mathcal{I}}_n(\tilde{\beta})^{-1} - \tilde{\mathcal{I}}_n(\beta_0)^{-1}\}\sqrt{n}\tilde{s}_n(\beta_0) = O_p(n^{-1/2}).$$

We now develop the expansions of the score function and the Hessian. By definition

$$\sqrt{n}\tilde{s}_n(\beta_0) = -\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\tilde{f}'(\varepsilon_i(\beta_0))}{\tilde{f}(\varepsilon_i(\beta_0))} x_i G_b(\tilde{f}_i).$$

For simplicity, we denote $G_b(\tilde{f}_i)$ as $\tilde{G}_i$. For the denominator

$$\frac{1}{\tilde{f}(\varepsilon_i)} = \frac{1}{f(\varepsilon_i)} - \frac{\tilde{f}(\varepsilon_i) - f(\varepsilon_i)}{f(\varepsilon_i)^2} + \frac{\{\tilde{f}(\varepsilon_i) - f(\varepsilon_i)\}^2}{f(\varepsilon_i)^3} + R_2,$$

where

$$R_2 = \frac{-\{\tilde{f}(\varepsilon_i) - f(\varepsilon_i)\}^3}{f^3(\varepsilon_i)\tilde{f}(\varepsilon_i)}.$$

Substituting $B_i + V_i$ for $\tilde{f}(\varepsilon_i) - f(\varepsilon_i)$ and $f'(\varepsilon_i) + B_i' + V_i'$ for $\tilde{f}'(\varepsilon_i(\beta_0))$ and Taylor expanding, we get

$$
\begin{aligned}
\sqrt{n}\tilde{s}_n(\beta_0) &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\tilde{f}'(\varepsilon_i(\beta_0))}{\tilde{f}(\varepsilon_i(\beta_0))}x_i G_b(\tilde{f}_i) \\
&= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{1}{f(\varepsilon_i)} - \frac{B_i+V_i}{f(\varepsilon_i)^2} + \frac{\{B_i+V_i\}^2}{f(\varepsilon_i)^3} + R_2\right] \\
&\quad \times [f'(\varepsilon_i) + B_i' + V_i']x_i G_b(\tilde{f}_i) \\
&= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f'(\varepsilon_i)}{f(\varepsilon_i)}x_i - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{B_i'}{f(\varepsilon_i)}x_i\tilde{G}_i \\
&\quad - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{V_i'}{f(\varepsilon_i)}x_i\tilde{G}_i + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f'(\varepsilon_i)}{f(\varepsilon_i)^2}x_i B_i\tilde{G}_i \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f'(\varepsilon_i)}{f(\varepsilon_i)^2}x_i V_i\tilde{G}_i - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f'(\varepsilon_i)}{f(\varepsilon_i)^3}x_i B_i^2\tilde{G}_i \\
&\quad - 2\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f'(\varepsilon_i)}{f(\varepsilon_i)^3}x_i B_i V_i\tilde{G}_i - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f'(\varepsilon_i)}{f(\varepsilon_i)^3}x_i V_i^2\tilde{G}_i \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{B_i'}{f(\varepsilon_i)^2}B_i x_i\tilde{G}_i + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{V_i'}{f(\varepsilon_i)^2}B_i x_i\tilde{G}_i \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{B_i'}{f(\varepsilon_i)^2}V_i x_i\tilde{G}_i + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{V_i'}{f(\varepsilon_i)^2}V_i x_i\tilde{G}_i \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f'(\varepsilon_i)}{f(\varepsilon_i)}x_i[1-\tilde{G}_i] + R_{n2} \\
&= \sum_{j=0}^{11}M_j + \text{Remainder terms,}
\end{aligned}
$$

where the remainder term equals

$$
R_{n2} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\tilde{f}'(\varepsilon_i)[\tilde{f}(\varepsilon_i) - f(\varepsilon_i)]^3}{f(\varepsilon_i)^3\tilde{f}(\varepsilon_i)}x_i\tilde{G}_i \qquad \textbf{(A.11)}
$$

and

$$M_0 = \frac{-1}{\sqrt{n}} \sum_{i=1}^{n} \frac{f'(\varepsilon_i)}{f(\varepsilon_i)} x_i; \qquad M_2 = \frac{-1}{\sqrt{n}} \sum_{i=1}^{n} \frac{V_i'}{f(\varepsilon_i)} x_i \widetilde{G}_i;$$

$$M_1 = \frac{-1}{\sqrt{n}} \sum_{i=1}^{n} \frac{B_i'}{f(\varepsilon_i)} x_i \widetilde{G}_i + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{f'(\varepsilon_i)}{f(\varepsilon_i)^2} x_i B_i \widetilde{G}_i;$$

$$M_3 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{f'(\varepsilon_i)}{f(\varepsilon_i)^2} x_i V_i \widetilde{G}_i; \qquad M_4 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{B_i'}{f(\varepsilon_i)^2} B_i x_i \widetilde{G}_i;$$

$$M_5 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{V_i'}{f(\varepsilon_i)^2} B_i x_i \widetilde{G}_i; \qquad M_6 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{B_i'}{f(\varepsilon_i)^2} V_i x_i \widetilde{G}_i;$$

$$M_7 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{V_i'}{f(\varepsilon_i)^2} V_i x_i \widetilde{G}_i; \qquad M_8 = \frac{-1}{\sqrt{n}} \sum_{i=1}^{n} \frac{f'(\varepsilon_i)}{f(\varepsilon_i)^3} x_i B_i^2 \widetilde{G}_i;$$

$$M_9 = \frac{-1}{\sqrt{n}} \sum_{i=1}^{n} \frac{f'(\varepsilon_i)}{f(\varepsilon_i)^3} x_i V_i^2 \widetilde{G}_i; \qquad M_{10} = \frac{-2}{\sqrt{n}} \sum_{i=1}^{n} \frac{f'(\varepsilon_i)}{f(\varepsilon_i)^2} x_i B_i V_i \widetilde{G}_i;$$

$$M_{11} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{f'(\varepsilon_i)}{f(\varepsilon_i)} x_i [1 - \widetilde{G}_i].$$

The leading term, $M_0 = (-1/\sqrt{n}) \sum_{i=1}^{n} (f'(\varepsilon_i)/f(\varepsilon_i)) x_i$, is of order $O_p(1)$. We now verify the orders of the other terms. First we show that $G_b(\tilde{f}_i)$ can be replaced by $G_b(f_i) = G_i$. If we perform a first-order expansion on $G_b(\tilde{f}_i)$ around f_i we obtain

$$\begin{aligned}
\max_{1\le i\le n} |G_b(\tilde{f}_i) - G_b(f_i)| &= \max_{1\le i\le n} |g_b(\bar{f}_i)(\tilde{f}_i - f_i)| \\
&\le \frac{1}{b} \max_{1\le i\le n} |\tilde{f}_i - f_i| 1(\bar{f}_i \ge b) \\
&\le \frac{1}{b} \max_{1\le i\le n} |\tilde{f}_i - f_i| 1(f_i \ge b)\{1 + o_p(1)\},
\end{aligned}$$

where $\bar{f}_i$ is an intermediate point between $\tilde{f}_i$ and f_i. Note that the second equality follows from Lemma 2, part 1, and Assumption A6. The last term is $o_p(1)$ under our conditions. We also need further expansions

$$G_b(\tilde{f}_i) - G_b(f_i) = \sum_{\ell=0}^{L-1} \frac{1}{\ell!} g_b^{(\ell)}(f_i)(\tilde{f}_i - f_i)^{\ell} + \frac{1}{L!} g_b^{(L)}(\bar{f}_i)(\tilde{f}_i - f_i)^{L},$$

whose properties can be similarly derived.

We verify that under Assumption A6, M_{11} is of order $O_p(b^{(\varrho-1)/2\varrho})$. We have

$$M_{11} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{f'(\varepsilon_i)}{f(\varepsilon_i)} x_i [1 - G_b(f_i)] + \frac{1}{\sqrt{n}} \sum_{\ell=0}^{L-1} \frac{1}{\ell!} \sum_{i=1}^{n} \frac{f'(\varepsilon_i)}{f(\varepsilon_i)} x_i g_b^{(\ell)}(f_i)(\tilde{f}_i - f_i)^{\ell} + \frac{1}{\sqrt{n}} \frac{1}{L!} \sum_{i=1}^{n} \frac{f'(\varepsilon_i)}{f(\varepsilon_i)} x_i g_b^{(L)}(\bar{f}_i)(\tilde{f}_i - f_i)^{L}.$$

It can be shown that under the assumptions,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{f'(\varepsilon_i)}{f(\varepsilon_i)} x_i [1 - G_b(f_i)] \tag{A.12}$$

is the leading term and is of order $O_p(b^{(\varrho-1)/2\varrho})$. Notice that x_i are mean zero and $\{x_i, \varepsilon_i\}_{i=1}^{n}$ are generated as an i.i.d. sample. We only need to calculate the second moment of (A.2), which is

$$E\left\{ \frac{1}{n} \sum_{i=1}^{n} \left[\frac{f'(\varepsilon_i)}{f(\varepsilon_i)} \right]^2 x_i x_i' [1 - G_b(f_i)]^2 \right\} = \Omega_x E\left\{ \left[\frac{f'(\varepsilon)}{f(\varepsilon)} \right]^2 [1 - G_b(f_i)]^2 \right\}.$$

Notice that Assumptions A2–A4 imply that in a small neighborhood around $\underline{a}$,

$$f(\varepsilon) \approx \frac{1}{\varrho!} f^{(\varrho)}(\underline{a})(\varepsilon - \underline{a})^{\varrho}, \tag{A.13}$$

$$f'(\varepsilon) \approx \frac{1}{(\varrho - 1)!} f^{(\varrho)}(\underline{a})(\varepsilon - \underline{a})^{\varrho-1} \tag{A.14}$$

for $\underline{a} \le \varepsilon \le \underline{a} + \delta$ and δ is small. Similar results hold in small neighborhoods around $\bar{a}$. Thus, if we use the trimming function given by (5), noting that $G_b(x)$ is a polynomial in $(x - b)/b$, it can be shown that

$$E\left\{ \left[\frac{f'(\varepsilon)}{f(\varepsilon)} \right]^2 [1 - G_b(f_i)]^2 \right\} \approx \int_{\underline{a}}^{\underline{a}+\underline{\delta}_1} c_0(\underline{a})(\varepsilon - \underline{a})^{\varrho-2} d\varepsilon + \int_{\underline{a}+\underline{\delta}_1}^{\underline{a}+\underline{\delta}_2} \left[\sum_{l=1}^{2k+2} c_l(\underline{a})(\varepsilon - \underline{a})^{l\varrho-2} b^{1-l} \right] d\varepsilon + \int_{\bar{a}-\bar{\delta}_1}^{\bar{a}} c_0(\bar{a})(\varepsilon - \bar{a})^{\varrho-2} d\varepsilon + \int_{\bar{a}-\bar{\delta}_2}^{\bar{a}-\bar{\delta}_1} \left[\sum_{l=1}^{2k+2} c_l(\bar{a})(\varepsilon - \bar{a})^{l\varrho-2} b^{1-l} \right] d\varepsilon,$$

where

$$\underline{\delta}_1 = \left(\frac{\varrho!}{f^{(\varrho)}(\underline{a})} \right)^{1/\varrho} b^{1/\varrho}; \qquad \underline{\delta}_2 = \left(\frac{2\varrho!}{f^{(\varrho)}(\underline{a})} \right)^{1/\varrho} b^{1/\varrho}, \tag{A.15}$$

$$\bar{\delta}_1 = \left(\frac{\varrho!}{f^{(\varrho)}(\bar{a})} \right)^{1/\varrho} b^{1/\varrho}; \qquad \bar{\delta}_2 = \left(\frac{2\varrho!}{f^{(\varrho)}(\bar{a})} \right)^{1/\varrho} b^{1/\varrho}, \tag{A.16}$$

and $c_l(a)$ are functions of $f^{(\varrho)}(a)$ and ϱ. By calculating the integrals, we obtain that

$$E\left\{\left[\frac{f'(\varepsilon)}{f(\varepsilon)}\right]^2[1-G_b(f_i)]^2\right\}\approx\varphi(\varrho,\underline{a},\bar{a},f)b^{(\varrho-1)/\varrho},$$

where

$$\varphi(\varrho,\underline{a},\bar{a},f)=c(\varrho)[f^{(\varrho)}(\underline{a})^{1/\varrho}+f^{(\varrho)}(\bar{a})^{1/\varrho}],$$

where $c(\varrho)$ is a coefficient that depends on ϱ. In general,

$$E\left\{\left[\frac{f'(\varepsilon)}{f(\varepsilon)}\right]^2[1-G_b(f_i)]^2\right\}\approx\varphi(\varrho,\underline{a},\bar{a},f)b^{(\varrho-1)/\varrho}.$$

We now calculate the magnitude of

$$\frac{1}{\sqrt{n}}\frac{1}{\ell!}\sum_{i=1}^n\frac{f'(\varepsilon_i)}{f(\varepsilon_i)}x_i\,g_b^{(\ell)}(f_i)(\tilde{f}_i-f_i)^\ell,\qquad \ell=0,1,\dots,L-1$$

and

$$\frac{1}{\sqrt{n}}\frac{1}{L!}\sum_{i=1}^n\frac{f'(\varepsilon_i)}{f(\varepsilon_i)}x_i\,g_b^{(L)}(\bar{f}_i)(\tilde{f}_i-f_i)^L.$$

The random variable

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{f'(\varepsilon_i)}{f(\varepsilon_i)}x_i\,g_b(f_i)(\tilde{f}_i-f_i)$$

is mean zero, and its second moment is

$$E\frac{1}{n}\sum_{i=1}^n\left[\frac{f'(\varepsilon_i)}{f(\varepsilon_i)}\right]^2x_ix_i'[g_b(f_i)]^2(\tilde{f}_i-f_i)^2=\Omega_xE\left[\frac{f'(\varepsilon_i)}{f(\varepsilon_i)}\right]^2[g_b(f_i)]^2(\tilde{f}_i-f_i)^2,$$

by iterated expectations. Notice that for small enough b

$$|g_b(f_i)|\le b^{-1},\quad \text{for any } i,$$

and $g_b(\cdot)$ has support on $[b,2b]$. By an application of Lemma 2, we have

$$E\left[\frac{f'(\varepsilon_i)}{f(\varepsilon_i)}\right]^2[g_b(f_i)]^2(\tilde{f}_i-f_i)^2$$
$$\le c_1(h^{2q}+n^{-1}h^{-1}\log n)b^{-2}E\left[\frac{f'(\varepsilon_i)}{f(\varepsilon_i)}\right]^2I(b\le f_i\le 2b)$$

for some constant c_1. Under Assumptions A2 and A4, we can use the approximations (A.13) and (A.14) and thus

$$E\left[\frac{f'(\varepsilon_i)}{f(\varepsilon_i)}\right]^2[g_b(f_i)]^2(\tilde{f}_i-f_i)^2\le c_2(h^{2q}+n^{-1}h^{-1}\log n)b^{-(\varrho+1)/\varrho}=o(b^{(\varrho-1)/\varrho}).$$

Thus

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{f'(\varepsilon_i)}{f(\varepsilon_i)} x_i \, g_b(f_i)(\tilde{f}_i - f_i) = o_p(b^{(\varrho-1)/2\varrho}).$$

Similarly we can show that

$$\frac{1}{\sqrt{n}} \frac{1}{\ell!} \sum_{i=1}^{n} \frac{f'(\varepsilon_i)}{f(\varepsilon_i)} x_i \, g_b^{(\ell)}(f_i)(\tilde{f}_i - f_i)^{\ell} = o_p(b^{(\varrho-1)/2\varrho}), \quad \text{for } \ell = 1, \dots, L-1.$$

For

$$\frac{1}{\sqrt{n}} \frac{1}{L!} \sum_{i=1}^{n} \frac{f'(\varepsilon_i)}{f(\varepsilon_i)} x_i \, g_b^{(L)}(\bar{f}_i)(\tilde{f}_i - f_i)^{L},$$

notice that

$$|g_b^{(L)}(y)| \le \frac{c_3}{b^L},$$

for some constant c_3,

$$\left\| \frac{1}{\sqrt{n}} \frac{1}{L!} \sum_{i=1}^{n} \frac{f'(\varepsilon_i)}{f(\varepsilon_i)} x_i \, g_b^{(L)}(\bar{f}_i)(\tilde{f}_i - f_i)^{L} \right\|$$
$$\le \frac{c}{b^{L+1}\sqrt{n}} \max_{1\le i\le n} |\tilde{f}_i - f_i|^L \sum_{i=1}^{n} \left| \frac{f'(\varepsilon_i)}{f(\varepsilon_i)} x_i \right| 1(b \le \bar{f}_i \le 2b).$$

Notice that $h/b \to 0$ as $n \to \infty$ and $\bar{f}_i$ is an intermediate point, by the result of Lemma 2,

$$\left\| \frac{1}{\sqrt{n}} \frac{1}{L!} \sum_{i=1}^{n} \frac{f'(\varepsilon_i)}{f(\varepsilon_i)} x_i \, g_b^{(L)}(\bar{f}_i)(\tilde{f}_i - f_i)^{L} \right\|$$
$$= \frac{c}{b^{L+1}} \max_{1\le i\le n} |\tilde{f}_i - f_i|^L \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| \frac{f'(\varepsilon_i)}{f(\varepsilon_i)} x_i \right| 1(b \le f_i \le 2b)\{1 + o_p(1)\},$$

and

$$\max_{1\le i\le n} |\tilde{f}_i - f_i|^L = O_p(h^{Lq} + n^{-L/2} h^{-L/2} \log^L n).$$

Thus, for $L > 4$,

$$\left\| \frac{1}{\sqrt{n}} \frac{1}{L!} \sum_{i=1}^{n} \frac{f'(\varepsilon_i)}{f(\varepsilon_i)} x_i \, g_b^{(L)}(\bar{f}_i)(\tilde{f}_i - f_i)^{L} \right\| = o_p(b^{(\varrho-1)/2\varrho}).$$

The next two terms are M_1 and M_2. Letting $z^{\otimes 2}$ denote zz^T for any vector z, we have

$$
\begin{aligned}
E(M_1 M_1^T) &\approx E\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{B_i'}{f_i}x_i G_i - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f_i' B_i}{f_i^2}x_i G_i\right]^{\otimes 2} \\
&= E\left[\frac{1}{n}\sum_{i,k=1}^{n}\frac{B_i' B_k'}{f_i f_k}x_i x_k^T G_i G_k + \frac{1}{n}\sum_{i,k=1}^{n}\frac{f_i' f_k' B_i B_k}{f_i^2 f_k^2}x_i x_k^T G_i G_k\right] \\
&\quad - E\left[\frac{2}{n}\sum_{i,k=1}^{n}\frac{f_k' B_i' B_k}{f_i f_k^2}x_i x_k^T G_i G_k\right] \\
&= E\left\{\frac{h_n^{2q}}{n}\sum_{i=1}^{n}\left[\frac{f_i^{(q+1)}}{f_i}\right]^2 x_i x_i^T G_i^2 \mu_q(K)^2\right\} \\
&\quad + E\left\{\frac{h_n^{2q}}{n}\sum_{i=1}^{n}\left[\frac{f_i' f_i^{(q)}}{f_i^2}\right]^2 x_i x_i^T G_i^2 \mu_q(K)^2\right\} \\
&\quad - E\left\{\frac{2h_n^{2q}}{n}\sum_{i=1}^{n}\left[\frac{f_i' f_i^{(q)} f_i^{(q+1)}}{f_i^3}\right]^2 x_i x_i^T G_i^2 \mu_q(K)^2\right\} + o(h_n^{2q}) \\
&= O(h_n^{2q}).
\end{aligned}
$$

For M_2, we write

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{V_i'}{f(\varepsilon_i)}x_i G_i = \frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j\neq i}\xi_{ij},
$$

where $\xi_{ij} = \{K_{h_n}'(\varepsilon_i - \varepsilon_j) - E_i K_{h_n}'(\varepsilon_i - \varepsilon_j)\}x_i G_i / f(\varepsilon_i)$ satisfies $E_i \xi_{ij} = 0$. Furthermore,

$$
\begin{aligned}
E[M_2 M_2^T] &\approx \frac{1}{n^3}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{j\neq i}\sum_{s\neq k}E[\xi_{ij}\xi_{ks}^T] \\
&\approx \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i}E[\xi_{ij}\xi_{ij}^T] + \text{smaller terms} \\
&\approx \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i}E[K_{h_n}'(\varepsilon_i - \varepsilon_j)]^2 G_i^2 x_i x_i^T f(\varepsilon_i)^{-2} + \text{smaller terms} \\
&\approx \Omega_x \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i}\iint h^{-4}K'\left(\frac{\varepsilon_i - \varepsilon_j}{h}\right)^2 f(\varepsilon_i)^{-2} f(\varepsilon_i) f(\varepsilon_j)\,d\varepsilon_i\,d\varepsilon_j \\
&= \Omega_x \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i}h^{-3}\iint K'(u)^2 f(\varepsilon_i)^{-1} f(\varepsilon_i - uh)\,du\,d\varepsilon_i \\
&= O\left(\frac{1}{nh^3}\right).
\end{aligned}
$$

The magnitudes of other terms follow by straightforward moment calculations, because the random variables are *U*-statistics, of various orders, whose moments exist. For example, for

$$M_3 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{f'(\varepsilon_i)}{f(\varepsilon_i)^2} x_i V_i G_i,$$

note that this is of the form $\sum\sum_{i \neq j} \varphi_n(Z_i, Z_j)$, where

$$\varphi_n(Z_i, Z_j) = \frac{1}{nh\sqrt{n}} \frac{f'(\varepsilon_i)}{f(\varepsilon_i)^2} x_i G_i \left[K\left(\frac{\varepsilon_i - \varepsilon_j}{h}\right) - E_i K\left(\frac{\varepsilon_i - \varepsilon_j}{h}\right)\right],$$

which satisfies $E_i \varphi_n(Z_i, Z_j) = 0$ and $E_j \varphi_n(Z_i, Z_j) = 0$. Furthermore,

$$\begin{aligned}
&E[\varphi_n(Z_i, Z_j)\varphi_n(Z_i, Z_j)^T] \\
&\quad = \frac{1}{n^3h^2} E\left(\left(\frac{f'(\varepsilon_i)}{f(\varepsilon_i)^2}\right)^2 x_i x_i^T G_i^2 \left[K\left(\frac{\varepsilon_i - \varepsilon_j}{h}\right) - E_i K\left(\frac{\varepsilon_i - \varepsilon_j}{h}\right)\right]^2\right) \\
&\quad = \frac{1}{n^3h^2} \Omega_x E\left(\left(\frac{f'(\varepsilon_i)}{f(\varepsilon_i)}\right)^2 \frac{1}{f(\varepsilon_i)^2} G_i^2 \left[K\left(\frac{\varepsilon_i - \varepsilon_j}{h}\right) - E_i K\left(\frac{\varepsilon_i - \varepsilon_j}{h}\right)\right]^2\right) \\
&\quad \leq \frac{c}{n^3h^2} \Omega_x E\left[\left(\frac{f'(\varepsilon_i)}{f(\varepsilon_i)}\right)^2 \frac{1}{f(\varepsilon_i)^2} G_i^2 K\left(\frac{\varepsilon_i - \varepsilon_j}{h}\right)^2\right] \\
&\quad = \frac{c}{n^3h^2} \Omega_x \iint \left(\frac{f'(\varepsilon_i)}{f(\varepsilon_i)}\right)^2 \frac{1}{f(\varepsilon_i)^2} G_i^2 K\left(\frac{\varepsilon_i - \varepsilon_j}{h}\right)^2 f(\varepsilon_i) f(\varepsilon_j) d\varepsilon_i \, d\varepsilon_j \\
&\quad = \frac{c}{n^3h} \Omega_x \int K(u)^2 du \int \left(\frac{f'(\varepsilon_i)}{f(\varepsilon_i)}\right)^2 G_i^2 d\varepsilon_i \\
&\quad = O\left(\frac{1}{n^3h} \times b^{-1/\varrho}\right),
\end{aligned}$$

where the last line follows because

$$\int_{b^{1/\varrho}}^{\infty} \varepsilon^{-2} d\varepsilon = O(b^{-1/\varrho}).$$

And thus

$$M_3 = o_p(n^{-1/2}h^{-3/2}).$$

Similarly, it can be verified that the rest of the terms are of order $o_p(n^{-1/2}h^{-3/2})$ or $o_p(h^q)$.

For R_{n2}, by the Cauchy–Schwarz inequality, we obtain that

$$\|R_{n2}\|^2 \le c\left\{\max_{1\le i\le n}\frac{|\tilde{f}_i - f_i|^6}{f_i^6}\tilde{G}_i\right\}^{1/2} \times \left\|\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\tilde{f}'(\varepsilon_i)}{\tilde{f}(\varepsilon_i)}\right\}^2 x_i x_i^T \tilde{G}_i\right\|^{1/2},$$

where c is a constant, and we have

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\tilde{f}'(\varepsilon_i)}{\tilde{f}(\varepsilon_i)}\right\}^2 x_i x_i^T \tilde{G}_i\right\| = O_p(1)$$

$$\max_{1\le i\le n}\frac{|\tilde{f}_i - f_i|^6}{f_i^6}\tilde{G}_i = O_p(b^{-6}h^{6q} + b^{-6}n^{-3}h^{-3}(\log n)^6).$$

Thus, $\|R_{n2}\|^2$ is of order $O_p(h^{3q}b^{-3} + n^{-3/2}h^{-3/2}b^{-3}(\log n)^3)$ and, under Assumption A6, is of smaller order of magnitude than $O_p(h^q + n^{-1/2}h^{-3/2})$.

We now turn to $\Delta_H = \tilde{\mathcal{I}}_n(\beta_0) - \mathcal{I}$. Notice that

$$\tilde{\mathcal{I}}_n(\beta_0) = -\frac{1}{n}\sum_{i=1}^{n} x_i G_b(\tilde{f}_i)\left[\frac{\partial\tilde{f}'(\varepsilon_i)/\partial\beta}{\tilde{f}_i} - \left(\frac{\tilde{f}'(\varepsilon_i)\partial\tilde{f}(\varepsilon_i)/\partial\beta}{\tilde{f}(\varepsilon_i)^2}\right)\right]\Big\downarrow_{\beta=\beta_0}.$$

Expanding each of the estimates in the square brackets to the third term, we have

$$\begin{aligned}
\tilde{\mathcal{I}}_n(\beta_0) = {} & -\frac{1}{n}\sum_{i=1}^{n} x_i[\bar{f}_i'' + \bar{B}_i'' + \bar{V}_i''']\left[\frac{1}{f_i} - \frac{B_i + V_i}{f_i^2} + \frac{(B_i+V_i)^2}{f_i^3}\right]G_b(\tilde{f}_i) \\
& + \frac{1}{n}\sum_{i=1}^{n} x_i\Bigg\{(f_i' + B_i' + V_i')(\bar{f}_i' + \bar{B}_i' + \bar{V}_i') \\
& \qquad \times\left[\frac{1}{f_i^2} - 2\frac{B_i+V_i}{f_i^3} + \frac{3(B_i+V_i)^2}{f_i^4}\right]\Bigg\}G_b(\tilde{f}_i) \\
= {} & -\frac{1}{n}\sum_{i=1}^{n} x_i G_b(\tilde{f}_i)\Bigg\{\frac{\bar{f}_i''}{f_i} + \frac{\bar{B}_i''}{f_i} + \frac{\bar{V}_i''}{f_i} - \frac{\bar{f}_i''}{f_i^2}B_i - \frac{\bar{B}_i''B_i}{f_i^2} - \frac{\bar{V}_i''B_i}{f_i^2} \\
& \qquad - \frac{\bar{f}_i''}{f_i^2}V_i - \frac{\bar{B}_i''V_i}{f_i^2} - \frac{\bar{V}_i''V_i}{f_i^2} + \frac{\bar{f}_i''}{f_i^3}B_i^2 \\
& \qquad + \frac{\bar{f}_i''}{f_i^3}V_i^2 + 2\frac{\bar{f}_i''}{f_i^3}B_iV_i\Bigg\} \\
& + \frac{1}{n}\sum_{i=1}^{n} x_i G_b(\tilde{f}_i)\Bigg\{\frac{f_i'\bar{f}_i'}{f_i^2} + \frac{\bar{f}_i'B_i'}{f_i^2} + \frac{\bar{f}_i'V_i'}{f_i^2} + \frac{f_i'\bar{B}_i'}{f_i^2} + \frac{B_i'\bar{B}_i'}{f_i^2} + \frac{\bar{B}_i'V_i'}{f_i^2} \\
& \qquad + \frac{f_i'\bar{V}_i'}{f_i^2} + \frac{B_i'\bar{V}_i'}{f_i^2} + \frac{V_i'\bar{V}_i'}{f_i^2} - 2\frac{f_i'\bar{f}_i'}{f_i^3}B_i - 2\frac{\bar{f}_i'B_i'B_i}{f_i^3} \\
& \qquad - 2\frac{\bar{f}_i'V_i'B_i}{f_i^3} - 2\frac{f_i'\bar{B}_i'B_i}{f_i^3} - 2\frac{f_i'B_i\bar{V}_i'}{f_i^3} - 2\frac{f_i'\bar{f}_i'}{f_i^3}V_i \\
& \qquad - 2\frac{\bar{f}_i'B_i'V_i}{f_i^3} - 2\frac{\bar{f}_i'V_i'V_i}{f_i^3} - 2\frac{f_i'\bar{B}_i'V_i}{f_i^3} - 2\frac{f_i'\bar{V}_i'V_i}{f_i^3} \\
& \qquad + 3\frac{f_i'\bar{f}_i'}{f_i^4}B_i^2 + 3\frac{f_i'\bar{f}_i'}{f_i^4}V_i^2 + 6\frac{f_i'\bar{f}_i'}{f_i^4}B_iV_i\Bigg\},
\end{aligned}$$

where we have written $f_i = f(\varepsilon_i)$, etc., for convenience. After collecting terms and dropping higher order terms,

$$\tilde{\mathcal{I}}_n(\beta_0) = \frac{1}{n}\sum_{i=1}^n x_i x_i^T\left[\frac{f_i''}{f_i} - \left\{\frac{f_i'}{f_i}\right\}^2\right] + \sum_{\ell=0}^{16} Z_\ell = \mathcal{I}_n + \sum_{\ell=0}^{16} Z_\ell,$$

where

$$Z_0 = -\frac{1}{n}\sum_{i=1}^n \bar{x}_{-i} x_i^T\left[\frac{f_i''}{f_i}\right] + \frac{1}{n}\sum_{i=1}^n \bar{x}_{-i} x_i^T\left[\frac{f_i'}{f_i}\right]^2,$$

$$Z_1 = -\frac{1}{n}\sum_{i=1}^n x_i \tilde{G}_i\left[\frac{\bar{B}_i''}{f_i}\right] + \frac{1}{n}\sum_{i=1}^n x_i \tilde{G}_i\left[\frac{\bar{f}_i''}{f_i^2} B_i\right],$$

$$+ \frac{1}{n}\sum_{i=1}^n x_i \tilde{G}_i\left[\frac{\bar{f}_i'}{f_i^2} B_i'\right] + \frac{1}{n}\sum_{i=1}^n x_i \frac{f_i'}{f_i^2}\bar{B}_i'\tilde{G}_i - \frac{2}{n}\sum_{i=1}^n x_i \tilde{G}_i\left[\frac{f_i'\bar{f}_i'}{f_i^3} B_i\right],$$

$$Z_2 = -\frac{1}{n}\sum_{i=1}^n x_i \tilde{G}_i\left[\frac{\bar{V}_i''}{f_i}\right],$$

$$Z_3 = \frac{1}{n}\sum_{i=1}^n x_i \tilde{G}_i\left[\frac{\bar{f}_i''}{f_i^2} V_i\right] - \frac{2}{n}\sum_{i=1}^n x_i \tilde{G}_i\left[\frac{f_i'\bar{f}_i'}{f_i^3} V_i\right],$$

$$Z_4 = \frac{1}{n}\sum_{i=1}^n x_i \tilde{G}_i\left[\frac{\bar{f}_i' V_i'}{f_i^2}\right], \qquad Z_5 = \frac{1}{n}\sum_{i=1}^n x_i \tilde{G}_i\left[\frac{f_i'\bar{V}_i'}{f_i^2}\right],$$

$$Z_6 = \frac{1}{n}\sum_{i=1}^n x_i \tilde{G}_i \frac{\bar{B}_i'' B_i}{f_i^2} + \frac{1}{n}\sum_{i=1}^n x_i \tilde{G}_i \frac{B_i'\bar{B}_i'}{f_i^2} + \frac{3}{n}\sum_{i=1}^n x_i \tilde{G}_i \frac{f_i'\bar{f}_i' B_i^2}{f_i^4}$$

$$- \frac{2}{n}\sum_{i=1}^n x_i \tilde{G}_i \frac{\bar{f}_i' B_i' B_i}{f_i^3} - \frac{1}{n}\sum_{i=1}^n x_i \tilde{G}_i \frac{\bar{f}_i'' B_i^2}{f_i^3} - \frac{2}{n}\sum_{i=1}^n x_i \tilde{G}_i \frac{f_i'\bar{B}_i' B_i}{f_i^3},$$

$$Z_7 = \frac{1}{n}\sum_{i=1}^n x_i \tilde{G}_i\left[\frac{\bar{V}_i'' B_i}{f_i^2}\right],$$

$$Z_8 = \frac{1}{n}\sum_{i=1}^n x_i \tilde{G}_i \frac{V_i'\bar{B}_i'}{f_i^2} - \frac{2}{n}\sum_{i=1}^n x_i \tilde{G}_i \frac{\bar{f}_i' V_i' B_i}{f_i^3},$$

$$Z_9 = \frac{1}{n}\sum_{i=1}^n x_i \tilde{G}_i \frac{B_i'\bar{V}_i'}{f_i^2} - \frac{2}{n}\sum_{i=1}^n x_i \tilde{G}_i \frac{f_i' B_i\bar{V}_i'}{f_i^3},$$

$$Z_{10} = \frac{1}{n}\sum_{i=1}^n x_i \tilde{G}_i\left[\frac{V_i\bar{B}_i''}{f_i^2}\right] - \frac{2}{n}\sum_{i=1}^n x_i \tilde{G}_i \frac{\bar{f}_i'' B_i V_i}{f_i^3}$$

$$- \frac{2}{n}\sum_{i=1}^n x_i \tilde{G}_i \frac{\bar{f}_i' V_i B_i'}{f_i^3} - \frac{2}{n}\sum_{i=1}^n x_i \tilde{G}_i \frac{f_i'\bar{B}_i' V_i}{f_i^3} + \frac{6}{n}\sum_{i=1}^n x_i \tilde{G}_i \frac{f_i'\bar{f}_i' B_i V_i}{f_i^4},$$

$$Z_{11} = \frac{1}{n}\sum_{i=1}^{n} x_i \tilde{G}_i \left[\frac{V_i' \bar{V}_i'}{f_i^2}\right], \qquad Z_{13} = \frac{1}{n}\sum_{i=1}^{n} x_i \tilde{G}_i \left[\frac{\bar{V}_i'' V_i}{f_i^2}\right],$$

$$Z_{12} = \frac{3}{n}\sum_{i=1}^{n} x_i \tilde{G}_i \left[\frac{f_i' \bar{f}_i'}{f_i^4} V_i^2\right] - \frac{1}{n}\sum_{i=1}^{n} x_i \tilde{G}_i \left[\frac{\bar{f}_i''}{f_i^3} V_i^2\right],$$

$$Z_{14} = -\frac{2}{n}\sum_{i=1}^{n} x_i \tilde{G}_i \frac{\bar{f}_i' V_i' V_i}{f_i^3}, \qquad Z_{15} = -\frac{2}{n}\sum_{i=1}^{n} x_i \tilde{G}_i \frac{f_i' \bar{V}_i' V_i}{f_i^3},$$

$$Z_{16} = -\frac{1}{n}\sum_{i=1}^{n} \bar{x}_{-i} x_i^T \left\{\frac{f_i''}{f_i} + \left[\frac{f_i'}{f_i}\right]^2\right\}\{1 - \tilde{G}_i\}.$$

Combining the expansions for $\tilde{s}_n(\beta_0)$ and $\tilde{\mathcal{I}}_n$, notice that $\mathcal{I}_n = \mathcal{I} + O_p(n^{-1/2})$. Dropping higher order terms, we get

$$\begin{aligned}
\sqrt{n}(\hat{\beta} - \beta_0) &= \mathcal{I}^{-1}\sum_{j=0}^{11} M_j + \mathcal{I}^{-1}\left(\sum_{\ell=0}^{16} Z_\ell\right)\mathcal{I}^{-1}\sum_{j=0}^{11} M_j \\
&= \mathcal{I}^{-1}M_0 + \mathcal{I}^{-1}M_1 + \mathcal{I}^{-1}M_{11} + \mathcal{I}^{-1}M_2 + \mathcal{I}^{-1}Z_1\mathcal{I}^{-1}M_0 \\
&\quad + \mathcal{I}^{-1}Z_{16}\mathcal{I}^{-1}M_0 + o_p(\delta_n) \\
&= X_0 + \mathcal{T} + h_n^q\mathcal{B} + \frac{1}{\sqrt{nh_n^3}}\mathcal{V} + o_p(\delta_n), \qquad \textbf{(A.17)}
\end{aligned}$$

where $\delta_n = \max\{h_n^q, n^{-1/2}h_n^{-3/2}\}$, whereas

$$X_0 = \mathcal{I}^{-1}M_0,$$

$$\mathcal{T} = \mathcal{I}^{-1}M_{11} + \mathcal{I}^{-1}Z_{16}\mathcal{I}^{-1}M_0,$$

$$\mathcal{B} = h_n^{-q}\{\mathcal{I}^{-1}M_1 + \mathcal{I}^{-1}\bar{Z}_1\mathcal{I}^{-1}M_0\},$$

$$\mathcal{V} = n^{1/2}h_n^{3/2}\mathcal{I}^{-1}M_2,$$

where $\bar{Z}_1 = E(Z_1)$, and (A.17) follows because $Z_1 = \bar{Z}_1 + O_p(h_n^q n^{-1/2})$ by a central limit theorem for independent random variables. Furthermore, note that the term $\mathcal{B}$, being a sum of mean zero independent random variables, satisfies a central limit theorem. For the trimming effect $\mathcal{T}$, as we shown in the expansion of the score function, the first term, $\mathcal{I}^{-1}M_{11}$, is of order $O_p(b^{(\varrho-1)/2\varrho})$. By a similar argument, it can be verified that $\mathcal{I}^{-1}Z_{16}\mathcal{I}^{-1}M_0$ is $O_p(b^{(\varrho-1)/\varrho})$ and thus of smaller order of magnitude than the leading term.

We have $E(\tau - \mathcal{T}) = 0$, and

$$\begin{aligned}
E(\tau - \mathcal{T})(\tau - \mathcal{T})^T &= h_n^{2q}E(\mathcal{B}\mathcal{B}^T) + \frac{1}{nh_n^3}E(\mathcal{V}\mathcal{V}^T) \\
&= \mathcal{I}^{-1}E(M_1M_1^T)\mathcal{I}^{-1} + \mathcal{I}^{-1}\bar{Z}_1\mathcal{I}^{-1}E(M_0M_0^T)\mathcal{I}^{-1}\bar{Z}_1^T\mathcal{I}^{-1} \\
&\quad + 2\mathcal{I}^{-1}E(M_1M_0^T)\mathcal{I}^{-1}\bar{Z}_1^T\mathcal{I}^{-1} + \mathcal{I}^{-1}E(M_2M_2^T)\mathcal{I}^{-1}.
\end{aligned}$$

Previous calculation gives that $E(M_1 M_1^T) = h_n^{2q}\mathcal{M}_1 + o(h_n^{2q})$. Also, we have $E(M_0 M_0^T) = \mathcal{I}$ and $E(M_1 M_0^T) = h_n^q \mathcal{M}_2 + o(h_n^q)$. Finally,

$$E(Z_1) = -\frac{1}{n}\sum_{i=1}^{n} x_i\left[\frac{B_i''}{f_i}\right] + \frac{1}{n}\sum_{i=1}^{n} x_i\left[\frac{\bar{f}_i''}{f_i^2}B_i\right] + \frac{1}{n}\sum_{i=1}^{n} x_i\left[\frac{\bar{f}_i'}{f_i^2}B_i'\right]$$

$$+ \frac{1}{n}\sum_{i=1}^{n} x_i \frac{f_i'}{f_i^2}\bar{B}_i' - \frac{2}{n}\sum_{i=1}^{n} x_i\left[\frac{f_i'\bar{f}_i'}{f_i^3}B_i\right]$$

$$= -h_n^q \mu_q(K)\Omega_x\left\{E\left[\frac{f(\varepsilon)^{(q+2)}}{f(\varepsilon)}\right] - 2E\left[\frac{f'(\varepsilon)f^{(q+1)}(\varepsilon)}{f(\varepsilon)^2}\right]\right.$$

$$\left. + 2E\left[\frac{f'(\varepsilon)^2 f^{(q)}(\varepsilon)}{f(\varepsilon)^3}\right] - E\left[\frac{f^{(2)}(\varepsilon)f^{(q)}(\varepsilon)}{f(\varepsilon)^2}\right]\right\}$$

$$+ o(h_n^q)$$

$$= -h_n^q \mu_q(K)\Omega_x \mathcal{D}_q \ell^{(2)}(\varepsilon) + o(h_n^q)$$

and

$$E[M_2 M_2^T] \approx \Omega_x \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i} h^{-3}\iint K'(u)^2 f(\varepsilon_i)^{-1} f(\varepsilon_i - uh)\,du\,d\varepsilon_i$$

$$= \frac{1}{nh^3}\Omega_x E f(\varepsilon)^{-1}\int K'(u)^2 du.$$

We now apply De Jong's (1987) central limit theorem for degenerate weighted U-statistics to the scalar quantity $c^T M_2$ for any vector c, and the result follows by an application of the Cramer–Wold device. ■

Proof of Theorem 2. For linear hypothesis $H_0 \colon c^T\beta = c_0$, the corresponding t-statistic is

$$\hat{t} = \frac{c^T\hat{\beta} - c_0}{\widehat{\mathrm{se}}(c^T\hat{\beta})} = \frac{c^T\hat{\beta} - c_0}{\sqrt{n^{-1}c^T\hat{\mathcal{I}}_n^{-1}c}},$$

where $\hat{\mathcal{I}}_n = -\tilde{s}'(\tilde{\beta})$ is the estimator of the information matrix. Under the null hypothesis that $c^T\beta = c_0$,

$$\hat{t} = \frac{c^T\sqrt{n}(\hat{\beta} - \beta)}{\sqrt{c^T\hat{\mathcal{I}}_n^{-1}c}}. \quad \textbf{(A.18)}$$

Under our conditions, $c^T\hat{\mathcal{I}}_n^{-1}c = c^T[-\tilde{s}'(\beta_0)]^{-1}c + O_p(n^{-1/2})$. Because $\bar{\mathcal{I}} = -\bar{s}'(\beta_0) = \mathcal{I} + O_p(n^{-1/2})$, the expansions of $[c^T\hat{\mathcal{I}}_n^{-1}c]^{-1/2}$ and $\sqrt{n}(\hat{\beta} - \beta)$ can be written as

$$[c^T\hat{\mathcal{I}}_n^{-1}c]^{-1/2} = [c^T\mathcal{I}^{-1}c]^{-1/2} - \frac{1}{2}[c^T\mathcal{I}^{-1}c]^{-3/2}c^T\mathcal{I}^{-1}\Delta_H\mathcal{I}^{-1}c + \text{higher order terms}$$

(A.19)

and

$$\sqrt{n}(\hat{\beta} - \beta) = [\mathcal{I}^{-1} + \mathcal{I}^{-1}\Delta_H \mathcal{I}^{-1}]\sqrt{n}\tilde{s}(\beta_0) + \text{higher order terms}, \tag{A.20}$$

where Δ_H is defined by (11). The expansions of Δ_H and $\sqrt{n}\tilde{s}(\beta_0)$ are given in the proof of Theorem 1. Substituting these expansions and (A.19) and (A.20) into (A.18), it can be verified that, after dropping higher order terms,

$$\begin{aligned}
\hat{t} &= [c^T\mathcal{I}^{-1}c]^{-1/2}c^T\mathcal{I}^{-1}\left(\sum_{i=0}^{11} M_i\right) + [c^T\mathcal{I}^{-1}c]^{-1/2}c^T\mathcal{I}^{-1}\left(\sum_{i=1}^{16} Z_i\right)\mathcal{I}^{-1}\left(\sum_{i=0}^{10} M_i\right) \\
&\quad - \frac{1}{2}[c^T\mathcal{I}^{-1}c]^{-3/2}\left[c^T\mathcal{I}^{-1}\left(\sum_{i=1}^{16} Z_i\right)\mathcal{I}^{-1}c\right]c^T\mathcal{I}^{-1}\left(\sum_{i=0}^{11} M_i\right) + \text{higher order terms} \\
&= [c^T\mathcal{I}^{-1}c]^{-1/2}c^T\mathcal{I}^{-1}\sqrt{n}\bar{s}(\beta_0) + [c^T\mathcal{I}^{-1}c]^{-1/2}c^T\mathcal{I}^{-1}M_{11} \\
&\quad + [c^T\mathcal{I}^{-1}c]^{-1/2}c^T\mathcal{I}^{-1}Z_{16}\mathcal{I}^{-1}M_0 - \frac{1}{2}[c^T\mathcal{I}^{-1}c]^{-3/2}[c^T\mathcal{I}^{-1}Z_{16}\mathcal{I}^{-1}c]c^T\mathcal{I}^{-1}M_0 \\
&\quad + [c^T\mathcal{I}^{-1}c]^{-1/2}c^T\mathcal{I}^{-1}M_1 + [c^T\mathcal{I}^{-1}c]^{-1/2}c^T\mathcal{I}^{-1}Z_1\mathcal{I}^{-1}M_0 \\
&\quad - \frac{1}{2}[c^T\mathcal{I}^{-1}c]^{-3/2}c^T\mathcal{I}^{-1}Z_1\mathcal{I}^{-1}cc^T\mathcal{I}^{-1}M_0 \\
&\quad + [c^T\mathcal{I}^{-1}c]^{-1/2}c^T\mathcal{I}^{-1}M_2 + \text{higher order terms},
\end{aligned}$$

where M_j and Z_j are defined in the proof of Theorem 1. Thus the t-statistic can be expanded as the sum of a leading term, $[c^T\mathcal{I}^{-1}c]^{-1/2}c^T\mathcal{I}^{-1}\sqrt{n}\bar{s}(\beta_0)$, and

$$\begin{aligned}
\tau_t &= [c^T\mathcal{I}^{-1}c]^{-1/2}c^T\mathcal{I}^{-1}M_{11} + [c^T\mathcal{I}^{-1}c]^{-1/2}c^T\mathcal{I}^{-1}Z_{16}\mathcal{I}^{-1}M_0 \\
&\quad - \frac{1}{2}[c^T\mathcal{I}^{-1}c]^{-3/2}[c^T\mathcal{I}^{-1}Z_{16}\mathcal{I}^{-1}c]c^T\mathcal{I}^{-1}M_0 \\
&\quad + [c^T\mathcal{I}^{-1}c]^{-1/2}c^T\mathcal{I}^{-1}M_1 + [c^T\mathcal{I}^{-1}c]^{-1/2}c^T\mathcal{I}^{-1}Z_1\mathcal{I}^{-1}M_0 \\
&\quad - \frac{1}{2}[c^T\mathcal{I}^{-1}c]^{-3/2}c^T\mathcal{I}^{-1}Z_1\mathcal{I}^{-1}cc^T\mathcal{I}^{-1}M_0 \\
&\quad + [c^T\mathcal{I}^{-1}c]^{-1/2}c^T\mathcal{I}^{-1}M_2.
\end{aligned}$$

The term τ_t can be decomposed into a trimming effect

$$\begin{aligned}
t_r &= [c^T\mathcal{I}^{-1}c]^{-1/2}c^T\mathcal{I}^{-1}M_{11} + [c^T\mathcal{I}^{-1}c]^{-1/2}c^T\mathcal{I}^{-1}Z_{16}\mathcal{I}^{-1}M_0 \\
&\quad - \frac{1}{2}[c^T\mathcal{I}^{-1}c]^{-3/2}[c^T\mathcal{I}^{-1}Z_{16}\mathcal{I}^{-1}c]c^T\mathcal{I}^{-1}M_0
\end{aligned}$$

and the nonparametric estimation biases and variances effect

$$\begin{aligned}
&[c^T\mathcal{I}^{-1}c]^{-1/2}c^T\mathcal{I}^{-1}M_1 + [c^T\mathcal{I}^{-1}c]^{-1/2}c^T\mathcal{I}^{-1}Z_1\mathcal{I}^{-1}M_0 \\
&\quad - \frac{1}{2}[c^T\mathcal{I}^{-1}c]^{-3/2}c^T\mathcal{I}^{-1}Z_1\mathcal{I}^{-1}cc^T\mathcal{I}^{-1}M_0 \\
&\quad + [c^T\mathcal{I}^{-1}c]^{-1/2}c^T\mathcal{I}^{-1}M_2,
\end{aligned}$$

which is mean zero, and it can be verified that

$$
\begin{aligned}
E(\tau_t - t_r)^2 = h^{2q}\{&[c^T\mathcal{I}^{-1}c]^{-1}(c^T\mathcal{I}^{-1}\mathcal{M}_1\mathcal{I}^{-1}c) \\
&+ [c^T\mathcal{I}^{-1}c]^{-1}(c^T\mathcal{I}^{-1}\mathcal{M}_3\mathcal{I}^{-1}\mathcal{M}_3\mathcal{I}^{-1}c) \\
&- \frac{3}{4}[c^T\mathcal{I}^{-1}c]^{-2}(c^T\mathcal{I}^{-1}\mathcal{M}_3\mathcal{I}^{-1}c)^2 \\
&+ 2[c^T\mathcal{I}^{-1}c]^{-1}(c^T\mathcal{I}^{-1}\mathcal{M}_2\mathcal{I}^{-1}\mathcal{M}_3\mathcal{I}^{-1}c) \\
&- [c^T\mathcal{I}^{-1}c]^{-2}(c^T\mathcal{I}^{-1}\mathcal{M}_2\mathcal{I}^{-1}c)(c^T\mathcal{I}^{-1}\mathcal{M}_3\mathcal{I}^{-1}c)\} \\
&+ n^{-1}h^{-3}[c^T\mathcal{I}^{-1}c]^{-1}(c^T\mathcal{I}^{-1}\mathcal{S}_1\mathcal{I}^{-1}c),
\end{aligned}
$$

giving the result in Theorem 2. ■

Proof of Theorem 3. The steps of expansions for the adaptive estimator of models (1) and (17) are very similar to those in the previous section. Notice that x_i^* and ε_j are i.i.d. and are mutually independent and that the analysis for the x_i^* part is basically the same as those given earlier in this Appendix. The only thing that is substantially different from the previous section lies on the part of $\sum_{k=1}^{\infty}\Psi_k\varepsilon_{t-k}$, which has serial correlation. Specifically, the expansion (A.17) holds with the same included terms and the same magnitude of approximation error. The two main differences arise in the properties of M_1 and M_2.

Note that $(f'(\varepsilon_t)/f(\varepsilon_t))x_t$ is a martingale difference sequence because

$$
E\left[\frac{f'(\varepsilon_t)}{f(\varepsilon_t)}\right] = 0
$$

for any regular density f. Therefore,

$$
E(M_0M_0^T) = \frac{1}{n}\sum_t E\left[\frac{f'(\varepsilon_t)}{f(\varepsilon_t)}x_t\right]^{\otimes 2} = I(f)E(x_tx_t^T)
$$

by stationarity and the assumption about the process x. Furthermore, $Z_1 = E(Z_1) + o_p(n^{-1/2})$, just as in the i.i.d. case.

Note that

$$
M_1 = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{B_i'}{f_i} - \frac{f_i'B_i}{f_i^2}\right\}x_iG_i\{1+o(1)\} = h_n^q\mu_q(K)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta(\varepsilon_i)x_iG_i\{1+o(1)\},
$$

where

$$
\eta(\varepsilon_i) = \frac{f^{(q+1)}(\varepsilon_i)}{f(\varepsilon_i)} - \frac{f'(\varepsilon_i)f^{(q)}(\varepsilon_i)}{f^2(\varepsilon_i)}
$$

is not a martingale, because $E[\eta(\varepsilon_i)] \neq 0$. However, when f is symmetric about zero it is a martingale because it is an odd function. In any case, $E(M_1) = 0$ and M_1 satisfies a central limit theorem for mixing processes. Indeed,

$$
\begin{aligned}
E(M_1M_1^T) &= h_n^{2q}\mu_q^2(K)E\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta(\varepsilon_i)x_iG_i\right]^{\otimes 2} \\
&\approx h_n^{2q}\mu_q^2(K)E\left[\frac{1}{n}\sum_{i,k=1}^{n}\eta(\varepsilon_i)x_i\eta(\varepsilon_k)x_k^T\right],
\end{aligned}
$$

where

$$E\left[\frac{1}{n}\sum_{i=1}^{n}\eta^2(\varepsilon_i)x_i x_i^T\right] = E[\eta^2(\varepsilon_i)]\Omega_x,$$

$$\begin{aligned}
E\left[\frac{1}{n}\sum\sum_{i\neq k}\eta(\varepsilon_i)x_i\eta(\varepsilon_k)x_k^T\right] &= \frac{2}{n}\sum_{i=1}^{n-1}\sum_{r=1}^{n-i}E[\eta(\varepsilon_i)x_i\eta(\varepsilon_{i+r})x_{i+r}^T] \\
&= \frac{2}{n}\sum_{i=1}^{n-1}\sum_{r=1}^{n-i}\sum_{j=0}^{\infty}\sum_{\ell=0}^{\infty}\Psi_j\Psi_\ell^T \\
&\quad\times E[\eta(\varepsilon_i)\varepsilon_{i-1-j}\eta(\varepsilon_{i+r})\varepsilon_{i+r-1-\ell}] \\
&= \frac{2}{n}\sum_{i=1}^{n-1}\sum_{r=1}^{n-i}\sum_{\ell=0}^{\infty}\Psi_{r+\ell}\Psi_\ell^T E[\eta(\varepsilon_i)]E[\eta(\varepsilon_{i+r})]E[\varepsilon_{i-1-\ell}^2] \\
&= 2E^2[\eta(\varepsilon_i)]\sigma_\varepsilon^2\frac{1}{n}\sum_{i=1}^{n-1}\left\{\sum_{r=1}^{n-i}\sum_{\ell=0}^{\infty}\Psi_{r+\ell}\Psi_\ell^T\right\} \\
&= 2E^2[\eta(\varepsilon_i)]\sigma_\varepsilon^2\sum_{r=1}^{\infty}\sum_{\ell=0}^{\infty}\Psi_{r+\ell}\Psi_\ell^T + o(1).
\end{aligned}$$

Similarly, we can show that

$$\begin{aligned}
E(M_1M_0^T) &= \frac{h_n^q}{n}\mu_q(K)\sum_{i=1}^{n}\sum_{k=1}^{n}E\left[\eta(\varepsilon_i)\frac{f'(\varepsilon_k)}{f(\varepsilon_k)}x_iG_ix_k^TG_k\right]\{1+o(1)\} \\
&= h_n^q\mu_q(K)\Omega_xE\left\{\frac{f_i^{(q+1)}}{f_i}\frac{f_i'}{f_i} - \frac{f_i'f_i^{(q)}}{f_i^2}\frac{f_i'}{f_i}\right\} + o(h_n^q).
\end{aligned}$$

The second main departure is in the term

$$\begin{aligned}
M_2 &= \frac{1}{(n-1)\sqrt{n}}\sum\sum_{j\neq i}\xi_{ij} \\
&= \frac{1}{(n-1)\sqrt{n}}\sum_{j\neq i}\{K_{h_n}'(\varepsilon_i-\varepsilon_j) - E_iK_{h_n}'(\varepsilon_i-\varepsilon_j)\}x_iG_i/f(\varepsilon_i).
\end{aligned}$$

Note that M_2 is not necessarily mean zero because when $j < i$, $K_{h_n}'(\varepsilon_i - \varepsilon_j)$ can be correlated with x_i.

However,

$$\begin{aligned}
E(M_2) &= \frac{1}{(n-1)\sqrt{n}}\sum_{i=2}^{n}\sum_{j=1}^{i-1}E[K_{h_n}'(\varepsilon_i-\varepsilon_j)x_iG_i/f(\varepsilon_i)] \\
&\approx \frac{1}{(n-1)\sqrt{n}}\sum_{i=2}^{n}\sum_{j=1}^{i-1}\sum_{\ell=1}^{\infty}\Psi_\ell E[K_{h_n}'(\varepsilon_i-\varepsilon_j)\varepsilon_{i-\ell}/f(\varepsilon_i)] \\
&= \frac{1}{(n-1)\sqrt{n}}\sum_{i=2}^{n}\sum_{j=1}^{i-1}\Psi_{i-j}E[K_{h_n}'(\varepsilon_i-\varepsilon_j)\varepsilon_j/f(\varepsilon_i)] \\
&= \frac{E[f'(\varepsilon_i)\varepsilon_i/f(\varepsilon_i)+O(h_n^q)]}{(n-1)\sqrt{n}}\sum_{i=2}^{n}\sum_{k=1}^{n-i}\Psi_k \\
&= O(n^{-1/2}).
\end{aligned}$$

Furthermore, $\operatorname{var}(M_2)$ is the same as for the associated i.i.d. sequence (see Fan and Li, 1996). Let $M_2^* = M_2 - E(M_2)$, where

$$M_2^* = \frac{1}{(n-1)\sqrt{n}} \sum_{i=1}^{n} \sum_{j \neq i} \xi_{ij}^*,$$

where $\xi_{ij}^* = \xi_{ij} - E(\xi_{ij})$. Therefore, it can be verified that the leading term in $E(M_2^* M_2^{*T}) = (1/(n-1)^2 n) \sum_{i=1}^{n} \sum_{j \neq i} \sum_{\ell=1}^{n} \sum_{s \neq \ell} E(\xi_{ij}^* \xi_{\ell s}^{*T})$ is given as follows:

$$\frac{1}{(n-1)^2 n} \sum_{i=1}^{n} \sum_{j \neq i} E[K'_{h_n}(\varepsilon_i - \varepsilon_j)]^2 \left(\sum_{k=1}^{\infty} \Psi_k \varepsilon_{i-k} \right) \left(\sum_{s=1}^{\infty} \Psi_s^T \varepsilon_{i-s} \right) f(\varepsilon_i)^{-2}$$

$$= \frac{1}{(n-1)^2 n} \sum_{i=1}^{n} \sum_{j \neq i} \sum_{k=1}^{\infty} E[K'_{h_n}(\varepsilon_i - \varepsilon_j)]^2 (\Psi_k \Psi_k^T \varepsilon_{i-k}^2) f(\varepsilon_i)^{-2}$$

$$= \frac{1}{(n-1)^2 n} \sum_{i=1}^{n} \sum_{j \neq i} E[K'_{h_n}(\varepsilon_i - \varepsilon_j)]^2 (\Psi_k \Psi_k^T \varepsilon_{i-k}^2) f(\varepsilon_i)^{-2}.$$

In this expectation, if $j \neq i - k$, we get $\sigma_\varepsilon^2 (\sum_{k=1}^{\infty} \Psi_k \Psi_k^T)(1/(n-1)^2 n) \sum_{i=1}^{n} \times \sum_{j \neq i} E[K'_{h_n}(\varepsilon_i - \varepsilon_j)]^2 f(\varepsilon_i)^{-2}$, and for $j = i - k$, there is correlation between ε_{i-k}^2 and $K'_h(\varepsilon_i - \varepsilon_j)$, and we get

$$\frac{1}{(n-1)^2 n} \sum_{i=1}^{n} \sum_{j \neq i} E[K'_{h_n}(\varepsilon_i - \varepsilon_j)]^2 (\Psi_{i-j} \Psi_{i-j}^T \varepsilon_j^2) f(\varepsilon_i)^{-2}.$$

For the first term,

$$\sigma_\varepsilon^2 \left(\sum_{k=1}^{\infty} \Psi_k \Psi_k^T \right) \frac{1}{(n-1)^2 n} \sum_{i=1}^{n} \sum_{j \neq i} E[K'_{h_n}(\varepsilon_i - \varepsilon_j)]^2 f(\varepsilon_i)^{-2}$$

$$= \sigma_\varepsilon^2 \left(\sum_{k=1}^{\infty} \Psi_k \Psi_k^T \right) \frac{1}{(n-1)^2 n} \sum_{i=1}^{n} \sum_{j \neq i} \iint h^{-4} K' \left(\frac{\varepsilon_i - \varepsilon_j}{h} \right)^2 f(\varepsilon_i)^{-2} f(\varepsilon_i) f(\varepsilon_j) d\varepsilon_i \, d\varepsilon_j$$

$$= \sigma_\varepsilon^2 \left(\sum_{k=1}^{\infty} \Psi_k \Psi_k^T \right) \frac{1}{(n-1)^2 n} \sum_{i=1}^{n} \sum_{j \neq i} h^{-3} \iint K'(u)^2 f(\varepsilon_i)^{-1} f(\varepsilon_i - uh) du d\varepsilon_i$$

$$\sim \sigma_\varepsilon^2 \left(\sum_{k=1}^{\infty} \Psi_k \Psi_k^T \right) \frac{1}{n^3 h^3} \sum_{i=1}^{n} \sum_{j \neq i} Ef(\varepsilon)^{-1} \int K'(u)^2 du$$

$$\sim \frac{1}{nh^3} \sigma_\varepsilon^2 \left(\sum_{k=1}^{\infty} \Psi_k \Psi_k^T \right) Ef(\varepsilon)^{-1} \int K'(u)^2 du.$$

Notice that $\sum_i \sum_{j \neq i} \Psi_{i-j} \Psi_{i-j}^T = O(n)$, and it follows that

$$\frac{1}{(n-1)^2 n} \sum_{i=1}^{n} \sum_{j \neq i} E[K'_{h_n}(\varepsilon_i - \varepsilon_j)]^2 (\Psi_{i-j} \Psi_{i-j}^T \varepsilon_j^2) f(\varepsilon_i)^{-2} = O(n^{-2} h^{-3}).$$

The central limit theorem follows from Theorem 2.1 of Fan and Li (1996). ■