Statistica Sinica **20** (2010), 1771-1787

ADAPTIVE GLOBAL CONFIDENCE BAND FOR NONPARAMETRIC REGRESSION: AN EMPIRICAL LIKELIHOOD METHOD

Li-Xing Zhu[1], Lu Lin[2] and Qiang Chen[2,3]

[1]*Hong Kong Baptist University,* [2]*Shandong University and* [3]*Qilu Securities Co., Ltd*

Abstract: In this paper, we construct adaptive global confidence bands for nonparametric regression functions by empirical likelihood (EL). First, we show that the size of the classical EL-based confidence region is not adaptive to the submodels of the function in rate-optimal way, that is, it is not model-adaptive. In contrast, the existing model-adaptive methods are not data-adaptive, that is, the shapes of the resulting confidence regions are not determined by data. Thus, we propose an EL-based method to construct model-data-adaptive global confidence bands for nonparametric regression models with some constraints. The key remark is that the size (radius) of the confidence region is not determined by the (asymptotic) distribution but by a U-statistic that is highly related to the smoothness of the submodels. The newly proposed confidence region has the model-data-adaptive property: the size adapts to the submodels in a rate-optimal way and its shape is determined by the data. Implementation issue is investigated, and simulations are carried out for illustration.

Key words and phrases: Adaptability, confidence region, empirical likelihood, nonparametric regression.

1. Introduction

Empirical likelihood (EL) is a powerful tool for statistical inference. This method does not require a full specification of distribution from which data are drawn, but only an unbiased estimating function. The fundamental notion stems from the seminal work of Owen (1988, 1990); since then it has been further developed for various parametric, semiparametric, and nonparametric statistical problems. For examples, amongst others, see Owen (1991), Qin and Lawless (1994), Chen and Qin (1993), Chen and Hall (1993), DiCiccio, Hall, and Romano (1991) Hall (1990), Shen, Shi, and Wong (1999), Kitamura, Tripathi, and Ahn (2004), Lin, Zhu, and Yuen (2005), Zhu and Xue (2006), Stute, Xue, and Zhu (2007), Xue and Zhu (2007), and Hjort, McKeague, and Keilegom (2009). Some comprehensive treatments can be found in Owen (2001). One of the most

significant advantages of EL is that the shape of a confidence region is automatically determined by the data, EL is adaptive to data. Consequently, with a given coverage probability, the size of a confidence region is in general smaller than that of common confidence regions, such as those constructed in terms of point estimation. This holds since confidence regions are usually designed as a standard ball, see Zhu and Xue (2006).

Model-adaptability is another important notion for especially semiparametric and nonparametric models. Consider constructing confidence region for nonparametric function, the commonly used Bonferroni region is obviously too conservative, see Xue and Zhu (2007). One way around this is first to express the function as an infinite series of parametric functions, and then to consider an approximation of the region for the nonparametric function by using part of this series. Obviously, we need way to define a "good" approximation. And model-adaptability is then raised. Here adaptation means that the inference procedure adjusts automatically to properties of submodels of interest under an optimality criterion. The common submodels are the regular ellipsoid, anisotropic class, and Besov body. For example, if we consider the sequence space $l_2 = \{\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots)' : \sum_{i=1}^{\infty} \theta_i^2 < \infty\}$, the regular ellipsoid is $S(\beta, L) = \{\boldsymbol{\theta} \in l_2 : \sum_{i=1}^{\infty} \theta_i^2 i^{2\beta} \leq L^2\}$ for parameters $\beta > 0$ and given $L > 0$. Then model-adaptive inference, say adaptive confidence regions given, should adapt automatically to the choices of parameters β and L in a rate-optimal way. Model-adaptability has attracted much attention; see for example Li (1989), Beran and Dümbgen (1998), Birgé (2002), Hoffmann and Lepski (2002), Juditsky and Lambert-Lacroix (2003), Baraud (2004), Cai and Low (2004, 2005, 2006a,b), Genovese and Wasserman (2005), and Robins and Vaart (2006). According to the existing theories, a model-adaptive confidence region satisfies at least the following conditions: it is honest on the model of interest and, at the same time, its size adapts to submodels in a rate-optimal way. For details see Robins and Vaart (2006), Cai and Low (2004, 2005, 2006a,b), and Li (1989), among others.

From the above description, existing data-adaptive and model-adaptive method have, respectively, their pros and cons. Model-adaptive confidence regions are not data-adaptive because their shapes are usually designed as a standard ball not determined by data; see for example Robins and Vaart (2006) and Cai and Low (2006a). For instance, let $\theta_{j,k}, j = 0, \ldots, J-1, k = 1, \ldots, 2^j$, be wavelet coefficients of a nonparametric regression function r and $\boldsymbol{\theta}_j = (\theta_{j,1}, \ldots, \theta_{j,2^j})'$ be the coefficient at level j in a wavelet expansion of r. For this model, Cai and Low (2006a) proposed a model-adaptive confidence ball for $\boldsymbol{\theta}_j$ as $\{\boldsymbol{\theta}_j : \|\boldsymbol{\theta}_j - \boldsymbol{\delta}\| < d_\alpha\}$ with an adaptive choice of radius d_α. This is a standard ball and the shape is not determined by data. In contrast, we verify in Section 2 that the size of EL-based confidence region is large and cannot tend to zero even

when the submodel is small and the size of sample tends to infinity. Then it cannot be model-adaptive because its size cannot be adjusted automatically to the submodels of target functions.

Thus, constructing a confidence region that is adaptive to both data and model is of great interest. To the best of our knowledge, there is no reference in the literature. We propose a simple approach to achieve this goal: combining EL and a model-adaptive method to construct model-data-adaptive confidence regions for nonparametric regression. As we know, the commonly used EL for nonparametric regression is based mainly on local smoothing; see for example Chen (1996), Chen and Qin (2000), but for global smoothing and model-adaptation, it cannot be applied because we want to use the information on the submodels. We solve this problem in the following sections. In our procedure, the key fact is that the size (radius) of the confidence region is not determined by the (asymptotic) distribution but by a U-statistic that is highly related to the smoothness of the submodels. Classical methods (including EL) have the size of confidence regions usually determined by the (asymptotic) distribution and cannot obtain the adaptive choice of size.

The paper is organized as follows. In Section 2, the model and the required conditions are presented. The properties of EL are re-examined to show why it is not model-adaptive. In Section 3, a procedure for constructing adaptive confidence region is introduced and its model-data-adaptability is obtained. In Section 4 an implementation procedure is proposed. It contains data-driven methods to select a trade-off parameter, and some simulations are performed for illustration. Proofs are presented in Section 5.

2. Model and Property of EL-based Confidence Region

In this section we define the model under study and re-examine EL-based confidence regions. We see that their size of EL-based confidence region is not adaptive to the submodels of regression functions.

We consider the following nonparametric regression model with random design:

$$Y = r(X) + \varepsilon, \tag{2.1}$$

where $r(x) = E(Y|X = x)$ is an unknown regression function, $X \in \mathcal{X} \subset R^p$ is a p-dimensional random covariate with known distribution $F(x)$, $Y \in \mathcal{Y} \subset R$ is a real-valued response variable, and the error term ε satisfies $E(\varepsilon\,|X = x) = 0$ and $Var(\varepsilon\,|X = x) = \sigma^2(x)$. Here we need the known distribution $F(X)$ to construct the basis functions below. If $F(x)$ is unknown, without loss of the model-adaptive property, we can use a consistent estimator $\hat{F}(x)$ in place of $F(x)$ because model-adaptability is an asymptotic criterion. Although we consider random design here, the method can be similarly employed to fixed design cases.

Let $L_2(\mathcal{X})$ denote the set of functions $f : \mathcal{X} \to R$ such that $E(f^2(X)) < \infty$. As is well known, when $r \in L_2(\mathcal{X})$, it can be expressed as

$$r(x) = \sum_{j=0}^{\infty} \theta_j p_j(x), \tag{2.2}$$

where $p_0(x) = 1$ and $\{1, p_j(x) : j = 1, 2, \ldots\}$ is a set of basis functions, such as polynomial bases, Fourier bases or B-splines, satisfying

$$E(p_j(X)) = 0,\ E(p_j^2(X)) = 1 \ \text{ and } \ E(p_j(X)p_k(X)) = 0 \ \text{ for } j \neq k. \tag{2.3}$$

Because the distribution or an estimated distribution of X is assumed known, such basis functions can be obtained. With these representations, (2.1) can be rewritten as

$$Y = \sum_{j=0}^{\infty} \theta_j p_j(X) + \varepsilon. \tag{2.4}$$

Here we consider the parameter and function spaces: $\Theta = \Big\{\boldsymbol{\theta} : \sum_{j=0}^{\infty} \theta_j^2 < \infty\Big\}$ and $\mathcal{R} = \Big\{r(x) : \|r\|_F^2 < \infty\Big\}$, where $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots)'$ and $\|r\|_F^2 = E(r^2(X))$.

For global smoothing (Huang, Wu, and Zhou (2002)), we first approximate $r(x)$ by $r(x) \approx \sum_{j=0}^{N} \theta_j p_j(x)$ and then (2.1) can be approximately expressed as

$$Y \approx \sum_{j=0}^{N} \theta_j p_j(X) + \varepsilon, \tag{2.5}$$

where N is a trade-off parameter depending on submodels of Θ or $\mathcal{R}$. The detailed discussion about selecting N will be presented in Sections 3 and 4. Let $(X_i, Y_i), i = 1, \ldots, n$, be i.i.d. observations of (X, Y) at (2.1). Based on the approximate model (2.5), the empirical likelihood ratio (ELR) for part parameter vector $\boldsymbol{\theta}_N$ is

$$\Re(\boldsymbol{\theta}_N) = \sup\Big\{\prod_{i=1}^{n} n w_i : w_i \geq 0, \sum_{i=1}^{n} w_i = 1, \sum_{i=1}^{n} w_i P_N(X_i)(Y_i - P_N'(X_i)\boldsymbol{\theta}_N) = 0\Big\}, \tag{2.6}$$

where $\boldsymbol{\theta}_N = (\theta_0, \theta_1, \ldots, \theta_N)'$ and $P_N(X_i) = (1, p_1(X_i), \ldots, p_N(X_i))'$. Note that N depends on n and tends to infinity as $n \to \infty$. As shown by Hjort, McKeague, and Keilegom (2009), the distribution of $-2\log\Re(\boldsymbol{\theta}_N)$ can be approximated by a normal distribution with mean N and variance $2N$. In this situation, an approximate confidence region for $\boldsymbol{\theta}_N$ can be expressed as

$$\Big\{\boldsymbol{\theta}_N : -U_{\alpha/2}(2N)^{1/2} \leq -2\log\Re(\boldsymbol{\theta}_N) - N \leq U_{\alpha/2}(2N)^{1/2}\Big\}, \tag{2.7}$$

where $U_{\alpha/2}$ is the quantile of standard normal distribution.

As mentioned in the Introduction, the EL-based confidence region given in (2.7) satisfies data-adaptability for part parameter vector $\boldsymbol{\theta}_N$. On the other hand, while we expect to obtain a small radius when the the submodel of functions is small, *the "radius" of* (2.7)*,* $U_{\alpha/2}(2N)^{1/2}$*, is large and cannot tend to zero even when the submodel is small and* n *tends to infinity.* Then confidence region (2.7) is too large to derive interesting information about $\boldsymbol{\theta}_N$ (or r) in a large sample, and the model-adaptive property is not achieved.

We propose a confidence region for the full parameter vector $\boldsymbol{\theta}$ of the form

$$\hat{C}_n = \Big\{\boldsymbol{\theta} : d \le -2\log \Re_*(\boldsymbol{\theta}) \le D\Big\} \quad \text{or} \quad \hat{C}_n = \Big\{\boldsymbol{\theta} : -2\log \Re_*(\boldsymbol{\theta}) \le D\Big\}, \tag{2.8}$$

where $\Re_*(\boldsymbol{\theta})$ is an extended version of ELR for the parameter vector $\boldsymbol{\theta}$. Here the essential difference from (2.7) is that the radiuses d and D in (2.8) are not determined by the normal distribution but are determined by an adaptive method such that they are highly related to the smoothness of the submodels of Θ (or $\mathcal{R}$).

3. Model-data-adaptive Confidence Region

3.1. Construction of adaptive confidence region

By the adaptability requirement and the version proposed by Robins and Vaart (2006) and Li (1989), for example, a model-adaptive confidence region $\hat{C}_n$ of $\boldsymbol{\theta}$ should satisfy the following.

(i) $\hat{C}_n$ is honest at the model Θ in the sense that $\liminf_{n\to\infty} \inf_{\boldsymbol{\theta}\in\Theta} P_{\boldsymbol{\theta}}\{\boldsymbol{\theta} \in \hat{C}_n\} \ge 1-\alpha$, for given confidence level $1-\alpha$ with $0 < \alpha < 1$.

(ii) $\hat{C}_n$ is centered at an estimator of $\boldsymbol{\theta}$, for example, an adaptive estimator.

(iii) The radius of $\hat{C}_n$ adapts to submodels of Θ or $\mathcal{R}$ in a rate-optimal way.

For more details about the descriptions of adaptive estimation and confidence regions see, for example, Nishii (1984), Birgé and Massart (2001), Barron, Birgé and Massart (1999) and Baraud (2000), Li (1989), Cai and Low (2004, 2006b), and Robins and Vaart (2006). Informally, an adaptive estimation procedure automatically adjusts to the smoothness properties of the underlying functions and a common way to evaluate such a procedure is to compute its maximum risk over a collection of parametric spaces, and to compare these values to the minimax risk over each of them; this adaptive estimation procedure does not necessarily influence the construction of an adaptive confidence region. Thus, the existing papers mainly focus on the construction of the radius for a confidence region and assume that an adaptive estimator is given. As is shown above, the

radius of an adaptive confidence region should adjust to submodels of interest in a rate-optimal way while maintaining a prespecified coverage probability. Here the optimal rate is in fact the minimax rate of estimation for the given submodel and the common submodels are chosen as the subsets of a regular ellipsoid, and anisotropic class and a Besov body; see for example Cai and Low (2006a) Robins and Vaart (2006), and Hoffmann and Lepski (2002).

Here we use EL to construct adaptive confidence region as determined by (2.8) and, thus, the estimation for the center is not required. To valuate such a confidence region, we only need condition (i) and an extended version of condition (iii):

(iii)' The "radiuses" d and D in (2.8) adapt to submodels of Θ or $\mathcal{R}$ in a rate-optimal way.

Motivated by Robins and Vaart (2006), to facilitate construction of a confidence region, we split the sample into two subsamples. The first half of the data is denoted by $(\mathbf{X}^1, \mathbf{Y}^1) = \{(X_i, Y_i) : i = 1, \ldots, [\gamma n]\}$. Here $0 < \gamma < 1$ is given and is usually chosen to be $1/2$. Hereafter, the superscript 1 indicates reliance on the first half. The first half of the data is used to construct EL and the second to construct radiuses d and D. Based on the first half, the ELR for the parameter vector $\boldsymbol{\theta}_N$ is

$$\Re^1(\boldsymbol{\theta}_N) = \sup\Big\{ \prod_{i=1}^{[\gamma n]} [\gamma n] w_i : w_i \geq 0, \sum_{i=1}^{[\gamma n]} w_i = 1, \sum_{i=1}^{[\gamma n]} w_i P_N(X_i)(Y_i - P_N'(X_i)\boldsymbol{\theta}_N) = 0 \Big\}.$$

Further, an extended version log-ELR for the full parameter $\boldsymbol{\theta}$ is

$$-2\log \Re_*^1(\boldsymbol{\theta}) \triangleq -2\log \Re^1(\boldsymbol{\theta}_N) + [\gamma n] \sum_{j=N+1}^{\infty} \theta_j^2. \tag{3.1}$$

Then, similar to (2.8), our goal is to construct confidence region as

$$\hat{C}_n = \Big\{ \boldsymbol{\theta} : d \leq -2\log \Re_*^1(\boldsymbol{\theta}) \leq D \Big\} \quad \text{or} \quad \hat{C}_n = \Big\{ \boldsymbol{\theta} : -2\log \Re_*^1(\boldsymbol{\theta}) \leq D \Big\} \tag{3.2}$$

with d and D being adaptive choices of the radiuses. Then the remaining work is to estimate d and D so that conditions (i) and (iii)' hold.

To determine d and D, set $N = O(n^\delta)$ for $\delta > 0$, and

$$\mathbf{P}_N^1(X) = (P_N(X_1), \ldots, P_N(X_{[\gamma n]})), \ \mathbf{Y}^1 = (Y_1, \ldots, Y_{[\gamma n]})',$$

$$M_N^1 = \Big(\Phi_N^{-1} \frac{1}{[\gamma n]} \sum_{i=1}^{[\gamma n]} (Y_i - P_N'(X_i)\boldsymbol{\theta}_N)^2 P_N(X_i) P_N'(X_i) \Phi_N^{-1} \Big)^{-1},$$

$$\hat{\boldsymbol{\theta}}_N^1 = \Phi_N^{-1} \frac{1}{[\gamma n]} \mathbf{P}_N^1(X) \mathbf{Y}^1,$$

where $\Phi_N = (1/[\gamma n])\mathbf{P}_N^1(X)(\mathbf{P}_N^1(X))'$. Although M_N^1 depends on $\boldsymbol{\theta}_N$, hereafter we suppress $\boldsymbol{\theta}_N$ for convenience.

Lemma 3.1. *When $0 < \delta < 1/3$, if $p_j(X_i)$ and Y_i are uniformly bounded and the eigenvalues of M_N^1 are bounded away from zero and infinity, then*

$$-2\log\Re^1(\boldsymbol{\theta}_N) = [\gamma n](\hat{\boldsymbol{\theta}}_N^1 - \boldsymbol{\theta}_N)' M_N^1(\hat{\boldsymbol{\theta}}_N^1 - \boldsymbol{\theta}_N) + o_p(n^{\delta/2}),$$

$$-2\log\Re_*^1(\boldsymbol{\theta}) = [\gamma n](\hat{\boldsymbol{\theta}}_N^1 - \boldsymbol{\theta}_N)' M_N^1(\hat{\boldsymbol{\theta}}_N^1 - \boldsymbol{\theta}_N) + [\gamma n]\sum_{j=N+1}^{\infty}\theta_j^2 + o_p(n^{\delta/2}).$$

The proof of the lemma is in Section 5. The condition on eigenvalues in Lemma 3.1 is commonly assumed for models with a high-dimensional parameter; see for example Fan and Peng (2004). The condition on the boundedness of $p_j(X_i)$ holds for the Fourier basis, but not for the polynomial basis, wavelets *etc.* However, the condition on the boundedness of $p_j(X_i)$ and Y_i can replaced by other conditions; see for example Chen and Peng (2007).

From Lemma 3.1, we can write the leading term of $-2\log\Re_*^1(\boldsymbol{\theta})$ as

$$R_N(\boldsymbol{\theta}_N) = [\gamma n](\hat{\boldsymbol{\theta}}_N^1 - \boldsymbol{\theta}_N)' M_N^1(\hat{\boldsymbol{\theta}}_N^1 - \boldsymbol{\theta}_N). \tag{3.3}$$

From (2.3) and (2.4), $E(Y) = \theta_0$ and $E(Yp_j(X)) = \theta_j$ for $j \geq 1$, or $E(YP_N(X)) = \boldsymbol{\theta}_N$. Then, given the first half of the data $(\mathbf{X}^1, \mathbf{Y}^1)$, we estimate $R(\boldsymbol{\theta}_N)$ by

$$\hat{R}_N = \frac{[\gamma n]}{2L_n(L_n-1)}\sum_{i=[\gamma n]+1}^{n}\sum_{k\neq i, k=[\gamma n]+1}^{n}(\hat{\boldsymbol{\theta}}_N^1 - Y_iP_N(X_i))' M_N^1(\hat{\boldsymbol{\theta}}_N^1 - Y_kP_N(X_k)),$$

where $L_n = n - [\gamma n]$. Here $\hat{R}_N$ depends on $\boldsymbol{\theta}_N$ because M_N^1 contains $\boldsymbol{\theta}_N$. Without confusion, hereafter we suppress $\boldsymbol{\theta}_N$ for the convenience. Note that, although $\hat{R}_N$ depends on $\boldsymbol{\theta}_N$, it is similar to a U-statistic of R_n. Using this "U-statistic" to define the radius, the resulting confidence region was some desirable properties. Lemma 3.2 is a key step toward our results, the proof is postponed to Section 5.

Lemma 3.2. *Given $(\mathbf{X}^1, \mathbf{Y}^1)$, $Var(\hat{R}_N|(\mathbf{X}^1, \mathbf{Y}^1)) \leq \tau_{N,n}^2(\boldsymbol{\theta}_N)$, where*

$$\begin{aligned}\tau_{N,n}^2(\boldsymbol{\theta}_N) = {} & \frac{[\gamma n]\left((N+1)\|M_N^1\|^2(\|r\|_\infty^2 + \|\sigma^2\|_\infty)^2 - (\boldsymbol{\theta}_N' M_N^1\boldsymbol{\theta}_N)^2\right)}{L_n(L_n-1)} \\ & + \frac{2[\gamma n](\|r\|_\infty^2 + \|\sigma^2\|_\infty)}{L_n}(\hat{\boldsymbol{\theta}}_N^1 - \boldsymbol{\theta}_N)'(M_N^1)^2(\hat{\boldsymbol{\theta}}_N^1 - \boldsymbol{\theta}_N) \\ & - \frac{2[\gamma n]}{L_n}\Big((\hat{\boldsymbol{\theta}}_N^1 - \boldsymbol{\theta}_N)' M_N^1\boldsymbol{\theta}_N\Big)^2,\end{aligned}$$

with $\|M_N^1\|$ the maximum eigenvalue of M_N^1, $\|r\|_\infty^2 = \sup_{x\in\mathcal{X}}|r(x)|^2$, and $\|\sigma\|_\infty^2 = \sup_{x\in\mathcal{X}}|\sigma(x)|^2$.

Note that $E(\hat{R}_N|(\mathbf{X}^1,\mathbf{Y}^1)) = R(\boldsymbol{\theta}_N)$, where $R(\boldsymbol{\theta}_N)$ is defined at (3.3). Then the lemma implies that $\sup_{\boldsymbol{\theta}\in\Theta} E\Big(\Big((\hat{R}_N - R(\boldsymbol{\theta}_N))/(\tau_{N,n}(\boldsymbol{\theta}_N))\Big)^2\Big|(\mathbf{X}^1,\mathbf{Y}^1)\Big) \le 1$. From Markov's inequality, it then follows that

$$\inf_{\boldsymbol{\theta}\in\Theta} P_{\boldsymbol{\theta}}\Big(\Big|\hat{R}_N - R(\boldsymbol{\theta}_N)\Big| \le z_\alpha \tau_{N,n}(\boldsymbol{\theta}_N)\Big|(\mathbf{X}^1,\mathbf{Y}^1)\Big) \ge 1-\alpha, \tag{3.4}$$

where $z_\alpha = 1/\sqrt{\alpha}$. Letting the dimension go to infinity, we set $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2_{M_N^1} = R(\boldsymbol{\theta}_N) + [\gamma n]\sum_{j=N+1}^{\infty}\theta_j^2$, and we get the following.

Lemma 3.3. *For Model* (2.1)*, if the conditions of Lemma* 3.1 *hold and* $\hat{\boldsymbol{\theta}} \in \Theta$*, then*

$$\Big\{\boldsymbol{\theta}\in\Theta : \sqrt{d^*} \le \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_{M_N^1} \le \sqrt{D^*}\Big\}$$

is an honest $(1-\alpha)$*-confidence region for* $\boldsymbol{\theta}\in\Theta$*, where* $d^* = (\hat{R}_N - z_\alpha\tau_{N,n}(\boldsymbol{\theta}_N))_+$ *,* $D^* = \Big((\hat{R}_N + z_\alpha\tau_{N,n}(\boldsymbol{\theta}_N))^{1/2} + 2\sqrt{[\gamma n]}B_N\Big)^2$*, and* $B_N^2 = \sup_{\boldsymbol{\theta}\in\Theta}\sum_{j=N+1}^{\infty}\theta_j^2$*.*

The proof of this lemma is in Section 5. Combining Lemma 3.1 and Lemma 3.3,

$$\Big\{\boldsymbol{\theta}\in\Theta : \sqrt{d^*} + a_n \le (-2\log\Re_*^1(\boldsymbol{\theta}))^{1/2} \le \sqrt{D^*} + b_n\Big\}$$

is an honest $(1-\alpha)$-confidence region for $\boldsymbol{\theta}\in\Theta$, where $a_n = o_p(n^{\delta/4})$ and $b_n = o_p(n^{\delta/4})$. Moreover as is shown by Li (1989) and Baraud (2004), for n-dimensional space, if the confidence region is designed as a ball with the center at an estimator of the parameter vector, the $n^{-1/4}$ lower bound of average confidence region cannot be improved upon without losing full honesty. Furthermore, Lemma 3.3 shows that $-2\log\Re_*^1(\boldsymbol{\theta})$ can be approximately expressed as an ellipse with the center at an estimator and the shape matrix M_N^1 independent of the second half of the data. Then a_n and b_n in the above confidence region can be ignored because, by averaging, $n^{-1/2}a_n$ and $n^{-1/2}b_n$ are asymptotically smaller than $n^{-1/4}$. On the other hand, to guarantee the honesty as defined in (i) when a_n and b_n are omitted from the confidence region, we need the condition that (X,Y) is continuous.

Theorem 3.1. *For Model* (2.1)*, if the conditions in Lemma* 3.3 *hold,*

$$\hat{C}_n = \Big\{\boldsymbol{\theta}\in\Theta : d^* \le -2\log\Re_*^1(\boldsymbol{\theta}) \le D^*\Big\} \tag{3.5}$$

is an honest $(1-\alpha)$*-confidence region for* $\boldsymbol{\theta}\in\Theta$*, with* d^* *and* D^* *defined as in Lemma* 3.3*.*

The proof is postponed to Section 5. By Theorem 3.1,

$$\hat{C}_n = \Big\{\boldsymbol{\theta}\in\Theta : -2\log\Re_*^1(\boldsymbol{\theta}) \le D^*\Big\} \tag{3.6}$$

is an honest $(1-\alpha)$-confidence region for $\boldsymbol{\theta} \in \Theta$. We usually choose (3.6) as a practicable confidence region for $\boldsymbol{\theta}$ because Lemma 3.1 shows that it can be approximately an ellipse. Finally, by (3.5) or (3.6), we can construct the adaptive global confidence band for regression function r as

$$\Big\{(x,y) : y = r(x), r(x) = \sum_{j=0}^{\infty} \theta_j p_j(x), x \in \mathcal{X}, \boldsymbol{\theta} \in \hat{C}_n\Big\}, \tag{3.7}$$

where $\hat{C}_n$ is determined by (3.5) or (3.6).

While the confidence region obtained above is of a nice form and a clear mathematical description, the radius still depends on unknown functions r and σ. A statistical implementation is desired. We delay this issue to Section 4. Then too, the continuity of (X, Y) in the theorem can be replaced by some weaker conditions, but we not discuss this issue further here.

3.2. Adaptability properties

We turn to the adaptability properties of the honest confidence region of (3.5) or (3.6). Obviously, it inherits the favorable property of EL in that its shape is automatically adaptive to data. Furthermore, (3.5), (3.6) and the representation of $\tau_{N,n}$ show that

$$\frac{1}{\sqrt{n}}\mathrm{rad}(\hat{C}_n^j) = O_p\Big(\Big(\frac{N}{n^2}\Big)^{1/4} + B_N + \Big(\frac{R_N}{n}\Big)^{1/2}\Big). \tag{3.8}$$

This is a standard convergence order in nonparametric estimation; for details, see Proposition 2.1 of Robins and Vaart (2006). The above formula shows that, based on the property of submodels, we can choose N to minimize the radius of the confidence region. The following examples show how the confidence region (3.5) or (3.6) adapts to the submodels in a rate-optimal way.

Example 3.1. Suppose the regression function r can be expressed by a finite number of basis functions, say m independent of n, so $r(x) = \sum_{j=0}^{m} \theta_j p_j(x)$. In this case the radius of the confidence region is of the order of the maximum of $n^{-1/2}$ and the average of estimation error $(R_N/n)^{1/2}$. For this standard parametric model, the common method can ensure $(R_N/n)^{1/2} = O_p(n^{-1/2})$, so a radius of order $n^{-1/2}$. It is known that $n^{-1/2}$ is the standard rate in finite-dimensional parameter models, and cannot be improved without losing full honesty.

Example 3.2. Suppose that the regression function r can be expressed as $r(x) = \sum_{j=0}^{n} \theta_j p_j(x)$. To avoid bias, N is chosen to be n. In this case, the radius of the confidence region is of the order of the maximum of $n^{-1/4}$ and the average of estimation error $(R_N/n)^{1/2}$. As shown by Li (1989) and Baraud

(2004), for n-dimensional space, the $n^{-1/4}$ lower bound cannot be improved upon without losing full honesty. To achieve this best rate, we have to choose a favorite estimator $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}'_N, 0, \ldots)'$.

Example 3.3. Assume that the regression function r can be expressed by $r(x) = \sum_{j=0}^{\infty} \theta_j p_j(x)$, where $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots)'$ is in the regular ellipsoid $\Theta = \{\boldsymbol{\theta} : \sum_{j=0}^{\infty} \theta_j^2 j^{2\beta} \leq L^2\}$ for $\beta, L > 0$ given. Then, by the definition, we have $B_N^2 \leq \sup_{\theta_{j\infty} \in \Theta_j} \sum_{j=N+1}^{\infty} \theta_j^2 (j/(N+1))^{2\beta} \leq L^2/(N+1)^{2\beta}$. This leads to the trade-off $N^{1/4}/n^{1/2} \sim L/N^{\beta}$, implying a cut-off of the order $N \sim L^{4/(4\beta+1)} n^{1/(2\beta+1/2)}$. With such a choice of N, we obtain a diameter of $\hat{C}_n$ of order equal to the maximum of $n^{-\beta/(2\beta+1/2)} L^{1/(4\beta+1)}$ and the average of estimation error $(R_N/n)^{1/2}$. As shown by Robins and Vaart (2006), this is the optimal rate.

4. Implementation and Simulations

4.1. Implementation

To implement our procedure for constructing confidence regions, we still need to deal with the following: properly selecting the trade-off parameter N, that is related to the submodels of interest; determining suitable estimators for $\|r^2\|_\infty$, $\|\sigma^2\|_\infty$, and B_N^2, because the ideal radiuses depend on them.

For the submodels in Examples 3.1−3.3, we see that the model-adaptive N can be selected in the scope of large samples. However, for finite samples, we need to select it so that the resulting confidence region is of model-data-adaptability. Thus, in addition to honesty, the resulting confidence region should have the smallest radius among the class of confidence regions related to given submodels. More precisely, for (3.5) or (3.6), the selected N should satisfy

$$N(\boldsymbol{\theta}_N) = \arg_N \min(D^*(\boldsymbol{\theta}_N) - d^*(\boldsymbol{\theta}_N)) \text{ or } N(\boldsymbol{\theta}_N) = \arg_N \min D^*(\boldsymbol{\theta}_N)$$
$$\text{s.t.} \quad P_{\boldsymbol{\theta}}\{\boldsymbol{\theta} \in \hat{C}_n\} \geq 1 - \alpha.$$

However, such a $N(\boldsymbol{\theta}_N)$ depends on $\boldsymbol{\theta}_N$. An implementable choice of $N(\boldsymbol{\theta}_N)$ is as follows. Let $\boldsymbol{\theta}_N^{(k)} = (\theta_0^{(k)}, \theta_1^{(k)}, \ldots, \theta_N^{(k)}, 0, \ldots)'$ be inner points in $\Theta, k = 1, \ldots, L$, and each component $\theta_j^{(k)}$ be randomly chosen in a working region to be specified. Then we choose $\hat{N}$ such that

$$\hat{N} = \arg_K \min\Big(\frac{1}{L} \sum_{k=1}^{L} (D^*(\boldsymbol{\theta}_N^{(k)}) - d^*(\boldsymbol{\theta}_N^{(k)}))\Big)$$
$$\text{or } \hat{N} = \arg_K \min \frac{1}{L} \sum_{k=1}^{L} D^*(\boldsymbol{\theta}_N^{(k)})$$
$$\text{s.t.} \quad P_{\boldsymbol{\theta}}\{\boldsymbol{\theta} \in \hat{C}_n\} \geq 1 - \alpha.$$

Such choices for N can guarantee the adaptive property if the working region for each component is large enough, for reasons as follows. The adaptive choice for N, by definition, results in a smallest radius of the confidence region in the sense of large-sample point, and the $\hat{N}$ given above leads to a smallest radius of the confidence region. Thus $\hat{N}$ is an adaptive choice by definition. From the simulation examples below, we see in Tables 1 and 2 that $(1/L)\sum_{k=1}^{L} D^*(\boldsymbol{\theta}_N^{(k)})$ is decreasing from $N = 0$ to $N = \hat{N}$, and is increasing thereafter, so such an optimal choice exists. We have *ad hoc* methods for the selection of N, but further investigation is needed.

Finally, we replace $\|r^2\|_\infty$ and $\|\sigma_\infty^2\|$ by consistent estimators. That can be obtained, respectively, by

$$\|\hat{r}^2\|_\infty = \max\{|\hat{r}^2(X_i)| : i = 1, \ldots, n\},$$
$$\|\hat{\sigma}^2\|_\infty = \max\{|\hat{\sigma}^2(X_i)| : i = 1, \ldots, n\},$$
$$\hat{B}_N^2 = \max\{|Y_i - \hat{r}(X_i)|^2 : i = 1, \ldots, n\},$$

where $\hat{r}^2$ and $\hat{\sigma}^2$ are adaptive estimators, respectively, of r^2 and σ^2. These may be simply constructed as the squares of the estimators of r and σ, but for adaptability and optimality reasons, we should use a particular method for estimating quadratic functionals; see Fan (1991), Laurent and Massart (2000), Low and Efromovich (1996), and Cai and Low (2006b).

Under some regularity conditions, we can obtain consistent estimators $\hat{r}^2$ and $\hat{\sigma}^2$. For example, kernel estimators $\hat{r}^2$ and $\hat{\sigma}^2$ are uniformly consistent and then $\|\hat{r}^2\|_\infty$ and $\|\hat{\sigma}^2\|_\infty$ are consistent, see Rao (1983). Thus, when $\|r^2\|_\infty$ and $\|\sigma_\infty^2\|$ are replaced, respectively, by $\|\hat{r}^2\|_\infty$ and $\|\hat{\sigma}^2\|_\infty$, the asymptotic order of radius of the confidence region cannot be reduced (see Lemma 3.1 and (3.7)) and the honesty defined in (i) still holds if (X, Y) is continuous.

4.2. Simulations

Examples are used to illustrate the new theory and to compare the adaptive global confidence band with the classical EL-based global confidence band. In our examples, the set of basis functions is $1, \sqrt{2}\cos(\pi u), \sqrt{2}\cos(2\pi u), \ldots$, the size of the samples is 2,000 and the nominal level is $1 - \alpha = 0.95$. In the simulations, the empirical coverage probabilities are based on the average of simulations of 100 samples. In the figures below, the curve "—" is the true regression curve, "$\cdot - \cdot$" indicates the boundaries of the adaptive global confidence bands and "- -" denotes the boundaries of the classical EL-based global confidence bands.

Example 4.1. Let

$$Y = \cos(2.5\pi X) + \varepsilon, \tag{4.1}$$

Table 1. The choice for K for Model (4.1).

N	1	2	3	4	5
average of D^*	1.8634	1.8923	1.0780	1.0780	1.0782
N	6	7	8	9	10
average of D^*	1.0840	1.0937	1.1026	1.1111	1.1194

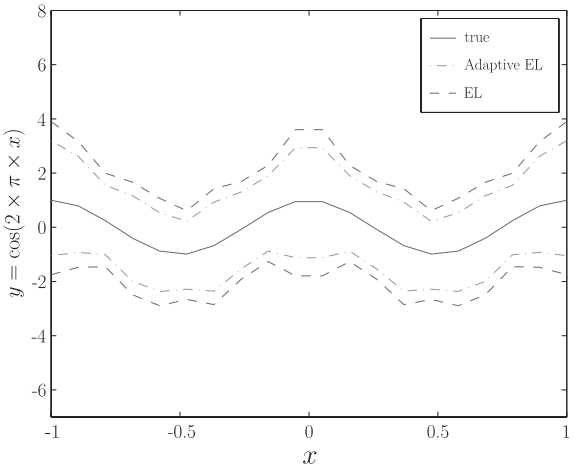

Figure 1. The confidence bands for the regression function in Model (4.1).

where X is uniformly distributed on $[-1, 1]$ and $\varepsilon \sim N(0, 0.25^2)$. Table 1 reports the average values of the criterion in 100 time repetitions, and then we see that the optimal choice for N is 3 or 4, $\hat{N} = 3$ was used. The working region for $\theta_j^{(k)}$ was $[-10, 10]$. Simulations showed that the two methods can achieve similar coverage probabilities 0.98, but the width of the adaptive global confidence band was significantly smaller than that of the classical EL-based global confidence band. Figure 1 reports one of the simulation results.

Example 4.2. Let

$$Y = -15x^6 + 15x^4 - x^2 + \varepsilon, \tag{4.2}$$

where X is uniformly distributed on $[-1, 1]$ and $\varepsilon \sim N(0, 0.25^2)$. Now the regression curve is not periodic, and is smoother than that in Example 4.1. Table 2 indicates that the optimal choice for N is 7 with the working region for $\theta_j^{(k)}$ of $[-20, 20]$. Simulations showed that the width of adaptive global confidence band was slightly smaller than that of the standard EL-based confidence band, with similar empirical coverage probabilities at 0.98. Figure 2 shows the result.

Based on the above limited simulation study, we suggest that the new method can perform better than the classical EL in the sense of that, with the same empirical coverage probability, the width of adaptive EL global confidence band

Table 2. The choice for K for Model (4.2).

N	1	2	3	4	5
average of D^*	1.6976	1.4523	1.4503	1.3277	1.2516
N	6	7	8	9	10
average of D^*	1.2475	1.2279	1.2307	1.2313	1.2343

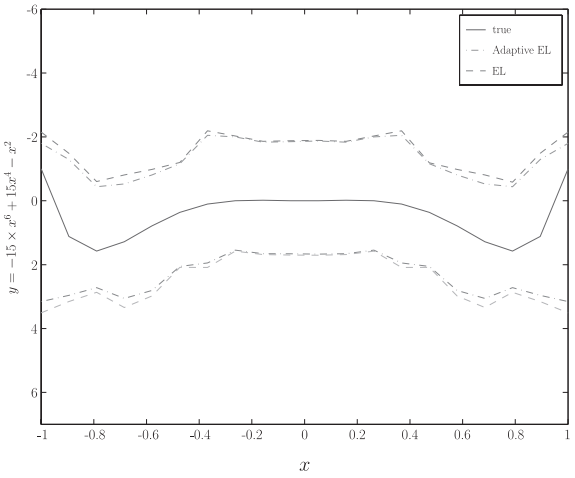

Figure 2. The confidence bands for the regression function in Model (4.2).

is smaller than that of the classical EL-based global confidence band. For a rough regression curve, this improvement can be significant.

5. Proofs

Proof of Lemma 3.1. In Theorem 4.1 of Hjort, McKeague, and Keilegom (2009), we set

$$X_{n,i} = P_N(X_i)(Y_i - P_N'(X_i)\boldsymbol{\theta}_N), \quad T_n = -2\log \Re_*^1(\boldsymbol{\theta}),$$
$$T_n^* = [\gamma n]\Big(\frac{1}{[\gamma n]}\sum_{i=1}^{[\gamma n]} X_{n,i}'\Big)\Big(\frac{1}{[\gamma n]}\sum_{i=1}^{[\gamma n]} X_{n,i}X_{n,i}'\Big)^{-1}\frac{1}{[\gamma n]}\sum_{i=1}^{[\gamma n]} X_{n,i}.$$

Then Lemma 3.1 follows.

Proof of Lemma 3.2. Note that the matrix M_N^1 is symmetric and its eigenvalues stay away form zero, so it is of full rank and non-degenerate. Further, $\hat{R}_N$ is a U-statistic of order 2 with kernel

$$h((Y_i, X_i), (Y_k, X_k)) = (\hat{\boldsymbol{\theta}}_N^1 - Y_i P_N(X_i))' M_N^1 (\hat{\boldsymbol{\theta}}_N^1 - Y_k P_K(X_k)).$$

Its Hoeffding decomposition is

$$\hat{R}_N = R(\boldsymbol{\theta}_N) + \frac{1}{NL_n}\sum_{i=L_n}^{n} P_1 h(Y_i, X_i) + \frac{1}{NL_n(L_n-1)}\sum_{i=L_n}^{n}\sum_{k\neq i, k=L_n}^{n} P_{1,2}h((Y_i,X_i),(Y_k,X_k)),$$

where $P_1 h(Y,X) = 2(\hat{\boldsymbol{\theta}}_N^1 - \boldsymbol{\theta}_N)' M_N^1 (\hat{\boldsymbol{\theta}}_N^1 - Y P_N(X))$ and

$$\begin{aligned} P_{1,2}h((Y_i,X_i),(Y_k,X_k)) &= (\boldsymbol{\theta}_N - Y_i P_N(X_i))' M_N^1 (\boldsymbol{\theta}_N - Y_k P_K(X_k)) \\ &= Y_i Y_k P_N'(X_i) M_N^1 P_N(X_k) \\ &\quad -\boldsymbol{\theta}_N' M_N^1 (Y_i P_N(X_i) + Y_k P_N(X_k)) + \boldsymbol{\theta}_N' M_N^1 \boldsymbol{\theta}_N. \end{aligned}$$

According to condition (2.3), we have

$$\begin{aligned} &\mathrm{Var}\,(P_1 h(Y,X)|(\mathbf{X}^1,\mathbf{Y}^1)) \\ &\quad = 4E\Big(E(Y^2|X)((\hat{\boldsymbol{\theta}}_N^1 - \boldsymbol{\theta}_N)' M_N^1 P_N(X))^2 |(\mathbf{X}^1,\mathbf{Y}^1)\Big) \\ &\qquad -4\Big(E((\hat{\boldsymbol{\theta}}_N^1 - \boldsymbol{\theta}_N)' M_N^1 E(Y|X) P_N(X)|(\mathbf{X}^1,\mathbf{Y}^1))\Big)^2 \\ &\quad = 4E\Big((r^2(X)+\sigma^2(X))((\hat{\boldsymbol{\theta}}_N^1 - \boldsymbol{\theta}_N)' M_N^1 P_N(X))^2 |(\mathbf{X}^1,\mathbf{Y}^1)\Big) \\ &\qquad -4\Big(E((\hat{\boldsymbol{\theta}}_N^1 - \boldsymbol{\theta}_N)' M_N^1 r(X) P_N(X)|(\mathbf{X}^1,\mathbf{Y}^1))\Big)^2 \\ &\quad \le 4(\|r\|_\infty^2 + \|\sigma^2\|_\infty) E\Big(((\hat{\boldsymbol{\theta}}_N^1 - \boldsymbol{\theta}_N)' M_N^1 P_N(X))^2 |(\mathbf{X}^1,\mathbf{Y}^1)\Big) \\ &\qquad -4\Big((\hat{\boldsymbol{\theta}}_N^1 - \boldsymbol{\theta}_N)' M_N^1 \boldsymbol{\theta}_N\Big)^2 \\ &\quad = 4(\|r\|_\infty^2 + \|\sigma^2\|_\infty)(\hat{\boldsymbol{\theta}}_N^1 - \boldsymbol{\theta}_N)'(M_N^1)^2(\hat{\boldsymbol{\theta}}_N^1 - \boldsymbol{\theta}_N) - 4\Big((\hat{\boldsymbol{\theta}}_N^1 - \boldsymbol{\theta}_N)' M_N^1 \boldsymbol{\theta}_N\Big)^2. \end{aligned}$$

Further, $(\boldsymbol{\theta}_N - Y_i P_N(X_i))' M_N^1 (\boldsymbol{\theta}_N - Y_k P_N(X_k))$ and $\boldsymbol{\theta}_N'(Y_i P_N(X_i) + Y_k P_N(X_k))$ are uncorrelated and their sum is $Y_i Y_k P_N'(X_i) M_N^1 P_N(X_k) + \boldsymbol{\theta}_N' M_N^1 \boldsymbol{\theta}_N$. Then

$$\begin{aligned} &\mathrm{Var}\,(P_{1,2}h((Y_i,X_i),(Y_k,X_k)))|(\mathbf{X}^1,\mathbf{Y}^1)) \\ &\quad \le \mathrm{Var}\,(Y_i Y_k P_N'(X_i) M_N^1 P_N(X_k)|(\mathbf{X}^1,\mathbf{Y}^1)) \\ &\quad = E\Big(E(Y_i^2|X_i)E(Y_k^2|X_k)(P_N'(X_i) M_N^1 P_N(X_k))^2 |(\mathbf{X}^1,\mathbf{Y}^1)\Big) \\ &\qquad -\Big(E(E(Y_i|X_i)E(Y_k|X_k) P_N'(X_i) M_N^1 P_N(X_k)|(\mathbf{X}^1,\mathbf{Y}^1))\Big)^2 \\ &\quad \le (\|r\|_\infty^2 + \|\sigma^2\|_\infty^2)^2 E(P_N'(X_i) M_N^1 P_N(X_k)|(\mathbf{X}^1,\mathbf{Y}^1))^2 - (\boldsymbol{\theta}_N' M_N^1 \boldsymbol{\theta}_N)^2 \\ &\quad = (K+1)\|M_N^1\|^2(\|r\|_\infty^2 + \|\sigma^2\|_\infty)^2 - (\boldsymbol{\theta}_N' M_N^1 \boldsymbol{\theta}_N)^2. \end{aligned}$$

From the above computation about the variance, the lemma follows.

Proof of Lemma 3.3. It can be easily verified from the definition of $\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|_{M_N^1}$ that

$$R^{1/2} \le \|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|_{M_N^1} \le R^{1/2} + 2\sqrt{[\gamma n]}B_N \tag{5.1}$$

provided $\hat{\boldsymbol{\theta}} \in \Theta$. Note that (3.4) is equivalent to

$$\inf_{\boldsymbol{\theta}\in\Theta} P_{\boldsymbol{\theta}}\Big\{(\hat{R}_N - \tau_{N,n})^{1/2} \le R^{1/2} \le (\hat{R}_N + z_\alpha \tau_{N,n})^{1/2} \,|(\mathbf{X}^1, \mathbf{Y}^1)\Big\} \ge 1-\alpha. \tag{5.2}$$

Then (5.1) and (5.2) imply that

$$\begin{aligned}
&\inf_{\boldsymbol{\theta}\in\Theta} P_{\boldsymbol{\theta}}\Big\{\sqrt{d^*} \le \|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|_{M_N^1} \le \sqrt{D^*}\,|(\mathbf{X}^1, \mathbf{Y}^1)\Big\} \\
&\quad = \inf_{\boldsymbol{\theta}\in\Theta} P_{\boldsymbol{\theta}}\Big\{\sqrt{d^*} \le \|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|_{M_N^1} \le \sqrt{D^*}, \\
&\qquad\qquad R^{1/2} \le \|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|_{M_N^1} \le R^{1/2} + 2[\gamma n]B_N \,|(\mathbf{X}^1, \mathbf{Y}^1)\Big\} \\
&\quad \ge \inf_{\boldsymbol{\theta}\in\Theta} P_{\boldsymbol{\theta}}\Big\{(\hat{R}_N - \tau_{N,n})^{1/2} \le R^{1/2}, \\
&\qquad\qquad R^{1/2} + 2[\gamma n]B_N \le (\hat{R}_N + z_\alpha \tau_{N,n})^{1/2} + 2[\gamma n]B_N \,|(\mathbf{X}^1, \mathbf{Y}^1)\Big\} \\
&\quad = \inf_{\boldsymbol{\theta}\in\Theta} P_{\boldsymbol{\theta}}\Big\{(\hat{R}_N - \tau_{N,n})^{1/2} \le R^{1/2} \le (\hat{R}_N + z_\alpha \tau_{N,n})^{1/2} \,|(\mathbf{X}^1, \mathbf{Y}^1)\Big\} \\
&\quad \ge 1-\alpha,
\end{aligned}$$

which leads to the theorem.

Proof of Theorem 3.1. Note that here the random variables (X, Y) are continuous. Then we can use the result of Lemma 3.3 and the continuity of probability to obtain the result.

Acknowledgement

The research was supported by NNSF project (10771123) of China, NBRP (973 Program 2007CB814901) of China, RFDP (20070422034) of China, NSF projects (Y2006A13 and Q2007A05) of Shandong Province of China and a grant from Research Grants Council of Hong Kong, Hong Kong, China. The authors thank the editors, the associate editor and two referees for their constructive comments and suggestions which led to an improvement of the early draft.

References

Baraud, Y. (2000). Model selection for regression on a fixed design. *Probab. Theory Related Fields* **117**, 467-493.

Baraud, Y. (2004). Confidence balls in Gaussian regression. *Ann. Statist.* **32**, 528-551.

Barron, A., Birgé, L. and Massart, P. (1999). Risk bound for model selection via penalization. *Probab. Theory Related Fields* **113**, 301-413.

Beran, R. and Dümbgen, L. (1998). Modulation of estimators and confidence sets. *Ann. Statist.* **26**, 1826-1856.

Birgé, L. (2002). Discussion of "Random rates in anisotropic regression," by M. Hoffman and O. Lepski. *Ann. Statist.* **30**, 359-363.

Birgé, L. and Massart, P. (2001). Gaussian model selection. *J. Eur. Math. Soc.* **3**, 203-268.

Cai, T. and Low, M. (2004). An adaptive theory for nonparametric confidence intervals. *Ann. Statist.* **32**, 1805-1850.

Cai, T. and Low, M. (2005). An adaptive estimation of linear functions. *Ann. Statist.* **33**, 2311-2343.

Cai, T. and Low, M. (2006a). Adaptive confidence balls. *Ann. Statist.* **34**, 202-228.

Cai, T. and Low, M. (2006b). Optimal adaptive estimation of a quadratic functional. *Ann. Statist.* **34**, 2298-2325.

Chen, S. X. (1996). Empirical likelihood confidence intervals for nonparametric density estimation. *Biometrika* **83**, 329-341.

Chen, S. X. and Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *Ann. Statist.* **21**, 1166-1181.

Chen, S. X. and Peng, L. (2007). Empirical likelihood for high dimension data.

Chen, J. H. and Qin, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika* **80**, 107-116.

Chen, S. X. and Qin, Y. S. (2000). Empirical likelihood confidence intervals for local linear smoothers. *Biometrika* **87**, 946-953.

DiCiccio, T. J., Hall, P., and Romano, J. P. (1991). Bartlett adjustment for empirical likelihood. *Ann. Statist.* **19**, 1053-1061.

Fan, J. (1991). On the estimation of quadratic functionals. *Ann. Statist.* **19**, 1273-1294.

Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *Ann. Statist.* **32**, 928-961.

Genovese, C. and Wasserman, L. (2005). Confidence sets for nonparametric wavelet regression. *Ann. Statist.* **33**, 698-729.

Hall, P. (1990). Pseudo-likelihood theory for empirical likelihood. *Ann. Statist.* **18**, 121-140.

Hjort, N. L., McKeague, L. W. and Keilegom, I. V. (2009). Extending the scope of empirical likelihood. *Ann. Statist.* **37**, 1079-1111.

Huang, J. Z., Wu, C. and Zhou, L. (2002). Vary-coefficient models and basis function approximations for the analysis of repeated measurement. *Biometrika* **89**, 111-128.

Hoffmann, M. and Lepski, O. (2002). Random rates in anisotropic regression. *Ann. Statist.* **30**, 325-396.

Juditsky, A. and Lambert-Lacroix, R. (2003). Nonparametric confidence set estimation. *Math. Methods Statist.* **12**, 410-428.

Kitamura, Y., Tripathi, G. and Ahn, H. (2004). Empirical likelihood-based inference in conditional moment restriction models. *Econometrica* **72**, 1667-1714.

Laurent, B. and Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28**, 1302-1338.

Li, K.-C. (1989). Honest confidence regions for nonparametric regression. *Ann. Statist.* **17**, 1101-1008.

Lin, L., Zhu, L. X. and Yuen, K. C. (2005). Profile empirical likelihood for parametric and semi-parametric models. *Ann. Inst. Statist. Math.* 57, 485-505.

Low, M. and Efromovich, S. (1996). On optimal adaptive estimation of a quadratic functional. *Ann. Statist.* **24**, 1106-1125.

Nishii, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *Ann. Statist.* **12**, 758-765.

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. *Ann. Statist.* **18**, 90-120.

Owen, A. (1991). Empirical likelihood for linear models. *Ann. Statist.* **19**, 1725-1747.

Owen, A. (2001). *Empirical likelihood.* Chapman and Hall/CRC.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 300-325.

Robins, J. and Vaart, A. V. D. (2006). Adaptive nonparametric confidence sets. *Ann. Statist.* **34**, 229-253.

Rao, B. L. S. P. (1983). *Nonparametric Functional Estimation.* Acadamic Press, London.

Shen, X., Shi, J. and Wong, W. H. (1999). Random sieve likelihood and general regression models. *J. Amer. Statist. Asscc.* **94**, 835-846.

Stute, W. Xue, L. G. and Zhu, L. X. (2007). Empirical likelihood inference in nonlinear errors-in-covariables Models With Validation Data. *J. Amer. Statist. Asscc.* **102**, 332-346.

Xue, L. G. and Zhu, L. X. (2007). Empirical likelihood for a varying coefficient model with longitudinal data. *J. Amer. Statist. Asscc.* **102**, 642-654.

Zhu, L. X. and Xue, L. G. (2006). Empirical likelihood confidence regions in a partially linear single-index model. *J. Roy. Statist. Soc. Ser. B* **86**, 549-570.

Hong Kong Baptist University, Hong Kong, China.

E-mail: lzhu@hkbu.edu.hk

School of Mathematical Sciences, Shandong University, Jinan 250100, China.

E-mail: linlu@sdu.edu.cn

School of Mathematical Sciences, Shandong University, Jinan 250100, China.

E-mail: chqsdu@mail.sdu.edu.cn

(Received April 2008; accepted June 2009)