Local Polynomial Fitting Based on Empirical Likelihood

Jian Zhang
EURANDOM, Den Dolech 2, 5612 AZ, Eindhoven
and Institute of Systems Science
The Chinese Academy of Sciences, Beijing 100080

Anna Liu
Department of Statistics and Applied Probability
University of California, Santa Barbara
California 93106-3110

September 1, 2000

Abstract

In this paper, a new nonparametric regression technique is proposed by extending the local polynomial fitting to the empirical likelihood context where the distribution of the stochastic error is not fully specificied. The aim of this extension is to reduce the possible modeling bias of parametric likelihood and to allow one to use the auxiliary information about the stochastic error in the local polynomial fitting. The asymptotic bias and variance, consistency and asymptotic distribution of the proposed estimators are established. The proposed estimators are shown to inherit the main advantage of the local polynomial estimator based on the parametric likelihood over the Nadaraya-Watson kernel estimator near the boundaries. Moreover, the proposed estimators can be more flexible and efficient than the parametric likelihood based local polynomial estimator when the distribution of the stochastic error is misspecified. The new method is illustrated with applications to some simulated and real data sets.

Key words and Phrases: Nonparametric regression, local polynomial, empirical likelihood.
AMS 1991 subject classification: 62G07, 62G10, 62J12.
Running title: Local empirical likelihood

1 Introduction

The method of empirical likelihood, introduced by Owen (1988), is commonly employed to deal with the possible modeling bias of parametric likelihood. In this paper, a new estimator for a nonparametric function is developed by incorporating such a method into the framework of local polynomial modeling. By local polynomial expansion we reduce the nonparametric function estimation problem to several parametric estimation problems. Then the empirical likelihood approach can be applied to each parametric problem. Unlike the parametric likelihood based estimators (here parametric likelihood means the likelihood based on the parametric model of the stochastic error in the regression case; see, for example, Fan and Gijbels, 1996), the new estimator only requires one to specify some conditional estimating equations rather than the full probabilistic mechanism for the observations. So it releases not only the assumptions imposed on the form of a regression function but also those imposed on the stochastic error.

To highlight the idea of our proposal, we consider the following regression model

$$Y = \theta(X) + \varepsilon$$

with response Y, covariate X, regression function θ, and stochastic error ε. Given X, ε is assumed to be symmetrically distributed, that is, $\theta(X)$ is the center of symmetry of Y. This model is just the symmetric location model when θ is restricted to a finite dimensional parametric space, which is well studied (see, e.g., Bickel et al., 1993, pp. 75 and pp. 400–405). Here we consider the nonparametric case that θ is a nonparametric function from $[0, 1]$ to R^1 with $(p+1)$ continuous derivatives. To use the information about ε, we let $0 = s_0 < s_1 < ... < s_{k_0}$ and $S_k = [s_{k-1}, s_k)$, $1 \le k \le k_0$. Set $H_k(y, \theta(x)) = I(y - \theta(x) \in S_k) - I(y - \theta(x) \in -S_k)$, $1 \le k \le k_0$, where $I(\cdot)$ is the indicator of a set. Let $H = (H_1, ..., H_{k_0})^\tau$. Then we have the conditional equations

$$E\{H_k(Y, \theta(X))\,|X\} = 0, \quad 1 \le k \le k_0 \tag{1.1}$$

for θ. Note that as $\max_{1\le k\le k_0}(s_k - s_{k-1}) \to 0$, $k_0 \to \infty$, these equations are asymptotically equivalent to the assumption that ε is symmetric. These kinds of constraints were introduced in Zhang and Gijbels (1998).

Let (x_i, y_i), $i = 1, ..., n$ be i.i.d. observations from the above model. Given $x_0 \in (0, 1)$, if we have n i.i.d. observations $y'_i, i = 1, ..., n$ with the same covariate x_0, then the conditional nonparametric likelihood at $\theta(x_0)$ is of the form $\prod_{i=1}^n p_i$ where p_i is the mass we place at point (x_0, y'_i). In practice, it is rare that we have the observations with the same covariate x_0. This problem can be solved by the local modeling technique (see, e.g., Fan and Gijbels, 1996): take all (x_i, y_i), weight the logarithm of the nonparametric likelihood in such a way that it places more emphasis on these observations with the covariates close to x_0, and at the same time approximate $\theta(x)$ in (1. 1) by its pth order Taylor expansion at x_0. More specifically, let $K(\cdot)$ be a bounded symmetric density function with support $[-1,1]$. Set $K_h(\cdot) = K(\cdot/h)/h$ and $\underline{X}(t) = (1, t, ..., t^p)^\tau$. Then the profile local polynomial empirical likelihood function at x_0 is defined as follows:

$$
\begin{aligned}
l(\beta) \;=\; & \sup\{\sum_{i=1}^n K_h(x_i - x_0) \log p_i | p_i \geq 0, 1 \leq i \leq n, \sum_{i=1}^n p_i = 1, \\
& \sum_{i=1}^n p_i \underline{H}(y_i, x_i, x_0, \beta) = 0\}
\end{aligned}
\tag{1.2}
$$

where $\otimes$ is the Kronecker product, $\beta = (\beta_0, ..., \beta_p)^\tau$, and

$$\underline{H}(y_i, x_i, x_0, \beta) = H(y_i, \underline{X}((x_i - x_0)/h)^\tau \beta) \otimes \underline{X}\,((x_i - x_0)/h)\,.$$

It is easily shown by the Lagrange multiplier method that

$$
\begin{aligned}
l(\beta) \;=\; & \sum_{i=1}^n K_h(x_i - x_0) \log[K_h(x_i - x_0) / \sum_{j=1}^n K_h(x_j - x_0)] \\
& - \sum_{i=1}^n K_h(x_i - x_0) \log\left(1 + \alpha_n(x_0, \beta)^\tau \underline{H}(y_i, x_i, x_0, \beta)\right),
\end{aligned}
$$

where $\alpha_n(x_0, \beta)$ satisfies

$$\sum_{i=1}^n K_h(x_i - x_0) \frac{\underline{H}(y_i, x_i, x_0, \beta)}{1 + \alpha_n(x_0, \beta)^\tau \underline{H}(y_i, x_i, x_0, \beta)} = 0. \tag{1.3}$$

Choose an appropriate set Θ_0. Let $\widehat{\beta} = (\widehat{\beta}_0, ..., \widehat{\beta}_p)^\tau$ be the maximum estimator over Θ_0 based on $l(\beta)$. Then the local polynomial empirical likelihood estimator of $\theta(x_0)$ is given by $\widehat{\theta}(x_0) = \widehat{\beta}_0$. Through the coefficients of the higher-order terms in the polynomial fit, $\widehat{\beta}$ also provides an estimator for the higher-order derivative $\theta^{(r)}(x_0)$, namely, $\widehat{\theta}_r(x_0) = r!\widehat{\beta}_r/h^r$.

In this paper, we study this kind of estimator under a more general set of conditional equations. Under some regularity conditions, the above estimator is proved to be consistent and asymptotically normal. The asymptotic bias and variance are also derived, which have the same performance as the parametric likelihood based local polynomial estimator near the boundaries. It is shown that the new estimator can be more flexible and efficient than the parametric likelihood based local polynomial estimator. Especially, in the setting of the symmetric location model, the new estimator is nearly adaptive with respect to the unknown density function of ε. That is, when the number of the equations in (1. 1) tends to infinity, we can estimate the regression function asymptotically equally well whether or not we know the density of ε. This implies the least squares based local polynomial estimator may be inefficient when the stochastic error is not normal. Note that the least squares based local polynomial estimator can be used under the assumption that the second moment of the stochastic error exists.

The idea of using the local polynomial fitting to the parametric likelihood based regression models appeared, for example, in Stone (1977), Cleveland (1979), Tibshirani and Hastie (1987), Fan and Gijbels (1996). Carroll, Ruppert and Welsh (1998) developed an alternative method called the local moment method. It is known that the empirical likelihood has certain advantages over the moment method (see Hanfelt and Liang, 1995, Kitamura, 1997, and Qin and Lawless, 1994). In the similar setting, Zhang and Gijbels (1998) introduced an approximate empirical likelihood for a nonparametric function and gave the global convergence rate of the corresponding maximum estimator. Unlike the above ones, our estimator is based on local weighting of logarithms of empirical likelihoods.

The remains of this paper proceed as follows. In Section 2 we investigate the asymptotic properties of the proposed estimator. Applications to both simulated and real data sets are presented in Section 3. The proofs of the main results can be found in the appendix.

2 Asymptotic Theory

In what follows, we consider a general nonparametric regression model with response Y and covariate X. Assume that the regression function $\theta(\cdot)$ has $(p+1)$ continuous derivatives. Adopt the same notations as in (1. 2) and the associated estimators but replace H (and $\underline{H}$) by a more general vector-valued function $G = (G_1, ..., G_{k_0})^\tau$ (and $\underline{G}$), which satisfies

$$E[G_k(Y, \theta(X))|X] = 0, \quad k = 1, 2, ..., k_0. \tag{2.1}$$

Note that the ordinary nonparametric mean regression model and median regression model are two particular examples if we set $G = Y - \theta(X)$ and $G = I\left(Y \le \theta(X)\right) - 1/2$, respectively. For the simplicity of the proofs, we assume that G has a continuous derivative with respect to θ below.

2.1 Estimation

Set

$$\mu_{j+l} = \int t^{j+l} K(t) dt, \quad \nu_{j+l} = \int t^{j+l} K^2(t) dt,$$
$$S = (\mu_{j+l})_{0 \le j,l \le p}, \quad S^* = (\nu_{j+l})_{0 \le j,l \le p},$$

$$\begin{aligned}
V_G(x_0) &= E[G\left(Y, \theta(x_0)\right) G^\tau \left(Y, \theta(x_0)\right) | X = x_0], \\
D_G(x_0) &= E[\partial G\left(Y, \theta(x_0)\right) / \partial \theta | X = x_0], \\
\lambda_0 &= (\theta(x_0), h\theta^{(1)}(x_0), ..., h^p \theta^{(p)}(x_0)/p!), \\
\beta(u, x_0) &= \underline{X}\left((u - x_0)/h\right)^\tau \lambda_0,
\end{aligned}$$

$$V_{\beta G}(x_0) = \frac{1}{f(x_0)}[D_G(x_0)^\tau V_G(x_0)^{-1} D_G(x_0)]^{-1} S^{-1} S^* S^{-1},$$

$$\begin{aligned} V_{\alpha G}(x_0) &= \frac{1}{f(x_0)}[V_G(x_0)^{-1} - V_G(x_0)^{-1} D_G(x_0) \left(D_G(x_0)^\tau V_G(x_0)^{-1} D_G(x_0)\right)^{-1} \\ &\quad \times D_G(x_0)^\tau V_G(x_0)^{-1}] \otimes S^{-1} S^* S^{-1}. \end{aligned}$$

Let f be the density of X. If f and $\theta^{(p+1)}$ have continuous derivatives, then define

$$\begin{aligned} \text{bias} &= h^{p+1} S^{-1} (\mu_{p+1}, ..., \mu_{2p+1})^\tau \frac{\theta^{(p+1)}(x_0)}{(p+1)!} \\ &\quad + h^{p+2} S^{-1} (\mu_{p+2}, ..., \mu_{2p+2})^\tau \\ &\quad \times \{\frac{\theta^{(p+2)}(x_0)}{(p+2)!} + \frac{\theta^{(p+1)}(x_0)}{(p+1)!} \frac{f'(x_0)}{f(x_0)}\}. \end{aligned}$$

Let $||\cdot||$ stand for the Euclidean norm and $\xrightarrow{\mathcal{L}}$ for convergence in distribution. Assume that $\lambda_0 \in \Theta_0$.

Theorems 1 and 2 below show that $\widehat{\theta}(x_0)$ is weakly consistent and asymptotically normal.

Theorem 1 *Under conditions (A1)~(A8) in the Appendix, for $2 < \alpha_1 \le \alpha_0$ (α_0 is defined in condition (A1)), as $h = h_n \to 0$, $hn^{1-2/\alpha_1}/\log n \to \infty$ and $h^{p+1} n^{1/\alpha_1} \to 0$, we have*

$$\widehat{\beta} - \lambda_0 = o_p(n^{-1/\alpha_1}), \quad \alpha_n(x_0, \widehat{\beta}) = o_p(n^{-1/\alpha_1}).$$

Theorem 2 *Suppose that conditions (A1)~(A8), (B1), and (B2) in the Appendix hold. Suppose that f and $\theta^{(p+1)}$ have continuous derivatives. Then as $h = h_n \to 0$, $hn^{1-/\alpha_0}/\log n \to \infty$ (α_0 is defined in condition (A1)) and $h^{p+1} n^{1/\alpha_0} \to 0$,*

$$\sqrt{nh} V_{\beta G}(x_0)^{-1/2} \{\widehat{\beta} - \lambda_0 - bias(1 + o(1))\} \xrightarrow{\mathcal{L}} N(0, I_{p+1}).$$

Furthermore, if $nh^{2p+3} \to 0$, then

$$\sqrt{nh} V_\alpha(x_0)^{-1/2} \widehat{\alpha}(x_0) \xrightarrow{\mathcal{L}} N(0, I_{k_0(p+1)}),$$

where I_{p+1} and $I_{k_0(p+1)}$ are the $p \times p$ and $k_0(p+1) \times k_0(p+1)$ unit matrices, and $N(0, I_{p+1})$ and $N(0, I_{k_0(p+1)})$ are normal distributions.

Remark 2.1 *The requirement that G is differentiable in θ can be relaxed by imposing some entropy condition on G (see condition (A4′) in the Appendix). Then Theorems 1 and 2 can cover the special* *example in (1. 1). For example, suppose G is bounded. Then, under the conditions (A1),(A4′), and (A5)∼(A8), as $h = h_n \to 0$, $hn/\log n \to \infty$,*

$$\hat{\beta} - \lambda_0 = o_p(1), \quad \alpha_n(x_0, \hat{\beta}) = o_p(1).$$

Furthermore, the asymptotic normality still holds if we impose the second order differentiablity on $E\{G(Y,t)|X\}$ with respect to t. Here $D_G(x_0)$ should be defined as $\partial E[G(Y,\theta(x_0))|X = x_0]/\partial t$. A rigorous justification of the statement is tedious but very similar to Zhang and Gijbels (1998) and is not pursued here.

Remark 2.2 *Surprisingly, the bias of $\hat{\beta}$ is asymptotically free of the constraint (2. 1). This leads to a simple criterion, the asymptotic covariance $V_{\beta G}(x_0)$, for the comparison of the efficiencies of the above local estimators derived from a class of constraint functions. Let $l(z, x_0) = \frac{\partial \log f_{\varepsilon|X=x_0}}{\partial z}$ where $f_{\varepsilon|X=x_0}$ is the conditional density of ε given $X = x_0$. It follows directly from Bhapkar (1991) that*

$$D_G(x_0)^\tau V_G(x_0)^{-1} D_G(x_0) \leq El(z, x_0)^2$$

for any estimating function G satisfying

$$E[G\left(Y, \theta(X)\right)|X = x_0] = 0, \quad V_G(x_0) < \infty, \quad \textit{and} \quad D_G(x_0) \quad \textit{exists}.$$

Thus

$$V_{\beta G}(x_0) \geq V_{\beta l(z,x_0)}.$$

Furthermore, in the setting of the symmetric location model mentioned in Section 1, the proposed estimator is shown to be nearly adaptive with respect to the unknown conditional density, $f_{\varepsilon|X=x_0}$, in the sense that there exists $(G^{(n)})_{n=1}^{\infty}$ such that $V_{\beta G^{(n)}}(x_0) \to V_{\beta l(z,x_0)}(x_0)$.

To see this, we calculate the asymptotic variance of $\hat{\beta}$ when $G = H$ defined in (1. 1). Under some mild regularity conditions, we have

$$\begin{aligned} D_H(x_0)^\tau V_H(x_0)^{-1} D_H(x_0) &= \sum_{k=1}^{k_0} 2 \frac{[f_{\varepsilon|X=x_0}(s_k) - f_{\varepsilon|X=x_0}(s_{k-1})]^2}{\left(F_{\varepsilon|X=x_0}(s_k) - F_{\varepsilon|X=x_0}(s_{k-1})\right)} \\ &\to El(z, x_0)^2 \end{aligned}$$

as $s_1 \to 0$, $\max_k(s_k - s_{k-1}) \to 0$ and $s_{k_0} \to \infty$, where $F_{\varepsilon|X}$ is the distribution function of the error. So

$$V_{\beta H} \to \frac{1}{f(x_0)}\{El(z, x_0)^2\}^{-1} S^{-1} S^* S^{-1}$$

which is just the asymptotic variance of the local polynomial estimator based on the local log-likelihood $\sum_{i=1}^{n} K_h(x_i - x_0) \log f_{\varepsilon|X=x_i}(y_i - \theta(x_i))$ (see Fan and Gijbels, 1996). This means that we can estimate the regression function asymptotically equally well whether or not we know the density of ε. In particular, we can construct an asymptotically better estimator than the smoothing spline estimator of θ because the latter is equivalent to a kernel regression estimator with some special kernel (see Silverman, 1984).

Note that $f_{\varepsilon|X=x_0}$ is symmetric if and only if for every $t > 0$

$$\int \sin(tz) f_{\varepsilon|X=x_0}(z)dz = 0.$$

Moreover, if $\theta_0(X)$ is the conditional symmetric center of Y given X, then

$$\int \sin(y - \theta(x_0)) f_{Y|X=x_0}(y - \theta_0(x_0)) = 0$$

if and only if $\theta(x_0) = \theta_0(x_0)$. This means we can choose a smooth and bounded G instead of $G = H$ in (2. 1). For example, for $0 < s_1 < s_2 < ... < s_{k_0} \to \infty$, let

$$G_k(y, \theta(x)) = \sin(s_k(y - \theta(x))), 1 \le k \le k_0.$$

Then these functions satisfy the equations in (2. 1). As $s_1 \to 0$, $s_{k_0} \to \infty$, and $\max_k(s_k - s_{k-1}) \to 0$, these equations are asymptotically equivalent to the requirement that $f_{\varepsilon|X=x_0}$ is symmetric for any x_0.

Remark 2.3 *Let e_{r+1} denote the unit vector with 1 in the $(r+1)$th position. Then, from Lemma 3.7 in the Appendix we obtain the asymptotic bias$\{\hat{\theta}_r(x_0)\}$ (defined as the leading term of the bias of $\hat{\theta}_r(x_0)$): For the odd $p - r$,*

$$\begin{aligned} bias\{\widehat{\theta}_r(x_0)\} &= e_{r+1}^{\tau} S^{-1}(\mu_{p+1}, ..., \mu_{2p+1})^{\tau} \frac{r!}{(p+1)!} \theta^{(p+1)}(x_0) h^{p+1-r} \\ &\quad \times (1 + o(1)); \end{aligned} \tag{2.2}$$

for the even $p - r$,

$$\begin{aligned} bias\{\widehat{\theta}_r(x_0)\} &= e_{r+1}^{\tau} S^{-1}(\mu_{p+2}, ..., \mu_{2p+2})^{\tau} \frac{r!}{(p+2)!} [\theta^{(p+2)}(x_0) \\ &\quad + (p+2)\theta^{(p+1)}(x_0) \frac{f'(x_0)}{f(x_0)}] h^{p+2-r}. \end{aligned} \tag{2.3}$$

We also have the asymptotic variance of $\hat{\theta}$:

$$Var\{\widehat{\theta}_r(x_0)\} = e_{r+1}^{\tau} S^{-1} S^* S^{-1} e_{r+1} \frac{(r!)^2 [D_G(x_0)^{\tau} V_G(x_0)^{-1} D_G(x_0)]^{-1}}{f(x_0) n h^{2r+1}} (1 + o(1)). \tag{2.4}$$

As a result of Theorem 2, we get

$$Var\{\hat{\theta}_r(x_0)\}^{-1/2} \left\{ \hat{\theta}_r(x_0) - \theta_r(x_0) - bias\{\hat{\theta}_r(x_0)\}(1 + o(1)) \right\} \xrightarrow{\mathcal{L}} N(0, 1).$$

Remark 2.4 *Although we have shown the asymptotic behavior of $\widehat{\theta}_r(x_0)$ for a general r, we are most interested in estimation of θ itself. Note that for a fixed sequence of bandwidths, the bias of a linear fit ($p = 1$) is of order h^2 [when $\theta^{(2)}(x_0) \neq 0$], the bias of a quadratic fit is of order h^4, and the bias of a cubic fit is of order h^4. So like the parametric likelihood based local polynomial fitting, when estimating θ at peaks and valleys, which means that $\theta^{(2)}(x_0) \neq 0$, there is a significant reduction in the biases for the quadratic and cubic fits compared with the linear fit, while the orders of the asymptotic variances are always $O(n^{-1}h^{-1})$.*

Remark 2.5 *Since we assume for simplicity that supp(f)=[0,1], then the left boundary points are of the form $x_0 = ch$ and the right ones are of the form $x_0 = 1 - ch$, with $c > 0$. When the kernel function K has support [-1,1], the real boundary points are those for which $c < 1$, whereas for $c > 1$ we have interior points. The asymptotic bias and variance expressions for the estimator when $x_0 = ch$ and $x_0 = 1 - ch$ are derived in a way analogous to those for interior points. Set*

$$\mu_{j,c} = \int_{-c}^{1} u^j K(u)du,$$

$$S_= (\mu_{j+l,c})_{0 \leq j,l \leq p},$$

$$K^*_{r,c}(t) = e^\tau_{r+1} S_c^{-1}(1, t, ..., t^p)^\tau K(t).$$

Then for $x_0 = ch$,

$$bias\{\widehat{\theta}_r(x_0)\} = \{\int_{-c}^{1} t^{p+1} K^*_{r,c}(t)dt\} \frac{r!}{(p+1)!} \theta^{(p+1)}(0+) h^{p+1-r}(1 + o(1)). \tag{2.5}$$

and

$$Var\{\widehat{\theta}_r(x_0)\} = \int_{-c}^{1} {K^*_{r,c}}^2(t)dt \frac{(r!)^2 [D(0+)^\tau V(0+)^{-1} D(0+)]^{-1}}{f(0+) n h^{2r+1}} (1 + o(1)). \tag{2.6}$$

For right boundary points $x_0 = 1 - ch$, the asymptotic bias and variance expressions are similar to those provided in (2. 5) and (2. 6), but with the integral interval [-c,1] replaced by [-1,c] and $0+$ by $1-$.

Remark 2.6 *The odd-degree fitting is better than the even-degree fitting. The reason is that for the even $p - r$ not only the unknown derivative $\theta^{(p+1)}(x_0)$ but also unknown $f'(x_0)$ and $\theta^{(p+1)}(x_0)$ are involved in the asymptotic bias. Moreover, for the even $p - r$ a comparison of (2. 3) and (2. 4) with (2. 5) and (2. 6) shows that the order of the asymptotic bias is different at the boundary and in the interior. In contrast, for the odd $p - r$ only $\theta^{(p+1)}(x_0)$ is unknown in the asymptotic bias. The asymptotic bias and variance are also of the same order at the boundary and in the interior. In another words, the proposed estimation procedure adapts automatically to the boundary of supp(f). This feature is parallel to that of the traditional local polynomial fitting.*

3 Numerical examples

3.1 Bandwidth selection

When we apply the local polynomial empirical likelihood estimator to a finite sample, we must first select the bandwidth. This smoothing parameter plays a very important role in the trade-off between reducing bias and variance. So we need to choose it carefully instead of randomly. There are different kinds of bandwidth selection methods (see Fan and Gijbels (1996) for details). We use the suggestion of Carroll, Ruppert and Welsh (1998) . The basic idea behind this proposal is that we view the mean squared error (MSE) as a function of h. Ideally we should choose the optimal bandwidth by minimizing the MSE function with respect to h, where

$$MSE(x_0, h) = var(x_0, h) + \mathrm{b}ias^2(x_0, h)$$

with $var(x_0, h)$ and bias(x_0, h) being the variance and bias of $\widehat{\theta}(x_0)$, respectively. In practice, the MSE is unknown and estimated by the empirical bias bandwidth selection (EBBS) method and sandwich method.

The basic idea for EBBS is as follows. For fixed x_0 and h_0, according to the asymptotic results in our asymptotic theories, $\mathrm{b}ias(x_0, h_0)$ should have a form like $\mathrm{b}ias(x_0, h_0) = \gamma_1 h_0^{p+1} + ... + \gamma_t h_0^{p+t}$, where $t \geq 1$; $\gamma = (\gamma_1, ..., \gamma_t)$ unknown. Later we shall denote it by $f(h_0, \gamma)$. The local polynomial estimator $\widehat{\theta}(x_0, h_0)$ should be well described by $\gamma_0 + f(h_0, \gamma) + o_p(h_0^{p+t})$, where $\gamma_0 = \theta(x_0)$ in the limit. Then let $(\widehat{\gamma}_0, \widehat{\gamma})$ minimize $\sum_{k=1}^K \{\widehat{\theta}(x_0, h_k) - (\widehat{\gamma}_0 + f(h_k, \widehat{\gamma}))\}^2$, in which $\{h_1, ..., h_K\}$ is a grid of bandwidths in a neighborhood H_0, of h_0 with $K \geq t+1$. It is obvious that if H_0 is small enough, the bias should be well estimated at h_0 by $f(h_0, \widehat{\gamma})$. In practice, we need to choose K and t. See Carroll, Ruppert, and Welsh (1998) for some specific selection technique. In our simulation and real data fitting, we take t=1 and K=3. We are most attracted by the EBBS property of avoiding the direct estimation of the higher-order derivatives arising in the asymptotic bias formulas, which might limit the range of applications because of its complications.

The sandwich formula for the asymptotic covariance matrix of β is analogous to that in Carroll, Ruppert, and Welsh (1998), that is,

$$\{\{\hat{D}(x_0)\}\{\hat{V}(x_0)\}^{-1}\{\hat{D}(x_0)\}\}^{-1},$$

where

$$\hat{D}(x_0) = \sum_{i=1}^n K_h(x_i - x_0)[\frac{\partial G(y_i, \underline{X}((x_i - x_0)/h)^\tau \widehat{\beta})}{\partial \theta} \otimes \underline{X}((x_i - x_0)/h)\underline{X}^\tau((x_i - x_0)/h)].$$

and

$$\begin{aligned} \hat{V}(x_0) &= \sum_{i=1}^n K_h^2(x_i - x_0)[G(y_i, \underline{X}((x_i - x_0)/h)^\tau \widehat{\beta}) G^\tau(y_i, \underline{X}((x_i - x_0)/h)^\tau \widehat{\beta}) \\ & \otimes \underline{X}((x_i - x_0)/h)\underline{X}^\tau((x_i - x_0)/h)]. \end{aligned}$$

It is easily seen from our asymptotic results that the sandwich formula provides consistent variance estimators.

3.2 Simulation

In the following examples the x_i's were generated from the uniform distribution on [0,1]. The local linear empirical likelihood fitting (i.e., $p = 1$) is used to estimate the regression functions.

Example 3.1 *The regression model is*

$$Y = 1 - 48X + 218X^2 - 315X^3 + 145X^4 + \varepsilon.$$

Given X, ε follows the t-distribution with 3 degrees of freedom and the constraint function is

$$G(y, \theta(x)) = y - \theta(x).$$

Generate a sample of size 200.

Example 3.2 *Adopt the same notations as in Example 3.1, except that we now assume that given X, ε follows the normal distribution $N(0, \sigma(X)^2), \sigma(X)^2 = 1 + X^2$. Generate a sample of size* 200.

Figures 3.1 and 3.2 show the performance of the local linear empirical likehood fitting when ε has heavy tails (Example 3.1) or when ε is heteroscedastic (Example 3.2).

3.3 Application

Example 3.3 *(Great Barrier Reef data). In a survey of the fauna on the sea bed in an area lying between the coast of northern Queensland and the Great Barrier Reef, data are collected at a number of locations. In view of the large numbers of types of species captured in the survey the response variable is expressed as a score, on a log weight scale, which combines information across species. The relationship between the catch score and the spatial coordinates, latitude and longitude was analyzed in Bowman ans Azzalini (1997, pp.53-55) via ordinary nonparametric regression. Here we use our proposed method to analyze these data. We let p=1, $\varepsilon(x) = y - \theta(x)$, and*

$$G(y, \theta(x)) = \varepsilon(x). \tag{3.1}$$

$$G(y, \theta(x)) = (\varepsilon(x), \varepsilon(x)^3)^\tau. \tag{3.2}$$

As an example, in Figure 3.3 we present the fitting results lsle, lele $- 1$, and lele $- 2$ only for the relationship between the catch score and the latitude, which are based on the least squares local linear fitting, the local linear empirical likelihood fittings with the restriction function in (3. 1), and with the restriction function in (3. 2), respectively. Note that from the proof of Theorem 2 it is *easily seen that using the least squares based local polynomial fitting is asymptotically equal to using normal likelihood based local polynomial fitting. So it is not surprising that lele − 1 is very close to lsle. However, lele − 2 is significantly different from both lele − 1 and lsle. It is natural to ask which one is better. To this aim, some goodness-of-fit tests for these restrictions are needed. The details can be found in Fan and Zhang (2000).*

Appendix: Proofs of theorems

We begin with some notations. Suppose there exists $Z(y,x)$ (independent of h) such that

$$Z(y,x) \geq \sup_{\beta\in\Theta_0} ||G(y, \underline{X}(\frac{x-x_0}{h})^\tau \beta)|| I(|x-x_0| \leq h).$$

Let $Z_i = Z(y_i, x_i), \quad 1 \leq i \leq n$. Denote

$$A_{n1}(x_0,\beta) = \frac{1}{n}\sum_{i=1}^n K_h(x_i-x_0)\underline{G}(y_i,x_i,x_0,\beta) I(Z_i \leq n^{1/\alpha_1}), \tag{3.3}$$

$$\begin{aligned} A_n(x_0,\beta) &= \frac{1}{n}\sum_{i=1}^n K_h(x_i-x_0)\underline{G}(y_i,x_i,x_0,\beta), \\ W_n(x_0,\beta) &= \frac{1}{n}\sum_{i=1}^n K_h(x_i-x_0)\underline{G}(y_i,x_i,x_0,\beta)\underline{G}^\tau(y_i,x_i,x_0,\beta), \\ W_{n1}(x_0,\beta) &= \frac{1}{n}\sum_{i=1}^n K_h(x_i-x_0)\underline{G}(y_i,x_i,x_0,\beta)\underline{G}^\tau(y_i,x_i,x_0,\beta) \\ &\quad \times I(Z_i \leq n^{1/\alpha_1}). \end{aligned}$$

To establish the consistency of $\widehat{\beta}$, we impose the following regulrarity conditions (A1)∼(A8) when $x_0 \in (0,1)$:

A1: There exists a constant c_0 such that for $x \in [0,1]$ and $x+\Delta \in [0,1]$

$$|f(x+\Delta) - f(x)| \leq c_0|\Delta|.$$

A2: For some $2 < \alpha_0 \leq \infty$,

$$\sup_{x\in[0,1]} E\{Z(Y,X)^{\alpha_0}|X=x\} < \infty.$$

Here $\alpha_0 = \infty$ means $Z(Y,X)$ is bounded by some constant.

A3: For $1 \leq j \leq k_0$, as $h = h_n \to 0$, uniformly for $\beta \in \Theta_0$ and $|t| \leq 1$,

$$E\{G_j^2(Y, \underline{X}(\frac{X-x_0}{h})^\tau \beta)|X = x_0 + th\} = O(1).$$

A4: There exists $\psi_{h1}(y,x)$ such that for $\beta_j \in \Theta_0$, $j=1,2$,

$$E\psi_{h1}(Y,X)K(\frac{X-x_0}{h}) = O(1), \quad EZ(Y,X)\psi_{h1}(Y,X)K(\frac{X-x_0}{h}) = O(1),$$

and for $|x-x_0| \le h$,

$$||G(y, \underline{X}(\frac{x-x_0}{h})^\tau \beta_1) - G(y, \underline{X}(\frac{x-x_0}{h})^\tau \beta_2|| \le \psi_{h1}(y,x)||\beta_1 - \beta_2||.$$

A5: The function θ has a $(p+1)$th continuous derivative and there exists $\psi_{h2}(x)$ such that

$$E\{K_h(X-x_0)\psi_{h2}(X)||\underline{X}(\frac{X-x_0}{h})||\} = O(1),$$

$$\begin{aligned} ||E\left\{[G(Y, \underline{X}(\frac{X-x_0}{h})^\tau \beta) - G(Y, \theta(X))]|X=x\right\}|| \quad &\le \quad \psi_{h2}(x)(||\beta - \lambda_0|| \\ &\quad + ||\beta(x,x_0) - \theta(x)||) \end{aligned}$$

for $\beta \in \Theta_0$, $|x-x_0| \le h$, where $\beta(x,x_0) = \underline{X}(\frac{x-x_0}{h})^\tau \lambda_0$.

A6: As $n \to \infty$, $h = h_n \to 0$,

$$P\{W_n(x_0,\beta) > 0, \beta \in \Theta_0\} \to 0,$$

where $W_n(x_0,\beta) > 0$ means: $W_n(x_0,\beta)$ is positive definite.

A7: For $1 \le k_1, j_1 \le k_0$, as $h = h_n \to 0$,

$$E\{G_{k_1}^2(Y, \beta(X,x_0))G_{j_1}^2(Y,\beta(X,x_0))|X = x_0 + th\} = O(1)$$

uniformly for $\beta \in \Theta_0$ and $|t| \le 1$. As $\delta \to 0$ and $h = h_n \to 0$, uniformly for $||\beta - \lambda_0|| \le \delta$,

$$||E\left\{G(Y, \underline{X}(\frac{X-x_0}{h})^\tau \beta)G^\tau(Y, \underline{X}(\frac{X-x_0}{h})^\tau \beta)|X = x_0 + th\right\} - V_G(x_0)||$$

is of order $o(1)$. Moreover, we suppose $V_G(x_0)$ and S are positive definite.

A8: For any fixed constant $\rho > 0$ there exists a positive constant $c(\rho)$ such that as $h = h_n \to 0$,

$$\inf_{||\beta - \lambda_0|| \ge \rho} ||EK_h(X-x_0)\underline{G}(Y,X,x_0,\beta)|| \ge c(\rho).$$

In addition, there exists a fixed positive constant c such that as $||\beta - \lambda_0|| + h \to 0, \beta \in \Theta_0$, $||EK_h(X-x_0)\underline{G}(Y,X,x_0,\beta)||$ is bounded below by $c||\beta - \lambda_0|| + O(h^{p+1})$, where c and $O(h^{p+1})$ are independent of β.

When G is not smooth, we need to replace condition (A4) by the following condition.

A4′: Set

$$g(y,x,\beta)=K(\frac{x-x_0}{h})G_{i_1}(y,\underline{X}(\frac{x-x_0}{h})^\tau\beta)(\frac{x-x_0}{h})^{j_1},$$

$$\mathcal{F}(i_1,j_1)=\{g(\cdot,\cdot,\beta):\beta\in\Theta_0\}.$$

$$\begin{aligned} g_1(y,x,\beta) &= K(\frac{x-x_0}{h})G_{i_1}(y,\underline{X}(\frac{x-x_0}{h})^\tau\beta)(\frac{x-x_0}{h})^{j_1} \\ &\quad \times G_{k_1}(y,\underline{X}(\frac{x-x_0}{h})^\tau\beta)(\frac{x-x_0}{h})^{s_1}, \end{aligned}$$

$$\mathcal{F}(i_1,j_1,k_1,s_1) = \{g_1(\cdot,\cdot,\beta):\beta\in\Theta_0\}.$$

For some positive constants c_1, c_2, w_1, and w_2 such that

$$\begin{aligned} N(\delta,L_2(P_n),\mathcal{F}(i_1,j_1)) &\le c_1\delta^{-w_1}, \\ N(\delta,L_2(P_n),\mathcal{F}(i_1,j_1,k_1,s_1)) &\le c_2\delta^{-w_2} \end{aligned}$$

where P_n is the empirical distribution of (x_i,y_i), $i=1,...,n$, and $N(d,L_2(P_n),\Upsilon)$ is called the covering number of Υ, which is defined in Pollard (1984).

To obtain asymptotic normality, we need two additional conditions.

B1: For the small $\delta_0>0$, there exists a function $U_1(y,x)$ satisfying

$$\begin{aligned} EK_h(X-x_0)U_1(Y,X) &= O(1), \\ EK_h(X-x_0)Z(Y,X)U_1(Y,X) &= O(1), \\ \sup_{||\beta-\lambda_0||\le\delta}||\frac{\partial\underline{G}(y,x,x_0,\beta)}{\partial\beta}||I(|x-x_0|\le h) &\le U_1(y,x). \end{aligned}$$

There exists a function $U_2(y,x)$ satisfying

$$\begin{aligned} EK_h(X-x_0)U_2(Y,X) &= O(1), \\ ||\frac{\partial\underline{G}(y,x,x_0,\beta)}{\partial\beta}-\frac{\partial\underline{G}(y,x,x_0,\lambda_0)}{\partial\beta}|| &\le U_2(y,x)||\beta-\lambda_0||. \end{aligned}$$

Furthermore,

$$\begin{aligned} EK_h(X-x_0)||\frac{\partial\underline{G}(Y,X,x_0,\lambda_0)}{\partial\beta^\tau}||^2 &= O(1), \\ EK_h(X-x_0)\frac{\partial\underline{G}(Y,X,x_0,\lambda_0)}{\partial\beta^\tau} &= f(x_0)D_G(x_0)\otimes S+o(1). \end{aligned}$$

B2: For some small $\delta_0 > 0$, there exists a function $U_3(y,x)$ such that

$$\begin{aligned} EK_h(X-x_0)U_3(Y,X) &= O(1), \\ \sup_{||\beta-\lambda_0||\leq\delta_0} ||\frac{\partial^2 \underline{G}(y,x,x_0,\beta)}{\partial\beta\partial\beta^\tau}|| &\leq U_3(y,x). \end{aligned}$$

For $U_1(y,x)$ defined in (B1)

$$EK_h(X-x_0)U_1(Y,X)^2 = O(1).$$

When G is not smooth, we need to impose some conditions similar to Zhang and Gijbels (1998). The details are not pursued here.

For $x_0 = 0$ or 1, the conditions similar to (A1)$\sim$(A8) and (B1)$\sim$(B2), denoted by the same notations, can be imposed by restricting the value of t (or $(x-x_0)/h$)) to $[0,1]$ or $[-1,0]$ in the above.

Lemmas 3.1$\sim$3.3 below will be used in the proof of Theorem 1.

Lemma 3.1 *Under conditions (A1)$\sim$(A4), for $2 < \alpha_1 \leq \alpha_0$, as $h = h_n \to 0$ and $hn^{1-2/\alpha_1}/\log n \to \infty$, there exists a sequence of constants $(d_{n1})_{n=1}^\infty$, $0 < d_{n1} \to 0$, such that uniformly for $\beta \in \Theta_0$,*

$$A_n(x_0,\beta) = EK_h(X-x_0)\underline{G}(Y,X,x_0,\beta) + o_p(n^{-1/\alpha_1})d_{n1}, \tag{3.4}$$

$$A_{n1}(x_0,\beta) = EK_h(X-x_0)\underline{G}(Y,X,x_0,\beta) + o_p(n^{-1/\alpha_1})d_{n1}. \tag{3.5}$$

Furthermore, under (A5),

$$EK_h(X-x_0)\underline{G}(Y,X,x_0,\beta) = O(h^{p+1} + ||\beta-\lambda_0||). \tag{3.6}$$

Proof. Without loss of generality, we assume $x_0 \in (0,1)$. Write $A_n(x_0,\beta)$ as

$$A_n(x_0,\beta) = A_{n1}(x_0,\beta) + A_{n2}(x_0,\beta)$$

with

$$A_{n2}(x_0,\beta) = \frac{1}{n}\sum_{i=1}^n K_h(x_i-x_0)\underline{G}(y_i,x_i,x_0,\beta)I(Z_i > n^{1/\alpha_1}).$$

It follows from (A1) and (A2) that, for $2 < \alpha_1 \leq \alpha_0$,

$$\begin{aligned} E\sup_{\beta\in\Theta_0} ||A_{n2}(x_0,\beta)|| &\leq EK_h(X-x_0)Z(Y,X)I(Z(Y,X) > n^{1/\alpha_1}) \\ &\quad \times\sqrt{p+1}f(x_0+th)dt \\ &= o(n^{-1/\alpha_1}) \end{aligned}$$

which implies

$$A_n(x_0,\beta) = EK_h(X-x_0)G(Y,X,x_0,\beta) + \frac{1}{nh}\sum_{i=1}^{n}\{f_{ni}(\beta)-Ef_{ni}(\beta)\} + o_p(n^{-1/\alpha_1}) \quad (3.7)$$

where

$$f_{ni}(\beta) = K(\frac{x_i-x_0}{h})\underline{G}(y_i,x_i,x_0,\beta)I(Z_i\le n^{1/\alpha_1}).$$

Set

$$\begin{aligned} g(y,x,\beta) &= n^{-1/\alpha_1}K(\frac{x-x_0}{h})G_{i_1}(Y,\underline{X}(\frac{x-x_0}{h})^\tau\beta) \\ &\quad \times(\frac{x-x_0}{h})^{j_1}I(Z(y,x)\le n^{1/\alpha_1}), \\ \mathcal{F}(i_1,j_1) &= \{g(\cdot,\cdot,\beta):\beta\in\Theta_0\}. \end{aligned}$$

Then by conditions (A1) and (A3) we have

$$\sup_{\beta\in\Theta_0} Eg^2(Y,X,\beta) = O(hn^{-2/\alpha_1}).$$

For $g(\beta_j)=g(y,x,\beta_j)\in\mathcal{F}(i_1,j_1)$, $\quad j=1,2$, by condition (A4), we have

$$|g(\beta_1)-g(\beta_2)| \le n^{-1/\alpha_1}K(\frac{x-x_0}{h})(\frac{x-x_0}{h})^{j_1}\psi_{h1}(y,x)||\beta_1-\beta_2||.$$

Let $u_n = hn^{1-2/\alpha_1}$ and $d_{n1}^2 = (\log n/u_n)^{1/2}$. By Lemma 7.2 in Zhang and Gijbels (1998) there exist positive constants c_j, $1\le j\le 4$, and w_0, such that for any positive constant M_0

$$\begin{aligned} &P\{\sup_{\beta\in\Theta_0}|\frac{1}{nh}\sum_{i=1}^{n}[g(y_i,x_i,\beta)-Eg(Y,X,\beta)]|\ge M_0n^{-1/\alpha_1}d_{n1}\} \\ &\le c_1(n^{1/\alpha_1}h^{-1}d_{n1})^{w_0}\exp\{-\frac{M_0^2nh^2n^{-4/\alpha_1}d_{n1}^2}{c_3hn^{-2/\alpha_1}}\} \\ &\quad + c_2(hn^{-2/\alpha_1})^{-w_0}\exp\{-c_4nhn^{-2/\alpha_1}\}. \end{aligned} \quad (3.8)$$

As $h=h_n\to 0$, $u_n/\log n\to\infty$, we have

$$\log u_n + \log n = o\{(u_n\log n)^{1/2}\} = o(u_nd_{n1}^2),$$

therefore, (3. 8) tends to zero. This together with (3. 7) completes the proofs for (3. 4) and (3. 5).

Finally, (3. 6) follows from condition (A5) and the equality

$$||EK_h(X-x_0)\underline{G}(Y,X,x_0,\beta)|| = ||E\left\{K_h(X-x_0)\Psi(\beta)\otimes\underline{X}(\frac{X-x_0}{h})\right\}||,$$

where

$$\Psi(\beta) = E[G(Y,\underline{X}(\frac{X-x_0}{h})^\tau\beta)-G(Y,\theta(X))|X].$$

The proof of Lemma 3.1 is finished.

Lemma 3.2 *Under conditions (A1) and (A2), as* $h = h_n \to 0$,

$$\sup_{\beta \in \Theta_0} ||W_{n1}(x_0, \beta)|| = O_p(1). \tag{3.9}$$

Under conditions (A1), (A2), (A4) and (A7), as $h = h_n \to 0$ *and* $nh \to \infty$,

$$W_n(x_0, \beta) = f(x_0) V_G(x_0) \otimes S + o_p(1) \tag{3.10}$$

uniformly for $||\beta - \lambda_0|| \leq \delta \to 0$.

Proof. Equation (3. 9) follows from the fact that, under conditions (A1) and (A2),

$$E \sup_{\beta \in \Theta_0} ||W_{n1}(x_0, \beta)|| \leq (p+1) E K_h(X - x_0) Z(Y, X)^2 = O(1).$$

Note that, by condition (A4),

$$\begin{aligned} ||W_n(x_0, \beta) - W_n(x_0, \lambda_0)|| &\leq \frac{1}{n} \sum_{i=1}^{n} K_h(x_i - x_0) Z_i \psi_{h1}(y_i, x_i) ||\beta - \lambda_0|| \\ &= O_p(1) ||\beta - \lambda_0||. \end{aligned}$$

In order to prove (3. 10), it suffices to show

$$W_n(x_0, \lambda_0) = f(x_0) V_G(x_0) \otimes S + o_p(1). \tag{3.11}$$

To this end, we calculate the mean and covariance of $W_n(x_0, \lambda_0)$. It is easily seen that, under conditions (A1) and (A7),

$$E W_n(x_0, \beta) = f(x_0) V_G(x_0) \otimes S + o_p(1). \tag{3.12}$$

For $k = (p+1)(k_1 - 1) + k_2$ and $j = (p+1)(j_1 - 1) + j_2$ with $1 \leq k_1, j_1 \leq k_0$, $1 \leq k_2, j_2 \leq p+1$, we obtain the variance of the $(k, j)-$th element of $W_n(x_0, \lambda_0)$ is smaller than or equal to

$$\begin{aligned} &\frac{1}{n} E K_h(X - x_0)^2 G_{k_1}^2(Y, \beta(X, x_0)) G_{k_2}^2(Y, \beta(X, x_0)) (\frac{X - x_0}{h})^{2(k_2 + j_2 - 2)} \\ = \quad &O(\frac{f(x_0)}{nh}) = o(1) \end{aligned}$$

by condition (A7). This, together with (3. 12), leads to (3. 11). The proof is completed.

Lemma 3.3 *Under conditions (A1)∼ (A7), if both* $V_G(x_0)$ *and* S *are positive definite, then, for any* $2 < \alpha_1 \leq \alpha_0$, *as* $h = h_n \to 0$, $h n^{1-2/\alpha_1} / \log n \to \infty$, $h^{p+1} n^{1/\alpha_0} = o(1)$, *there exists a sequence of constants* $(d_{n1})_{n=1}^{\infty}$, $0 < d_{n1} \to 0$, *such that*

$$\alpha_n(x_0, \beta) = o_p(n^{-1/\alpha_1}) d_{n1} + O(h^{p+1} + ||\beta - \lambda_0||)$$

uniformly for $||\beta - \lambda_0|| \leq O(n^{-1/\alpha_0})$.

Proof. Without loss of generality, we assume $x_0 \in (0,1)$. By (A2) we have

$$\max_{1\le i\le n} Z_i = o_p(n^{1/\alpha_0}).$$

It follows from Lemma 3.1 that, as $h = h_n \to 0$, $hn^{1-2/\alpha_1}/\log n \to \infty$,

$$||A_n(x_0,\beta)|| = o_p(n^{-1/\alpha_1})d_{n1} + O(h^{p+1} + ||\beta - \lambda_0||) \tag{3.13}$$

for some $0 < d_{n1} \to 0$ and uniformly for $||\beta - \lambda_0|| \le O(n^{-1/\alpha_0})$. Thus, we have

$$||A_n(x_0,\beta) \max_{1\le i\le n} Z_i\sqrt{p+1}|| = o_p(1) \tag{3.14}$$

uniformly for $||\beta - \lambda_0|| \le O(n^{-1/\alpha_0})$. It follows from Lemma 2 that there exists a positive constant c such that, as $h = h_n \to 0$, $hn \to \infty$, and $\delta \to 0$,

$$\sup\{\rho_n(x_0,\beta) : ||\beta - \lambda_0|| \le \delta\} \ge c \tag{3.15}$$

where $\rho_n(x_0,\beta)$ is the minimum eigenvalue of $W_n(x_0,\beta)$. Finally, by (3. 13), (3. 14), (3. 15) and by using the technique of Owen (1988), we have

$$\begin{aligned} ||\alpha_n(x_0,\beta)|| &\le \frac{||A_n(x_0,\beta)||}{\rho_n(x_0,\beta) - ||A_n(x_0,\beta)|| \max_{1\le i\le n} Z_i\sqrt{p+1}} \\ &= O_p(||A_n(x_0,\beta)||) \\ &= o_p(n^{-1/\alpha_1})d_{n1} + O(h^{p+1} + ||\beta - \lambda_0||) \end{aligned}$$

uniformly for $||\beta - \lambda_0|| \le O(n^{-1/\alpha_0})$. The proof is completed.

Proof of Theorem 1. Without loss of generality, we assume $x_0 \in (0,1)$. We first establish some facts. Let $2 < \alpha_1 \le \alpha_0$, $d_{n2} \ge 0$, $d_{n2}^2 = h^{p+1}n^{1/\alpha_1}$, and $d_n = \max\{d_{n1}, d_{n2}\}$, where d_{n1} is defined in Lemma 3.3. Then, by Lemmas 3.1 and 3.3, we have

$$\alpha_n(x_0,\lambda_0) = o_p(n^{-1/\alpha_1})d_n, \quad A_n(x_0,\lambda_0) = o_p(n^{-1/\alpha_1})d_n.$$

We have the first fact:

$$\begin{aligned} 0 &\ge -\frac{1}{n}\sum_{i=1}^{n} K_h(x_i - x_0)\log(1 + \alpha_n(x_0,\lambda_0)^\tau \underline{G}(y_i,x_i,x_0,\lambda_0)) \\ &\ge -\alpha_n(x_0,\lambda_0)^\tau A_n(x_0,\lambda_0) \\ &= -|o_p(n^{-2/\alpha_1})|d_n^2. \end{aligned} \tag{3.16}$$

Let $u_0 = u_0(\beta) \in R^{k_0(p+1)}$, $||u_0|| = 1$, satisfying

$$u_0||EK_h(X - x_0)\underline{G}(Y,X,x_0,\beta)|| = EK_h(X - x_0)\underline{G}(Y,X,x_0,\beta).$$

Denote

$$T_{n1} = \frac{1}{n}\sum_{i=1}^{n} K_h(x_i - x_0)\log(1 + n^{-1/\alpha_1} d_n u_0^\tau \underline{G}(y_i, x_i, x_0, \beta))I(Z_i \le n^{1/\alpha_1}).$$

Then we have

$$T_{n1} = n^{-1/\alpha_1} d_n u_0^\tau A_{n1}(x_0, \beta) - n^{-2/\alpha_1} d_n^2 W_{n1}^*(x_0, \beta). \tag{3.17}$$

Here $A_{n1}(x_0, \beta)$ is in (3. 3) and

$$W_{n1}^*(x_0, \beta) = \frac{1}{n}\sum_{i=1}^{n} K_h(x_i - x_0)\frac{1}{2(1+t_i)^2}(u_0^\tau \underline{G}(y_i, x_i, x_0, \beta))^2 I(Z_i \le n^{1/\alpha_1})$$

and for $1 \le i \le n$, t_i lies between 0 and $n^{-1/\alpha_1} d_n u_0^\tau \underline{G}(y_i, x_i, x_0, \beta)$. When $\max_i Z_i \le n^{1/\alpha_1}$, $\max_i |t_i| \le \sqrt{p+1} d_n$ uniformly in β. This leads to

$$W_{n1}^*(x_0, \beta) \le \frac{1}{2(1 - \sqrt{p+1} d_n)^2} u_0^\tau W_{n1} u_0.$$

By Lemma 2 we obtain that $W_{n1}^*(x_0, \beta)$ is uniformly bounded in $\beta \in \Theta_0$. This, in conjunction with (3. 17) and Lemma 3.1, leads to the second fact, namely that, uniformly for $\beta \in \Theta_0$,

$$\begin{aligned} T_{n1} &= n^{-1/\alpha_1} d_n u_0^\tau E\{K_h(X - x_0)\underline{G}(Y, X, x_0, \beta)\} \\ &\quad + O_p(n^{-1/\alpha_1}) d_n^2. \end{aligned} \tag{3.18}$$

Furthermore,

$$P(\max_i Z_{hi} > n^{1/\alpha_1}) = o(1). \tag{3.19}$$

Denote

$$\begin{aligned} T_n(x_0, \beta) &= \frac{1}{n}\sum_{i=1}^{n} K_h(x_i - x_0)\log(1 + \alpha_n(x_0, \beta)^\tau \underline{G}(y_i, x_i, x_0, \beta)), \\ \Xi &= \{\alpha : 1 + \alpha^\tau \underline{G}(y_i, x_i, x_0, \beta) > 0, 1 \le i \le n\}. \end{aligned}$$

Then, when

$$\frac{1}{n}\sum_{i=1}^{n} K_h(x_i - x_0)\underline{G}(y_i, x_i, x_0, \beta)\underline{G}^\tau(y_i, x_i, x_0, \beta) > 0,$$

we have

$$-T_n(x_0, \beta) = \min_{\alpha \in \Xi} -\frac{1}{n}\sum_{i=1}^{n} K_h(x_i - x_0)\log(1 + \alpha^\tau \underline{G}(y_i, x_i, x_0, \beta)). \tag{3.20}$$

Now combining the facts (3. 16), (3. 18), (3. 19), (3. 20) and condition (A8), we obtain that, for any fixed positive constant ρ, as $n \to \infty$,

$$
\begin{aligned}
&P\left\{\sup_{||\beta-\lambda_0||\geq\rho}(-T_n(x_0,\beta)) > -T_n(x_0,\lambda_0)\right\}\\
&\quad\leq P\left\{\sup_{||\beta-\lambda_0||\geq\rho}(-T_n(x_0,\beta)) > -|o_p(n^{-2/\alpha_1})|d_n^2\right\}\\
&\quad\leq P\left\{\sup_{||\beta-\lambda_0||\geq\rho}(-T_{n1}) > -|o_p(n^{-2/\alpha_1})|d_n^2\right\}\\
&\qquad+P\{\max_i Z_i > n^{1/\alpha_1}\} + o(1)\\
&\quad\leq P\left\{c\inf_{||\beta-\lambda_0||\geq\rho}||EK_h(X-x_0)\underline{G}(Y,X,x_0,\beta)|| \leq |O_p(n^{-1/\alpha_1})d_n|\right\}\\
&\qquad+o(1)
\end{aligned}
$$

which implies

$$
||\widehat{\beta}-\lambda_0|| = o_p(1). \tag{3.21}
$$

Similarly, for any constants $0<\rho_n\to 0$ and δ small enough, we have

$$
\begin{aligned}
&P\{\delta\geq||\widehat{\beta}-\lambda_0||\geq\rho_n\}\\
&\leq P\{c\inf_{\delta\geq||\beta-\lambda_0||\geq\rho_n}||\beta-\lambda_0|| + O(h^{p+1}) \leq |O_p(n^{-1/\alpha_1}d_n|\}\\
&\quad+o(1).
\end{aligned} \tag{3.22}
$$

It follows from (3. 21) and (3. 22) that

$$
\begin{aligned}
\hat{\beta}-\lambda_0 &= O_p(n^{-1/\alpha_1})d_n + O(h^{p+1})\\
&= o_p(n^{-1/\alpha_1}).
\end{aligned}
$$

Using Lemma 3.1 again, we obtain $\alpha_n(x_0,\widehat{\beta}) = o_p(n^{-1/\alpha_1})$. The proof is completed.

We now turn to some technical lemmas for the proof of Theorem 2. For this purpose, we first introduce some additional notations. Let

$$
\begin{aligned}
B_{n1}(\beta,\alpha) &= \frac{1}{n}\sum_{i=1}^n K_h(x_i-x_0)\frac{\underline{G}(y_i,x_i,x_0,\beta)}{1+\alpha^\tau\underline{G}(y_i,x_i,x_0,\beta)},\\
B_{n2}(\beta,\alpha) &= \frac{1}{n}\sum_{i=1}^n K_h(x_i-x_0)\frac{\alpha^\tau\partial\underline{G}(y_i,x_i,x_0,\beta)/\partial\beta^\tau}{1+\alpha^\tau\underline{G}(y_i,x_i,x_0,\beta)}.
\end{aligned}
$$

$$
\begin{aligned}
C_{n11}(\beta,\alpha) &= \frac{\partial B_{n1}(\beta,\alpha)}{\partial\alpha^\tau},\quad C_{n12}(\beta,\alpha) = \frac{\partial B_{n1}(\beta,\alpha)}{\partial\beta^\tau},\\
C_{n21}(\beta,\alpha) &= \frac{\partial B_{n2}(\beta,\alpha)}{\partial\alpha^\tau},\quad C_{n22}(\beta,\alpha) = \frac{\partial B_{n2}(\beta,\alpha)}{\partial\beta^\tau}.
\end{aligned}
$$

Lemma 3.4 *Under conditions (A1), (A2), (A4) and (A7), as $h = h_n \to 0$, and $nh \to \infty$, for any random vectors $\zeta_1 = \lambda_0 + o_p(1)$ and $\alpha_1 = o_p(1)$, we have*

$$C_{n11}(\zeta_1, \alpha_1) = -f(x_0) V_G(x_0) \otimes S + o_p(1).$$

Proof. Note that

$$C_{n11}(\zeta_1, \alpha_1) = -W_n(x_0, \zeta_1) + R_{n11},$$

where

$$\begin{aligned} R_{n1} &= \frac{1}{n} \sum_{i=1}^{n} K_h(x_i - x_0) \frac{\alpha_1^\tau \underline{G}(y_i, x_i, x_0, \zeta_1)(2 + \alpha_1^\tau \underline{G}(y_i, x_i, x_0, \zeta_1))}{(1 + \alpha_1^\tau \underline{G}(y_i, x_i, x_0, \zeta_1))^2} \\ &\quad \times \underline{G}(y_i, x_i, x_0, \zeta_1) \underline{G}^\tau(y_i, x_i, x_0, \zeta_1). \end{aligned}$$

Note that, under condition (A2),

$$\max_i Z_i = O_p(n^{-1/\alpha_0}),$$

which implies

$$\max_i ||\alpha_1^\tau \underline{G}(y_i, x_i, x_0, \zeta_1)|| = o_p(1)$$

by the assumption that $\alpha_1 = o_p(n^{-1/\alpha_0})$. Therefore,

$$\begin{aligned} ||R_{n11}|| &\leq \frac{(p+1)|o_p(1)|(2 + |o_p(1)|)}{(1 - |o_p(1)|)^2} \frac{1}{n} \sum_{i=1}^{n} K_h(x_i - x_0) Z_i^2 \\ &= o_p(1). \end{aligned}$$

Now by Lemma 3.2 and the assumption that $\zeta_1 = \lambda_0 + o_p(1)$ we complete the proof.

Lemma 3.5 *Under conditions (A1), (A2) and (B1), as $h = h_n \to 0$ and $nh \to \infty$, for any random vectors $\zeta_1 = \lambda_0 + o_p(1)$ and $\alpha_1 = o_p(n^{-1/\alpha_0})$, we have*

$$\begin{aligned} C_{n12}(\zeta_1, \alpha_1) &= f(x_0) D_G(x_0) \otimes S + o_p(1), \\ C_{n21}(\zeta_1, \alpha_1)^\tau &= f(x_0) D_G(x_0) \otimes S + o_p(1). \end{aligned}$$

Proof. We only need to consider $C_{n12}(\zeta_1, \alpha_1)$ because $C_{n12}(\zeta_1, \alpha_1) = C_{n21}(\zeta_1, \alpha_1)^\tau$. For simplicity, we write $\underline{G}(y_i, x_i, x_0, \zeta_1)$ as $\underline{G}_i$. Note that

$$C_{n12}(\zeta_1, \alpha_1) = D_n(\zeta_1) + R_{n12},$$

where

$$\begin{aligned} D_n(\beta) &= \frac{1}{n} \sum_{i=1}^{n} K_h(x_i - x_0) \frac{\partial \underline{G}(y_i, x_i, x_0, \beta)}{\partial \beta}, \\ R_{n12} &= -\frac{1}{n} \sum_{i=1}^{n} K_h(x_i - x_0) \frac{\alpha_1^\tau \underline{G}_i}{1 + \alpha_1^\tau \underline{G}_i} \frac{\partial \underline{G}_i}{\partial \beta^\tau} \\ &\quad -\frac{1}{n} \sum_{i=1}^{n} \frac{\underline{G}_i \alpha_1^\tau \partial \underline{G}_i / \partial \beta^\tau}{(1 + \alpha_1^\tau \underline{G}_i)^2}. \end{aligned}$$

By condition (B1), we have, as $h = h_n \to 0$ and $nh \to \infty$,

$$\begin{aligned} ||D_n(\zeta_1) - D_n(\lambda_0)|| &\leq \frac{1}{n}\sum_{i=}^{n} K_h(x_i - x_0)U_2(y_i, x_i)||\zeta_1 - \lambda_0|| \\ &= O_p(||\zeta_1 - \lambda_0||) = o_p(1) \end{aligned} \tag{3.23}$$

and

$$D_n(\lambda_0) = f(x_0)D_G(x_0) \otimes S + o_p(1). \tag{3.24}$$

Observe that under condition (A2) and the assumption that $\alpha_1 = o_p(n^{-1/\alpha_0})$, we have

$$\max_i |\alpha_1^\tau \underline{G}_i| = o_p(1)$$

which, with condition (B1), implies

$$\begin{aligned} ||R_{n12}|| &\leq \frac{|o_p(1)|}{1 - |o_p(1)|}\frac{1}{n}\sum_{i=1}^{n} K_h(x_i - x_0)U_1(y_i, x_i) \\ &\quad + \frac{|o_p(1)|}{(1 - |o_p(1)|)^2}\frac{1}{n}\sum_{i=1}^{n} Z_i U_1(y_i, x_i) \\ &= o_p(1). \end{aligned} \tag{3.25}$$

Now combining (3. 23), (3. 24) and (3. 25), we get the desired result.

Lemma 3.6 *Under conditions (A1), (A2), and (B2), as $h = h_n \to 0$, for any random vectors $\zeta_1 = \lambda_0 + o_p(1)$ and $\alpha_1 = o_p(n^{-1/\alpha_0})$, we have*

$$C_{n22}(\zeta_1, \alpha_1) = o_p(1).$$

Proof. It is similar to the proof of Lemma 3.5 and thus omitted.

Denote

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}, \quad C_{22.1} = C_{22} - C_{21}C_{11}^{-1}C_{12},$$

where

$$\begin{aligned} C_{11} &= -f(x_0)V_G(x_0) \otimes S, \quad C_{22} = 0, \\ C_{12} &= C_{21}^\tau = f(x_0)D_G(x_0) \otimes S. \end{aligned}$$

Lemma 3.7 *Suppose conditions (A1), (A4), (A7), (B1), and (B2) hold. Then, as* $h = h_n \to 0$, *we have*

$$nh\mathrm{Var}(B_{n1}(\lambda_0, 0)) = f(x_0)V_G(x_0) \otimes S^* + o(1).$$

If $\theta(x)$ *has a* $(p+1)-th$ *continuous derivative* $\theta^{(p+1)}(x)$, *then*

$$C_{22.1}^{-1}C_{21}C_{11}^{-1}EB_{n1}(\lambda_0, 0) = bias^*(1 + o(1))$$

with

$$bias^* = h^{p+1}S^{-1}(\mu_{p+1}, ..., \mu_{2p+1})^\tau \theta^{(p+1)}(x_0)/(p+1)!.$$

In addition, if f *and* $\theta^{(p+1)}(x)$ *have continuous derivatives, then*

$$C_{22.1}^{-1}C_{21}C_{11}^{-1}EB_{n1}(\lambda_0, 0) = bias(1 + o(1))$$

where bias is defined in Section 2.

Proof. Note that

$$\begin{aligned} EB_{n1}(\lambda_0, 0) &= EK_h(X - x_0)E[G(Y, \theta(X) - (X - x_0)^{p+1} \\ &\quad \times \frac{\theta^{(p+1)}(x_0)}{(p+1)!} - o_p(h^{p+1}))|X] \otimes \underline{X}(\frac{X - x_0}{h}) \\ &= -EK_h(X - x_0)E\left\{\frac{\partial G(Y, \theta(X))}{\partial \theta}|X\right\}(X - x_0)^{p+1} \\ &\quad \times \frac{\theta^{(p+1)}(x_0)}{(p+1)!} \otimes \underline{X}(\frac{X - x_0}{h}) + o_p(h^{p+1}) \\ &= -f(x_0)D_G(x_0)h^{p+1} \otimes (\mu_{p+1}, ..., \mu_{2p+1})^\tau \frac{\theta^{(p+1)}(x_0)}{(p+1)!} \\ &\quad \times(1 + o_p(1)). \end{aligned}$$

Note that K is symmetric and the $(r+1)-$th element of $S^{-1}(\mu_{p+1}, ..., \mu_{2p+1})^\tau$ is zero. To obtain the non zero bias when $p - r$ is even, we expand $EB_{n1}(\lambda_0, 0)$ up to order h^{p+2} :

$$\begin{aligned} EB_{n1}(\lambda_0, 0) &= EK_h(X - x_0)E[G(Y, \theta(X) - (X - x_0)^{p+1}\frac{\theta^{(p+1)}(x_0)}{(p+1)!} \\ &\quad -(X - x_0)^{p+2}\frac{\theta^{(p+2)}(x_0)}{(p+2)!} - o_p(h^{p+2}))|X] \otimes \underline{X}(\frac{X - x_0}{h}) \\ &= -f(x_0)D_G(x_0) \otimes [(\mu_{p+1}, ..., \mu_{2p+1})^\tau \frac{\theta^{(p+1)}(x_0)}{(p+1)!}h^{p+1} \\ &\quad +[(\mu_{p+1}, ..., \mu_{2p+1})^\tau \frac{\theta^{(p+1)}(x_0)}{(p+1)!}\frac{f'(x_0)}{f(x_0)} \\ &\quad +(\mu_{p+2}, ..., \mu_{2p+2})^\tau \frac{\theta^{(p+2)}(x_0)}{(p+2)!}]h^{p+2}](1 + o_p(1)). \end{aligned}$$

Similarly, we have

$$\begin{aligned} Cov(B_{n1}(\lambda_0, 0)) &= \frac{1}{n}[EK_h^2(X - x_0)\underline{G}(Y, X, x_0, \lambda_0)\underline{G}^\tau(Y, X, x_0, \lambda_0)) \\ &\quad -EK_h(X - x_0)\underline{G}(Y, X, x_0, \lambda_0) \\ &\quad \times EK_h(X - x_0)\underline{G}^\tau(Y, X, x_0, \lambda_0)] \\ &= \frac{f(x_0)}{nh}\{V_G(x_0) \otimes S^* + O(h^{2p+3})\}. \end{aligned}$$

The proof is completed.

Proof of Theorem 2. Write $\hat{\alpha} = \alpha_n(x_0, \widehat{\beta})$. Then applying Theorem 1, we have

$$\widehat{\beta} - \lambda_0 = o_p(1), \quad \hat{\alpha} = o_p(1),$$

which, by the assumption, implies that as n is large, $\widehat{\beta}$ is an inner point of Θ_0. Since $\widehat{\beta}$ is the maximum estimator, we have

$$B_{n1}(\widehat{\beta}, \hat{\alpha}) = 0, \quad B_{n2}(\widehat{\beta}, \hat{\alpha}) = 0.$$

By virtue of a Taylor expansion, they become

$$0 = B_{n1}(\lambda_0, 0) + C_{n11}(\zeta_1, \alpha_1)\hat{\alpha} + C_{n12}(\zeta_1, \alpha_1)(\widehat{\beta} - \lambda_0), \tag{3.26}$$

$$0 = B_{n2}(\lambda_0, 0) + C_{n21}(\zeta_1, \alpha_1)\hat{\alpha} + C_{n22}(\zeta_1, \alpha_1)(\widehat{\beta} - \lambda_0), \tag{3.27}$$

where (ζ_j, α_j), $j = 1, 2$ are between $(\widehat{\beta}, \hat{\alpha})$ and $(\lambda_0, 0)$. Write

$$C_n = \begin{pmatrix} C_{n11}(\zeta_1, \alpha_1) & C_{n12}(\zeta_1, \alpha_1) \\ C_{n21}(\zeta_1, \alpha_1) & C_{n22}(\zeta_1, \alpha_1) \end{pmatrix}.$$

Applying Lemmas 3.4, 3.5 and 3.6, we have

$$C_n(\zeta_1, \alpha_1) = C + o_p(1),$$

which, in conjunction with (3. 26), (3. 27), implies that

$$\begin{aligned} \begin{pmatrix} \hat{\alpha} \\ \widehat{\beta} - \lambda_0 \end{pmatrix} &= -C_n^{-1}\begin{pmatrix} B_{n1}(\lambda_0, 0) \\ 0 \end{pmatrix} \\ &= -C^{-1}\begin{pmatrix} B_{n1}(\lambda_0, 0) \\ 0 \end{pmatrix}(1 + o_p(1)). \end{aligned}$$

Combining this with (3. 26) and (3. 27), we have

$$\sqrt{nh}(\widehat{\beta} - \lambda_0) = C_{22.1}^{-1}C_{21}C_{11}^{-1}\sqrt{nh}B_{n1}(\lambda_0, 0)(1 + o_p(1)),$$

$$\sqrt{nh}\hat{\alpha} = -(C_{11}^{-1} + C_{11}^{-1}C_{12}C_{22.1}^{-1}C_{21}C_{11}^{-1})\sqrt{nh}B_{n1}(\lambda_0, 0)(1 + o_p(1)).$$

Finally, according to the Cramér-Wold device and Lemma 3.7, to establish the asymptotic normality of $\widehat{\beta}$, it suffices to check Lyapounov's condition for any one-dimensional projection of $C_{22.1}^{-1}C_{21}C_{11}^{-1}\sqrt{nh} \times B_{n1}(\lambda_0, 0)$, which can be easily proved.

Analogously, we can prove the result for $\hat{\alpha}$. The proof is completed.

Acknowledgements

J. Zhang's research is partially supported by the National Natural Science Foundation of China and a grant from the research programme in EURANDOM, Netherlands. The authors would like to thank Professors A. J. Lenstra and R. Gill for their comments that helped to improve the presentation.

References

Bhapkar, V. P. (1991). Sufficiency, ancillarity, and information in estimating functions. in *Estimating Functions,* Edited by Godambe, V. P., pp. 241–254, Oxford University Press, Oxford.

Bickel, P.J., Klaassen, A.A.J., Ritov, Y., and Wellner, J.A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* The Johns Hopkins Press, Baltimore and London.

Bowman, A.W. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis.* Oxford University Press, Oxford.

Carroll, R.J., Ruppert, D., and Welsh, A.H. (1998). Local estimating equations. *J. Amer. Statist. Assoc.,* **93,** 214-227.

Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.,* **74,** 829-836.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and its Applications.* Chapman and Hall, London.

Hanfelt, J.J. and Liang, K.(1995). Approximate likelihood ratios for general estimating functions. *Biometrika,* **82,** 461-477.

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Ann. Statist.* **25,** 2084-2102.

Owen, A.(1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika,* **75,** 237-249.

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

Qin, J. and Lawless (1994). Empirical likelihood and general estimating equations. *Ann. Statist.,* **23,** 300-325.

Silverman, B.W. (1984). Spline smoothing: the equivalent variable kernel method. *Ann. Statist.,* **12,** 898-916.

Stone, C.J. (1977). Consistent nonparametric regression. *Ann. Statist.,* **5**, 595-645.

Tibshirani, R. and Hastie, T. (1987). Local likelihood estimation. *J. Amer. Statist. Assoc.,* **82,** 559-567.

Zhang, J. and Gijbels, I. (1998). Sieve empirical likelihood and extensions of the generalized least squares. *Technical Report,* Institut de Statistique, Universite Catholique de Louvain.

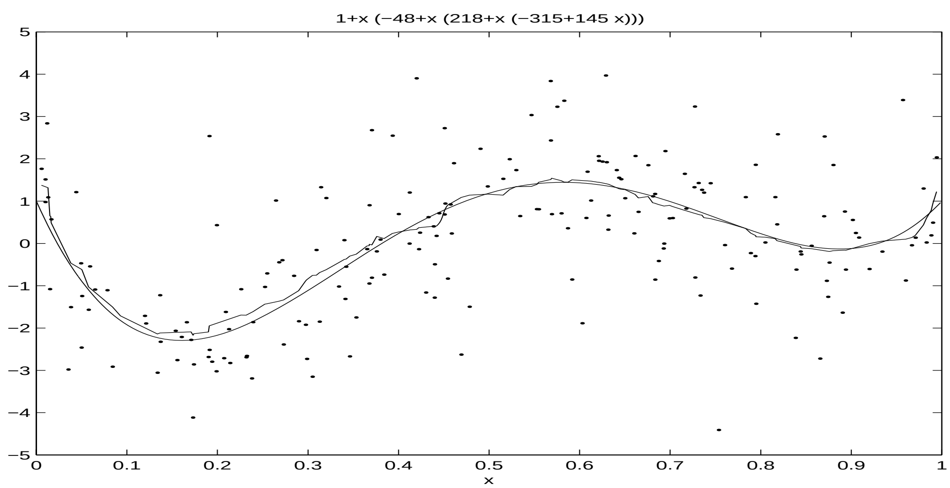

Figure 3.1: The smooth curve denotes the underlying regression function while the other is the estimated curve derived from the local linear empirical likelihood fitting under the first moment constraint. Given X, $\varepsilon \sim t_3$, the t-distribution with 3 degrees of freedom.

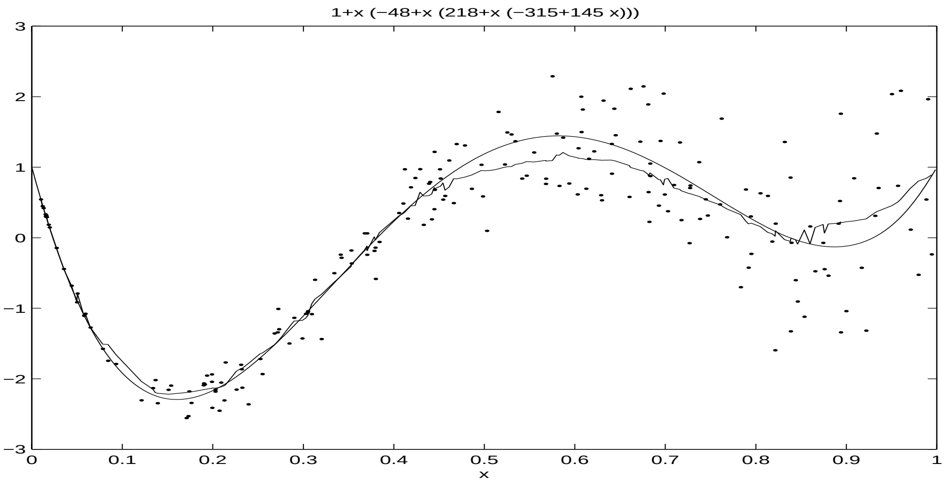

Figure 3.2: The smooth curve denotes the underlying regression function while the other is the estimated curve derived from the local linear empirical likelihood fitting under the first moment constraint. Given X, $\varepsilon \sim N(0, \sigma(X)^2)$ with $\sigma(X)^2 = 1 + X^2$.

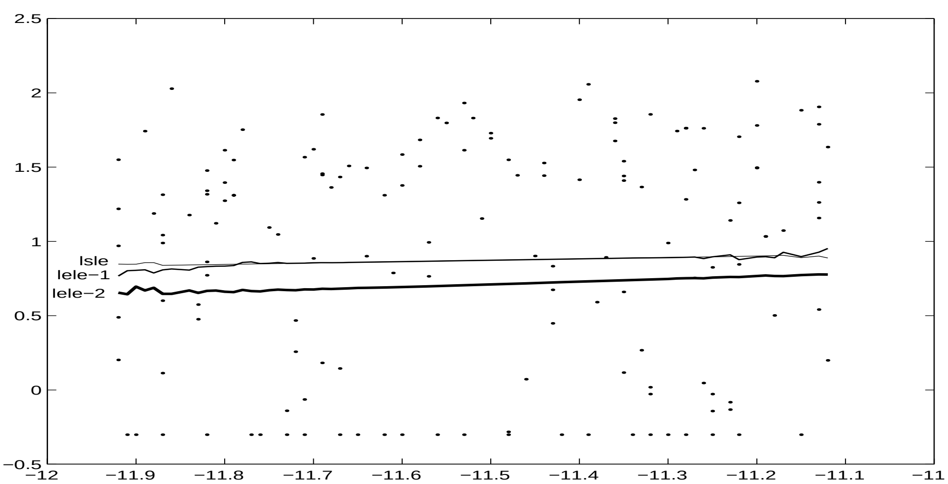

Figure 3.3: $lsle$, $lele-1$ and $lele-2$ denote the estimators derived, respectively, from the local linear least squares fitting, the local linear empirical likelihood fitting under the first moment constraint, and the local linear empirical likelihood fitting under the first and third moment constraints.