Journal of Multivariate Analysis 112 (2012) 172–182

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Empirical likelihood inferences for the semiparametric additive isotonic regression

Guang Cheng [a,*], Yichuan Zhao [b], Bo Li [a]

[a] Department of Statistics, Purdue University, West Lafayette, IN 47906, United States
[b] Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, United States

ARTICLE INFO

Article history:
Received 29 July 2011
Available online 1 July 2012

AMS subject classifications:
62G15
62G08
62F30

Keywords:
Confidence region
Empirical likelihood
Isotonic regression
Semiparametric additive model

ABSTRACT

We consider the (profile) empirical likelihood inferences for the regression parameter (and its any sub-component) in the semiparametric additive isotonic regression model where each additive nonparametric component is assumed to be a monotone function. In theory, we show that the empirical log-likelihood ratio for the regression parameters weakly converges to a standard chi-squared distribution. In addition, our simulation studies demonstrate the empirical advantages of the proposed empirical likelihood method over the normal approximation method in Cheng (2009) [4] in terms of more accurate coverage probability when the sample size is small. It is worthy pointing out that we can construct the empirical likelihood based confidence region without the hassle of tuning any smoothing parameter due to the shape constraints assumed in this paper.

Published by Elsevier Inc.

1. Introduction

The semiparametric additive isotonic regression model takes the form

$$Y = \boldsymbol{X}'\boldsymbol{\beta} + \sum_{j=1}^{J} h_j(W_j) + \epsilon, \tag{1.1}$$

where $\boldsymbol{X} \in \mathbb{R}^p$, $W_j \in \mathbb{R}^1$, $\boldsymbol{\beta} \in \mathbb{R}^p$ is a non-random parameter and each function h_j is assumed to be monotone, for $j = 1, \ldots, J$. For simplicity, we assume that the error term ϵ has mean zero and is independent of the covariate $(\boldsymbol{X}', \boldsymbol{W}')$, where $\boldsymbol{W} = (W_1, \ldots, W_J)'$. The model (1.1) covers the partly linear isotonic regression studied in [8], i.e., $J = 1$, the additive isotonic regression studied in [12], i.e., without the parametric covariate, and also covers the possibility of using a (known) link function, which is the case presented in [1,14]. The model (1.1) has found wide applications in econometrics and epidemiology areas. For example, Morton-Jones et al. [14] employed (1.1) to model the effect of the father's paternal preconceptional radiation dose on the sex ratio of children. In their studies, the response of interest Y is the log sex ratio. The dose received in the 90 days prior to conception and the total doses received prior to that 90 days period are treated as the isotonic variables $\boldsymbol{W}$. The other explanatory variables $\boldsymbol{X}$ are either found linear relationship with the response, e.g., paternal age, or of categorical type, e.g., the social class of fathers.

The model (1.1) was initially investigated in [4] using the least squares estimation, which leads to the so called isotonic regression. Under the minimal smoothness assumption on h_j, it is shown that $\widehat{\boldsymbol{\beta}}_n$ is $\sqrt{n}$-consistent and asymptotically normal based on which the confidence region for $\boldsymbol{\beta}$ is constructed. However, our simulations reveal that this normal approximation approach yields confidence regions with biased coverage probability when the sample size is small. This discovery motivates

* Corresponding author.
E-mail addresses: chengg@stat.purdue.edu (G. Cheng), yichuan@gsu.edu (Y. Zhao), boli@stat.purdue.edu (B. Li).

0047-259X/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.jmva.2012.06.003

us to propose the empirical likelihood (EL) based confidence region, which was a nonparametric approach introduced by Owen [15,16], without estimating the asymptotic covariance. Similar to the bootstrap and jackknife methods, the EL method does not require knowing the corresponding semiparametric likelihood. Furthermore, it holds some excellent properties, such as range respecting and asymmetric confidence interval, etc. In addition, the proposed EL procedure enables us to obtain confidence regions for any sub-components or any linear combination of $\boldsymbol{\beta}$.

In this paper, we generalize the basic EL theorem by allowing for an infinite dimensional plug-in estimate $\widehat{h}_j(\cdot;\boldsymbol{\beta})$ in the estimating equations of $\boldsymbol{\beta}$. This is distinct from the general plug-in EL theorem [6] that is only valid for some nonparametric estimate $\widetilde{h}(\cdot)$ not depending on $\boldsymbol{\beta}$. Moreover, the monotonicity of $h_j(\cdot)$'s assumed in this paper enables us to apply the isotonic approach that can automatically "regularize" the estimation process *without penalization or kernel smoothing*. In other words, the EL confidence region studied in this paper can be easily obtained in practice without the hassle of tuning any smoothing parameter. This is the key difference from the empirical likelihood literature that assumes the smoothness of the nonparametric function, e.g., [6]. In contrast with other semiparametric models in which the EL inferences are applied, e.g., [21,25], our model (1.1) is allowed to have more than one nonparametric component. Some other related work include linear models [3], general estimating equations [17], confidence bands with right censoring [7,5,10,13,21], among others.

The rest of the paper is organized as follows. In Section 2, we review the normal approximation method for construction of confidence regions for regression parameters. In Section 3, we construct empirical likelihood confidence regions for regression parameters. In addition, we propose a profile EL to construct the confidence interval for the sub-component of regression parameters. In Section 4, simulation studies are conducted to demonstrate the empirical advantages of the proposed EL method over the normal approximation method in terms of coverage probability when the sample size is small. Section 5 applied our proposed approach to a real data set. We discuss our future work in the last section. The proof is postponed to the Appendix. Without loss of generality, we assume h_j's to be increasing from now on.

2. The normal approximation method

The isotonic estimate $(\widehat{\boldsymbol{\beta}}_n, \widehat{h}_1, \ldots, \widehat{h}_J)$ is defined as the minimizer of

$$S_n(\boldsymbol{\beta}, h_1, \ldots, h_J) = n^{-1}\sum_{i=1}^{n}\left(Y_i - \boldsymbol{X}_i'\boldsymbol{\beta} - \sum_{j=1}^{J} h_j(W_{ij})\right)^2 \tag{2.1}$$

subject to the restrictions that $\boldsymbol{\beta}$ belongs to some convex subset $\mathcal{B} \in \mathbb{R}^p$ and h_j belongs to a class of strictly increasing and uniformly bounded functions defined on $\mathbb{R}^1$, denoted as $\mathcal{H}_j$, for $j = 1, \ldots, J$. The solution of (2.1) is well defined and uniquely determined since $\mathcal{B}$ is a convex subset and the class of increasing functions forms a closed convex cone. We assume that the true function $h_{j0}(\cdot)$ is strictly increasing and bounded and that $\boldsymbol{X}$ is in some compact set. Without loss of generality, we assume that $\boldsymbol{X} \in [-1, 1]^p$ and $E\boldsymbol{X} = 0$. We also assume the norming condition that $\int h_j(w_j)dw_j = 0$ for the parameter identifiability.

Cheng [4] proposes the following estimating equation for $\boldsymbol{\beta}$:

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \boldsymbol{X}_i'\boldsymbol{\beta} - \sum_{j=1}^{J}\widehat{h}_j(W_{ij};\boldsymbol{\beta})\right)\left(\boldsymbol{X}_i - \sum_{j=1}^{J}\widehat{h}_j(W_{ij};\boldsymbol{\beta})\right) = 0, \tag{2.2}$$

where $(\widehat{h}_1(\cdot;\boldsymbol{\beta}), \ldots, \widehat{h}_J(\cdot;\boldsymbol{\beta}))$ is the minimizer of (2.1) for any fixed $\boldsymbol{\beta}$; see the proof of his Theorems 1 and 1.3.2 in [18]. In fact, we can express $\widehat{h}_j(\cdot;\boldsymbol{\beta})$ via the following max–min formula:

$$\widehat{h}_j(w_{(i)j};\boldsymbol{\beta}) = \max_{s\le i}\min_{t\ge i}\frac{\sum_{l=s}^{t}\left(y_{[l]} - \boldsymbol{x}_{[l]}'\boldsymbol{\beta} - \sum_{L\neq j}\widehat{h}_L(w_{[l]L};\boldsymbol{\beta})\right)}{t-s+1}, \tag{2.3}$$

where $w_{(i)j}$ is the i-th ordered w_{ij}'s and $(y_{[l]}, x_{[l]}, w_{[l]L})$ is the observation corresponding to the l-th ordered w_{ij}'s for $j = 1, \ldots, J$ and $L = 1, \ldots, j-1, j+1, \ldots, J$. Note that $\widehat{h}_j(\cdot;\widehat{\boldsymbol{\beta}}_n) = \widehat{h}_j(\cdot)$. A fast algorithm for computing $\widehat{h}_j(\cdot;\boldsymbol{\beta})$ and $\widehat{\boldsymbol{\beta}}_n$ by iterating between a CPAV procedure based on (2.3) [1] and solving a standard OLS problem has been established in Section 4 of [4]. Furthermore, Cheng [4] establishes that $\widehat{\boldsymbol{\beta}}_n$ is asymptotically normal with $\sqrt{n}$-rate as summarized in the following lemma.

The outer product $\boldsymbol{V}^{\otimes 2}$ is defined as $\boldsymbol{V}\boldsymbol{V}'$.

Lemma 2.1. *Suppose that Conditions* (S1)–(S3) *in the Appendix hold and* $\boldsymbol{W}$ *is pairwise independent. If* $E(\boldsymbol{X} - \sum_{j=1}^{J} E(\boldsymbol{X}|W_j))^{\otimes 2}$ *is strictly positive and* ϵ *satisfies sub-exponential tail, i.e.,* $E(\exp(\gamma|\epsilon|)) < C$ *for some* $\gamma, C > 0$*, we have*

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, \boldsymbol{\Sigma}), \tag{2.4}$$

where $\boldsymbol{\Sigma} = \sigma^2[E(\boldsymbol{X} - \sum_{j=1}^{J} E(\boldsymbol{X}|W_j))^{\otimes 2}]^{-1}$ *and* $\sigma^2 = Var(\epsilon)$*.*

According to (2.4), the asymptotic $100(1-\alpha)\%$ confidence region is constructed as follows

$$\mathcal{R}_1 = \left\{ \boldsymbol{\beta} : n(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})' \widehat{\boldsymbol{\Sigma}}^{-1} (\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \le \chi_p^2(\alpha) \right\}, \tag{2.5}$$

where $\chi_p^2(\alpha)$ is the upper α-quantile of the distribution χ_p^2 and $\widehat{\boldsymbol{\Sigma}}$ is a consistent estimate for $\boldsymbol{\Sigma}$. In view of the form of $\boldsymbol{\Sigma}$, we can build $\widehat{\boldsymbol{\Sigma}}$ upon the consistent estimate of $E(\boldsymbol{X}|W_j = w_{ij})$, denoted as $\widehat{E}(\boldsymbol{X}|W_j = w_{ij})$, i.e.,

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^n \left(y_i - \boldsymbol{x}_i' \widehat{\boldsymbol{\beta}}_n - \sum_{j=1}^J \widehat{h}_j(w_{ij}) \right)^2 \left(\frac{1}{n} \sum_{i=1}^n (\boldsymbol{x}_i - \sum_{j=1}^J \widehat{E}(\boldsymbol{X}|W_j = w_{ij}))^{\otimes 2} \right)^{-1}. \tag{2.6}$$

Any reasonable nonparametric approach, e.g., cubic smoothing spline, can be used to obtain $\widehat{E}(\boldsymbol{X}|W_j = w_{ij})$. However, when $\boldsymbol{X}$ is a categorical type of data, we may use the kernel method as in [19]. For example, we assume that X is a dichotomous variable indicating two treatment groups with $P(X = 1) = \gamma$ and $P(X = 0) = 1 - \gamma$. Then $\widehat{E}(X|W_j = w_{ij}) = (\widehat{\gamma}\widehat{p}_1(w_{ij}))/\widehat{p}_{W_j}(w_{ij})$, where $\widehat{\gamma}$ is the proportion of subjects in the treatment group with $X = 1$ in all the observations, $\widehat{p}_{W_j}$ and $\widehat{p}_1$ are the corresponding consistent kernel estimates for the density of W_j and the conditional density of W_j given $X = 1$.

3. The empirical likelihood confidence region

In this section, we show that the Wilks theorem (which states that the empirical log-likelihood ratio converges in distribution to the standard chi-square distribution) is valid for the semiparametric additive isotonic regression models. And then we develop the profile EL (PEL) ratio based confidence region. This approach enables us to make inferences for any linear combination of regression coefficients such as a single coefficient, a subset of coefficients, and linear contrasts. The PEL approach was first proposed by Subramanian [20] in studying the censored median regression models. More recently, Yu et al. [25] have applied this idea to the linear transformation model.

Motivated by the estimating equation (2.2), we define

$$\boldsymbol{W}_{ni}(\boldsymbol{\beta}) = \left(Y_i - \boldsymbol{X}_i' \boldsymbol{\beta} - \sum_{j=1}^J \widehat{h}_j(W_{ij}; \boldsymbol{\beta}) \right) \left(\boldsymbol{X}_i - \sum_{j=1}^J \widehat{h}_j(W_{ij}; \boldsymbol{\beta}) \right).$$

The estimated empirical likelihood at the value $\boldsymbol{\beta}$ is given by

$$L_n(\boldsymbol{\beta}) = \sup \left\{ \prod_{i=1}^n p_i : \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \boldsymbol{W}_{ni}(\boldsymbol{\beta}) = 0, p_i \ge 0 \right\}.$$

Note that $\prod_{i=1}^n p_i$ attains its maximum at $p_i = 1/n$. Thus, the empirical likelihood ratio at $\boldsymbol{\beta}$ is actually

$$R_n(\boldsymbol{\beta}) = \sup \left\{ \prod_{i=1}^n n\, p_i : \sum p_i = 1, \sum_{i=1}^n p_i \boldsymbol{W}_{ni}(\boldsymbol{\beta}) = 0,\ p_i \ge 0 \right\}.$$

Define the log-empirical likelihood ratio $l_n(\boldsymbol{\beta}) = -2 \log R_n(\boldsymbol{\beta})$. The Lagrange multiplier method implies that

$$l_n(\boldsymbol{\beta}) = 2 \sum_{i=1}^n \log\{1 + \boldsymbol{\lambda}(\boldsymbol{\beta})' \boldsymbol{W}_{ni}(\boldsymbol{\beta})\}, \tag{3.1}$$

where $\boldsymbol{\lambda}(\boldsymbol{\beta})$ satisfies the equation

$$\frac{1}{n} \sum_{i=1}^n \frac{\boldsymbol{W}_{ni}(\boldsymbol{\beta})}{1 + \boldsymbol{\lambda}(\boldsymbol{\beta})' \boldsymbol{W}_{ni}(\boldsymbol{\beta})} = 0. \tag{3.2}$$

Theorem 3.1. *Suppose that the conditions in Lemma 2.1 hold. Under the null hypothesis $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, we have*

$$l_n(\boldsymbol{\beta}_0) \xrightarrow{d} \chi_p^2, \tag{3.3}$$

where χ_p^2 is the chi-square random variable with p degrees of freedom.

Thus, we can construct the asymptotic $100(1-\alpha)\%$ EL based confidence region as

$$\mathcal{R}_2 = \left\{ \boldsymbol{\beta} : l_n(\boldsymbol{\beta}) \le \chi_p^2(\alpha) \right\}. \tag{3.4}$$

In practice, we may be interested in constructing the confidence region (or testing hypothesis) for some sub-component of $\boldsymbol{\beta}$. For example, we may want to construct the EL based confidence region for a q-dim subvector $\boldsymbol{\beta}^{(1)}$. A natural idea for that is to profile out the nuisance parameter $\boldsymbol{\beta}^{(2)}$ from the EL as in [17,20,25]. Thus, we propose the PEL ratio at $\boldsymbol{\beta}^{(1)}$ as $l_n^*(\boldsymbol{\beta}^{(1)}) = \min_{\beta^{(2)}} l_n((\boldsymbol{\beta}^{(1)'}, \boldsymbol{\beta}^{(2)'})')$. Note that the profile likelihood method is widely used in (semi)parametric models for dealing with nuisance parameters and the profile likelihood ratio usually has a chi-square limit distribution. Similar phenomenon as the Wilks theorem still holds for the proposed PEL ratio as shown in Theorem 3.2. Moreover, its excellent performance is demonstrated in the simulation studies.

Theorem 3.2. *Suppose that the conditions in Lemma 2.1 hold. Under the null hypothesis $\boldsymbol{\beta}^{(1)} = \boldsymbol{\beta}_0^{(1)}$, we have*

$$l_n^*(\boldsymbol{\beta}_0^{(1)}) \xrightarrow{d} \chi_q^2. \tag{3.5}$$

An asymptotic $100(1-\alpha)\%$ PEL based confidence region for $\boldsymbol{\beta}^{(1)}$ is thus defined as

$$\mathcal{R}_3 = \left\{\boldsymbol{\beta}^{(1)} : l_n^*(\boldsymbol{\beta}^{(1)}) \leq \chi_q^2(\alpha)\right\}. \tag{3.6}$$

It is known that the computation of the proposed $l_n^*(\cdot)$ is a difficult problem. When the dimension of $\boldsymbol{\beta}^{(2)}$ is not high, a useful method is to apply the grid search method to do the corresponding constrained maximization of the empirical likelihood.

The proposed PEL method can also be used to make inference on a linear combination of regression parameters, denoted as $\boldsymbol{r}'\boldsymbol{\beta}$. Without loss of generality, we assume $\boldsymbol{r} = (r_1, \ldots, r_p)'$ with nonzero r_1. Note that we can express $\boldsymbol{\beta}'\boldsymbol{X}_i$ as follows.

$$\boldsymbol{\beta}'\boldsymbol{X}_i = \boldsymbol{\beta}'\boldsymbol{r}\frac{X_{i1}}{r_1} + \beta_2\left(X_{i2} - \frac{r_2X_{i1}}{r_1}\right) + \cdots + \beta_p\left(X_{ip} - \frac{r_pX_{i1}}{r_1}\right).$$

By introducing $\theta^* = \boldsymbol{\beta}'\boldsymbol{r}$, $\theta_j = \beta_j$, $X_{i1}^* = X_{i1}/r_1$, $X_{ij}^* = X_{ij} - (r_jX_{i1})/r_1$ for $j = 2, \ldots, p$, we can re-express (1.1) as

$$Y = (\boldsymbol{X}^*)'\boldsymbol{\theta} + \sum_{j=1}^{J} h_j(W_j) + \epsilon, \tag{3.7}$$

where $\boldsymbol{\theta} = (\theta^*, \theta_2, \ldots, \theta_p)'$ and $\boldsymbol{X}^* = (X_1^*, \ldots, X_p^*)'$. Under (3.7) we can define the profile EL ratio $l_n^*(\theta^*)$ as before. Our Theorem 3.2 immediately implies the following useful corollary.

Corollary 3.1. *Suppose that the conditions in Lemma 2.1 hold. The test statistic $l_n^*(\theta_0^*)$, where $\theta_0^* = \boldsymbol{\beta}_0'\boldsymbol{r}$, converges to χ_1^2 in distribution.*

An asymptotic $100(1-\alpha)\%$ EL confidence interval for $\boldsymbol{\beta}'\boldsymbol{r}$ is thus given by

$$\mathcal{I} = \{\boldsymbol{\beta}'\boldsymbol{r} : l_n^*(\boldsymbol{\beta}'\boldsymbol{r}) \leq \chi_1^2(\alpha)\}.$$

4. Simulation study

In this section, we compare the empirical performances of the proposed EL confidence regions/intervals with the normal approximation (NA) based confidence regions/intervals, in terms of their coverage probability. We consider the following model

$$Y = \boldsymbol{X}'\boldsymbol{\beta} + h_1(W_1) + h_2(W_2) + \epsilon,$$

where $\boldsymbol{X} = (X_1, X_2)'$, $\boldsymbol{\beta} = (\beta_1, \beta_2)'$, and we set $\boldsymbol{\beta}_0 = (1, 1)'$, $h_{10}(w_1) = w_1\exp(-w_1^2/2)$ and $h_{20}(w_2) = \sin(\pi w_2/2)$ for $w_1, w_2 \in [-1, 1]$. In order to compare the NA and EL based confidence regions/intervals under different scenarios, we assume two types of errors separately: $\epsilon \sim N(0, 1)$ and $\epsilon \sim$ skewed normal distribution with shape, scale and location parameters as 4, 1.580, −1.223, respectively. The choice of the parameters for the skewed normal distribution make the variance of errors to be one.

To examine the effect of correlations among $\boldsymbol{X}$, W_1 and W_2 on the accuracy of confidence regions, we respectively generate X_1, X_2, W_1 and W_2 under two completely different setups.

(a) $X_1, W_1 \sim \mathit{Unif}[-1, 1]$ and $X_2, W_2 \sim$ truncated normal within the interval $[-1, 1]$.
(b) $X_1 \sim \mathit{Unif}[-1, 1]$, $X_2 \sim$ truncated normal on $[-1, 1]$, and then $W_1 = X_1 + \epsilon_0$, $W_2 = X_2 + \epsilon_0$ for $\epsilon_0 \sim N(0, 1)$.

The setup (a) will yield independent X_1, X_2, W_1 and W_2, while the setup (b) will introduce correlation between X_1 and W_1, X_2 and W_2, and W_1 and W_2.

Table 1
Coverage probability for confidence regions/intervals under setup (a).

Sample size	α	Normal errors				Skewed normal errors			
		R_1	$R_1(\beta_1)$	R_2	$R_3(\beta_1)$	R_1	$R_1(\beta_1)$	R_2	$R_3(\beta_1)$
30	0.90	393	560	772	812	401	553	769	805
	0.95	461	627	827	873	479	637	837	871
50	0.90	575	664	839	847	524	638	825	843
	0.95	660	753	895	909	610	723	886	911
70	0.90	661	735	835	849	676	766	853	834
	0.95	742	798	904	914	765	827	914	917
100	0.90	718	767	852	862	726	805	871	868
	0.95	807	852	920	921	823	864	921	935

The numbers are 1000 times the probability. R_1 is from normal approximation, and R_2 and R_3 are from empirical likelihood.

Table 2
Coverage probability for confidence regions/intervals under setup (b).

Sample size	α	Normal errors				Skewed normal errors			
		R_1	$R_1(\beta_1)$	R_2	$R_3(\beta_1)$	R_1	$R_1(\beta_1)$	R_2	$R_3(\beta_1)$
30	0.90	340	510	840	916	318	530	853	903
	0.95	410	595	892	961	415	603	911	953
50	0.90	421	624	851	931	395	570	847	928
	0.95	488	704	909	962	482	652	909	967
70	0.90	435	655	853	940	438	649	870	938
	0.95	536	727	926	966	539	731	924	969
100	0.90	470	664	865	926	456	687	838	940
	0.95	560	749	931	971	550	758	916	979

The numbers are 1000 times the probability. R_1 is from normal approximation, and R_2 and R_3 are from empirical likelihood.

We compute the confidence regions for $\boldsymbol{\beta}$ as well as the confidence intervals for its sub-component, β_1, based on both NA and EL methods. We use the cubic smoothing spline to estimate $E(\boldsymbol{X}|W_j = w_j)$ in the EL approach. All comparisons are made at four different sample sizes, 30, 50, 70 and 100, to examine the performance of the two different methods for small, moderate and large samples. Each coverage probability is computed over 1000 simulation runs, and at each run $\boldsymbol{\beta}$ are estimated by 20 iterations as we find that the backfitting algorithm converges quickly. We take 0.90 and 0.95 as the nominal confidence level, respectively.

The results under setup (a) are summarized in Table 1, and under setup (b) in Table 2. We observe very similar patterns in both tables. Regardless of error type, at each nominal level the coverage accuracies for both methods increase as the sample size increases. However, it is seen that both $\mathcal{R}_1$ and $\mathcal{R}_1(\beta_1)$ computed based on NA have poorer coverage probability than $\mathcal{R}_2$ and $\mathcal{R}_3(\beta_1)$ based on EL. Even with sample size 100, $\mathcal{R}_1$ and $\mathcal{R}_1(\beta_1)$ still largely underestimate the true probability. The coverage probabilities of $\mathcal{R}_2$ and $\mathcal{R}_3(\beta_1)$ in general show a nice pattern that they are low with small sample sizes but then approach the nominal level as the sample size increases, except that $\mathcal{R}_3(\beta_1)$ under setup (b) seems slightly over-coverage. In all circumstances, the $\mathcal{R}_2$ and $\mathcal{R}_3(\beta_1)$ apparently outperform the regions/intervals of $\mathcal{R}_1$ and $\mathcal{R}_1(\beta_1)$.

Although here we only show the coverage probabilities of confidence intervals for $\boldsymbol{\beta}_1$, we indeed also examined those for $\boldsymbol{\beta}_2$ using two different methods, and we found that the conclusions are very similar to that for $\boldsymbol{\beta}_1$. Overall, our simulation study demonstrates that for both normal errors and skewed normal errors, the EL method yields more accurate confidence regions/intervals for the entire set or components of regression parameters than the NA method.

5. Application to a real data

To illustrate the difference between two distinct methods of constructing confidence regions, we use the "cars" data from the 1983 ASA Data Exposition to show the two 95% joint confidence regions for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ based on $\mathcal{R}_1$ and $\mathcal{R}_2$, respectively. This data can be accessed at the StatLib Website at Carnegie Mellon University (http://lib.stat.cmu.edu/datasets/cars.data). This data set was also analyzed in [24] for constructing the penalized B-spline estimate of h_j. The response variable (Y) is the fuel efficiency difference between different cars and the most efficient car in this data set (46.6 miles per gallon). Two discrete variables are used as X_1 and X_2. In particular, X_1 is the cylinder and X_2 the model year. Following [24], we choose the displacement and the horsepower as W_1 and W_2, respectively.

Since this application mainly serves for the illustrative purpose, for computational ease, we randomly pick 30 data points and then construct the confidence regions for β_1 and β_2 only based on this relatively small data set. We further replace the original X_1 and X_2 by their two principal components to remove the collinearity and scale the data sets simply for regulation without loss of generality. Fig. 1 shows the contour of 95% confidence regions for β_1 and β_2 using both $\mathcal{R}_1$ and $\mathcal{R}_2$. Apparently,

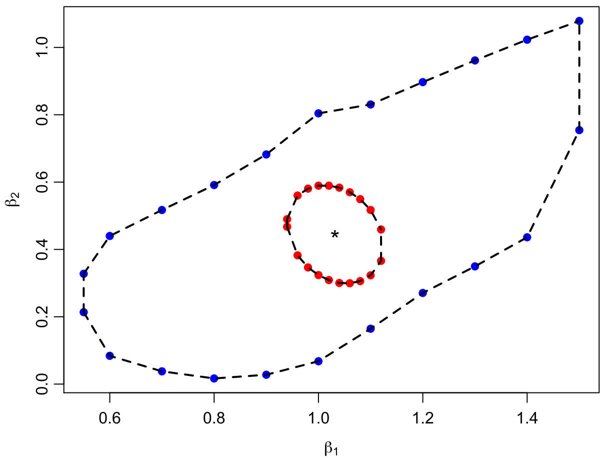

Fig. 1. The contour of a 95% confidence region using $\mathcal{R}_1$ (normal approximation) and $\mathcal{R}_2$ (empirical likelihood) based on the car data set. The star in the center represents the estimated values for β_1 and β_2.

the NA based confidence region is smaller than that of EL. This agrees with the pattern seen in Tables 1 and 2 that NA has the smaller coverage probability.

6. Conclusion

In this paper, we propose an empirical likelihood ratio method to make inferences for $\boldsymbol{\beta}$ as well as its sub-component in the semiparametric additive isotonic regression model. In comparison with the normal approximation method, the proposed EL method does not require estimating the asymptotic covariance of the limit distribution. In addition, there is no need for the EL method to solve any estimating equations while making inferences. The simulation results show that our proposed EL methods perform well in terms of coverage probability. In addition, the NA based method does not always work well in that it produces under-coverage confidence regions/intervals for small samples. One reason could be due to the fact that the NA method needs to estimate $\boldsymbol{\Sigma}$, while this variance estimate is not very stable. However, the proposed EL confidence regions/intervals hold an excellent property which is demonstrated in the simulation study. In the future, we will investigate the EL confidence interval for increasing functions $h_j(t)$ at any fixed point t in the isotonic regression type model; see [2]. We may also work on the increasing dimension scenario later.

Acknowledgments

The first author was partially sponsored by NSF Grant DMS-0906497, CAREER Award DMS-1151692. The second author was partially sponsored by Research Initiation Grant, Summer Research Grant in Georgia State University and the NSA Grant. The third author was partially sponsored by NSF Grant DMS-1007686.

Appendix

Regularity conditions S1–S3

The following assumptions hold for any $j = 1, \ldots, J$.

S1. The function h_j satisfies the condition that

$$\inf_{|w_j - w_j'| \geq \delta} |h_j(w_j) - h_j(w_j')| \geq C_1 \delta^{\gamma} \tag{A.1}$$

for any $\delta > 0$ and some constants $C_1, \gamma > 0$.

S2. The density for W_j, denoted as p_{W_j}, is assumed to be bounded away from zero and infinity, and fulfills the below Lipschitz condition

$$\sup_{w_j, w_j' \in \mathbb{R}^1} |p_{W_j}(w_j) - p_{W_j}(w_j')| \leq M |w_j - w_j'|^{\rho} \tag{A.2}$$

for some constants $M, \rho > 0$.

S3. The function $\zeta_j(w) \equiv E(\boldsymbol{X}|W_j = w)$ satisfies the condition

$$\|\zeta_j(w_j) - \zeta_j(w_j')\| \le C_2|w_j - w_j'| \tag{A.3}$$

for some constant $C_2 > 0$.

To facilitate the proofs in Appendix, we introduce another form of the estimation equation for $\boldsymbol{\beta}$:

$$\widetilde{\boldsymbol{W}}_{ni}(\boldsymbol{\beta}) = \left(Y_i - \boldsymbol{X}_i'\boldsymbol{\beta} - \sum_{j=1}^{J}\widehat{h}_j(W_{ij};\boldsymbol{\beta})\right)\left(\boldsymbol{X}_i - \sum_{j=1}^{J}\psi_j(\widehat{h}_j(W_{ij};\boldsymbol{\beta}))\right), \tag{A.4}$$

where $\psi_j = \zeta_j \circ h_{j0}^{-1}$ is the composite function of ζ_j and the inverse function of h_{j0}. This is because $\psi_j(\widehat{h}_j(\cdot))$ has exactly the same jump points as $\widehat{h}_j(\cdot)$ by the characterization of the solution to the isotonic regression problem; see p. 346 of [8].

We first state the following lemma in order to prove Theorems 3.1 and 3.2. Let

$$\boldsymbol{\Gamma} = \sigma^2 E\left(\boldsymbol{X} - \sum_{j=1}^{J} E(\boldsymbol{X}|W_j)\right)^{\otimes 2}.$$

Lemma A.1. *Suppose that the conditions in Lemma 2.1 hold. We have*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) \xrightarrow{d} N(0,\boldsymbol{\Gamma}), \tag{A.5}$$

and

$$\widehat{\boldsymbol{\Gamma}}_n \equiv \frac{\sum_{i=1}^{n}\boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)\boldsymbol{W}_{ni}'(\boldsymbol{\beta}_0)}{n} \xrightarrow{P} \boldsymbol{\Gamma}. \tag{A.6}$$

Proof. Based on (A.4), we can write

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{\boldsymbol{W}}_{ni}(\boldsymbol{\beta}_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\boldsymbol{X}_i - \sum_{j=1}^{J}E(\boldsymbol{X}_i|W_{ij})\right)\epsilon_i \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\boldsymbol{X}_i - \sum_{j=1}^{J}\psi_j(\widehat{h}_j(W_{ij};\boldsymbol{\beta}_0))\right)\left[\sum_{j=1}^{J}\left(h_{j0}(W_{ij}) - \widehat{h}_j(W_{ij;};\boldsymbol{\beta}_0)\right)\right] \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\sum_{j=1}^{J}E(\boldsymbol{X}_i|W_{ij}) - \sum_{j=1}^{J}\psi_j(\widehat{h}_j(W_{ij};\boldsymbol{\beta}_0))\right)\epsilon_i \\
&= I + II + III.
\end{aligned}$$

A direct application of CLT implies

$$I \xrightarrow{d} N(0,\boldsymbol{\Gamma})$$

since ϵ is assumed to be independent of $(\boldsymbol{X},\boldsymbol{W})$ and have mean zero.

We next show $II = o_P(1)$ and $III = o_P(1)$. Construct the class of uniformly bounded functions

$$\mathcal{F}_n = \left\{(\log n)^{-1}\left(\boldsymbol{x} - \sum_{j=1}^{J}\psi_j(h_j(w_j))\right)\left[\sum_{j=1}^{J}(h_{j0} - h_j)(w_j)\right] : h_j \in \mathcal{G}_n\right\},$$

where $\mathcal{G}_n = \{f : f \text{ is increasing with } \|f\|_\infty = O(\log n)\}$. The notation $\lesssim$ means $\le$ up to a universal constant. Note that we have

$$\begin{aligned}
\|\psi_j(h_{j1}(w_j)) - \psi_j(h_{j2}(w_j))\| &= \|E(\boldsymbol{X}|W_j = h_{j0}^{-1}(h_{j1}(w_j))) - E(\boldsymbol{X}|W_j = h_{j0}^{-1}(h_{j2}(w_j)))\| \\
&\le C_2|h_{j0}^{-1}(h_{j1}(w_j)) - h_{j0}^{-1}(h_{j2}(w_j))| \\
&\lesssim |h_{j1}(w_j)/\log n - h_{j2}(w_j)/\log n| \times \log n
\end{aligned} \tag{A.7}$$

based on Conditions S1 and S3 due to the observation that $h_{j0}[h_{j0}^{-1}(h_{ji}(w_j))] = h_{ji}(w_j)$ for $i = 1, 2$. It is easy to calculate that the δ-bracketing entropy of $\mathcal{F}_n$ is $1/\delta$ in terms of the $L_2(P)$-norm according to [22], Lemma 9.25 in [9] and Theorem 2.7.11 in [23]. Let $\mathbb{G}_n \equiv \sqrt{n}(\mathbb{P}_n - P)$ be the empirical processes for the observations and $\|\mathbb{G}_n\|_{\mathcal{F}_n} \equiv \sup_{f\in\mathcal{F}_n}|\mathbb{G}_n f|$. Then

Lemma 3.4.2 in [23] implies that the outer expectation $E^*\|\mathbb{G}_n\|_{\mathcal{F}_n} = O(n^{-1/6})$ since $\|f\|_\infty < \infty$ and $Pf^2 \lesssim n^{-2/3}$ for every $f \in \mathcal{F}_n$. Considering the known boundary effect of the isotonic estimator, i.e., $\|\widehat{h}_j(\boldsymbol{\beta}_0)\|_\infty = O_P(\log n)$; see [4], we thus have

$$\mathbb{G}_n\left(\left[\boldsymbol{X} - \sum_{j=1}^J \psi_j(\widehat{h}_j(W_j;\boldsymbol{\beta}_0))\right]\left[\sum_{j=1}^J \left(h_{j0}(W_j) - \widehat{h}_j(W_j;\boldsymbol{\beta}_0)\right)\right]\right) = O_P(n^{-1/6}\log n).$$

Then, we have

$$\begin{aligned} II &= \sqrt{n}P\left[\boldsymbol{X} - \sum_{j=1}^J \psi_j(\widehat{h}_j(W_j;\boldsymbol{\beta}_0))\right]\left[\sum_{j=1}^J \left(h_{j0}(W_j) - \widehat{h}_j(W_j;\boldsymbol{\beta}_0)\right)\right] + O_P(n^{-1/6}\log n) \\ &= II' + O_P(n^{-1/6}\log n). \end{aligned}$$

We note that

$$\begin{aligned} \|II'\| &= \sqrt{n}\left\|\sum_{j=1}^J P\left[\widehat{h}_j(W_j;\boldsymbol{\beta}_0) - h_{j0}(W_j)\right]\left[\boldsymbol{X} - \sum_{l=1}^J \psi_l(\widehat{h}_l(W_l;\boldsymbol{\beta}_0))\right]\right\| \\ &= \sqrt{n}\left\|\sum_{j=1}^J P\left[\widehat{h}_j(W_j;\boldsymbol{\beta}_0) - h_{j0}(W_j)\right]\left[E(\boldsymbol{X}|W_j) - \psi_j(\widehat{h}_j(W_j;\boldsymbol{\beta}_0))\right]\right\| \\ &\le \sqrt{n}\sum_{j=1}^J \|\widehat{h}_j(\boldsymbol{\beta}_0) - h_{j0}\|_2 \|E(\boldsymbol{X}|h_{j0}^{-1}(h_{j0}(W_j))) - \psi_j(\widehat{h}_j(W_j;\boldsymbol{\beta}_0))\|_2 \\ &\lesssim \sqrt{n}\sum_{j=1}^J \|\widehat{h}_j(\boldsymbol{\beta}_0) - h_{j0}\|_2^2, \end{aligned} \tag{A.8}$$

where $\|\cdot\|_2$ is the $L_2(P)$ norm, due to the pairwise independence assumption on $\boldsymbol{W}$, $E\boldsymbol{X} = 0$ and similar analysis on (A.7). Following similar arguments in proving Proposition 2.1 in [4], we can show

$$\max_{1\le j\le J}\|\widehat{h}_j(\boldsymbol{\beta}_0) - h_{j0}\|_2 = O_P(n^{-1/3}\log n). \tag{A.9}$$

This implies that $II = o_P(1)$.

We next apply Theorem 2.2 in [11, MV97] to show $III = o_P(1)$. Since we assume ϵ to have sub-exponential tail, which corresponds to (A0) in MV97, we need to calculate the bracketing entropy number on the class of function $\mathcal{A}_n \equiv \left\{(\log n)^{-1}\sum_{j=1}^J(E(\boldsymbol{X}|w_j) - \psi_j(h_j(w_j))) : h_j \in \mathcal{G}_n\right\}$. According to similar analysis on $\mathcal{F}_n$, we obtain that $v = 1$ in their Condition (2.2), and thus give

$$III = O_P\left(\max_{1\le j\le J}\|\widehat{h}_j(\boldsymbol{\beta}_0) - h_{j0}\|_2^{1/2}\log n \vee n^{-1/6}\log n\right) = o_P(1) \tag{A.10}$$

based on (A.9) and the asymptotic equivalence between the empirical L_2 norm and the L_2 norm given in Theorem 2.3 of MV97 (under the same set of conditions).

In view of the above analysis, we have shown

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{\boldsymbol{W}}_{ni}(\boldsymbol{\beta}_0) \xrightarrow{d} N(0, \boldsymbol{\Gamma}).$$

To complete the proof of (A.5), it suffices to show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \left[\widetilde{\boldsymbol{W}}_{ni}(\boldsymbol{\beta}_0) - \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)\right] = o_P(1). \tag{A.11}$$

Similarly, we can decompose (A.11) as the sum of IV and V, where

$$\begin{aligned} IV &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\sum_{j=1}^J (h_{j0}(W_{ij}) - \widehat{h}_j(W_{ij};\boldsymbol{\beta}_0))\right)\left[\sum_{j=1}^J (\widehat{h}_j(W_{ij};\boldsymbol{\beta}_0) - \psi_j(\widehat{h}_j(W_{ij};\boldsymbol{\beta}_0)))\right], \\ V &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\sum_{j=1}^J (\widehat{h}_j(W_{ij};\boldsymbol{\beta}_0) - \psi_j(\widehat{h}_j(W_{ij};\boldsymbol{\beta}_0)))\right]\epsilon_i. \end{aligned}$$

By applying similar techniques in analyzing II and III, we can show that $IV = o_P(1)$ and $V = o_P(1)$. This completes the proof of (A.5).

We next consider (A.6). Write $I = n^{-1/2}\sum \boldsymbol{R}_{i1}$, $II = n^{-1/2}\sum \boldsymbol{R}_{i2}$ and $III = n^{-1/2}\sum \boldsymbol{R}_{i3}$. Then, $\widetilde{\boldsymbol{W}}_{ni}(\boldsymbol{\beta}_0)$ can be expressed as $\boldsymbol{R}_{i1} + \boldsymbol{R}_{i2} + \boldsymbol{R}_{i3}$. Let $\boldsymbol{\Gamma}_n = \sum_{i=1}^n \boldsymbol{R}_{i1}\boldsymbol{R}_{i1}'/n$ and $\widetilde{\boldsymbol{\Gamma}}_n = \sum_{i=1}^n \widetilde{\boldsymbol{W}}_{ni}(\boldsymbol{\beta}_0)\widetilde{\boldsymbol{W}}_{ni}'(\boldsymbol{\beta}_0)/n$. Applying the Law of Large Number to $\boldsymbol{\Gamma}_n$ and considering (A.11), it remains to show that $\widetilde{\boldsymbol{\Gamma}}_n = \boldsymbol{\Gamma}_n + o_P(1)$. By some algebra, we have

$$\widetilde{\boldsymbol{\Gamma}}_n - \boldsymbol{\Gamma}_n = \frac{1}{n}\sum_{i=1}^n \left[(\boldsymbol{R}_{i2} + \boldsymbol{R}_{i3})(\boldsymbol{R}_{i2} + \boldsymbol{R}_{i3})' + \boldsymbol{R}_{i1}(\boldsymbol{R}_{i2} + \boldsymbol{R}_{i3})' + (\boldsymbol{R}_{i2} + \boldsymbol{R}_{i3})\boldsymbol{R}_{i1}'\right]. \tag{A.12}$$

By the Law of Large Number, we can easily show that $\sum_i \boldsymbol{R}_{i1}\boldsymbol{R}_{i1}'/n = O_P(1)$. As for $\sum_i \boldsymbol{R}_{i2}\boldsymbol{R}_{i2}'/n$, we apply similar analysis on $\mathcal{F}_n$ to obtain that $\|\mathbb{G}_n\|_{\mathcal{F}_n^2} = O_P(n^{-1/6})$, where $\mathcal{F}_n^2 \equiv \{f^{\otimes 2} : f \in \mathcal{F}_n\}$, which implies that

$$\sum_i \boldsymbol{R}_{i2}\boldsymbol{R}_{i2}'/n = P\left[\left(\boldsymbol{X}_i - \sum_{j=1}^J \psi_j(\widehat{h}_j(W_{ij};\boldsymbol{\beta}_0))\right)^{\otimes 2}\left(\sum_{j=1}^J \left(h_{j0}(W_{ij}) - \widehat{h}_j(W_{ij;};\boldsymbol{\beta}_0)\right)\right)^2\right] + O_P(n^{-2/3}(\log n)^2).$$

Considering (A.9), bounded $\boldsymbol{X}$ and the Cauchy–Schwarz inequality, i.e., $\sum_i a_i b_i \le \sqrt{\sum_i a_i^2} \times \sqrt{\sum_i b_i^2}$, we have $\sum_i \boldsymbol{R}_{i2}\boldsymbol{R}_{i2}'/n = o_P(1)$. Applying similar analysis also gives that $\sum_i \boldsymbol{R}_{i3}\boldsymbol{R}_{i3}'/n = o_P(1)$. Again, applying the Cauchy–Schwarz inequality to each component of the matrix on the right hand side of (A.12), we know that $\widetilde{\boldsymbol{\Gamma}}_n - \boldsymbol{\Gamma}_n = o_P(1)$ since it is shown that $\sum_i \boldsymbol{R}_{i1}\boldsymbol{R}_{i1}'/n = O_P(1)$, $\sum_i \boldsymbol{R}_{i2}\boldsymbol{R}_{i2}'/n = o_P(1)$ and $\sum_i \boldsymbol{R}_{i3}\boldsymbol{R}_{i3}'/n = o_P(1)$. This completes the proof of (A.6). □

Proof of Theorem 3.1. We first show that

$$\max_{1\le i\le n}\|\boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)\| = o_P(n^{1/2}). \tag{A.13}$$

Note that $\boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) = \boldsymbol{R}_{i1} + \bar{\boldsymbol{R}}_{i2} + \bar{\boldsymbol{R}}_{i3}$, where

$$\bar{\boldsymbol{R}}_{i2} = \left(\boldsymbol{X}_i - \sum_{j=1}^J \widehat{h}_j(W_{ij};\boldsymbol{\beta}_0)\right)\left[\sum_{j=1}^J \left(h_{j0}(W_{ij}) - \widehat{h}_j(W_{ij;};\boldsymbol{\beta}_0)\right)\right],$$

$$\bar{\boldsymbol{R}}_{i3} = \left(\sum_{j=1}^J E(\boldsymbol{X}_i|W_{ij}) - \sum_{j=1}^J \widehat{h}_j(W_{ij};\boldsymbol{\beta}_0)\right)\epsilon_i.$$

Following the same arguments of [15], we have $\max_{1\le i\le n}\|\boldsymbol{R}_{i1}\| = o_P(n^{1/2})$ since $\boldsymbol{R}_{i1}$'s are i.i.d. random variables with finite second moment. From the proof of Lemma A.1, we have that

$$\max_{1\le i\le n}\|\bar{\boldsymbol{R}}_{i2}\| \lesssim \max_{1\le j\le J}\|\widehat{h}_j(\boldsymbol{\beta}_0)\|_\infty^2 \le O_P((\log n)^2)$$

and

$$\max_{1\le i\le n}\|\bar{\boldsymbol{R}}_{i3}\| \lesssim \max_{1\le i\le n}|\epsilon_i| \max_{1\le j\le J}\|\widehat{h}_j(\boldsymbol{\beta}_0)\|_\infty \le O_P((\log n)^2)$$

since ϵ is assumed to have the subexponential tail. This completes the proof of (A.13).

Denote $\boldsymbol{\lambda}(\boldsymbol{\beta}_0) = \boldsymbol{\lambda}_0$. Recall $\widehat{\boldsymbol{\Gamma}}_n = \boldsymbol{\Gamma} + o_P(1)$ (cf. Lemma A.1). Note Lemma A.1. Then, it follows from (3.2) that

$$\|\boldsymbol{\lambda}_0\| = O_P(n^{-1/2}). \tag{A.14}$$

By (3.2), (A.5), (A.14), and Lemma A.1, it follows that

$$\boldsymbol{\lambda}_0 = \left\{\sum_{i=1}^n \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)\boldsymbol{W}_{ni}'(\boldsymbol{\beta}_0)\right\}^{-1}\sum_{i=1}^n \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) + O_P(n^{-1/2}). \tag{A.15}$$

We apply Taylor's expansion to (3.1). By (A.15) and (A.6), we have

$$
\begin{aligned}
l_n(\boldsymbol{\beta}_0) &= \sum_{i=1}^{n} \boldsymbol{\lambda}_0' \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) + o_P(1) \\
&= \left\{ \frac{\sum_{i=1}^{n} \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)}{\sqrt{n}} \right\}' \left\{ \frac{\sum_{i=1}^{n} \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)\boldsymbol{W}_{ni}'(\boldsymbol{\beta}_0)}{n} \right\}^{-1} \left\{ \frac{\sum_{i=1}^{n} \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)}{\sqrt{n}} \right\} + o_P(1) \\
&= \left\{ \boldsymbol{\Gamma}^{-1/2} n^{-1/2} \sum_{i=1}^{n} \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) \right\}' \left\{ \boldsymbol{\Gamma}^{1/2} \widehat{\boldsymbol{\Gamma}}_n^{-1} \boldsymbol{\Gamma}^{1/2} \right\} \left\{ \boldsymbol{\Gamma}^{-1/2} \frac{\sum_{i=1}^{n} \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0)}{\sqrt{n}} \right\} + o_P(1).
\end{aligned}
$$

By Lemma A.1, we complete the proof of Theorem 3.1. □

Proof of Theorem 3.2. We follow the similar arguments as in [25]. We denote $\boldsymbol{X} = (\boldsymbol{X}_1', \boldsymbol{X}_2')'$ corresponding to $(\boldsymbol{\beta}^{(1)'}, \boldsymbol{\beta}^{(2)'})'$. Define

$$
\mathbf{A}_1 = E\left[\boldsymbol{X} - \sum_{j=1}^{J} E(\boldsymbol{X}|W_j) \right] \left[\boldsymbol{X}_2 - \sum_{j=1}^{J} E(\boldsymbol{X}_2|W_j) \right]',
$$

$$
\mathbf{A} = E\left[\boldsymbol{X} - \sum_{j=1}^{J} E(\boldsymbol{X}|W_j) \right]^{\otimes 2}.
$$

Since $\mathbf{A}$ is positive definite, $\mathbf{A}_1$ is a $p \times (p-q)$ dimensional matrix of full rank. Let $\widetilde{\boldsymbol{\beta}}^{(2)} = \arg\min_{\beta^{(2)}} l_n((\boldsymbol{\beta}_0^{(1)'}, \boldsymbol{\beta}^{(2)'})')$. Similar to the proof of Theorem 1 in the Appendix of [4,17], we can show that

$$
\sqrt{n}(\widetilde{\boldsymbol{\beta}}^{(2)} - \boldsymbol{\beta}_0^{(2)}) = (\mathbf{A}_1' \boldsymbol{\Gamma}^{-1} \mathbf{A}_1)^{-1} \mathbf{A}_1' \boldsymbol{\Gamma}^{-1} n^{-1/2} \sum_{i=1}^{n} \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) + o_P(1),
$$

$$
\sqrt{n}\boldsymbol{\lambda}(([\boldsymbol{\beta}_0^{(1)}]', [\widetilde{\boldsymbol{\beta}}^{(2)}]')') = \{I - \boldsymbol{\Gamma}^{-1}\mathbf{A}_1(\mathbf{A}_1'\boldsymbol{\Gamma}^{-1}\mathbf{A}_1)^{-1}\mathbf{A}_1'\}\boldsymbol{\Gamma}^{-1} n^{-1/2} \sum_{i=1}^{n} \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) + o_P(1),
$$

where $\boldsymbol{\lambda}(([\boldsymbol{\beta}_0^{(1)}]', [\widetilde{\boldsymbol{\beta}}^{(2)}]')')$ is the corresponding Lagrange multiplier. Similarly as [17,25], by Taylor's expansion, we have that

$$
\begin{aligned}
l_n^*(\boldsymbol{\beta}_0^{(1)}) &= \sum_{i=1}^{n} \boldsymbol{\lambda}'(([\boldsymbol{\beta}_0^{(1)}]', [\widetilde{\boldsymbol{\beta}}^{(2)}]')')\boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) + o_P(1) \\
&= n^{-1/2} \sum_{i=1}^{n} \boldsymbol{W}_{ni}'(\boldsymbol{\beta}_0)\{\boldsymbol{\Gamma}^{-1} - \boldsymbol{\Gamma}^{-1}\mathbf{A}_1(\mathbf{A}_1'\boldsymbol{\Gamma}^{-1}\mathbf{A}_1)^{-1}\mathbf{A}_1'\boldsymbol{\Gamma}^{-1}\} n^{-1/2} \sum_{i=1}^{n} \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) + o_P(1) \\
&= \left\{ n^{-1/2} \boldsymbol{\Gamma}^{-1/2} \sum_{i=1}^{n} \boldsymbol{W}_{ni}'(\boldsymbol{\beta}_0) \right\} S \left\{ n^{-1/2} \boldsymbol{\Gamma}^{-1/2} \sum_{i=1}^{n} \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) \right\} + o_P(1),
\end{aligned}
$$

where

$$
S = \mathbf{I} - \boldsymbol{\Gamma}^{-1/2}\mathbf{A}_1(\mathbf{A}_1'\boldsymbol{\Gamma}^{-1}\mathbf{A}_1)^{-1}\mathbf{A}_1'\boldsymbol{\Gamma}^{-1/2}
$$

is an idempotent matrix with $trace(S) = q$. By Lemma A.1, we have

$$
n^{-1/2}\boldsymbol{\Gamma}^{-1/2} \sum_{i=1}^{n} \boldsymbol{W}_{ni}(\boldsymbol{\beta}_0) \xrightarrow{d} N(0, \mathbf{I}).
$$

Thus, Theorem 3.2 is proved. □

References

[1] P. Bacchetti, Additive isotonic models, Journal of the American Statistical Association 84 (1989) 289–294.
[2] M. Banerjee, J.A. Wellner, Likelihood ratio tests for monotone functions, Annals of Statistics 29 (2001) 1699–1731.
[3] S.X. Chen, Empirical likelihood confidence intervals for linear regression coefficients, Journal of Multivariate Analysis 49 (1994) 24–40.

[4] G. Cheng, Semiparametric additive isotonic regression, Journal of Statistical Planning and Inference 139 (2009) 1980–1991.
[5] J.H. Einmahl, I.W. McKeague, Confidence tubes for multiple quantile plots via empirical likelihood, Annals of Statistics 27 (1999) 1348–1367.
[6] N.L. Hjort, I.W. McKeague, I. Van Keilegom, Extending the scope of empirical likelihood, Annals of Statistics 37 (2009) 1079–1111.
[7] M. Hollander, I.W. McKeague, J. Yang, Likelihood ratio-based confidence bands for survival functions, Journal of the American Statistical Association 92 (1997) 215–226.
[8] J. Huang, A note on estimating a partly linear model under monotonicity constraints, Journal of Statistical Planning and Inference 107 (2002) 345–351.
[9] M.R. Kosorok, Introduction to Empirical Processes and Semiparametric Inference, Springer, New York, 2008.
[10] G. Li, I. Van Keilegom, Likelihood ratio confidence bands in nonparametric regression with censored data, Scandinavian Journal of Statistics 29 (2002) 547–562.
[11] E. Mammen, S. van de Geer, Penalized quasi-likelihood estimation in partial linear models, Annals of Statistics 25 (1997) 1014–1035.
[12] E. Mammen, K. Yu, Additive Isotone Regression, in: IMS Lecture Notes – Monograph Series, 55, 2007, pp. 179–195.
[13] I.W. McKeague, Y. Zhao, Simultaneous confidence bands for ratios of survival functions via empirical likelihood, Statistics & Probability Letters 60 (2002) 405–415.
[14] T. Morton-Jones, P. Diggle, L. Parker, H.O. Dickinson, K. Binks, Additive isotonic regression models in epidemiology, Statistics in Medicine 19 (2000) 849–859.
[15] A.B. Owen, Empirical likelihood ratio confidence intervals for a single functional, Biometrika 75 (1988) 237–249.
[16] A.B. Owen, Empirical likelihood and confidence regions, Annals of Statistics 18 (1990) 90–120.
[17] J. Qin, J.F. Lawless, Empirical likelihood and general estimating equations, Annals of Statistics 22 (1994) 300–325.
[18] T. Robertson, F.T. Wright, R.L. Dykstra, Order Restricted Statistical Inference, Wiley, New York, 1988.
[19] B.W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman and Hall, New York, 1986.
[20] S. Subramanian, Censored median regression and profile empirical likelihood, Statistical Methodology 4 (2007) 493–503.
[21] Y. Sun, R. Sundaram, Y. Zhao, Empirical likelihood inference for the Cox model with time-dependent coefficients via local partial likelihood, Scandinavian Journal of Statistics 36 (2009) 444–462.
[22] S.A. van de Geer, The entropy bound for monotone functions. Tech. Report 91-10, University of Leiden, 1991.
[23] A.W. Van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes: With Applications to Statistics, Springer, New York, 1996.
[24] X. Wang, J. Shen, A class of grouped Brunk estimators and penalized spline estimators for monotone regression, Biometrika 97 (2010) 585–601.
[25] W. Yu, Y. Sun, M. Zheng, Empirical likelihood method for linear transformation models, Annals of the Institute of Statistical Mathematics 63 (2011) 331–346.