*Econometric Theory*, **31**, 2015, 1153–1191.
doi:10.1017/S0266466614000450

# ADAPTIVE NONPARAMETRIC REGRESSION WITH CONDITIONAL HETEROSKEDASTICITY

SAINAN JIN AND LIANGJUN SU
*Singapore Management University*

ZHIJIE XIAO
*Boston College*

In this paper, we study adaptive nonparametric regression estimation in the presence of conditional heteroskedastic error terms. We demonstrate that both the conditional mean and conditional variance functions in a nonparametric regression model can be estimated adaptively based on the local profile likelihood principle. Both the one-step Newton–Raphson estimator and the local profile likelihood estimator are investigated. We show that the proposed estimators are asymptotically equivalent to the infeasible local likelihood estimators [e.g., Aerts and Claeskens (1997) *Journal of the American Statistical Association* 92, 1536–1545], which require knowledge of the error distribution. Simulation evidence suggests that when the distribution of the error term is different from Gaussian, the adaptive estimators of both conditional mean and variance can often achieve significant efficiency over the conventional local polynomial estimators.

## 1. INTRODUCTION

We consider the following regression model:

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

where $\varepsilon_i$ is independent and identically distributed (IID hereafter) with mean zero and variance one, $X_i$ is a $d \times 1$ IID independent variable, and $m(\cdot)$ and $\sigma^2(\cdot)$ are assumed to be unknown smooth functions. For simplicity, we will assume that $\varepsilon_i$ is independent of $X_i$ and has an unknown density function. We are interested in estimating the infinite dimensional parameters $m(\cdot)$ and $\sigma^2(\cdot)$ adaptively in the sense that they are asymptotically equivalent to the infeasible likelihood estimators which require knowledge of the error distribution.

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

© Cambridge University Press 2014 

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

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

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

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

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

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

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

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

## 2. THE MODEL AND ESTIMATOR

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

### 2.1. The Multiparameter Likelihood Model

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

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

For a kernel function $K$ and a bandwidth parameter $h$, let $K_h(\cdot) \equiv K(\cdot/h)/h^d$. Following the notation of Masry (1996a, 1996b), let $N_l \equiv (l+d-1)!/(l!(d-1)!)$ be the number of distinct $d$-tuples $\mathbf{j}$ with $|\mathbf{j}| = l$. Arrange the $N_l$ $d$-tuples as a sequence in a lexicographical order (with highest priority to last position so that $(0, 0, \ldots, l)$ is the first element in the sequence and $(l, 0, \ldots, 0)$ is the last element), and let $\phi_l^{-1}$ denote this one-to-one map. For each $\mathbf{j}$ with $0 \le |\mathbf{j}| \le 2p$,

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

let $\mu_{\mathbf{j}}(K) \equiv \int_{\mathbb{R}^d} u^{\mathbf{j}} K(u)du, \nu_{\mathbf{j}}(K) \equiv \int_{\mathbb{R}^d} u^{\mathbf{j}} K^2(u)du$, and define the $N \times N$ dimensional matrices $M$ and $\Gamma$ and $N \times N_{p+1}$ matrix $B$, where $N \equiv \sum_{l=0}^{p} N_l$, by

$$M \equiv \begin{bmatrix} M_{0,0} & M_{0,1} & \dots & M_{0,p} \\ M_{1,0} & M_{1,1} & \dots & M_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ M_{p,0} & M_{p,1} & \dots & M_{p,p} \end{bmatrix}, \qquad \Gamma \equiv \begin{bmatrix} \Gamma_{0,0} & \Gamma_{0,1} & \dots & \Gamma_{0,p} \\ \Gamma_{1,0} & \Gamma_{1,1} & \dots & \Gamma_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{p,0} & \Gamma_{p,1} & \dots & \Gamma_{p,p} \end{bmatrix},$$

$$B \equiv \begin{bmatrix} M_{0,p+1} \\ M_{1,p+1} \\ \vdots \\ M_{p,p+1} \end{bmatrix}, \tag{2.1}$$

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

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

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

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

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

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

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

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

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

where $\beta_{1s}^0(x) = m^{(s)}(x)$, and $\beta_{2s}^0(x) = a^{(s)}(x)$ for $s = 0, 1, \ldots, p$ where $a(x) = \varphi^{-1}\left(\sigma^2(x)\right)$, $\varphi^{-1}(\cdot)$ is the inverse function of $\varphi(\cdot)$, and $a^{(s)}$ is analogously defined as $m^{(s)}(x)$ with $a(x)$ replacing $m(x)$. In particular, $\beta_{10}^0(x) = m(x)$ and $\beta_{20}^0(x) = \varphi^{-1}(\sigma^2(x))$. We will frequently suppress the dependence of $\beta_r(x)$ and $\beta_r^0(x)$ on $x$ for $r = 1, 2$. Let $P_i\left(\beta_r\right) \equiv \sum_{0 \le |\mathbf{j}| \le p} \beta_{r\mathbf{j}}(x)(X_i - x)^{\mathbf{j}}$. Apparently, $P_i\left(\beta_1^0(x)\right)$ and $P_i\left(\beta_2^0(x)\right)$ are $p$-th order Taylor expansions of $m(X_i)$ and $\varphi^{-1}(\sigma^2(X_i))$ around $x$, respectively.

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

$$\mathcal{L}_n\left(\beta_1, \beta_2\right) \equiv \frac{1}{n}\sum_{i=1}^n \log f\left(Y_i; P_i\left(\beta_1\right), P_i\left(\beta_2\right)\right) K_h(x - X_i), \tag{2.3}$$

with respect to $(\beta_1^\top, \beta_2^\top) = (\beta_{10}, \beta_{11}^\top, \ldots, \beta_{1p}^\top, \beta_{20}, \beta_{21}^\top, \ldots, \beta_{2p}^\top)$.

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

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

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

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

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

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

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

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

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

Assumptions A1 and A2 are parallel to the Assumptions A1 and A2 in Linton and Xiao (2007). The main difference is that our Assumption A2 is stronger than theirs. In addition to the conditions on the score function $\psi(\cdot)$ for the conditional mean parameter, we also impose conditions on the score function for the conditional variance parameter that is associated with $\rho(\cdot)$. It is easy to verify that the normal or Student's $t$ distributions with degrees of freedom larger than three will satisfy A2(i). A2(ii) ensures the positive definiteness of certain information matrix. A3 mainly specifies conditions on the density of $X_i$ and the smoothness of functions of interest. A4 and A5 impose conditions on the kernel function and bandwidth parameter, respectively.

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

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

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

PROPOSITION 2.1. *Suppose that Assumptions A1–A5 hold. Then*

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

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

Proposition 2.1 complements existing results in the literature. Staniswalis (1989) studies the local constant ($p = 0$) estimation of a single location parameter by maximizing a kernel-weighted likelihood function. Fan et al. (1995) discuss the local polynomial estimator of a single location parameter when the error density belongs to a one-parameter exponential family. Aerts and Claeskens (1997) study the local polynomial estimation in multiparameter-likelihood models where $d = 1$ and $X_i$ is nonrandom. Claeskens and Van Keilegom (2003) study the construction of confidence bands via local polynomial estimation. One can apply their method to construct the confidence bands for both the conditional mean and variance function estimates in our framework. In addition, under appropriate conditions, the local MLE is equivalent to nonparametric estimation by treating the locally weighted score functions as estimating equations. So we can obtain the solution to the local MLE through solving the estimating equations. See Claeskens and Aerts (2000) for such a local polynomial estimation setup.

If $\varphi(u) = u$, then $\varphi'\left(\beta_2^0(x)\right) = 1$ and Proposition 2.1 holds with $\varphi'\left(\beta_2^0(x)\right)$ being replaced by 1 in the definition of $\mathcal{I}_{\beta^0}(x)$. If, in addition, the density $f$ of $\varepsilon_i$ is symmetric about zero, then $\psi(\cdot)$ is an odd equation, implying that

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

$E[\psi(\varepsilon)\rho(\varepsilon)] = 0$ under Assumption A2. In this case, $\mathcal{I}_{\beta^0}^{-1}(x)$ is a diagonal matrix and thus the estimation of conditional variance is not affected by the estimation of conditional mean. In particular, we have

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

and

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

where $[C]_1$ denotes the first element of $C$, and $[A]_{i,i}$ signifies the $(i,i)$th element of $A$. This indicates that in the case of symmetric error density, the asymptotic biases of $\overline{m}(x)$ and $\overline{\sigma}^2(x)$ are the same as those of the local polynomial estimators of $m(x)$ and $\sigma^2(x)$ when they are estimated separately (e.g., Ruppert et al., 1997; Fan and Yao, 1998). The asymptotic variance of $\overline{m}(x)$ is smaller than that of the conventional local polynomial estimator of $m(x)$, which is $\sigma^2(x)\left[M^{-1}\Gamma M^{-1}\right]_{1,1}/f_X(x)$. Similarly, the asymptotic variance of $\overline{\sigma}^2(x)$ is smaller than that of the conventional local polynomial estimator of $\sigma^2(x)$, which is $\sigma^4(x)E\left(\varepsilon_1^2 - 1\right)^2\left[M^{-1}\Gamma M^{-1}\right]_{1,1}/f_X(x)$. In the special case where the error term $\varepsilon$ is proportional to a Student's $t$ random variable, let $\varepsilon^*$ have the Student's $t$ distribution with $\gamma > 2$ degrees of freedom, we can normalize $\varepsilon^*$ so that $\varepsilon = \sqrt{(\gamma-2)/\gamma}\,\varepsilon^*$ has variance one. Table 1 lists the relative asymptotic efficiency ratio of $\left(\overline{m}(x), \overline{\sigma}^2(x)\right)$ over the conventional local polynomial estimators of $\left(m(x), \sigma^2(x)\right)$ in terms of asymptotic variance for the approximately $t(\gamma)$ distributed $\varepsilon$. Smaller number means that larger asymptotic gain can be achieved by using the local likelihood approach. Table 1 indicates that a large efficiency gain can be achieved by using the local likelihood approach when the error distribution is far away from normality.

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

**TABLE 1.** Relative efficiency of the local likelihood estimators over the conventional local polynomial estimators

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

In this case, Proposition 2.1 implies that

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

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

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

where

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

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

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

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

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

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

and the smoothed information matrix

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

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

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

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

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

## 2.2. The Adaptive Nonparametric Regression Estimators

To obtain a feasible analog of $\bar{\boldsymbol{\beta}}(x)$, we need to replace $f$ in (2.11) by a nonparametric estimate, say, $\tilde{f}$. Since $\tilde{f}$ appears as a random denominator, it causes technical difficulty when it is small. For this reason, we propose to trim out small values of $\tilde{f}$ as do Bickel (1982), Kreiss (1987), and Linton and Xiao (2007). In particular, we consider the following smoothed trimming, which has been used by Andrews (1995), Ai (1997), and Linton and Xiao (2007). Let $g(\cdot)$ be a density function with support [0,1], $g(0) = g(1) = 0$. Let $b$ be the trimming parameter that goes to zero at a certain rate as $n \to \infty$. Define $g_b(z) = \frac{1}{b} g\left(\frac{z}{b} - 1\right)$. Clearly, $g_b(z)$ has support on $[b, 2b]$. Defining $G_b(z) = \int_{-\infty}^{z} g_b(t)\, dt$, we have

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

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

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

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

2. Obtain a consistent estimator for the error density and its derivatives by the leave-one-out kernel method:

$$\tilde{f}_i(e_i) \equiv \frac{1}{nh_0}\sum_{j\neq i} k_0\left(\frac{e_i-\tilde{\varepsilon}_j}{h_0}\right), \quad \text{and} \quad \tilde{f}_i'(e_i) \equiv \frac{1}{nh_0^2}\sum_{j\neq i} k_0'\left(\frac{e_i-\tilde{\varepsilon}_j}{h_0}\right), \tag{2.13}$$

where $k_0$ and $h_0$ are the univariate kernel and bandwidth parameter, respectively.

3. Define the trimmed and recentered local score function

$$\tilde{S}_n\left(\boldsymbol{\beta};\tilde{f}\right) \equiv \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\tilde{s}_{n,i}(\boldsymbol{\beta})\otimes\tilde{\mathbf{X}}_i, \tag{2.14}$$

and the trimmed local information matrix

$$\tilde{I}_n\left(\boldsymbol{\beta};\tilde{f}\right) \equiv \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) G_b\left(\tilde{f}_i(\varepsilon_i(\boldsymbol{\beta}))\right)\left[\tilde{s}_i(\boldsymbol{\beta})\tilde{s}_i(\boldsymbol{\beta})^{\top}\right]\otimes(\tilde{\mathbf{X}}_i\tilde{\mathbf{X}}_i^{\top}), \tag{2.15}$$

where $\tilde{s}_{n,i}(\boldsymbol{\beta}) \equiv G_b(\tilde{f}_i(\varepsilon_i(\boldsymbol{\beta})))\tilde{s}_i(\boldsymbol{\beta}) + \tilde{s}_i^*(\tilde{\boldsymbol{\beta}})$,

$$\tilde{s}_i(\boldsymbol{\beta}) \equiv \begin{pmatrix}\tilde{q}_{1,i}(\boldsymbol{\beta})\\ \tilde{q}_{2,i}(\boldsymbol{\beta})\end{pmatrix} \equiv -\begin{pmatrix}\frac{\tilde{\psi}_i(\varepsilon_i(\boldsymbol{\beta}))}{\varphi(P_i(\boldsymbol{\beta}_2))^{1/2}}\\ \frac{\varphi'(P_i(\boldsymbol{\beta}_2))\tilde{\rho}(\varepsilon_i(\boldsymbol{\beta}))}{2\varphi(P_i(\boldsymbol{\beta}_2))}\end{pmatrix}, \tag{2.16}$$

$$\begin{aligned}\tilde{s}_i^*(\boldsymbol{\beta}) &\equiv \begin{pmatrix}\tilde{q}_{1,i}^*(\boldsymbol{\beta})\\ \tilde{q}_{2,i}^*(\boldsymbol{\beta})\end{pmatrix} \\ &\equiv -\log\left(\frac{\tilde{f}_i(\varepsilon_i(\boldsymbol{\beta}))}{\varphi\left(P_i\left(\boldsymbol{\beta}_2\right)\right)^{1/2}}\right) g_b\left(\tilde{f}_i(\varepsilon_i(\boldsymbol{\beta}))\right)\tilde{f}_i'(\varepsilon_i(\boldsymbol{\beta}))\begin{pmatrix}\frac{1}{\varphi(P_i(\boldsymbol{\beta}_2))^{1/2}}\\ \frac{\varphi'(P_i(\boldsymbol{\beta}_2))\varepsilon_i(\boldsymbol{\beta})}{2\varphi(P_i(\boldsymbol{\beta}_2))}\end{pmatrix},\end{aligned} \tag{2.17}$$

and $\tilde{\psi}_i(\cdot) \equiv \tilde{f}_i'(\cdot)/\tilde{f}_i(\cdot)$, $\tilde{\rho}(\varepsilon_i(\boldsymbol{\beta})) = \left[\tilde{\psi}_i(\varepsilon_i(\boldsymbol{\beta}))\varepsilon_i(\boldsymbol{\beta}) + 1\right]$. The proposed one-step adaptive estimator can then be calculated by

$$\hat{\boldsymbol{\beta}}(x) \equiv \tilde{\boldsymbol{\beta}}(x) + \tilde{I}_n\left(\tilde{\boldsymbol{\beta}}(x);\tilde{f}\right)^{-1}\tilde{S}_n\left(\tilde{\boldsymbol{\beta}}(x);\tilde{f}\right). \tag{2.18}$$

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

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

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

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

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

and that

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

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

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

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

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

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

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

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

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

## 3. THE MAIN RESULTS

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

### 3.1. The One-Step Newton–Raphson Estimator

To proceed, we add the following assumptions.

**A 6.** (i) The kernel $k_0$ has compact support and is symmetric about zero and satisfies $\int k_0(u)\,du = 1$, $\int u^j k_0(u)\,du = 0$ for $j = 1, \ldots, p$, and $\int u^{p+1} k_0(u)\,du \neq 0$. (ii) $k_0$ is three times differentiable on its support with $k'(0) = 0$. In addition, $k_0^{(j)}(u)$, $j = 1, 2, 3$, is Lipschitz continuous and $\left| k_0'(u)\,u \right|$ is uniformly bounded.

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

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

In this paper we only focus on the case where the error density $f$ has an unbounded support in order to apply some uniform convergence results for the

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

kernel estimates of density function and its derivatives (e.g., Hansen, 2008) and avoid the well-known boundary bias problems for kernel density estimates in the case of compact support. Linton and Xiao (2007) also consider the case of bounded support for the error density where special attention is needed. In particular, they assume that the density $f$ vanishes at the boundary at a sufficiently fast rate so that the properties of regular density estimation can hold.

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

THEOREM 3.1. *Suppose that Assumptions A1–A7 hold. Then*

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

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

## 3.2. The LPL Estimator

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

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

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

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

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

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

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

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

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

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

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

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

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

Note that $\overline{\mathcal{I}}_{\beta^0}(x)$ is given by (2.7). So Theorem 3.3 indicates the asymptotic equivalence between $\boldsymbol{\beta}^+(x)$ and $\hat{\boldsymbol{\beta}}(x)$ when $\varphi(u) = \exp(u)$ is used to obtain the latter estimate. As expected, the local profile likelihood estimators $m^+(x)$ and

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

$\sigma^{+2}(x)$ share the same asymptotic properties of the one-step NR adaptive estimators $\hat{m}(x)$ and $\hat{\sigma}^2(x)$ when $\varphi(u) = \exp(u)$. Despite the need to estimate certain density derivatives, the computation for the NR adaptive estimators is not very heavy. By contrast, even though the local profile likelihood estimates $m^+(x)$ and $\sigma^{+2}(x)$ only require estimation of the density function, they are computationally more demanding because certain optimization routine is needed. In the next section, we shall evaluate the finite sample performance of these estimators in Matlab.

## 4. MONTE CARLO SIMULATIONS

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

### 4.1. Data Generating Processes

We generate data from (1.1) with different specifications of the conditional mean and variance functions and different choices of the error distributions. In all cases, the regressor $X_i$ is independently and uniformly distributed on $[-2, 2]$. The conditional mean and variance functions are specified as follows:

DGP 1: $m(x) = 1 + x + x^2$, $\sigma^2(x) = 0.1 + x^2$;
DGP 2: $m(x) = 1 + x + x^2$, $\sigma^2(x) = \exp(2x)$;
DGP 3: $m(x) = 4\sin(x)$, $\sigma^2(x) = 0.1 + x^2$;
DGP 4: $m(x) = 4\sin(x)$, $\sigma^2(x) = \exp(2x)$.

For the error term $\varepsilon_i$, we consider two distributions: $t(4.1)$ and Beta$(2, 3)$. Note that the $t(4.1)$ distribution is symmetric around zero with variance 41/21 and the Beta $(2, 3)$ distribution is asymmetric with mean 2/5 and variance 1/25. For each case, we first generate $\varepsilon_i^*$ independently according to the specified distribution and then normalize it to have mean zero and variance one (e.g., let $\varepsilon_i = 5\left(\varepsilon_i^* - 2/5\right)$ for the Beta(2,3) case). In DGPs 1–4, we consider cases where the error term $\varepsilon_i^*$ are IID $t(4.1)$. DGPs 5–8 are specified as DGPs 1–4, respectively, but with $\varepsilon_i^*$ being generated from Beta $(2, 3)$.[2]

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

DGP 9: $m(x_1, x_2) = 1 + (x_1 + x_2)^2$, $\sigma^2(x_1, x_2) = 0.5\left(1 + x_1^2 + x_2^2\right)$;
DGP 10: $m(x_1, x_2) = 1 + (x_1 + x_2)^2$, $\sigma^2(x_1, x_2) = \exp(x_1 + x_2 - 0.5)$;
DGP 11: $m(x_1, x_2) = \cos(x_1 + x_2) + x_1^2 + 2x_2^2$, $\sigma^2(x_1, x_2) = 0.5\left(1 + x_1^2 + x_2^2\right)$;
DGP 12: $m(x_1, x_2) = \cos(x_1 + x_2) + x_1^2 + 2x_2^2$, $\sigma^2(x_1, x_2) = \exp(x_1 + x_2 - 0.5)$.

We generate the IID error terms $\varepsilon_i$ from normalized $\chi^2(6)$ distribution, namely, $\varepsilon_i = \left(\varepsilon_i^* - 6\right)/\sqrt{12}$ where $\varepsilon_i^*$ are IID $\chi^2(6)$.

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

### 4.2. Implementation

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

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

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

### 4.3. Results

Tables 2 and 3 report the results for estimating the regression mean and variance functions at $x = -1.2, -0.6, 0, 0.6$, and 1.2, respectively for DGPs 1–8. Table 2

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

| DGP | $x$ | Variance | | | MSE | | | Efficiency ratio | |
|---|---|---|---|---|---|---|---|---|---|
| | | LL (1) | NR (2) | LPL (3) | LL (4) | NR (5) | LPL (6) | $\frac{(5)}{(4)}$ | $\frac{(6)}{(4)}$ |
| 1 | −1.2 | 0.049 | 0.050 | 0.053 | 0.073 | 0.067 | 0.072 | 0.924 | 0.989 |
| | −0.6 | 0.015 | 0.017 | 0.017 | 0.055 | 0.036 | 0.045 | 0.648 | 0.816 |
| | 0 | 0.007 | 0.007 | 0.007 | 0.047 | 0.024 | 0.029 | 0.519 | 0.626 |
| | 0.6 | 0.015 | 0.017 | 0.017 | 0.052 | 0.032 | 0.039 | 0.618 | 0.754 |
| | 1.2 | 0.051 | 0.050 | 0.054 | 0.071 | 0.065 | 0.073 | 0.914 | 1.031 |

*Table continues on overleaf*

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

**TABLE 2.** *continued*

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

suggests that both the one-step NR adaptive estimator and the LPL estimator generally have lower MSE than the LL estimators for the regression mean. Somewhat surprisingly, in terms of MSE the efficiency gains for the LPL estimators are not

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press



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

| DGPs | $x$ | Variance | | | | MSE | | | | Efficiency ratio | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | LL(1) | ET(2) | NR(3) | LPL(4) | LL(5) | ET(6) | NR(7) | LPL(8) | $\frac{(6)}{(5)}$ | $\frac{(7)}{(5)}$ | $\frac{(8)}{(5)}$ |
| 1 | −1.2 | 2.178 | 2.233 | 0.590 | 0.663 | 2.198 | 2.264 | 0.695 | 0.725 | 1.030 | 0.316 | 0.330 |
| | −0.6 | 0.466 | 0.135 | 0.055 | 0.061 | 0.513 | 0.135 | 0.064 | 0.064 | 0.263 | 0.124 | 0.124 |
| | 0 | 0.036 | 0.037 | 0.012 | 0.029 | 0.083 | 0.051 | 0.015 | 0.036 | 0.608 | 0.177 | 0.426 |
| | 0.6 | 0.098 | 0.091 | 0.062 | 0.070 | 0.130 | 0.091 | 0.070 | 0.072 | 0.700 | 0.542 | 0.558 |
| | 1.2 | 1.048 | 0.741 | 0.490 | 0.474 | 1.074 | 0.786 | 0.607 | 0.547 | 0.732 | 0.565 | 0.509 |
| 2 | −1.2 | 0.011 | 0.007 | 0.002 | 0.004 | 0.015 | 0.007 | 0.002 | 0.004 | 0.478 | 0.147 | 0.235 |
| | −0.6 | 0.090 | 0.087 | 0.030 | 0.049 | 0.117 | 0.087 | 0.032 | 0.050 | 0.743 | 0.272 | 0.424 |
| | 0 | 0.694 | 0.687 | 0.345 | 0.420 | 0.905 | 0.687 | 0.366 | 0.424 | 0.760 | 0.405 | 0.469 |
| | 0.6 | 5.185 | 4.808 | 3.366 | 3.544 | 6.986 | 4.890 | 3.796 | 3.742 | 0.700 | 0.543 | 0.536 |
| | 1.2 | 101.78 | 56.50 | 36.08 | 34.72 | 117.19 | 57.00 | 39.49 | 36.49 | 0.486 | 0.337 | 0.311 |
| 3 | −1.2 | 2.188 | 2.235 | 0.476 | 0.562 | 2.227 | 2.244 | 0.564 | 0.599 | 1.008 | 0.253 | 0.269 |
| | −0.6 | 0.469 | 0.136 | 0.056 | 0.088 | 0.541 | 0.140 | 0.061 | 0.089 | 0.259 | 0.113 | 0.165 |
| | 0 | 0.038 | 0.041 | 0.012 | 0.026 | 0.094 | 0.055 | 0.015 | 0.033 | 0.590 | 0.160 | 0.350 |
| | 0.6 | 0.099 | 0.092 | 0.059 | 0.075 | 0.151 | 0.093 | 0.065 | 0.076 | 0.617 | 0.430 | 0.502 |
| | 1.2 | 1.060 | 0.728 | 0.526 | 0.548 | 1.099 | 0.744 | 0.606 | 0.587 | 0.677 | 0.552 | 0.534 |
| 4 | −1.2 | 0.013 | 0.009 | 0.003 | 0.006 | 0.028 | 0.017 | 0.004 | 0.010 | 0.612 | 0.149 | 0.342 |
| | −0.6 | 0.092 | 0.090 | 0.028 | 0.042 | 0.138 | 0.094 | 0.029 | 0.043 | 0.681 | 0.208 | 0.309 |
| | 0 | 0.709 | 0.714 | 0.312 | 0.397 | 0.935 | 0.714 | 0.339 | 0.404 | 0.763 | 0.363 | 0.432 |
| | 0.6 | 5.192 | 4.817 | 2.864 | 3.238 | 7.129 | 4.876 | 3.406 | 3.487 | 0.684 | 0.478 | 0.489 |
| | 1.2 | 102.19 | 56.67 | 34.60 | 32.97 | 117.95 | 57.04 | 38.32 | 34.86 | 0.484 | 0.325 | 0.296 |

*Table continues on overleaf*

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press



**TABLE 3.** *continued*

| DGPs | $x$ | Variance | | | | MSE | | | | Efficiency ratio | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | LL(1) | ET(2) | NR(3) | LPL(4) | LL(5) | ET(6) | NR(7) | LPL(8) | $\frac{(6)}{(5)}$ | $\frac{(7)}{(5)}$ | $\frac{(8)}{(5)}$ |
| 5 | −1.2 | 0.135 | 0.174 | 0.163 | 0.170 | 0.143 | 0.214 | 0.215 | 0.201 | 1.492 | 1.501 | 1.402 |
| | −0.6 | 0.021 | 0.024 | 0.022 | 0.022 | 0.060 | 0.024 | 0.026 | 0.022 | 0.402 | 0.423 | 0.370 |
| | 0 | 0.005 | 0.011 | 0.006 | 0.007 | 0.051 | 0.027 | 0.011 | 0.017 | 0.529 | 0.220 | 0.328 |
| | 0.6 | 0.024 | 0.024 | 0.020 | 0.021 | 0.060 | 0.024 | 0.024 | 0.022 | 0.396 | 0.399 | 0.356 |
| | 1.2 | 0.147 | 0.201 | 0.196 | 0.200 | 0.157 | 0.249 | 0.262 | 0.244 | 1.590 | 1.676 | 1.556 |
| 6 | −1.2 | 0.001 | 0.001 | 0.001 | 0.001 | 0.005 | 0.002 | 0.001 | 0.001 | 0.327 | 0.227 | 0.250 |
| | −0.6 | 0.009 | 0.010 | 0.009 | 0.011 | 0.031 | 0.010 | 0.010 | 0.011 | 0.343 | 0.341 | 0.351 |
| | 0 | 0.101 | 0.124 | 0.091 | 0.101 | 0.308 | 0.125 | 0.106 | 0.103 | 0.408 | 0.344 | 0.334 |
| | 0.6 | 1.260 | 1.379 | 1.243 | 1.202 | 3.281 | 1.420 | 1.472 | 1.268 | 0.433 | 0.449 | 0.387 |
| | 1.2 | 13.27 | 15.71 | 14.32 | 14.93 | 23.63 | 16.73 | 16.90 | 15.66 | 0.708 | 0.715 | 0.663 |
| 7 | −1.2 | 0.138 | 0.147 | 0.154 | 0.168 | 0.159 | 0.157 | 0.180 | 0.176 | 0.983 | 1.128 | 1.107 |
| | −0.6 | 0.022 | 0.025 | 0.021 | 0.023 | 0.083 | 0.029 | 0.022 | 0.023 | 0.348 | 0.265 | 0.279 |
| | 0 | 0.005 | 0.013 | 0.007 | 0.009 | 0.061 | 0.027 | 0.010 | 0.016 | 0.441 | 0.162 | 0.258 |
| | 0.6 | 0.026 | 0.029 | 0.022 | 0.028 | 0.085 | 0.032 | 0.026 | 0.028 | 0.380 | 0.309 | 0.326 |
| | 1.2 | 0.143 | 0.164 | 0.155 | 0.165 | 0.160 | 0.177 | 0.177 | 0.181 | 1.106 | 1.108 | 1.132 |
| 8 | −1.2 | 0.002 | 0.003 | 0.002 | 0.002 | 0.016 | 0.011 | 0.004 | 0.007 | 0.687 | 0.242 | 0.434 |
| | −0.6 | 0.010 | 0.015 | 0.011 | 0.012 | 0.049 | 0.017 | 0.011 | 0.013 | 0.349 | 0.229 | 0.260 |
| | 0 | 0.105 | 0.133 | 0.097 | 0.102 | 0.332 | 0.135 | 0.123 | 0.107 | 0.407 | 0.370 | 0.323 |
| | 0.6 | 1.277 | 1.485 | 1.130 | 1.194 | 3.453 | 1.511 | 1.501 | 1.275 | 0.438 | 0.435 | 0.369 |
| | 1.2 | 12.97 | 15.21 | 13.15 | 14.00 | 23.58 | 16.00 | 16.66 | 15.25 | 0.679 | 0.707 | 0.647 |

as large as the case of the one-step NR adaptive estimators. For the estimation of the variance functions, Table 3 suggests that the conventional LL estimator is typically outperformed by the ET estimator, which is in turn outperformed by the NR and LPL estimators. Exceptions may occur when $x$ moves toward the boundary points.

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

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

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

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

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

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

## 5. CONCLUSIONS

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

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

## NOTES

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

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

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

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

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

## REFERENCES


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

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

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

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

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

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

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

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

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

Claeskens, G. & I. Van Keilegom (2003) Bootstrap confidence bands for regression curves and their derivatives. *Annals of Statistics* 31, 1852–1884.

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

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

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

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


https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

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

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

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

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

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

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

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

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

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

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

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

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

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

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

# APPENDIX

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

## A. Notation and Some Preliminary Results

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

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

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

$$\bar{\varepsilon}_i \equiv \varepsilon_i\left(\boldsymbol{\beta}^0\right) = \frac{Y_i - P_i\left(\boldsymbol{\beta}_1^0\right)}{\sqrt{\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)}} = \frac{u_i + \delta_i}{\sqrt{\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)}}, \quad \text{and}$$

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

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

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

Let $\tilde{f}_i^{(s)}(e_i) = \frac{1}{nh_0^{s+1}}\sum_{j\neq i} k_0^{(s)}\left(\frac{e_i - \tilde{\varepsilon}_j}{h_0}\right)$, $s = 0,1,2,3$, where $k_0^{(0)} = k_0$. We first study the uniform consistency of $\tilde{f}_i^{(s)}\left(\overrightarrow{\tilde{\varepsilon}}_i\right)$ with $f^{(s)}(\varepsilon_i)$ by the following two lemmas.

LEMMA A.1. *Suppose that Assumptions A1–A7 hold. Then*

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

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

LEMMA A.2. *Suppose that Assumptions A1–A7 hold. Then*

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

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

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

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

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

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

## B. Proofs of the Main Results

**B.1 Proof of Proposition 2.1.** Let $\alpha_n = 1/\sqrt{nh^d}$ and $\boldsymbol{\beta}^* \equiv (\boldsymbol{\beta}_1^{*\top}, \boldsymbol{\beta}_2^{*\top})^\top$ with $\boldsymbol{\beta}_r^* \equiv \sqrt{nh^d} H \left(\boldsymbol{\beta}_r - \boldsymbol{\beta}_r^0\right)$ for $r = 1, 2$. If $\bar{\boldsymbol{\beta}} \equiv (\bar{\boldsymbol{\beta}}_1^\top, \bar{\boldsymbol{\beta}}_2^\top)^\top$ maximizes (2.3), then $\bar{\boldsymbol{\beta}}^* \equiv (\bar{\boldsymbol{\beta}}_1^{*\top}, \bar{\boldsymbol{\beta}}_2^{*\top})^\top$ maximizes

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

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

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

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

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

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

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

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

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

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

By Taylor series expansions,

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

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

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

$$
A_n \equiv \begin{pmatrix} A_{n,11} & A_{n,12} \\ A_{n,21} & A_{n,22} \end{pmatrix},
$$

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

By Taylor series expansions,

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

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

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

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

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

In addition, by Taylor expansions and Assumption A2,

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

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

**B.2 Proof of Theorem 3.1.** For notational simplicity, denote $\tilde{I}_n(\boldsymbol{\beta}(x); \tilde{f})$ as $\tilde{I}_n(\boldsymbol{\beta})$ and $\tilde{S}_n(\boldsymbol{\beta}(x); \tilde{f})$ as $\tilde{S}_n(\boldsymbol{\beta})$. We frequently suppress the dependence of $\hat{\boldsymbol{\beta}}(x)$, $\tilde{\boldsymbol{\beta}}(x)$, $\boldsymbol{\beta}(x)$, etc., on $x$. Denote $G_b(\tilde{f}_i(\varepsilon_i(\tilde{\boldsymbol{\beta}})))$ as $\tilde{G}_i$ and $G_b(f(\bar{\varepsilon}_i))$ as $G_i$, where $\bar{\varepsilon}_i \equiv \varepsilon_i\left(\boldsymbol{\beta}^0\right)$. Write

$$\sqrt{nh^d}\begin{pmatrix} H(\hat{\beta}_1 - \beta_1^0) \\ H(\hat{\beta}_2 - \beta_2^0) \end{pmatrix} = \sqrt{nh^d}\begin{pmatrix} H(\tilde{\beta}_1 - \beta_1^0) \\ H(\tilde{\beta}_2 - \beta_2^0) \end{pmatrix} + \sqrt{nh^d} \bar{H} \tilde{I}_n(\tilde{\boldsymbol{\beta}})^{-1} \tilde{S}_n(\tilde{\boldsymbol{\beta}}).$$

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

$$\bar{S}_n\left(\boldsymbol{\beta},\tilde{\boldsymbol{\beta}}\right) \equiv \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\tilde{s}_{n,i}\left(\boldsymbol{\beta}\right)\otimes\tilde{\mathbf{X}}_i,$$

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

$$\bar{R}_{2n}\left(\boldsymbol{\beta}^*,\tilde{\boldsymbol{\beta}}\right) \equiv \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\tilde{G}_i R_i\left(\boldsymbol{\beta}^*\right)\otimes\tilde{\mathbf{X}}_i,$$

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

It follows that

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

It suffices to prove the theorem by showing that

$$\bar{H}^{-1}\left[\tilde{I}_n(\tilde{\boldsymbol{\beta}})-I_n\left(\boldsymbol{\beta}^0,f\right)\right]\bar{H}^{-1} = O_p\left(h^\epsilon\right) \text{ for some } \epsilon>0, \tag{B.1}$$

$$\sqrt{nh^d}\left[\bar{H}^{-1}\bar{S}_n(\boldsymbol{\beta}^0;\tilde{\boldsymbol{\beta}})-S_n\left(\boldsymbol{\beta}^0,f\right)\right] = o_p\left(1\right), \tag{B.2}$$

$$\bar{H}^{-1}\bar{R}_{1n}(\boldsymbol{\beta}^0,\tilde{\boldsymbol{\beta}})\bar{H}^{-1} = o_p\left(1\right), \tag{B.3}$$

$$\sqrt{nh^d}\bar{H}^{-1}\bar{R}_{2n}(\boldsymbol{\beta}^*,\tilde{\boldsymbol{\beta}}) = o_p\left(1\right). \tag{B.4}$$

### B.2.1 Proof of (B.1)

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

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

and a typical element of $\bar{H}^{-1}\tilde{I}_n(\tilde{\boldsymbol{\beta}})\bar{H}^{-1}$ is $\frac{1}{nh^d}\sum_{i=1}^n K_{ix,\mathbf{j}}\tilde{G}_i\tilde{q}_{r,i}(\tilde{\boldsymbol{\beta}})\tilde{q}_{s,i}(\tilde{\boldsymbol{\beta}})$, where $K_{ix,\mathbf{j}}$ is defined in (A.1), $r,s=1,2$, and $0\le|\mathbf{j}|\le 2p$. It suffices to show $T_{n\mathbf{j}}(r,s)\equiv\frac{1}{nh^d}\sum_{i=1}^n K_{ix,\mathbf{j}}\{\tilde{G}_i\tilde{q}_{r,i}(\tilde{\boldsymbol{\beta}})\tilde{q}_{s,i}(\tilde{\boldsymbol{\beta}})-q_r(Y_i;P_i(\boldsymbol{\beta}_1^0),P_i(\boldsymbol{\beta}_2^0))q_s(Y_i;P_i(\boldsymbol{\beta}_1^0),P_i(\boldsymbol{\beta}_2^0))\}=O_p\left(h^\epsilon\right)$ for $r,s=1,2$, and $0\le|\mathbf{j}|\le 2p$. Noting that $\tilde{q}_r\tilde{q}_s-q_rq_s=q_r(\tilde{q}_s-q_s)+q_s(\tilde{q}_r-q_r)+(\tilde{q}_r-q_r)(\tilde{q}_s-q_s)$, we have $T_{n\mathbf{j}}(r,s)=T_{1n\mathbf{j}}(r,s)+T_{2n\mathbf{j}}(r,s)+T_{3n\mathbf{j}}(r,s)$, where

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

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

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

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

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

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

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

### B.2.2 The proof of (B.2)

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

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

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

$$S_{n\mathbf{j},2}=\frac{1}{\sqrt{nh^d}}\sum_{i=1}^n K_{ix,\mathbf{j}}\left\{q_{2,i}\left(\boldsymbol{\beta}^0\right)[\tilde{G}_i-1]-\log\left(f(\bar{\varepsilon}_i)\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)\right)^{-1/2}\right.$$
$$\left.\times\, g_b(f(\bar{\varepsilon}_i))f'(\bar{\varepsilon}_i)\frac{\bar{\varepsilon}_i\varphi'\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)}{2\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)}\right\}$$

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

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

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

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

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

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

### B.2.3 The proof of (B.3)

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

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

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

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

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

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

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

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

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

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

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

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

Recall that $\varepsilon_i$ is independent of $X_i$ and has PDF $f(\cdot)$. So the conditional density of $Y_i^\dagger$ given $X_i$ is $f_x^\dagger(\boldsymbol{\beta}; Y_i, X_i) \equiv f\left(\varepsilon_i^\dagger(\boldsymbol{\beta})\right)/\sqrt{\exp\left(P_i\left(\boldsymbol{\beta}_2\right)\right)}$, where

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

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

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

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

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

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

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

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

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

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

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

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

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

### B.2.4 The proof of (B.4)

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

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

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

we can show that uniformly in $\boldsymbol{\beta} \in \mathcal{B}_0$, $||\tilde{G}_i \partial^2 \tilde{q}_{r,i}(\boldsymbol{\beta}) / (\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top)|| = O_p(1 + b^{-1} v_{3n,3} + b^{-2} v_{3n,2} + b^{-3} v_{3n,1}) = o_p(\sqrt{nh^d})$. ■

**B.3 Proof of Theorem 3.2.** Let $\tilde{Q}_n(x, \boldsymbol{\beta})$ and $Q(x, \boldsymbol{\beta})$ be as defined in (2.21) and (3.3). By White (1994, Thm. 3.4), it suffices to show

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

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

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

We first show (B.6). Write

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

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

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

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

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

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

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

Similarly,

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

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

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

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

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

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

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

**B.4 Proof of Theorem 3.3.** Given Assumption A3, an element-by-element mean value expansion of $\partial\tilde{Q}_n(x, \boldsymbol{\beta}^+)/\partial\boldsymbol{\beta}$ about $\boldsymbol{\beta}^0$ gives

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

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

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

The proof is completed by showing that

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

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

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

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

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

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

By a geometric expansion, we can write

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

where

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

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

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

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

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

and

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

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

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

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

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

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

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

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

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

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

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

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

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

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

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

$$\begin{aligned}
&\frac{1}{\sqrt{nh^d}}\sum_{i=1}^n \begin{pmatrix} \psi(\bar{\varepsilon}_i)/\sqrt{\exp\left(P_i\left(\beta_2^0\right)\right)} \\ \frac{1}{2}\left[\psi(\bar{\varepsilon}_i)\bar{\varepsilon}_i + 1\right] \end{pmatrix} \\
&\quad \otimes \mathbf{Z}_i g_b(f_i)\frac{1}{nh_0^2}\sum_{j\neq i} k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\frac{m(X_j) - \tilde{m}(X_j)}{\sigma(X_j)} K_{ix} \\
&\quad + \frac{1}{\sqrt{nh^d}}\sum_{i=1}^n \begin{pmatrix} \psi(\bar{\varepsilon}_i)/\sqrt{\exp\left(P_i\left(\beta_2^0\right)\right)} \\ \frac{1}{2}\left[\psi(\bar{\varepsilon}_i)\bar{\varepsilon}_i + 1\right] \end{pmatrix} \\
&\quad \otimes \mathbf{Z}_i g_b(f_i)\frac{1}{nh_0^2}\sum_{j\neq i} k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\varepsilon_j \frac{\sigma(X_j) - \tilde{\sigma}(X_j)}{\sigma(X_j)} K_{ix} \\
&\quad + \frac{1}{\sqrt{nh^d}}\sum_{i=1}^n \begin{pmatrix} \psi(\bar{\varepsilon}_i)/\sqrt{\exp\left(P_i\left(\beta_2^0\right)\right)} \\ \frac{1}{2}\left[\psi(\bar{\varepsilon}_i)\bar{\varepsilon}_i + 1\right] \end{pmatrix} \otimes \mathbf{Z}_i g_b(f_i)\mathcal{V}(\bar{\varepsilon}_i) K_{ix} \\
&\quad + \frac{1}{\sqrt{nh^d}}\bar{H}^{-1}\sum_{i=1}^n \begin{pmatrix} \psi(\bar{\varepsilon}_i)/\sqrt{\exp\left(P_i\left(\beta_2^0\right)\right)} \\ \frac{1}{2}\left[\psi(\bar{\varepsilon}_i)\bar{\varepsilon}_i + 1\right] \end{pmatrix} \otimes \tilde{\mathbf{X}}_i g_b(f_i)\mathcal{B}(\bar{\varepsilon}_i) K_{ix} + o_p(1) \\
&\equiv \mathcal{J}_{12a} + \mathcal{J}_{12b} + \mathcal{J}_{12c} + \mathcal{J}_{12d} + o_p(1),
\end{aligned}$$

where

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

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

For $\mathcal{J}_{12a}$ and $\mathcal{J}_{12b}$, we can write them as the sum of a third order $U$-statistic and a term that is asymptotically negligible. Using the standard theory for third order

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

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

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

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

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

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press

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

$$\frac{1}{2^r\sqrt{nh^d}}\bar{H}^{-1}\sum_{i=1}^{n}K_{ix,\mathbf{j}}f^{-1}(\bar{\varepsilon}_i)\frac{1}{nh_0^3}\sum_{j\neq i}k_0''\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\frac{e_1^\top M_n^{-1}(X_j)U_{1,n}(X_j)}{\sigma(X_j)}\bar{\varepsilon}_i^r$$
$$\times\left[\exp\left(P_i\left(\beta_2^0\right)\right)\right]^{-\frac{2-r}{2}}G_i+\frac{1}{2^r\sqrt{nh^d}}\bar{H}^{-1}\sum_{i=1}^{n}K_{ix,\mathbf{j}}f^{-1}(\bar{\varepsilon}_i)$$
$$\times\frac{1}{nh_0^3}\sum_{j\neq i}k_0''\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\frac{e_1^\top M_n^{-1}(X_j)B_{1,n}(X_j)}{\sigma(X_j)}\bar{\varepsilon}_i^r\left[\exp\left(P_i\left(\beta_2^0\right)\right)\right]^{-\frac{2-r}{2}}G_i.$$

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

$$\frac{1}{2^r\sqrt{nh^d}}\bar{H}^{-1}\sum_{i=1}^{n}K_{ix,\mathbf{j}}f^{-1}(\bar{\varepsilon}_i)\left\{\frac{1}{nh_0^2}\sum_{j\neq i}\left\{k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)-E_j\left[k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\right]\right\}\right\}$$
$$\times\bar{\varepsilon}_i^r\left[\exp\left(P_i\left(\beta_2^0\right)\right)\right]^{-\frac{2-r}{2}}G_i+\frac{1}{2^r\sqrt{nh^d}}\bar{H}^{-1}\sum_{i=1}^{n}K_{ix,\mathbf{j}}f^{-1}(\bar{\varepsilon}_i)$$
$$\times\left\{\frac{1}{nh_0^2}\sum_{j\neq i}E_j\left[k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\right]-f'(\bar{\varepsilon}_i)\right\}\bar{\varepsilon}_i^r\left[\exp\left(P_i\left(\beta_2^0\right)\right)\right]^{-\frac{2-r}{2}}G_i.$$

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

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

$$-\frac{1}{2^r\sqrt{nh^d}}\bar{H}^{-1}\sum_{i=1}^{n}K_{ix,\mathbf{j}}f^{-1}(\bar{\varepsilon}_i)\tilde{f}_i'(\bar{\varepsilon}_i)\left(\tilde{f}_i(\bar{\varepsilon}_i)-f(\bar{\varepsilon}_i)\right)\bar{\varepsilon}_i^r\left[\exp\left(P_i\left(\beta_2^0\right)\right)\right]^{-\frac{2-r}{2}}$$
$$G_b\left(\tilde{f}_i(\bar{\varepsilon}_i)\right)$$
$$\simeq-\frac{1}{2^r\sqrt{nh^d}}\bar{H}^{-1}\sum_{i=1}^{n}K_{ix,\mathbf{j}}f^{-1}(\bar{\varepsilon}_i)\tilde{f}_i'(\bar{\varepsilon}_i)\left(\tilde{f}_i(\bar{\varepsilon}_i)-f(\bar{\varepsilon}_i)\right)\bar{\varepsilon}_i^r$$
$$\times\left[\exp\left(P_i\left(\beta_2^0\right)\right)\right]^{-\frac{2-r}{2}}G_b(f_i),$$

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

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

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

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

https://doi.org/10.1017/S0266466614000450 Published online by Cambridge University Press