Ann. Inst. Statist. Math.
Vol. 57, No. 3, 485–505 (2005)

PROFILE EMPIRICAL LIKELIHOOD FOR PARAMETRIC AND SEMIPARAMETRIC MODELS*

LU LIN[1], LIXING ZHU[2,3] AND K. C. YUEN[4]

[1] *School of Mathematics and System Sciences, Shandong University, Jinan, 250100, Shandong Province, China,* e-mail: linlu@sdu.edu.cn
[2] *Department of Mathematics, Hong Kong Baptist University, Hong Kong, China*
[3] *Department of Statistics, East China Normal University, Shanghai, 200062, China*
[4] *Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China*

(Received February 10, 2004; revised June 11, 2004)

Abstract. This paper introduces a profile empirical likelihood and a profile conditionally empirical likelihood to estimate the parameter of interest in the presence of nuisance parameters respectively for the parametric and semiparametric models. It is proven that these methods propose some efficient estimators of parameters of interest in the sense of least-favorable efficiency. Particularly, for the decomposable semiparametric models, an explicit representation for the estimator of parameter of interest is derived from the proposed nonparametric method. These new estimations are different from and more efficient than the existing estimations. Some examples and simulation studies are given to illustrate the theoretical results.

Key words and phrases: Profile likelihood, empirical likelihood, efficiency, parametric and semiparametric models.

1. Introduction

Profile likelihood has received much attention in the literature. Severini and Wong (1992) outlined the general profile likelihood and proposed a profile conditional likelihood to estimate the parameter of interest under the condition that the data come from a known class of distributions. These methods are based on the idea of estimating a one-dimensional subproblem of the original problem so that the obtained estimator is least-favorable in the sense of Stein (1956). Severini (1998, 1999, 2002) constructed some modified profile likelihood functions, or some approximations to the modified profile likelihood functions through known distribution of data, which yield some estimating functions for the parameter of interest satisfying approximate unbiasedness. Small and McLeish (1994) in Chapter 5 of their book summed up some Hilbert space methods to obtain the estimating function for the parameter of interest, which are based on the version of parameter orthogonality and use the projection of an estimating function for full parameters onto the E-ancillary subspace of estimating functions to make the

*The first author is supported by NNSF projects (10371059 and 10171051) of China. The second author is supported by a grant from The Research Grants Council of the Hong Kong Special Administrative Region, China (#HKU7060/04P). The third author is supported by the University Research Committee of the University of Hong Kong and a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. HKU7323/01M).

estimating function insensitive to the change in the nuisance parameter. Motivated by Severini and Wong (1992), Lin and Zhang (2002) proposed a profile quasi likelihood for the parametric models under the condition that the information about distribution of data relates only to the first two moments and the obtained estimator of parameter of interest is least-favorable efficient.

In this paper, we assume that the information about the distribution is available in the form of functionally independent unbiased estimating functions. The main purpose of this paper is to estimate the parameters of interest under this general condition of data. The basic goal is to estimate parameters of interest efficiently in the sense of least-favorable efficiency for the parametric and semiparametric models. By extending the profile likelihood and profile quasi likelihood to empirical likelihood, this paper introduces a profile empirical likelihood to estimate the parameters of interest in the presence of finite-dimensional nuisance parameter. When the nuisance parameter is infinite-dimensional, an empirical form of score function for the nuisance parameter is constructed and then a profile conditionally empirical likelihood for the parameter of interest is obtained. In the decomposable semiparametric models, the proposed nonparametric method leads to some explicit representations for the estimators of parameter of interest and the nonparametric component. These new explicit representations are very different from the existing estimations and are useful tools for statistical inference in the decomposable semiparametric models. Theoretical and simulation results show that these estimations are more efficient than the existing estimations.

2. Profile empirical likelihood for parametric models

In this section we first introduce a criterion, called the empirical Fisher information, to assess the empirical likelihood for parametric models and then construct a profile empirical likelihood for the parameter of interest. The obtained estimator of parameter of interest is asymptotically optimal under this criterion. Another goal of this section is to motivate some basic versions for profile empirical likelihood, which will be used in the next section.

In order to define a profile empirical likelihood for parametric models, we now outline the original empirical likelihood as proposed by Owen (1988, 1990), Qin and Lawless (1994) and so on. Let $y_1, \ldots, y_N$ be independent observations with an unknown d-variate distribution $F(y, \theta)$, where θ is a p-dimensional column vector of unknown parameters. We assume that the information about $F(y, \theta)$ is available in the form of unbiased estimating function $u(y, \theta) = (u_1(y, \theta), \ldots, u_r(y, \theta))'$, $r \geq p$, i.e., known function vector $u(y, \theta)$ satisfies $\mathrm{E}(u(y, \theta)) = 0$. In this case the empirical likelihood function is defined as

$$L(\theta) = \sup \left\{ \log \prod_{i=1}^{N} N p_i \;\middle|\; \sum_{i=1}^{N} p_i = 1, p_i \geq 0, i = 1, \ldots, N, \sum_{i=1}^{N} p_i u(y_i, \theta) = 0 \right\}.$$

By Lagrange multipliers, the empirical likelihood function can be expressed as

$$L(\theta) = \sum_{i=1}^{N} \log(1 + t'(\theta) u(y_i, \theta)), \tag{2.1}$$

where $t(\theta) = (t_1(\theta), \ldots, t_r(\theta))'$ satisfies

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{1+t'(\theta)u(y_i,\theta)}u(y_i,\theta) = 0.$$

On the other hand, if $\log\prod_{i=1}^{N} p_i$ is replaced by Euclidean distance $-\frac{1}{2}\sum_{i=1}^{N}(p_i - \frac{1}{N})^2$, the empirical Euclidean likelihood function is expressed as

$$L_E(\theta) = \sup\left\{-\frac{1}{2}\sum_{i=1}^{N}(Np_i - 1)^2 \,\middle|\, \sum_{i=1}^{N} p_i = 1, p_i \geq 0, i = 1, \ldots, N, \sum_{i=1}^{N} p_i u(y_i,\theta) = 0\right\}. \tag{2.2}$$

By Lagrange multipliers, the empirical Euclidean likelihood function has the form of

$$L_E(\theta) = -\frac{N}{2}\bar{u}'(\theta)S^{-1}(\theta)\bar{u}(\theta), \tag{2.3}$$

where $\bar{u}(\theta) = \frac{1}{N}\sum_{i=1}^{N} u(y_i,\theta)$, $S(\theta) = \frac{1}{N}\sum_{i=1}^{N}(u(y_i,\theta) - \bar{u}(\theta))(u(y_i,\theta) - \bar{u}(\theta))'$.

The two empirical likelihood functions $L(\theta)$ and $L_E(\theta)$ have the same asymptotic behaviours such as $L(\theta) = L_E(\theta) + o_P(1)$ (Owen (1990)). So, in what follows we focus only on the empirical Euclidean likelihood function $L_E(\theta)$.

We use $\hat{\theta}$ to denote the empirical Euclidean likelihood estimator obtained from the empirical Euclidean likelihood (2.3), i.e., $\hat{\theta} = \arg\sup_{\theta\in\Theta} L_E(\theta)$. Under the usual kinds of limiting conditions (Qin and Lawless (1994), Luo (1994)), $\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} N(0, i_\theta^{-1})$, where the asymptotic covariance matrix of $\hat{\theta}$ is

$$i_\theta^{-1} = \left\{E_0\left(\frac{\partial u(y,\theta_0)}{\partial\theta'}\right)' V^{-1}(\theta)E_0\left(\frac{\partial u(y,\theta_0)}{\partial\theta'}\right)\right\}^{-1}, \tag{2.4}$$

$V(\theta) = E_0(u(y,\theta_0)u'(y,\theta_0))$ and the subscript 0 denotes the evaluation at the true state. The version of equation (2.3) is similar to that of the Generalized Method of Moments (GMM). According to GMM, i_θ^{-1} provides a lower bound to the asymptotic covariances of all regularity estimators of θ (Chamberlain (1987)). Then, in this paper, we call i_θ the empirical Fisher information for θ.

We now introduce profile empirical likelihood. In this paper, parameter vector θ is decomposed as $\theta = (\alpha, \beta')'$ and then the parameter space Θ is also decomposed as $\Theta = \mathcal{A}\times\mathcal{B}$, in which $\alpha \in \mathcal{A}$ is a real-valued parameter of interest and $\beta \in \mathcal{B}$ is a $(p-1)$-dimensional column vector of nuisance parameters, and Θ, $\mathcal{A}$ and $\mathcal{B}$ are all open sets. To estimate the parameter α of interest, as mentioned by Severini and Wong (1992) and Lin and Zhang (2002), we assume that there is a curve in the parameter space such as $\alpha \mapsto (\alpha, \beta_\alpha')'$ with $\beta_{\alpha_0} = \beta_0$. The empirical Fisher information for estimating α along the subproblem defined by this curve is given by

$$E_0\left(\frac{\partial u(y,\alpha,\beta_\alpha)}{\partial\alpha} + U\right)' V^{-1}(\alpha,\beta_\alpha)E_0\left(\frac{\partial u(y,\alpha,\beta_\alpha)}{\partial\alpha} + U\right)\Bigg|_{\alpha=\alpha_0},$$

where $U = \frac{\partial u(y,\alpha,\beta_\alpha)}{\partial \beta'} \frac{d\beta_\alpha}{d\alpha}$. The minimum empirical Fisher information for α over all possible subproblems are given by

$$(2.5)\quad E_0\left(\frac{\partial u(y,\alpha,\beta_\alpha)}{\partial \alpha}+U^*\right)' V^{-1}(\alpha,\beta_\alpha)E_0\left(\frac{\partial u(y,\alpha,\beta_\alpha)}{\partial \alpha}+U^*\right)\Bigg|_{\alpha=\alpha_0}$$
$$=\inf_{U\in\mathcal{F}} E_0\left(\frac{\partial u(y,\alpha,\beta_\alpha)}{\partial \alpha}+U\right)' V^{-1}(\alpha,\beta_\alpha)E_0\left(\frac{\partial u(y,\alpha,\beta_\alpha)}{\partial \alpha}+U\right)\Bigg|_{\alpha=\alpha_0},$$

where $\mathcal{F} = \operatorname{span}\{\frac{\partial u(y,\alpha,\beta_\alpha)}{\partial \beta'}|_{\alpha=\alpha_0}\}$ and $U^* = \frac{\partial u(y,\alpha,\beta_\alpha)}{\partial \beta'}\frac{d\beta^*_\alpha}{d\alpha}$. It can be verified that $-U^*|_{\alpha=\alpha_0}$ is the projection of $\frac{\partial u(y,\alpha,\beta_\alpha)}{\partial \alpha}|_{\alpha=\alpha_0}$ onto $\mathcal{F}$ based on the inner product defined by

$$\langle A,B\rangle = E_0(A')V^{-1}(\alpha,\beta_\alpha)E_0(A)\,|_{\alpha=\alpha_0}\,.$$

It follows immediately from the previous characterization that

$$(2.6)\qquad i_\alpha = E_0\left(\frac{\partial u(y,\alpha,\beta_\alpha)}{\partial \alpha}+U^*\right)' V^{-1}(\alpha,\beta_\alpha)E_0\left(\frac{\partial u(y,\alpha,\beta_\alpha)}{\partial \alpha}+U^*\right)\Bigg|_{\alpha=\alpha_0}$$

provides another useful interpretation of i_α as the minimum empirical Fisher information over all possible smooth one-dimensional subproblems. Then we call i_α the empirical Fisher information (EFI) for α. Note that EFI for α associated with a curve $\alpha \mapsto (\alpha,(\beta^*_\alpha)')'$ depends on the curve only through the tangent vector $\frac{\partial\beta^*_\alpha}{\partial\alpha}$ at the true value point α_0. Similar to the terminology used in recent literature (Severini and Wong (1992), Lin and Zhang (2002)), in this paper, we call such a curve $\alpha \mapsto (\alpha,(\beta^*_\alpha)')'$ the empirical least-favorable curve (ELFC), call the tangent vector $\frac{\partial\beta^*_\alpha}{\partial\alpha}$ the empirical least-favorable direction (ELFD), and then call

$$(2.7)\qquad L_E(\alpha,\beta^*_\alpha) = -\frac{N}{2}\bar{u}'(\alpha,\beta^*_\alpha)S^{-1}(\alpha,\beta^*_\alpha)\bar{u}(\alpha,\beta^*_\alpha)$$

the empirical least-favorable Euclidean likelihood.

Suppose that we are able to identify some curves $\alpha \mapsto (\alpha,\beta'_\alpha)'$ for $\alpha\in(a,b)$ satisfying $\beta_{\alpha_0}=\beta_0$. A least-favorable-efficient selection for empirical likelihood of α is $L_E(\alpha,\beta^*_\alpha)$ as defined in (2.7). Of course in practice such a likelihood is not available because β^*_α depends on unknown expectation E_0 and unknown parameter α_0 and then $L_E(\alpha,\beta^*_\alpha)$ depends on them as well. However, if we are able to obtain a suitable estimator $\hat\beta_\alpha$ of a ELFC β^*_α, we may then obtain an estimate of α by solving a substitute equation $\frac{\partial L_E(\alpha,\hat\beta_\alpha)}{\partial\alpha}=0$.

From the characterization of U^* as a projection, it follows immediately that a necessary and sufficient condition for $\frac{\partial\beta^*_\alpha}{\partial\alpha}$ to be ELFD is that

$$(2.8)\qquad E_0\left(\frac{\partial u(y,\alpha,\beta_\alpha)}{\partial \alpha}+\frac{\partial u(y,\alpha,\beta_\alpha)}{\partial \beta'}\left(\frac{d\beta^*_\alpha}{d\alpha}\right)\right)' V^{-1}(\alpha,\beta_\alpha)$$
$$\times E_0\left(\frac{\partial u(y,\alpha,\beta_\alpha)}{\partial \beta'}\right)\Bigg|_{\alpha=\alpha_0}=0.$$

According to the orthogonality as defined above, an estimator $\hat{\beta}_\alpha$ is said to be consistent estimator of a ELFC, if $\hat{\beta}_\alpha$ converges in probability to a constant β_α and satisfies that

$$\left(\frac{\partial \bar{u}(\alpha,\beta_\alpha)}{\partial \alpha} + \frac{\partial \bar{u}(\alpha,\beta_\alpha)}{\partial \beta'}\left(\frac{d\hat{\beta}_\alpha}{d\alpha}\right)\right)' S^{-1}(\alpha,\beta_\alpha)\left(\frac{\partial \bar{u}(\alpha,\beta_\alpha)}{\partial \beta'}\right)\Bigg|_{\alpha=\alpha_0} = o_P(1). \tag{2.9}$$

To carry out this approach, we need to construct a consistent estimator of a ELFC. For any fixed α, let $\hat{\beta} = \arg\sup_\beta L_E(\alpha,\beta)$, equivalently, $\hat{\beta}_\alpha$ satisfies

$$\frac{\partial L_E(\alpha,\beta)}{\partial \beta'}\Bigg|_{\beta=\hat{\beta}_\alpha} = 0. \tag{2.10}$$

Then from this estimator we can obtain an estimator $\hat{\alpha}$ by maximizing $L_E(\alpha,\hat{\beta}_\alpha)$, equivalently, $\hat{\alpha}$ satisfies

$$\frac{\partial L_E(\alpha,\hat{\beta}_\alpha)}{\partial \alpha}\Bigg|_{\alpha=\hat{\alpha}} = 0. \tag{2.11}$$

The following condition helps us get the asymptotic properties of $\hat{\beta}_\alpha$:

(B_1) In the parametric model,

$$\frac{\partial E_0(u(\alpha,\beta))}{\partial \beta} = E_0\left(\frac{\partial u(\alpha,\beta)}{\partial \beta}\right), \qquad \frac{\partial E_0(u(\alpha,\beta)u'(\alpha,\beta))}{\partial \beta} = E_0\left(\frac{\partial (u(\alpha,\beta)u'(\alpha,\beta))}{\partial \beta}\right).$$

Let

$$\zeta(\alpha,\beta) = E_0(u'(\alpha,\beta))(E_0(u(\alpha,\beta)u'(\alpha,\beta)))^{-1}E_0(u(\alpha,\beta))$$

and β_α denote a solution to the equation $\partial\zeta(\alpha,\beta)/\partial\beta = 0$ with respect to β for fixed α. Assume that β_α is unique and that for any $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\sup_\alpha \left|\frac{\partial \zeta(\alpha,\tilde{\beta}_\alpha)}{\partial \beta}\right| \le \delta$$

implies

$$\sup_\alpha |\tilde{\beta}_\alpha - \beta_\alpha| \le \varepsilon.$$

LEMMA 2.1. *Assume that Condition* (A_1) *listed in the Appendix and Condition* (B_1) *hold. Then* $\hat{\beta}_\alpha$, *the root of the equation* (2.10), *satisfies*

$$\sup_\alpha |\hat{\beta}_\alpha - \beta_\alpha| = o_P(1), \tag{2.12}$$

$$\hat{\beta}_{\alpha_0} - \beta_0 = O_P(N^{-1/2}) \tag{2.13}$$

and

$$\left(\frac{\partial \bar{u}(\alpha,\beta_\alpha)}{\partial \alpha} + \frac{\partial \bar{u}(\alpha,\beta_\alpha)}{\partial \beta'}\left(\frac{d\hat{\beta}_\alpha}{d\alpha}\right)\right)' S^{-1}(\alpha,\beta_\alpha)\left(\frac{\partial \bar{u}(\alpha,\beta_\alpha)}{\partial \beta'}\right)\Bigg|_{\alpha=\alpha_0} = O_P(N^{-1/2}). \tag{2.14}$$

Remark 2.1. Lemma 2.1 shows that $\hat{\beta}_\alpha$ is a consistent estimator of ELFC. The approach to constructing the estimator $\hat{\beta}_\alpha$ of ELFC is of course common; in fact, for fixed α, $\hat{\beta}_\alpha$ is exactly the empirical likelihood estimator. This fact is similar to that, for fixed α, the maximum likelihood estimator of β is a consistent estimator of a least-favorable curve when the distribution $F(y,\theta)$ of y is known (Severini and Wong (1992)). Furthermore the result (2.14) implies (2.9), both describing the asymptotic orthogonality of the direction $d\hat{\beta}_\alpha/d\alpha$. On the other hand, the proof of this lemma motivates an idea to construct an empirical form of score function for the nonparametric component of the semiparametric model (for detail see the next section).

THEOREM 2.1. *If Condition* (A_1) *listed in the Appendix and Condition* (B_1) *hold, then* $\hat{\alpha}$, *the root of the equation* (2.11), *satisfies that*

$$\sqrt{N}(\hat{\alpha}-\alpha_0) \xrightarrow{\mathcal{D}} N(0,(i_\alpha)^{-1}),$$

where i_α *is defined in* (2.6).

Remark 2.2. The theorem shows that the estimator $\hat{\alpha}$ obtained by (2.11) is asymptotically least-favorable efficient.

3. Profile empirical likelihood for semiparametric models

Generally, the semiparametric model is parameterized by a parameter of interest taking values in a finite-dimensional space and a nuisance parameter taking values in an infinite-dimensional space. As mentioned previously, when the nuisance parameter is finite-dimensional, the profile empirical likelihood estimator $\hat{\alpha}$ obtained by (2.11) is asymptotically least-favorable efficient. However, if the nuisance parameter is infinite-dimensional, we are unable to get a root from the equation (2.10) and then we are unable to get an estimator of α by profile empirical likelihood as presented before.

3.1 *Semiparametric regression model*

In this subsection, we introduce a profile conditionally empirical likelihood for the following semiparametric regression model:

$$E(y \mid x,z) = h(\alpha,\beta,z), \tag{3.1}$$

where $h(\alpha,\beta,z) \in R$ is a known function of α, β and z, parameter β depends on the value of x in the sense of $\beta=\lambda(x)$ for some unknown smooth function λ. The model (3.1) can be rewritten as

$$E(y \mid x,z) = h(\alpha,\lambda(x),z).$$

For simplicity, we assume that $x \in [0,1]$ and $z \in [0,1]$. The model contains both parametric and nonparametric components, the nonparametric component $\lambda(x)$ playing a role of nuisance parameter. This model is generally semiparametric regression, including semiparametric linear regression, semiparametric nonlinear regression, semiparametric generalized linear regression and so on.

Similar to Severini and Wong (1992), we can easily obtain the partial derivative of $h(\alpha,\lambda(x),z)$ with respect to function $\lambda(x)$ as that, for any continuous function $v(x)$ on $[0,1]$,

$$\frac{h(\alpha,\lambda(x),z)}{\partial\lambda}(v) = \frac{h(\alpha,\beta,z)}{\partial\beta}\bigg|_{\beta=\lambda(x)} v(x),$$

which can be regarded as a linear transformation on the tangent space of the space of possible λ. By the idea of Section 2, for given x and z, here we regard β as a function of α, denoted by $\beta = \beta_\alpha$, and consequently $\lambda(x)$ is also a function of α, denoted by $\lambda(x) = \lambda_\alpha(x)$. The first inference problem changes to be the subproblem of constructing suitable estimator of $\lambda(x)$ in the class $\{\lambda_\alpha(x) : \alpha \in \mathcal{A}\}$. In this case, the unbiased estimating function of α along the curve $\alpha \mapsto (\alpha, \lambda_\alpha(x))$ is

$$u(y, \alpha, \lambda_\alpha(x), z) = y - h(\alpha, \lambda_\alpha(x), z).$$

With the similar argument as in the previous section and Severini and Wong (1992), we call that a curve $\lambda^*_\alpha(x)$ is a conditionally empirical least-favorable curve (CELFC), equivalently, $\partial\lambda^*_\alpha(x)/\partial\alpha$ is a conditionally empirical least-favorable direction (CELFD) if and only if

$$\begin{aligned}(3.2)\quad E_0\Bigg\{\frac{1}{\sigma^2(x,z)}\Bigg(&\frac{\partial h(\alpha_0, \lambda_0(x), z)}{\partial\alpha}\\ &+\frac{\partial h(\alpha_0,\beta,z)}{\partial\beta}\bigg|_{\beta=\lambda_0(x)}\frac{d\lambda^*_{\alpha_0}(x)}{d\alpha}\Bigg)\frac{\partial h(\alpha_0,\beta,z)}{\partial\beta}\bigg|_{\beta=\lambda_0(x)}v(x)\Bigg\}\\ = 0 \quad &\text{for any continuous function } v(x) \text{ on } [0,1],\end{aligned}$$

where $\sigma^2(x,z) = E_0(y - h(\alpha_0, \lambda_0(x), z) \mid x, z)^2$. Similar to (2.6), for given x and z, the empirical Fisher information (EFI) for α is defined by

$$i_\alpha = E_0\left[\left(\frac{\partial h(\alpha_0, \lambda_0(x), z)}{\partial\alpha} + U_0^*\right)^2 \Big/ \sigma^2(x,z)\right], \tag{3.3}$$

where $U_0^* = \frac{\partial h(\alpha_0,\beta,z)}{\partial\beta}|_{\beta=\lambda_0(x)}\frac{d\lambda^*_{\alpha_0}(x)}{d\alpha}$.

In order to carry out an estimation procedure, an estimator of $\lambda^*_\alpha(x)$ must be available. Since $\lambda_\alpha(x)$ is a function of x, the equation (2.10) in Section 2 can not be used to get the estimator of $\lambda^*_\alpha(x)$. So a new design is desired.

From the proof of Lemma 2.1, fortunately, we can see that, if β is a parameter, the main part of score function $\partial L_E(\alpha,\beta)/\partial\beta$ for β is $(N/2)(\partial\bar{u}'/\partial\beta)S^{-1}\bar{u}$. Motivated by this fact and the empirical method, for fixed α, here we consider an empirical form of score function for β as

$$\begin{aligned}(3.4)\quad q_\beta(\alpha,\beta,x,z) &= \frac{N}{2}\sum_{i=1}^{N} E\left(\frac{\partial u(y_i,\alpha,\beta,z_i)}{\partial\beta}\,\bigg|\, x_i, z_i\right)u(y_i,\alpha,\beta,z_i)K_{Ni}(x,z)\\ &= -\frac{N}{2}\sum_{i=1}^{N}\frac{\partial h(\alpha,\beta,z_i)}{\partial\beta}(y_i - h(\alpha,\beta,z_i))K_{Ni}(x,z),\end{aligned}$$

where $K_{Ni}(x,z) = K_{Ni}((x,z);(x_1,z_1),\ldots,(x_N,z_N))$ are weight functions satisfying the following conditions:

(i) $K_{Ni}(x,z) \geq 0$ for $x, z \in [0,1]$.

(ii) $\sum_{i=1}^N K_{Ni}(x,z) = 1$ for $x, z \in [0,1]$.

(iii) $\lim_{N\to\infty}\sup_{x,z\in[0,1]}\sum_{i=1}^N K_{Ni}(x,z)I(\|(x_i,z_i)-(x,z)\| > \delta) = 0$ for any $\delta > 0$, where $I(S)$ is an indicator function of set S.

(iv) $\sup_{x,z\in[0,1]} \sum_{i=1}^{N} K_{Ni}^2(x,z) = O(n^{-1})$ for $x, z \in [0,1]$.

For fixed α, x and z, from the equation

$$q_\beta(\alpha, \beta, x, z) = 0 \tag{3.5}$$

with respect to β, we get a solution denoted by $\hat{\lambda}_\alpha$. Note that, given $(x_1, z_1), \ldots, (x_N, z_N)$, weights $K_{Ni}(x,z)$ are the functions of x and z, which are different from the existing weights that are the functions of x only (Severini and Staniswalis (1994), Shi (2001)). As a result the solution $\hat{\lambda}_\alpha$ is also a function of x and z: $\hat{\lambda}_\alpha = \hat{\lambda}_\alpha(x,z)$. The following Lemmas 3.1 and 3.2 will show that, at the true state, $\hat{\lambda}_{\alpha_0}(x,z)$ tends to CELFC in probability. This is a new idea of using a function of 2-variables to approximate a function of single variable. Although this method is different from that of Severini and Staniswalis (1994), it is coincident with that of Severini and Wong (1992) in the sense of that the estimator of nonparametric component depends on all design variables. The purpose of this special construction is to make $\hat{\lambda}_{\alpha_0}(x,z)$ to be a consistent estimator of CELFC as described in the following Lemmas 3.1 and 3.2. On the other hand, our method here is somewhat different from the smoothed empirical likelihood. Although the two methods are based on smoothing technique, the former uses smoothing technique to construct estimating equation for CELFC and then solves a nonparametric problem, the latter employs smoothing technique to estimate full parameters and then solves a parametric problem.

For understanding how to construct $\hat{\lambda}_\alpha(x,z)$ and $\hat{\alpha}$ by the equation (3.5) and the empirical likelihood, we first consider the following examples.

Example 1. Consider a common semiparametric regression model, in which $h(\alpha, \lambda(x), z)$ has the decomposable form of

$$h(\alpha, \lambda(x), z) = g(\alpha, z) + \lambda(x) \tag{3.6}$$

for a known function $g(\alpha, z)$ and an unknown function $\lambda(x)$. In the case we can verify that the solution obtained from the equation (3.5) has a explicit representation as

$$\hat{\lambda}_\alpha(x,z) = \sum_{i=1}^{N} (y_i - g(\alpha, z_i)) K_{ni}(x,z). \tag{3.7}$$

Obviously, it is a common kernel method except that the weights are the functions of 2-variables. From this estimator, we can obtain an estimator $\hat{\alpha}$ by minimizing $L_E(\alpha, \hat{\lambda}_\alpha(x,z))$, or $\hat{\alpha}$ satisfies

$$\left.\frac{\partial L_E(\alpha, \hat{\lambda}_\alpha(x,z))}{\partial \alpha}\right|_{\alpha=\hat{\alpha}} = 0. \tag{3.8}$$

Particularly, in the decomposable model (3.6), if $g(\alpha, z)$ is a linear function of α, i.e., $g(\alpha, z) = \alpha z$, then we can get a conditionally empirical likelihood estimator $\hat{\alpha}$ of α by equation (3.8). The estimator has an explicit representation

$$\hat{\alpha} = \frac{\sum_{i=1}^{N}(y_i - \sum_{j=1}^{N} K_{Nj}(x_i, z_i) y_j)(z_i - \sum_{j=1}^{N} K_{Nj}(x_i, z_i) z_j)}{\sum_{i=1}^{N}(z_i - \sum_{j=1}^{N} K_{Nj}(x_i, z_i) z_j)^2}, \tag{3.9}$$

and then, when $\alpha = \hat{\alpha}$, the solution $\lambda_\alpha(x,z)$ also has an explicit representation

$$\hat{\lambda}_{\hat{\alpha}}(x,z) = \sum_{i=1}^{N} K_{Ni}(x,z)(y_i - \hat{\alpha} z_i). \tag{3.10}$$

Example 2. Consider the following semiparametric generalized linear model:

$$E(y \mid x, z) = \exp\{\alpha z + \lambda(x)\}.$$

From the equation (3.5), fixed α, we get a solution as

$$\hat{\lambda}_\alpha(x,z) = \log\left(\sum_{i=1}^{N} y_i e^{\alpha z_i} K_{Ni}(x,z)\right) - \log\left(\sum_{i=1}^{N} e^{2\alpha z_i} K_{Ni}(x,z)\right).$$

In this case, however, we can not get an explicit representation for α and then, for solving the equation $\partial L_E(\alpha, \hat{\lambda}_\alpha(x,z))/\partial\alpha = 0$, a numerical solution is necessary.

Let

$$\xi(\alpha,\beta,x,z) = E_0\left(\frac{\partial u(y,\alpha,\beta,z)}{\partial\beta} u(y,\alpha,\beta,z) \mid x, z\right),$$

which equals $\frac{\partial h(\alpha,\beta,z)}{\partial\beta}(h(\alpha_0,\lambda_0(x),z) - h(\alpha,\beta,z))$ in the semiparametric regression model (3.1). Since $\xi(\alpha_0,\beta_0,x,z) = 0$, we give the following regularity condition:

(B_2) In the semiparametric model, let $\lambda_\alpha(x,z)$ denote a solution to the equation $\xi(\alpha,\beta,x,z) = 0$ with respect to β for fixed α and (x,z). Assume that $\lambda_\alpha(x,z)$ is unique and that for any $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\sup_\alpha \sup_{(x,z)} |\xi(\alpha, \tilde{\lambda}_\alpha(x,z), x, z)| \le \delta$$

implies that

$$\sup_\alpha \sup_{(x,z)} |\tilde{\lambda}_\alpha(x,z) - \lambda_\alpha(x,z)| \le \varepsilon.$$

LEMMA 3.1. *For the semiparamtric regression model* (3.1), *given x and z, in addition to Conditions* (A_2) *listed in the Appendix, if Condition* (B_2) *holds, then*

$$\sup_\alpha \sup_{(x,z)} |\hat{\lambda}_\alpha(x,z) - \lambda_\alpha(x,z)| = o_P(1), \tag{3.11}$$

$$\hat{\lambda}_{\alpha_0}(x,z) - \lambda_0(x) = O_P(N^{-1/2}) \tag{3.12}$$

and

$$\left.\frac{d\hat{\lambda}_\alpha(x,z)}{d\alpha}\right|_{\alpha=\alpha_0} = -\frac{\sum_{i=1}^N \left.\frac{\partial h(\alpha_0,\beta,z_i)}{\partial\beta}\right|_{\beta=\lambda_0(x)} \frac{\partial h(\alpha_0,\lambda_0(x),z_i)}{\partial\alpha} K_{Ni}(x,z)}{\sum_{i=1}^N \left(\left.\frac{\partial h(\alpha_0,\beta,z_i)}{\partial\beta}\right|_{\beta=\lambda_0(x)}\right)^2 K_{Ni}(x,z)} + o_P(1) \tag{3.13}$$

$$= -\frac{\sum_{i=1}^{N} \frac{\partial u(y_i,\alpha_0,\beta,z_i)}{\partial \beta}\Big|_{\beta=\lambda_0(x)} \frac{\partial u(y_i,\alpha_0,\lambda_0(x),z_i)}{\partial \alpha} K_{Ni}(x,z)}{\sum_{i=1}^{N}\left(\frac{\partial u(y_i,\alpha_0,\beta,z_i)}{\partial \beta}\Big|_{\beta=\lambda_0(x)}\right)^2 K_{Ni}(x,z)}$$
$$+ o_P(1).$$

LEMMA 3.2. *For the semiparamtric regression model* (3.1), *if the conditions of Lemma* 3.1 *hold, then,*

$$\left(\frac{\partial \bar u}{\partial \alpha} + \frac{\partial \bar u}{\partial \beta}\Big|_{\beta=\lambda_0(x)} \frac{d\hat\lambda_\alpha(x,z)}{d\alpha}\Big|_{\alpha=\alpha_0}\right) S^{-1}(\alpha_0,\lambda_0(x)) \frac{\partial \bar u}{\partial \beta}\Big|_{\beta=\lambda_0(x)} v(x) \tag{3.14}$$
$$= o_P(1) \quad \textit{for any continuous function } v(x) \textit{ on } [0,1],$$

where

$$\frac{\partial \bar u}{\partial \alpha} = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial u(y_i,\alpha_0,\lambda_0(x_i),z_i)}{\partial \alpha}, \quad \frac{\partial \bar u}{\partial \beta}\Big|_{\beta=\lambda_0(x)} = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial u(y_i,\alpha_0,\beta,z_i)}{\partial \beta}\Big|_{\beta=\lambda_0(x_i)},$$
$$\frac{\partial \bar u}{\partial \beta}\Big|_{\beta=\lambda_0(x)} \frac{d\hat\lambda_\alpha(x,z)}{d\alpha}\Big|_{\alpha=\alpha_0} = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial u(y_i,\alpha_0,\beta,z_i)}{\partial \beta}\Big|_{\beta=\lambda_0(x_i)} \frac{d\hat\beta_\alpha(x_i,z_i)}{d\alpha}\Big|_{\alpha=\alpha_0}$$

and

$$S(\alpha_0,\lambda_0(x)) = \frac{1}{N}\sum_{i=1}^{N}(u(y_i,\alpha_0,\lambda_0(x_i),z_i) - \bar u)^2.$$

Remark 3.1. The lemmas above imply that $\lambda_{\alpha_0}(x,z)$ is just a CELFC with $\lambda_{\alpha_0}(x,z) = \lambda_0(x)$, $\hat\lambda_{\alpha_0}(x,z)$ is a consistent estimator of this curve. Furthermore, the direction $d\hat\lambda_\alpha(x,z)/d\alpha|_{\alpha=\alpha_0}$ of CELFD has an approximately explicit representation (3.13) and the equation (3.14) describes the asymptotic orthogonality of the direction $d\hat\lambda_\alpha(x,z)/d\alpha$, which just gives an empirical form of (3.2).

THEOREM 3.1. *If the conditions in Lemma* 3.2 *holds, then*

$$\sqrt{N}(\hat\alpha - \alpha_0) \xrightarrow{\mathcal{D}} N(0,(i_\alpha)^{-1}),$$

where i_α is defined in (3.3).

Remark 3.2. The theorem shows that the estimator $\hat\alpha$ obtained by (3.8) is least-favorable efficient. If we use the existing nonparametric method to estimate nonparametric component $\lambda(x)$ (Severini and Staniswalis (1994), Shi (2000)), the estimator is a function of univariable x only. For example, in the decomposable model (3.6), the estimator is

$$\check\lambda_\alpha(x) = \sum_{i=1}^{N} \check K_{Ni}(x)(y_i - g(\alpha,z_i)),$$

where $\check{K}_{Ni}(x) = \check{K}_{Ni}(x; x_1, \ldots, x_N)$ are weight functions depending only on univariable x. This results in the estimators of α and $\beta(x)$ respectively as

$$\check{\alpha} = \frac{\sum_{i=1}^N (y_i - \sum_{j=1}^N \check{K}_{Nj}(x_i)y_j)(z_i - \sum_{j=1}^N \check{K}_{Nj}(x_i)z_j)}{\sum_{i=1}^N (z_i - \sum_{j=1}^N \check{K}_{Nj}(x_i)z_j)^2} \tag{3.15}$$

and

$$\check{\lambda}_{\check{\alpha}}(x) = \sum_{i=1}^N \check{K}_{Ni}(x)(y_i - g(\check{\alpha}, z_i)). \tag{3.16}$$

However, $\check{\lambda}_\alpha(x)$ is not a consistent estimator of a CELFC because $\frac{\partial \check{\lambda}_\alpha(x)}{\partial \alpha}$ has not the orthogonality as described by Lemma 3.2 and then $\check{\alpha}$ is not a least-efficient estimation. The simulations in Section 4 also support this theory because the variance of $\sqrt{n}\hat{\alpha}$ is significantly smaller than that of $\sqrt{n}\check{\alpha}$.

3.2 *Generalized semiparametric models*

In this subsection we consider the generalized semiparametric model defined by the general unbiased estimating function $u(y, \alpha, \beta, z)$, which satisfies

$$E(u(y, \alpha, \beta, z) \mid x, z) = 0 \tag{3.17}$$

and β depends on the value of x in the sense of $\beta = \lambda(x)$ for some function λ. This model includes the semiparametric regression defined in previous subsection as a special case. As noted previously, the key problem is how to construct the score function for β. Motivated by (3.4), fixed α, the score function for β is defined by

$$q_\beta(\alpha, \beta, x, z) = \frac{N}{2}\sum_{i=1}^N E\left(\frac{\partial u(y, \alpha, \beta, z_i)}{\partial \beta} \,\middle|\, z_i\right) u(y_i, \alpha, \beta, z_i) K_{Ni}(x, z). \tag{3.18}$$

In the definition above, however, expectation $M(\alpha, \beta, z) = E(\frac{\partial u(y,\alpha,\beta,z)}{\partial \beta} \mid z)$ may be unknown and then a consistent estimator is desired. An efficient method is designed by the empirical notion as follows:

$$M_N(\alpha, \beta, z) = \sum_{j=1}^N \frac{\partial u(y_j, \alpha, \beta, z_j)}{\partial \beta} W_{Nj}(z), \tag{3.19}$$

where the weight functions $W_{Nj}(z) = W_{Nj}(z; z_1, \ldots, z_N)$ satisfy the regularity conditions as (i)–(iv) in Subsection 3.1 if z and z_i take the places respectively of (x, z) and (x_i, z_i). For example, the condition (iii) becomes what follows:

(iii)' $\lim_{n\to\infty} \sup_{z\in[0,1]} \sum_{j=1}^N W_{Nj}(z) I(|z_j - z| > \delta) = 0$ for any $\delta > 0$.

LEMMA 3.3. *If* $\max_j W_{Nj}^2(z) N \log\log N \to 0$ *uniformly for* $z \in [0,1]$, $E(\frac{\partial u(y,\alpha,\beta,z)}{\partial \beta} \mid z)$ *is a continuous function of* $z \in [0,1]$, *and Condition* (A_2) *of the Appendix holds, then*

$$M_N(\alpha, \beta, z) \overset{\text{a.s.}}{\to} E\left(\frac{\partial u(y, \alpha, \beta, z)}{\partial \beta} \,\middle|\, z\right).$$

From Lemma 3.3 we can ensure that

$$\hat{q}_\beta(\alpha,\beta,x,z) = \frac{N}{2}\sum_{i=1}^{N} M_N(\alpha,\beta,z_i)u(y_i,\alpha,\beta,z_i)K_{Ni}(x,z) \tag{3.20}$$

is an efficient approximation to the score function defined by (3.18). The approximate score function above is asymptotically unbiased. An solution, denoted by $\hat{\lambda}_\alpha(x,z)$, can be obtained from the equation $\hat{q}_\beta(\alpha,\beta,x,z)=0$ with respect to β, and then maximizing $L_E(\alpha,\hat{\lambda}_\alpha(x,z))$ yields an estimator of α denoted by $\hat{\alpha}$.

THEOREM 3.2. *For the generalized semiparamtric regression model* (3.17), *if the conditions in Lemmas* 3.1 *and* 3.3 *hold, then*

$$\sqrt{N}(\hat{\alpha}-\alpha_0) \xrightarrow{\mathcal{D}} N(0,(i_\alpha)^{-1}),$$

where i_α is defined in (3.3) *with $h(\alpha,\beta,z)$ being replaced by $u(y,\alpha,\beta,z)$.*

Remark 3.3. The theorem indicates that the estimator $\hat{\alpha}$ is least-favorable efficient for generalized semiparametric models.

Example 3. This example seems to be artificial, but it can show how to construct $\hat{\lambda}_\alpha(x,z)$ and $\hat{\alpha}$ in the generalized semiparametric models. In this example, the unbiased function has the form of $u(y,\alpha,\lambda(x),z)=y^2-\alpha z-yz\lambda(x)$. Then

$$M_N(\alpha,\beta,z) = -\sum_{j=1}^{N} y_j z_j W_{Nj}(z).$$

From the equation $\hat{q}_\beta(\alpha,\beta,x,z)=0$ with respect to β, we get

$$\hat{\lambda}_\alpha(x,z) = \frac{\sum_{i=1}^N\sum_{j=1}^N y_j z_j (y_i^2-\alpha z_i)K_{Ni}(x,z)W_{Nj}(z_i)}{\sum_{i=1}^N\sum_{j=1}^N y_j z_j y_i z_i K_{Ni}(x,z)W_{Nj}(z_i)}.$$

By maximizing $L_E(\alpha,\hat{\beta}_\alpha(x,z))$ we get an estimator of α as

$$\hat{\alpha} = \frac{\sum_{k=1}^N(y_k^2 - y_k z_k \sum_{i=1}^N\sum_{j=1}^N y_j z_j y_i^2 K_{Ni}(x_k,z_k)W_{Nj}(z_i)/s_k)}{\sum_{k=1}^N(z_k - y_k z_k \sum_{i=1}^N\sum_{j=1}^N y_j z_j z_i K_{Ni}(x_k,z_k)W_{Nj}(z_i)/s_k)},$$

where $s_k=\sum_{i=1}^N\sum_{j=1}^N y_j z_j y_i z_i K_{Ni}(x_k,z_k)W_{Nj}(z_i)$.

4. Simulations

We now carry out some simulation studies to illustrate the efficiency of the proposed method and compare the difference between the proposed method and existing method. Consider the following model:

$$y_i = 10z_i + a\sin(bx_i) + e_i, \qquad i=1,\ldots,N,$$

Table 1. $a = 1$, $b = 3$.

n	mean of $\hat{\alpha}$	mean of $\check{\alpha}$	variance of $\sqrt{n}\hat{\alpha}$	variance of $\sqrt{n}\check{\alpha}$	mean of $\hat{R}$	mean of $\check{R}$
100	9.987	9.985	0.099	0.103	0.090	0.091
200	10.000	10.002	0.062	0.071	0.094	0.094
300	9.999	9.995	0.033	0.037	0.084	0.084
400	9.998	10.000	0.032	0.035	0.078	0.078
500	9.994	9.997	0.031	0.036	0.076	0.076

Table 2. $a = 2$, $b = 3$.

n	mean of $\hat{\alpha}$	mean of $\check{\alpha}$	variance of $\sqrt{n}\hat{\alpha}$	variance of $\sqrt{n}\check{\alpha}$	mean of $\hat{R}$	mean of $\check{R}$
100	9.975	9.970	0.314	0.332	0.303	0.304
200	10.004	10.008	0.175	0.222	0.302	0.305
300	9.998	9.991	0.096	0.123	0.261	0.262
400	9.999	10.001	0.042	0.104	0.247	0.247
500	9.990	9.995	0.041	0.100	0.238	0.239

Table 3. $a = 3$, $b = 2$.

n	mean of $\hat{\alpha}$	mean of $\check{\alpha}$	variance of $\sqrt{n}\hat{\alpha}$	variance of $\sqrt{n}\check{\alpha}$	mean of $\hat{R}$	mean of $\check{R}$
100	9.986	9.971	0.154	0.235	0.413	0.414
200	10.001	10.006	0.140	0.144	0.305	0.306
300	9.997	9.994	0.124	0.075	0.255	0.259
400	9.995	10.001	0.123	0.062	0.260	0.257
500	10.001	9.996	0.121	0.053	0.228	0.228

Table 4. $a = 3$, $b = 3$.

n	mean of $\hat{\alpha}$	mean of $\check{\alpha}$	variance of $\sqrt{n}\hat{\alpha}$	variance of $\sqrt{n}\check{\alpha}$	mean of $\hat{R}$	mean of $\check{R}$
100	9.962	9.955	0.667	0.710	0.661	0.662
200	10.008	10.014	0.367	0.463	0.647	0.654
300	9.997	9.987	0.207	0.267	0.554	0.557
400	9.998	10.002	0.164	0.216	0.529	0.528
500	9.987	9.993	0.158	0.203	0.507	0.509

where (x_i, z_i) and e_i are independent with $(x_i, z_i) \sim U[0,1]^2$ and $e_i \sim U[-0.25, 0.25]$, a and b are constants. Obviously a and b can reorganize respectively the nonparametric and nonlinear influences because the larger $|a|$ is, the more nonparametric influence there is, and the larger $|b|$ is, the more nonlinear influence there is in this model.

To construct the estimators proposed by (3.9) and (3.10), the weigh functions are chosen as

$$K_{ni}(x, z) = \frac{G((x, z), (x_i, z_i))}{\sum_{j=1}^{N} G((x, z), (x_j, z_j))},$$

where

$$G((x,z),(x_i,z_i)) = \frac{1}{h_1h_2\sqrt{2\pi}} \exp\left\{-\frac{(x-x_i)^2}{2h_1^2} - \frac{(z-z_i)^2}{2h_2^2}\right\},$$

where h_1 and h_2 are bandwidth satisfying $h_1, h_2 = O(N^{-1/6})$. In the classical estimators of (3.15) and (3.16), the weigh functions are chosen as

$$\breve{K}_{ni}(x) = \frac{H(x,x_i)}{\sum_{j=1}^N H(x,x_j)},$$

where

$$H(x,x_i) = \frac{1}{h_3\sqrt{2\pi}} \exp\left\{-\frac{(x-x_i)^2}{2h_3^2}\right\},$$

where h_3 is bandwidth satisfying $h_3 = O(N^{-1/5})$.

Tables 1–4 report the simulated values of estimators $\hat{\alpha}$ and $\breve{\alpha}$, where $\hat{\alpha}$ is the proposed estimator based on weight functions $K_{Ni}(x,z)$ and $\breve{\alpha}$ is the classical estimator based on weight functions $\breve{K}_{Ni}(x)$. These results are the average of total 200 simulations, the sample size n is ranged from 100 to 500. According to the numerical results, we can see that, the variance of $\sqrt{n}\hat{\alpha}$ is significantly smaller than that of $\sqrt{n}\breve{\alpha}$ for the models with larger a and b. This implies that, to induce the variance, if the model has strongly nonparametric and nonlinear influences, the proposed method is better than the existing method. At the same time, $|\hat{\alpha}-10|$ and $\hat{R}$ are respectively smaller than $|\breve{\alpha}-10|$ and $\breve{R}$ slightly, where $\hat{R}$ and $\breve{R}$ are the residual sums of squares caused respectively by $\hat{\alpha}$ and $\breve{\alpha}$. These results are expected in our theory.

Acknowledgements

Authors would like to thank the Associate Editor and the two referees for their constructive suggestions.

Appendix

Theorem 1 and Theorem 2 need respectively the following conditions.

(A_1) In parametric model, $E_\theta(u(y,\theta_0)u'(y,\theta_0))$ is a positive definite matrix, $\partial^3 u(y,\theta)/\partial\theta^3$ is continuous in a neighborhood of θ_0, $\|\partial u(y,\theta)/\partial\theta\|$ and $\|u(y,\theta)\|^3$ are bounded by some integrable function $G(y)$ in the neighborhood, the rank of $E(\partial u(y,\theta)/\partial\theta)$ is p. Furthermore, the following integrations exist:

$$E((\partial u(y,\theta)/\partial\alpha)u'(y,\theta)), \quad E((\partial u(y,\theta)/\partial\beta_i)u'(y,\theta)), \quad E(\partial u(y,\theta)/\partial\alpha\partial\beta_i),$$
$$E((\partial u(y,\theta)/\partial\alpha\partial\beta_i)u'(y,\theta)), \quad E((\partial u(y,\theta)/\partial\alpha)\partial u'(y,\theta)/\partial\beta_i).$$

(A_2) In the semiparametric model, the condition is the same as (A_1) if all expectations above are replaced by conditional expectations given x and z.

PROOF OF LEMMA 2.1. For simplicity, we assume that β is one-dimensional parameter. By Condition (A_1), we can verify that

$$\sup_{(\alpha,\beta)} \left| -\frac{2}{N}\frac{\partial L_E(\alpha,\beta)}{\partial\beta} - \frac{\partial\zeta(\alpha,\beta)}{\partial\beta}\right| = o_P(1).$$

Then under Condition (B_1), for any $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$\begin{aligned} P\left\{\sup_{\alpha}|\hat{\beta}_\alpha - \beta_\alpha| \geq \varepsilon\right\} &\leq P\left\{\sup_{\alpha}\left|\frac{\partial\zeta(\alpha,\hat{\beta}_\alpha)}{\partial\beta}\right| \geq \delta\right\} \\ &= P\left\{\sup_{\alpha}\left|-\frac{2}{N}\frac{\partial L_E(\alpha,\hat{\beta}_\alpha)}{\partial\beta} - \frac{\partial\zeta(\alpha,\hat{\beta}_\alpha)}{\partial\beta}\right| \geq \delta\right\} \\ &\leq P\left\{\sup_{(\alpha,\beta)}\left|-\frac{2}{N}\frac{\partial L_E(\alpha,\beta)}{\partial\beta} - \frac{\partial\zeta(\alpha,\beta)}{\partial\beta}\right| \geq \delta\right\} \\ &= o(1), \end{aligned}$$

implying (2.12).

Expanding $\frac{\partial L_E(\alpha_0,\hat{\beta}_{\alpha_0})}{\partial\beta} = 0$ at β_0, we get

$$\begin{aligned} 0 = {} & \frac{\partial\bar{u}'}{\partial\beta}S^{-1}\bar{u} - \bar{u}'S^{-1}\frac{\partial S}{\partial\beta}S^{-1}\bar{u} + \bar{u}'S^{-1}\frac{\partial\bar{u}}{\partial\beta} \\ & + \left\{-\frac{\partial\bar{u}'}{\partial\beta}S^{-1}\frac{\partial S}{\partial\beta}S^{-1}\bar{u} + 2\frac{\partial\bar{u}'}{\partial\beta}S^{-1}\frac{\partial\bar{u}}{\partial\beta}\right. \\ & - \frac{\partial\bar{u}'}{\partial\beta}S^{-1}\frac{\partial S}{\partial\beta}S^{-1}\bar{u} + 2\bar{u}'S^{-1}\frac{\partial S}{\partial\beta}S^{-1}S^{-1}\frac{\partial S}{\partial\beta}S^{-1}\bar{u} - \bar{u}'S^{-1}\frac{\partial^2 S}{\partial\beta^2}S^{-1}\bar{u} \\ & \left. - 2\bar{u}'S^{-1}\frac{\partial S}{\partial\beta}S^{-1}\frac{\partial\bar{u}}{\partial\beta} + \bar{u}S^{-1}\frac{\partial^2\bar{u}}{\partial\beta^2}\right\}(\hat{\beta}_{\alpha_0} - \beta_0) + o_P(\|\hat{\beta}_{\alpha_0} - \beta_0\|), \end{aligned} \tag{A.1}$$

where all terms are valued at (α_0, β_0). It follows from Condition (A_1) and $E_0(u) = 0$ that

$$\begin{aligned} &\bar{u} = O_P(N^{-1/2}), \quad S = O_P(1), \quad \frac{\partial\bar{u}}{\partial\beta} = O_P(1), \\ &\frac{\partial S}{\partial\beta} = O_P(1), \quad \frac{\partial^2\bar{u}}{\partial\beta^2} = O_P(1), \quad \frac{\partial^2 S}{\partial\beta^2} = O_P(1). \end{aligned} \tag{A.2}$$

By comparing the convergence orders of all terms in (A.1), we can get (2.13).

Furthermore, from (2.11), we get

$$\frac{\partial}{\partial\alpha}\left\{\frac{\partial L_E(\alpha,\beta_\alpha)}{\partial\beta}\bigg|_{\beta_\alpha=\hat{\beta}_\alpha}\right\}_{\alpha=\alpha_0} = 0,$$

i.e.,

$$\begin{aligned} & \frac{\partial^2\bar{u}'}{\partial\alpha\partial\beta}S^{-1}\bar{u} - \frac{\partial\bar{u}'}{\partial\beta}S^{-1}\frac{\partial S}{\partial\alpha}S^{-1}\bar{u} + 2\frac{\partial\bar{u}'}{\partial\beta}S^{-1}\frac{\partial\bar{u}}{\partial\alpha} - \frac{\partial\bar{u}'}{\partial\alpha}S^{-1}\frac{\partial S}{\partial\beta}S^{-1}\bar{u} \\ & + 2\bar{u}'S^{-1}\frac{\partial S}{\partial\alpha}S^{-1}\frac{\partial S}{\partial\beta}S^{-1}\bar{u} - \bar{u}'S^{-1}\frac{\partial^2 S}{\partial\alpha\partial\beta}S^{-1}\bar{u} - \bar{u}'S^{-1}\frac{\partial S}{\partial\beta}S^{-1}\frac{\partial\bar{u}}{\partial\alpha} \\ & - \bar{u}'S^{-1}\frac{\partial\bar{u}'}{\partial\alpha}S^{-1}\frac{\partial\bar{u}}{\partial\beta} + \bar{u}'S^{-1}\frac{\partial^2\bar{u}'}{\partial\alpha\partial\beta} = 0, \end{aligned} \tag{A.3}$$

where all terms are valued at $\beta = \hat{\beta}_\alpha$ and $\alpha = \alpha_0$. Similar to (A.2),

$$\frac{\partial^2 \bar{u}}{\partial\beta\partial\alpha} = O_P(1), \qquad \frac{\partial^2 S}{\partial\beta\partial\alpha} = O_P(1). \tag{A.4}$$

By above result, (A.2) and comparing the convergence orders of all terms in (A.3), we have

$$\frac{\partial \bar{u}'}{\partial\beta} S^{-1} \frac{\partial \bar{u}}{\partial\alpha} + \frac{\partial \bar{u}'}{\partial\alpha} S^{-1} \frac{\partial \bar{u}}{\partial\beta} = O_P(N^{-1/2}),$$

i.e.,

$$\left(\frac{\partial \bar{u}}{\partial\alpha} + \frac{\partial \bar{u}}{\partial\beta} \frac{d\hat{\beta}_\alpha}{d\alpha} \right)' S^{-1} \frac{\partial \bar{u}}{\partial\beta} = O_P(N^{-1/2}), \tag{A.5}$$

implying (2.14). The proof is completed.

Proof of Theorem 2.1. It suffices to prove that as $N \to \infty$,

$$\frac{1}{\sqrt{N}} \frac{\partial L_E(\alpha, \hat{\beta}_\alpha)}{\partial\alpha} \bigg|_{\alpha=\alpha_0} = \frac{1}{\sqrt{N}} \frac{\partial L_E(\alpha, \beta_\alpha)}{\partial\alpha} \bigg|_{\alpha=\alpha_0} + o_P(1)$$

and

$$\sup_\alpha \left| \frac{1}{N} \frac{\partial^2 L_E(\alpha, \hat{\beta}_\alpha)}{\partial\alpha^2} - \frac{1}{N} \frac{\partial^2 L_E(\alpha, \beta_\alpha)}{\partial\alpha^2} \right| = o_P(1).$$

By Taylor expansion, (A.2) and (A.4),

$$\begin{aligned}
&\frac{1}{\sqrt{N}} \frac{\partial L_E(\alpha, \hat{\beta}_\alpha)}{\partial\alpha} \bigg|_{\alpha=\alpha_0} - \frac{1}{\sqrt{N}} \frac{\partial L_E(\alpha, \beta_\alpha)}{\partial\alpha} \bigg|_{\alpha=\alpha_0} \\
&\quad = \frac{1}{\sqrt{N}} \left\{ \frac{\partial^2 L_E(\alpha, \beta_\alpha)}{\partial\beta\partial\alpha} \right\}_{\alpha=\alpha_0} (\hat{\beta}_{\alpha_0} - \beta_0) + o_P(1) \\
&\quad = \frac{\sqrt{N}}{2} \left\{ \left(\frac{\partial \bar{u}}{\partial\alpha} + \frac{\partial \bar{u}}{\partial\beta} \frac{d\hat{\beta}_\alpha}{d\alpha} \right)' S^{-1} \frac{\partial \bar{u}}{\partial\beta} \right\}_{\alpha=\alpha_0} (\hat{\beta}_{\alpha_0} - \beta_0) + o_P(1).
\end{aligned}$$

From the result above, Lemma 2.1 and (A.5) it follows that the first term in right-hand side of above expression is asymptotically zero in probability. Then the first result holds. By (2.12) and the same methods above we can prove the second result. The proof is completed.

Proof of Lemma 3.1. By the weight conditions (i)–(iv), we can prove that

$$\sup_{(\alpha,\beta)} \sup_{(x,z)} \left| -\frac{2}{N} q_\beta(\alpha, \beta, x, z) - \xi(\alpha, \beta, x, z) \right| = o_P(1).$$

Then under Condition (B_2), for any $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$P \left\{ \sup_\alpha \sup_{(x,z)} |\hat{\lambda}_\alpha(x, z) - \lambda_\alpha(x, z)| \geq \varepsilon \right\}$$

$$\le P\left\{\sup_{\alpha}\sup_{(x,z)}|\xi(\alpha,\hat{\lambda}_\alpha(x,z),z)|\ge\delta\right\}$$
$$= P\left\{\sup_{\alpha}\sup_{(x,z)}\left|-\frac{2}{N}q_\beta(\alpha,\hat{\lambda}_\alpha(x,z),x,z)-\xi(\alpha,\hat{\lambda}_\alpha(x,z),z)\right|\ge\delta\right\}$$
$$\le P\left\{\sup_{\alpha}\sup_{(x,z)}\sup_{\beta}\left|-\frac{2}{N}q_\beta(\alpha,\beta,x,z)-\xi(\alpha,\beta,x,z)\right|\ge\delta\right\}$$
$$= o(1),$$

implying (3.11).

Expanding $q_\beta(\alpha_0,\hat{\lambda}_{\alpha_0}(x,z))=0$ at $\lambda_0(x)$, we get

$$\begin{aligned}0 &= \sum_{i=1}^N \frac{\partial h(\alpha_0,\beta,z_i)}{\partial\beta}\bigg|_{\beta=\lambda_0(x)}(y_i-h(\alpha_0,\lambda_0(x),z_i))K_{Ni}(x,z)\\ &\quad+\sum_{i=1}^N K_{Ni}(x,z)\left\{\frac{\partial^2 h(\alpha_0,\beta,z_i)}{\partial\beta^2}\bigg|_{\beta=\lambda_0(x)}(y_i-h(\alpha_0,\lambda_0(x),z_i))\right.\\ &\qquad\left.-\left(\frac{\partial h(\alpha_0,\beta,z_i)}{\partial\beta}\right)^2\bigg|_{\beta=\lambda_0(x)}\right\}(\hat{\lambda}_{\alpha_0}(x,z)-\lambda_0(x))\\ &\quad+r_N,\end{aligned}\tag{A.6}$$

where the convergence order of r_N is smaller than that of the other terms in the right-hand side of (A.6). It can be verified that, given (x,z),

$$\begin{aligned}&\sum_{i=1}^N \frac{\partial h(\alpha_0,\beta,z_i)}{\partial\beta}\bigg|_{\beta=\lambda_0(x)}(y_i-h(\alpha_0,\lambda_0(x),z_i))K_{Ni}(x,z)\\ &\quad\xrightarrow{P} E_0\left(\frac{\partial h(\alpha_0,\beta,z)}{\partial\beta}\bigg|_{\beta=\lambda_0(x)}(y-h(\alpha_0,\lambda_0(x),z))\mid x,z\right)\\ &\quad=0\end{aligned}$$

and

$$\begin{aligned}&\mathrm{Var}_0\left(\sum_{i=1}^N \frac{\partial h(\alpha_0,\beta,z_i)}{\partial\beta}\bigg|_{\beta=\lambda_0(x)}(y_i-h(\alpha_0,\lambda_0(x),z_i))K_{Ni}(x,z)\right)\\ &\quad=O\left(\sum_{i=1}^N K_{Ni}^2(x,z)\right).\end{aligned}$$

Then

$$\begin{aligned}&\sum_{i=1}^N \frac{\partial h(\alpha_0,\beta,z_i)}{\partial\beta}\bigg|_{\beta=\lambda_0(x)}(y_i-h(\alpha_0,\lambda_0(x),z_i))K_{Ni}(x,z)\\ &\quad=O_P\left(\left(\sum_{i=1}^N K_{Ni}^2(x,z)\right)^{1/2}\right)=O_P(N^{-1/2}).\end{aligned}\tag{A.7}$$

Similarly,

$$\sum_{i=1}^{N} \frac{\partial^2 h(\alpha_0,\beta,z_i)}{\partial\beta^2}\bigg|_{\beta=\lambda_0(x)} (y_i - h(\alpha_0,\lambda_0(x),z_i))K_{Ni}(x,z) = O_P(N^{-1/2}).$$

However,

$$\sum_{i=1}^{N} \left(\frac{\partial h(\alpha_0,\beta,z_i)}{\partial\beta}\right)^2 \bigg|_{\beta=\lambda_0(x)} K_{Ni}(x,z) = O_P(1).$$

By comparing the convergence orders of all terms in (A.6), we can ensure that (3.12) holds.

Furthermore, expanding $\frac{\partial}{\partial\alpha} q_\beta(\alpha,\hat\lambda_\alpha(x,z))|_{\alpha=\alpha_0} = 0$ at $\lambda_0(x)$, by (3.12) and the same argument as proving (A.7), we get

$$\begin{aligned}
0 &= \frac{\partial}{\partial\alpha} q_\beta(\alpha,\hat\lambda_\alpha(x,z)) \,|_{\alpha=\alpha_0}\\
&= \sum_{i=1}^{N}\left(\frac{\partial^2 h(\alpha_0,\beta,z_i)}{\partial\alpha\partial\beta}\bigg|_{\beta=\lambda_0(x)} + \frac{\partial^2 h(\alpha_0,\beta,z_i)}{\partial\beta^2}\bigg|_{\beta=\lambda_0(x)} \frac{d\hat\lambda_\alpha(x,z)}{d\alpha}\bigg|_{\alpha=\alpha_0}\right)\\
&\quad \times (y_i - h(\alpha_0,\lambda_0(x),z_i))K_{Ni}(x,z)\\
&\quad - \sum_{i=1}^{N} \frac{\partial h(\alpha_0,\beta,z_i)}{\partial\beta}\bigg|_{\beta=\lambda_0(x)}\\
&\quad \times \left(\frac{\partial h(\alpha_0,\lambda_0(x),z_i)}{\partial\alpha} + \frac{\partial h(\alpha_0,\beta,z_i)}{\partial\beta}\bigg|_{\beta=\lambda_0(x)} \frac{d\hat\lambda_\alpha(x,z)}{d\alpha}\bigg|_{\alpha=\alpha_0}\right) K_{Ni}(x,z)\\
&\quad + o_P(1)\\
&= -\sum_{i=1}^{N} \frac{\partial h(\alpha_0,\beta,z_i)}{\partial\beta}\bigg|_{\beta=\lambda_0(x)}\\
&\quad \times \left(\frac{\partial h(\alpha_0,\lambda_0(x),z_i)}{\partial\alpha} + \frac{\partial h(\alpha_0,\beta,z_i)}{\partial\beta}\bigg|_{\beta=\lambda_0(x)} \frac{d\hat\lambda_\alpha(x,z)}{d\alpha}\bigg|_{\alpha=\alpha_0}\right) K_{Ni}(x,z)\\
&\quad + o_P(1),
\end{aligned}$$

implying (3.13).

PROOF OF LEMMA 3.2. Under the weight conditions, for any $\varepsilon > 0$, when n is large enough, there is a $\delta > 0$ such that

$$\begin{aligned}
&\Bigg|\sum_{j=1}^{N} \frac{\partial u(y_i,\alpha_0,\lambda_0(x_i),z_i)}{\partial\alpha}\left(\frac{\partial u(y_j,\alpha_0,\beta,z_j)}{\partial\beta}\bigg|_{\beta=\lambda_0(x_i)}\right)^2 K_{nj}(x_i,z_i)\\
&\qquad - \sum_{j=1}^{N} \frac{\partial u(y_j,\alpha_0,\beta,z_j)}{\partial\beta}\bigg|_{\beta=\lambda_0(x_i)}\\
&\qquad \times \frac{\partial u(y_j,\alpha_0,\lambda_0(x_i),z_j)}{\partial\alpha}\frac{\partial u(y_j,\alpha_0,\beta,z_j)}{\partial\beta}\bigg|_{\beta=\lambda_0(x_i)} K_{nj}(x_i,z_i)\Bigg|
\end{aligned}$$

$$
\begin{aligned}
\le \sum_{j=1}^{N} \Bigg| & \frac{\partial u(y_i,\alpha_0,\lambda_0(x_i),z_i)}{\partial \alpha} \left(\frac{\partial u(y_j,\alpha_0,\beta,z_j)}{\partial \beta} \Bigg|_{\beta=\lambda_0(x_i)} \right)^2 \\
& - \frac{\partial u(y_j,\alpha_0,\beta,z_j)}{\partial \beta} \Bigg|_{\beta=\lambda_0(x_i)} \\
& \times \frac{\partial u(y_j,\alpha_0,\lambda_0(x_i),z_j)}{\partial \alpha} \frac{\partial u(y_j,\alpha_0,\beta,z_j)}{\partial \beta} \Bigg|_{\beta=\lambda_0(x_i)} \Bigg| \\
& \times K_{nj}(x_i,z_i) I(\|(x_i,z_i)-(x_j,z_j)\| > \delta) \\
+ \sum_{j=1}^{N} \Bigg| & \frac{\partial u(y_i,\alpha_0,\lambda_0(x_i),z_i)}{\partial \alpha} \left(\frac{\partial u(y_j,\alpha_0,\beta,z_j)}{\partial \beta} \Bigg|_{\beta=\lambda_0(x_i)} \right)^2 \\
& - \frac{\partial u(y_j,\alpha_0,\beta,z_j)}{\partial \beta} \Bigg|_{\beta=\lambda_0(x_i)} \\
& \times \frac{\partial u(y_j,\alpha_0,\lambda_0(x_i),z_j)}{\partial \alpha} \frac{\partial u(y_j,\alpha_0,\beta,z_j)}{\partial \beta} \Bigg|_{\beta=\lambda_0(x_i)} \Bigg| \\
& \times K_{nj}(x_i,z_i) I(\|(x_i,z_i)-(x_j,z_j)\| \le \delta) \\
\le C_2 \varepsilon & \sup_{x,z\in[0,1]} \frac{\partial u(y,\alpha_0,\lambda_0(x),z)}{\partial \alpha} \left(\frac{\partial u(y,\alpha_0,\beta,z)}{\partial \beta} \Bigg|_{\beta=\lambda_0(x)} \right)^2 + C_2\varepsilon.
\end{aligned}
$$

By the result above and Lemma 3.1, we get that, for any continuous function $v(x)$ on $[0,1]$,

$$
\begin{aligned}
& \left(\frac{\partial \bar u}{\partial \alpha} + \frac{\partial \bar u}{\partial \beta} \Bigg|_{\beta=\lambda_0(x)} \frac{d\hat\lambda_\alpha(x,z)}{d\alpha} \Bigg|_{\alpha=\alpha_0} \right) S^{-1}(\alpha_0,\beta_0(x)) \frac{\partial \bar u}{\partial \beta} \Bigg|_{\beta=\lambda_0(x)} v(x) \\
&= \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\partial u(y_i,\alpha_0,\lambda_0(x_i),z_i)}{\partial \alpha} - \frac{\partial u(y_i,\alpha_0,\beta,z_i)}{\partial \beta} \Bigg|_{\beta=\lambda_0(x_i)} \right. \\
& \qquad \left. \times \frac{\sum_{j=1}^{N} \frac{\partial u(y_j,\alpha_0,\lambda_0(x_i),z_j)}{\partial \alpha} \frac{\partial u(y_j,\alpha_0,\beta,z_j)}{\partial \beta} \Big|_{\beta=\lambda_0(x_i)} K_{nj}(x_i,z_i)}{\sum_{j=1}^{N} \left(\frac{\partial u(y_j,\alpha_0,\beta,z_j)}{\partial \beta} \Big|_{\beta=\lambda_0(x_i)} \right)^2 K_{nj}(x_i,z_i)} \right) \\
& \qquad \times S^{-1}(\alpha_0,\beta_0(x)) \frac{\partial \bar u}{\partial \beta} \Bigg|_{\beta=\beta_0(x)} v(x) + o_P(1) \\
&= o_P(1).
\end{aligned}
$$

The proof is completed.

PROOF OF THEOREM 3.1. The note after Lemma 3.1 shows that $\lambda_\alpha(x,z)$ is a conditionally least-favorable curve. Then, to prove the Theorem 3.1, it is sufficient to verify that

$$\text{(A.8)} \qquad \frac{1}{\sqrt{N}} \frac{\partial L_E(\alpha, \hat{\lambda}_\alpha(x,z))}{\partial \alpha}\bigg|_{\alpha=\alpha_0} = \frac{1}{\sqrt{N}} \frac{\partial L_E(\alpha, \lambda_\alpha(x,z))}{\partial \alpha}\bigg|_{\alpha=\alpha_0} + o_P(1).$$

and

$$\text{(A.9)} \qquad \sup_{\alpha \in \alpha} \left| \frac{1}{N} \frac{\partial L_E(\alpha, \hat{\lambda}_\alpha(x,z))}{\partial \alpha}\bigg|_{\alpha=\alpha_0} - \frac{1}{N} \frac{\partial L_E(\alpha, \lambda_\alpha(x,z))}{\partial \alpha}\bigg|_{\alpha=\alpha_0} \right| = o_P(1).$$

In fact, by Lemma 3.1, Lemma 3.2, $\lambda_{\alpha_0}(x,z) = \lambda_0(x)$ and a similar argument to proof of Lemma 2.1, the Taylor expansion leads to

$$\begin{aligned}
\text{(A.10)} \quad & \frac{1}{\sqrt{N}} \frac{\partial L_E(\alpha, \hat{\lambda}_\alpha(x,z))}{\partial \alpha}\bigg|_{\alpha=\alpha_0} - \frac{1}{\sqrt{N}} \frac{\partial L_E(\alpha, \lambda_\alpha(x,z))}{\partial \alpha}\bigg|_{\alpha=\alpha_0} \\
&= \frac{1}{\sqrt{N}} \frac{\partial L_E(\alpha_0, \beta)}{\partial \beta}\bigg|_{\beta=\hat{\lambda}_{\alpha_0}(x,z)} (\hat{\lambda}_{\alpha_0}(x,z) - \lambda_0(x)) + o_P(1) \\
&= \frac{1}{\sqrt{N}} \left(\frac{\partial \bar{u}}{\partial \alpha} + \frac{\partial \bar{u}}{\partial \beta}\bigg|_{\beta=\lambda_0(x)} \frac{d\hat{\lambda}_\alpha(x,z)}{d\alpha}\bigg|_{\alpha=\alpha_0} \right) S^{-1}(\alpha_0, \beta_0(x)) \frac{\partial \bar{u}}{\partial \beta}\bigg|_{\beta=\lambda_0(x)} \\
&\quad \times (\hat{\lambda}_{\alpha_0}(x,z) - \lambda_0(x)) + o_P(1) \\
&= o_P(1),
\end{aligned}$$

which implies (A.8). By (3.11) and the same methods above we can prove (A.9).

Proof of Lemma 3.3. Denote $E(\frac{\partial u(y,\alpha,\beta,z)}{\partial \beta} \mid z)$ by $M(\alpha, \beta, z)$. For any $\varepsilon > 0$, when N is large enough, there is a $\delta > 0$ such that

$$\begin{aligned}
&\sup_{z, z_j \in [0,1]} |E(M_N(\alpha, \beta, z) \mid z, z_j) - M(\alpha, \beta, z)| \\
&\quad \le \sup_{z, z_j \in [0,1]} \sum_{j=1}^N W_{Nj}(z) |M(\alpha, \beta, z_j) - M(\alpha, \beta, z)| I(|z_j - z| > \delta) \\
&\qquad + \sup_{z, z_j \in [0,1]} \sum_{j=1}^N W_{Nj}(z) |M(\alpha, \beta, z_j) - M(\alpha, \beta, z)| I(|z_j - z| \le \delta) \\
&\quad = 2 \sup_{z \in [0,1]} |M(\alpha, \beta, z)| \varepsilon + \varepsilon.
\end{aligned}$$

Then

$$\text{(A.11)} \qquad \lim_{N \to \infty} \sup_{z, z_j \in [0,1]} |E(M_N(\alpha, \beta, z) \mid z, z_i) - M(\alpha, \beta, z)| = 0.$$

Furthermore, by the law of iterated logarithm, we have

$$\begin{aligned}
\text{(A.12)} \qquad & |E(M_N(\alpha, \beta, z) \mid z, z_i) - M_N(\alpha, \beta, z)| \\
&= O\left(\left(\sup_i K_{Ni}^2(x,z) N \log\log N \right)^{1/2} \right) \\
&\to 0.
\end{aligned}$$

Therefore (A.11) and (A.12) ensure that

$$M_N(\alpha,\beta,z) \overset{\text{a.s.}}{\to} E\left(\frac{\partial u(y,\alpha,\beta,z)}{\partial\beta} \;\middle|\; z\right).$$

This finishes the proof.

PROOF OF THEOREM 3.2. It is similar to the proof of Theorem 2.

REFERENCES

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions, *Journal of Econometrics*, **34**, 305–334.

Lin, L. and Zhang, R. C. (2002). Profile quasi-likelihood, *Statistics and Probability Letters*, **56**, 147–154.

Luo, X. (1994). Large sample properties of the empirical Euclidean likelihood estimation for semiparametric model, *Chinese Journal of Applied Probability and Statistics*, **10**, 344–352.

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

Owen, A. (1990). Empirical likelihood ratio confidence intervals regions, *The Annals of Statistics*, **18**, 90–120.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations, *The Annals Statistics*, **22**, 300–325.

Severini, T. (1998). An approximation to the modified profile likelihood function, *Biometrika*, **85**, 403–411.

Severini, T. (1999). An empirical adjustment to the likelihood ratio statistic, *Biometrika*, **86**, 235–247.

Severini, T. (2002). Modified estimating functions, *Biometrika*, **89**, 333–343.

Severini, T. and Staniswalis, G. (1994). Quasi-likelihood estimation in semiparametric models, *Journal of the American Statistical Association*, **89**, 501–511.

Severini, T. and Wong, W. H. (1992). Profile likelihood and conditionally parametric models, *The Annals of Statistics*, **20**, 1768–1802.

Shi, J. (2000). Empirical likelihood for partially linear Models, *Journal of Multivariate Analysis*, **72**, 132–148.

Small, C. G. and Mcleish, D. L. (1994). *Hilbert Space Methods in Probability and Statistical Inference*, John Wiley and Sons, New York.

Stein, C. (1956). Efficient nonparametric testing and estimating, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1, 187–196, University of California Press, Berkeley.