Journal of Multivariate Analysis 121 (2013) 22–32

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Empirical likelihood for partially linear proportional hazards models with growing dimensions

Xingyu Tang [a], Jianbo Li [b], Heng Lian [a,*]

[a] Division of Mathematical Sciences, SPMS, Nanyang Technological University, Singapore, 637371, Singapore

[b] School of Mathematics and Statistics, Jiangsu Normal University, Xuzhou, Jiangsu, 221116, China

ARTICLE INFO

Article history:
Received 24 May 2012
Available online 26 June 2013

AMS subject classifications:
62N02
62G20

Keywords:
Confidence interval
Coverage probability
Empirical likelihood
Growing dimensions

ABSTRACT

Empirical-likelihood-based inferences for the linear part in a partially linear Cox's proportional hazards model are investigated. It was shown in some previous studies, for some related but different semiparametric models, that if there is no bias correction, the limit distribution of the empirical likelihood ratio statistic is not a standard chi-square distribution. In some previous studies, the bias correction is achieved by subtracting a conditional expectation of a predictor from itself. In proportional hazards models, the situation is different and it is not clear how to do so. Motivated from the form of the asymptotic variance of the parameters, the bias-corrected empirical likelihood ratio is proposed, with a standard χ^2 limit. The demonstrated asymptotics even apply to models with growing dimensions. For computational simplicity, we use polynomial splines to approximate the nonparametric component so that the computations involved are similar to those for the parametric model. Some simulations are carried out to study the performance of bias-corrected empirical likelihood ratio.

1. Introduction

In this paper, we consider proportional hazards models with a partially linear risk score

$$\lambda(t|W, X) = \lambda_0(t)\exp\{\phi_0(W) + X^T\beta_0\}, \tag{1}$$

where the model contains both the nonparametric component $\phi_0(W)$ and the parametric component $X^T\beta_0$, W is q-dimensional and X is p-dimensional. Although $q > 1$ is possible, in practice only $q = 1$ is popular to avoid curse of dimensionality. Thus we will only consider $q = 1$ here. This model combines the flexibility of nonparametric modeling and parsimony and easy interpretability of parametric modeling. In particular, it avoids the curse of dimensionality of a purely nonparametric model [4,15].

Model (1) has been previously considered, for example, in [3,5,7,14,19,20], several of which are concerned with the more general partially linear additive models. Our investigations here can also be easily extended to relative risk with a partially linear additive structure.

Based on asymptotic normality of the linear part, inferences can be performed using a sandwich formula, which was advocated in [3,7], among others. In this paper, we will consider using empirical likelihood ratio for inferences on the linear part for partially linear proportional hazards models, which has not been considered before. Empirical likelihood was proposed first in [16,17] and has become very popular ever since. The advantages of a confidence interval/region constructed

* Corresponding author.
E-mail address: henglian@ntu.edu.sg (H. Lian).

0047-259X/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmva.2013.06.002

based on an empirical likelihood ratio include that it has a data-adaptive shape, it typically has better coverage, and it can easily incorporate some auxiliary information [18]. Many investigations of inferences on semiparametric models exist including Zhu and Xue [29], Xue and Zhu [24], You and Zhou [27], Xue and Zhu [25,26], Li et al. [13].

Nowadays, more and more researchers are concerned with data analysis tasks in which a large number of predictors/features are used. This is due to the fact that, in a study where there are limited previous experiences, it is hard to identify a small number of predictors such that it is believed only these variables contribute to the response of interest. Thus a large number of predictors suspected to be related to responses need to be collected to avoid model misspecification. It turns out that the additional efforts required to deal with partially linear models with growing dimensions are minimal, and thus we will accommodate growing dimension in this work. Empirical likelihood with growing dimensions has been recently studied in [2,6,12,22], for example. Our study is directly motivated by Li et al. [12] which considered semiparametric varying-coefficient models with growing dimensions. In their work, the bias correction is achieved by subtracting the conditional expectation of the predictor from itself in the estimating equation. For Cox models, such a simple strategy is not valid. Instead we use the estimated estimating equations obtained from the efficient score equation which was presented in [7] for the fixed dimensional case. Other related works include Sun et al. [21] who studied Cox models with varying coefficients and empirical likelihood is used for inference of the nonparametric coefficients. We focus on the linear coefficients for which the score functions are different from that used in [21]. Huang and Zhang [9,10] considered an empirical likelihood method for single-index models which is different from Cox models. Zhu et al. [28] considered empirical likelihood for an uncensored semiparametric problem. Although bias-correction was used in all these works, it is different to that used here due to the fact that we consider a different model. Also, dimensions in these works are fixed.

The rest of the article is organized as follows. In the next section, we present the asymptotic properties of the maximum partial likelihood estimator with growing dimensions. In Section 3, a bias-corrected empirical likelihood ratio for β is proposed and it is shown that the statistic has a standard chi-square limit. In Section 4, simulation studies are carried out to assess the performance of the proposed method in comparison with intervals constructed from the sandwich formula. We conclude the paper in Section 5 with a short discussion. Finally, the Appendix contains all the technical proofs.

2. Maximum partial likelihood estimator

Let T^e and T^c be the event time and the censoring time respectively, where the hazard function of T^e is given by (1). Assume that T^e and T^c are independent given the covariates. The true nonparametric functions and parameters will be denoted using a subscript 0. The observable random variables are (T, Δ, W, X) where $T = \min\{T^e, T^c\}$ and $\Delta = I\{T^e \le T^c\}$ ($I\{\cdot\}$ is the indicator function), $W \in R$ and $X = (X_1, \ldots, X_p)^T \in R^p$ are the covariates in the nonparametric part and the parametric part respectively. Note that ϕ_0 is identifiable only up to a constant and thus we assume $E\Delta\phi_0(W) = 0$. We make n i.i.d. observations $(T_i, \Delta_i, W_i, X_i)$.

We use polynomial splines to approximate the nonparametric component. Let $\tau_0 = 0 < \tau_1 < \cdots < \tau_{K'} < 1 = \tau_{K'+1}$ be a partition of $[0, 1]$ into subintervals $[\tau_k, \tau_{k+1}), k = 0, \ldots, K'$ with K' internal knots. A polynomial spline of order r is a function whose restriction to each subinterval is a polynomial of degree $r - 1$ and globally $r - 2$ times continuously differentiable on $[0, 1]$. The collection of splines with a fixed sequence of knots has a normalized B-spline basis $\{\tilde{B}_1(x), \ldots, \tilde{B}_{\tilde{K}}(x)\}$ with $\tilde{K} = K' + r$. Because of the centering constraint $E\Delta\phi_0(W) = 0$, we instead focus on the subspace of spline functions $S^0 := \{s : s = \sum_{k=1}^{\tilde{K}} a_k\tilde{B}_k(x), \sum_{i=1}^{n} \Delta_i s(W_i) = 0\}$ with basis $\{B_k(x) = \sqrt{K}(\tilde{B}_k(x) - \sum_{i=1}^{n} \Delta_i\tilde{B}_k(W_i)/n), k = 1, \ldots, K = \tilde{K} - 1\}$ (the subspace is $K = \tilde{K} - 1$ dimensional due to the empirical version of the constraint). The multiplicative constant $\sqrt{K}$ is incorporated in the basis definition to simplify some expressions later in the proofs, as done in [23]. Using spline expansions, we can approximate the nonparametric component by $\phi_0(x) \approx \sum_k a_kB_k(x)$.

Using spline expansion introduced above, the problem of estimating ϕ_0 is then transformed to the problem of estimating the coefficients $a = (a_1, \ldots, a_K)^T$. Let $Y_i(t) = 1\{T_i \ge t\}$. We can estimate (ϕ, β) as the maximizer of the (log-)partial likelihood

$$l(\phi, \beta) = \sum_{i=1}^{n} \Delta_i \left\{\phi(W_i) + X_i^T\beta - \log \sum_{k=1}^{n} Y_k(T_i) \exp[\phi(W_k) + X_k^T\beta]\right\}.$$

Using the notation

$$Z_i = (B_1(W_i), \ldots, B_K(W_i))^T,$$

the partial likelihood is written equivalently as

$$l(a, \beta) = \sum_{i=1}^{n} \Delta_i \left\{Z_i^T a + X_i^T\beta - \log \sum_{k=1}^{n} Y_k(T_i) \exp[Z_k^T a + X_k^T\beta]\right\}. \tag{2}$$

We have chosen to use B-splines to estimate the nonparametric components, although other estimation approaches can be applied such as with smoothing splines as in [3]. Using B-splines has the distinct advantage that it can be implemented as in the parametric case after spline approximation. If inferences on β and ϕ are desired, one can use the sandwich estimator involving the observed information matrix.

Even though our main interest is in empirical likelihood, we first study the maximum partial likelihood estimator for Cox models with growing dimensions. The results, in particular the score functions used in the proof, will motivate our bias-corrected estimating equation later. The efforts we make in the proof will also shorten our later proofs for empirical likelihood, since the proofs have many common features. Finally, the theoretical study of the maximum partial likelihood estimator is of independent interest in the case of growing dimensions.

It is sometimes more convenient to consider the counting process representation of the partial likelihood. Let $N_i(t) = 1\{T_i \le t, \delta = 1\}$ be the right continuous counting process and $Y_i(t) = 1\{T_i \ge t\}$ be the left-continuous at-risk process for the ith individual. Denote the true risk score by $m_0(W, X) = \phi_0(W) + X^T\beta_0$. Let $R = (W, X)$ be all the covariates. Denote by g, h any functions of R (h can be vector-valued). Define

$$S_n^{(0)}(g,t) = n^{-1}\sum_{i=1}^n Y_i(t)\exp[g(R_i)],$$

$$S_n^{(1)}(g,t)[h] = n^{-1}\sum_{i=1}^n Y_i(t)h(R_i)\exp[g(R_i)],$$

$$S_n^{(2)}(g,t)[h] = n^{-1}\sum_{i=1}^n Y_i(t)h(R_i)^{\otimes 2}\exp[g(R_i)],$$

$$G_n(g,t)[h] = S_n^{(1)}(g,t)[h]/S_n^{(0)}(g,t),$$

$$V_n(g,t)[h] = S_n^{(2)}(g,t)[h]/S_n^{(0)}(g,t) - G_n(g,t)[h]G_n^T(g,t)[h],$$

where for any vector ξ, $\xi^{\otimes 2}$ simply means $\xi\xi^T$.

Let $S^{(j)}(g,t) = E(S_n^{(j)}(g,t)), j = 0, 1, 2, G(g,t)[h] = S^{(1)}(g,t)[h]/S^{(0)}(g,t), V(g,t)[h] = S^{(2)}(g,t)[h]/S^{(0)}(g,t) - G(g,t)[h]G^T(g,t)[h]$. We also let $P_{\Delta n}$ be the empirical measure of $(T_i, \Delta_i = 1, R_i)$, that is for any function f of (T, Δ, R), $P_{\Delta n}f = n^{-1}\sum_i \Delta_i f(T_i, \Delta_i, R_i)$, and let P_Δ be the measure of $(T, \Delta = 1, R)$.

The partial likelihood can be rewritten as

$$l(a,\beta) = \sum_{i=1}^n \int_0^\tau \{Q_i^T b - \log(S_n^{(0)}(g,t))\}dN_i(t),$$

where $b = (a^T, \beta^T)^T, Q_i = (Z_i^T, X_i^T)^T$, and g is the function of R defined by (a, β): $g(R) = a^T B(W) + X^T\beta, B(W) = (B_1(W), \ldots, B_K(W))^T$. Note that as usual we only consider events over a finite interval $[0, \tau]$. The score function is given by

$$U(a,\beta) = \sum_{i=1}^n \int_0^\tau \{Q_i - G_n(g,t)[Q]\}dN_i(t).$$

Define $\mathcal{H}_d$ as the collection of all functions on support $[0, 1]$ whose mth order derivative satisfies the Hölder condition of order r with $d \equiv m+r$. That is, for each $h \in \mathcal{H}_d$, there exists a constant $M_0 \in (0,\infty)$ such that $|h^{(m)}(s) - h^{(m)}(t)| \le M_0|s-t|^r$, for any $s, t \in [0, 1]$.

The following technical conditions are used in the study of asymptotics.

(C1) The covariate vector R has a bounded support (without loss of generality the support is assumed to be $[0, 1]^{(p+1)}$), with the marginal density of each covariate being continuous and bounded away from zero and infinity.

(C2) Only observations with censored event times in a finite interval $[0, \tau]$ are used in the partial likelihood. $P(\Delta = 1|R)$ and $P(T^c > \tau|R)$ are both bounded away from zero with probability one.

(C3) $\phi_0(x) \in \mathcal{H}_d$, for some $d > 1/2$. The order of the spline satisfies $r > d + 1/2$.

(C4) Let $\Sigma = \int_0^\tau V(m_0,t)[Q]S^{(0)}(m_0,t)\lambda_0(t)dt$. The eigenvalues of Σ are bounded away from zero and infinity.

Assumptions (C1)–(C4) are similar to those contained in [7] or [1]. Boundedness of covariates is assumed for simplicity and might be relaxed to some moment conditions. The term QQ^T appears in the definition of Σ. Under mild assumptions, Huang et al. [8] showed that eigenvalues of EZZ^T are bounded and bounded away from zero. Thus it is expected that eigenvalues of EQQ^T are bounded and bounded away from zero if eigenvalues of EXX^T are so and Z and X are not linear dependent. This can in turn be guaranteed with assumptions on the density of (T, W, X) as in assumptions (B5) and (B6) of Huang [7]. Note that Σ is the population information matrix of the model and thus its positive-definiteness is a reasonable assumption.

Theorem 1 (*Convergence Rates*). *Under conditions* (C1)–(C4), *assume that* $K \to \infty$, $(K+p)^2/n \to 0$, *and let* $(\hat{a}, \hat{\beta})$ *be the maximizer of* $l(a,\beta)$ *in* (2) *and* $\hat{\phi} = \sum_k \hat{a}_k B_k$. *We have*

$$\|\hat{\phi} - \phi_0\| + \|\hat{\beta} - \beta_0\| = O_p\left(\sqrt{\frac{K+p}{n}} + \frac{1}{K^d}\right).$$

From the rates, it is seen that the optimal choice of K is of order $O(n^{1/(2d+1)})$ as usual. The following theorem shows that in fact the linear coefficients are asymptotically normal and thus converge at the $\sqrt{n}$ rate. For this further notations and assumptions are necessary.

Let a^* and h^* be R^p-valued L_2 functions that minimize $E\Delta\|X-a(T)-h(W)\|^2$. Denote $\tilde{\Sigma} = E\Delta(X-a^*(T)-h^*(W))^{\otimes 2}$. By direct calculations, it can be easily verified that $\tilde{\Sigma}$ can also be written as $\int_0^\tau V(m_0,t)[X-h^*(W)]S^{(0)}(m_0,t)\lambda_0(t)dt$ and is thus the information matrix for the linear part after taking into account the effect of the nonparametric part. The p components of h^* are denoted by h_j^*, $1 \le j \le p$.

(C5) All h_j^*, a_j^*, $1 \le j \le p$, are in $\mathcal{H}_d$. The eigenvalues of $\tilde{\Sigma}$ are bounded and bounded away from zero.

Theorem 2 (*Asymptotic Normality*)**.** *Under the same conditions as assumed in Theorem 1 and in addition* (C5) *holds and $p = o(n^{1/3})$, then for any unit p-vector ν_n, we have*

$$\sqrt{n}\nu_n^T\tilde{\Sigma}^{1/2}(\hat{\beta}-\beta_0) \to N(0,1).$$

Thus, informally, we can say $\hat{\beta}$ is asymptotically normal with asymptotic variance $\tilde{\Sigma}^{-1}/n$.

3. Empirical likelihood

If the nonparametric part ϕ is known, we can use the following random vector

$$\int_0^\tau X_i - \frac{S_n^{(1)}(\phi(W)+X\beta,t)[X]}{S_n^{(0)}(\phi(W)+X\beta,t)}dN_i(t), \quad i=1,\dots,n,$$

which is almost the same as the parametric case, to define the empirical likelihood. When ϕ is unknown, the most obvious strategy is to plug in some estimator $\hat{\phi}$ of ϕ. In particular we will use the estimator $\hat{\phi}(w)$ which is obtained from (2). However, as noted in [12], the estimating equation with the nonparametric part plugged in is not longer unbiased, which in turn will lead to the fact that the limit distribution of the empirical likelihood ratio is non-standard. Motivated from the score function contained in the proof of Theorem 2 (see Eq. (11) there), we define the bias-corrected random vector

$$\eta_i(\beta) = \int_0^\tau X_i - h^*(W) - \frac{S_n^{(1)}(\phi(W)+X\beta,t)[X-h^*]}{S_n^{(0)}(\phi(W)+X\beta,t)}dN_i(t), \quad i=1,\dots,n,$$

and its estimated version

$$\hat{\eta}_i(\beta) = \int_0^\tau X_i - \hat{h}^*(W) - \frac{S_n^{(1)}(\hat{\phi}(W)+X\beta,t)[X-\hat{h}^*]}{S_n^{(0)}(\hat{\phi}(W)+X\beta,t)}dN_i(t), \quad i=1,\dots,n.$$

By the definition of h^* we can estimate it based on the minimization of

$$\sum_{i=1}^n \Delta_i\|X_i - h(W_i) - a(T_i)\|^2$$

over h and a simultaneously and both are approximated using splines. Under the smoothness assumption of h^* and a^* in (C5), it is easy to show that $\|\hat{h}^* - h^*\| = O_p(\sqrt{K/n} + K^{-d})$. To see this, note that this is just a weighted additive model with weights given by Δ_i. The loss function is weighted least squares loss and no partial likelihood is used and thus the proof of convergence is quite standard.

Summarizing the above, a bias-corrected empirical likelihood ratio is defined as

$$R(\beta) = \max\left\{\prod_{i=1}^n (nw_i) : w_i \ge 0, \sum_{i=1}^n w_i = 1, \sum_{i=1}^n w_i\hat{\eta}_i(\beta) = 0\right\}.$$

By the Lagrange multiplier, $-2\log R(\beta)$ can be represented in the dual form

$$-2\log R(\beta) = 2\sum_{i=1}^n \log(1+\lambda^T\hat{\eta}_i(\beta)),$$

where $\lambda \in R^p$ is the solution of

$$\frac{1}{n}\sum_{i=1}^n \frac{\hat{\eta}_i(\beta)}{1+\lambda\hat{\eta}_i(\beta)} = 0.$$

The subtraction of h^* (or its estimator) introduces (approximate) conditional orthogonality into the estimating equation and helps to get a faster rate of the bias to zero (see Eq. (14) in the Appendix, for example). It is worth noting that even in the fixed p case, our results below are of interest and not found in the existing literature. Our main result is the following Wilk's theorem.

Theorem 3. *Under the assumptions* (C1)–(C5), *and that* $p \to \infty$, $(K+p)^2/n \to 0$, $p^3/n \to 0$. *Let* β_0 *be the true value of the parameter vector, we have*

$$\frac{-2\log R(\beta_0) - p}{\sqrt{2p}} \xrightarrow{d} N(0, 1).$$

Remark 1. This theorem indicates that the distribution of $-2\log R(\beta_0)$ is asymptotically close to $N(p, 2p)$. Also, an asymptotically valid confidence region of β is $\{\beta : -2\log R(\beta) \le p + z_\alpha\sqrt{2p}\}$ where z_α is the upper α-quantile of the standard normal distribution.

In practice, it is rarely the case that confidence regions for the entire parameter vector will be sought, since as soon as $p > 3$, it is hard to visualize or represent the region. Typically one will only be interested in a one- or two-dimensional subvector of β. Suppose $\beta = (\beta^{(1)T}, \beta^{(2)T})^T$ where $\beta^{(1)}$ is l-dimensional (l is fixed and does not diverge with n) for which confidence interval/region is to be constructed. Similarly, the true parameter vector can be partitioned as $\beta_0 = (\beta_0^{(1)T}, \beta_0^{(2)T})^T$. Other vectors such as X_i and h^* can be similarly partitioned.

With fixed $\beta^{(1)}$ we can obtain estimators for the rest of the parameters and the nonparametric function from (2), denoted by $\hat{\beta}^{(2)}$ and $\hat{\phi}$ respectively.

Define

$$\hat{\hat{\eta}}_i(\beta) = \int_0^\tau X_i^{(1)} - \hat{h}^{*(1)}(W) - \frac{S_n^{(1)}(\hat{\phi}(W) + X^{(1)}\beta^{(1)} + X^{(2)}\hat{\beta}^{(2)}, t)[X^{(1)} - \hat{h}^{*(1)}]}{S_n^{(0)}(\hat{\phi}(W) + X\beta, t)} dN_i(t).$$

The bias-corrected empirical likelihood for $\beta^{(1)}$ is

$$R(\beta^{(1)}) = \max\left\{\prod_{i=1}^n (nw_i) : w_i \ge 0, \sum_{i=1}^n w_i = 1, \sum_{i=1}^n w_i \hat{\hat{\eta}}_i(\beta) = 0\right\}.$$

Theorem 4. *Under the same assumptions as in Theorem 3, we have*

$$-2\log R(\beta_0^{(1)}) \xrightarrow{d} \chi_l^2,$$

where χ_l^2 *denotes the chi-square distribution with l degrees of freedom.*

Based on the above theorem, the confidence interval or region for $\beta^{(1)}$ can be easily constructed.

To implement the procedures, we also need to choose the number of spline basis K. To ease the computational burden, we use cubic splines with $K = 6$ following Huang et al. [8]. This choice of K is small enough to avoid overfitting in typical problems with sample size not too small, and big enough to flexibly approximate many smooth functions. We also find the results are very similar for K ranging from 5 to 8 in our simulations, and thus we only report the results obtained with $K = 6$ later.

4. Simulations

In this section, we conduct some Monte Carlo studies to demonstrate the finite sample properties of the empirical likelihood inferences. We generate our data from the Cox model with hazard function given by

$$\lambda(t|W, X) = \exp\left\{\phi_0(W) + \sum_{j=1}^p X_j\beta_{0j} - 5\right\},$$

where the nonparametric component is $\phi_0(t) = \sin(2\pi t)$, and the coefficients in the linear part are $\beta = (5, 4.5, 4, \ldots, 5.5 - p/2)$. The covariates $R = (W, X)$ were generated as follows. We first generate a multivariate $(p+1)$-dimensional Gaussian vector $(V_1, \ldots, V_{p+1})$ with covariance given by $\mathrm{Cov}(V_j, V_{j'}) = (0.2)^{|j-j'|}$. Then the cumulative distribution function of the standard normal distribution is applied to each component to map the components to the range (0, 1), to obtain $W, X_1, \ldots, X_p$. The censoring times are independently generated from an exponential distribution with mean 200. Under our simulation setup, about 25% observations are censored in the generated datasets. We use $n = 200, 400, 600$ and $p = 10, 20, 30$. We use cubic splines with $K = 6$.

We consider an empirical-likelihood-based 95% confidence interval for β_1 and results for other parameters are similar and not presented here. In each case we repeat the simulation 1000 times. For comparison, we also construct confidence intervals using the sandwich formula. The results are presented in Table 1. We find that empirical-likelihood-based intervals (EL) consistently have better coverage probability (CP) with slightly longer average length (AL) than the sandwich-formula-based intervals (SF).

In Table 2, we further consider confidence regions for (β_1, β_2) constructed by empirical likelihood and the sandwich formula where we show the coverage of the constructed regions. The message is similar as before and we see that empirical likelihood performs much better.

Table 1
The coverage probability (CP) and average length (AL) on β_1. We compare empirical-likelihood-based intervals (EL) with the sandwich-formula-based intervals (SF).

Method	n	$p = 10$		$p = 20$		$p = 30$	
		CP	AL	CP	AL	CP	AL
EL	200	0.937	2.02	0.934	2.40	0.924	2.38
	400	0.945	1.31	0.941	1.45	0.935	1.49
	600	0.947	1.04	0.947	1.12	0.937	1.13
SF	200	0.906	1.93	0.861	2.16	0.803	1.99
	400	0.921	1.21	0.904	1.40	0.852	1.47
	600	0.930	0.99	0.927	1.09	0.861	1.12

Table 2
The coverage probability (CP) of confidence regions for (β_1, β_2). We compare empirical-likelihood-based intervals (EL) with the sandwich-formula-based intervals (SF).

Method	n	$p = 10$	$p = 20$	$p = 30$
EL	200	0.925	0.920	0.920
	400	0.928	0.926	0.925
	600	0.939	0.934	0.935
SF	200	0.884	0.848	0.772
	400	0.882	0.857	0.804
	600	0.887	0.852	0.811

5. Concluding remarks

In this article, we studied the empirical likelihood ratio for inferences on the parametric part of partially linear Cox models. A simple bias correction method is proposed so that the asymptotic distribution of the statistic is a standard chi-square distribution.

When p is large, it is also interesting to perform variable selection to identify a small number of significant predictors. Recently, Tang and Leng [22], [11] have considered combining penalized variable selection with empirical likelihood. It would be interesting to consider a similar strategy for partially linear Cox models.

One could also consider the case where W is high-dimensional. However, using a partially linear model with high-dimensional W leads to the worry of the curse of dimensionality. To address this problem, one might consider a partially linear additive model as in [7]. A detailed study of this is outside the scope of the current paper.

Appendix

Proof of Theorem 1. The strategy of proof is similar to Bradic et al. [1] with the main difference being that the nonparametric components need to be appropriately dealt with in spline approximation. Let a_0 be a K dimensional vector that satisfies $\|\phi_0 - a_0^T B\|_\infty = O(K^{-d})$ (such approximation rates are possible due to our smoothness assumption (C2) and well-known approximation properties of B-splines). Denote $b = (a, \beta)$ and $b_0 = (a_0, \beta_0)$. Let $\gamma_n = C(\sqrt{(K+p)/n} + K^{-d})$ and $u \in R^{K+p}$ with $\|u\| = 1$.

It is sufficient to show that for any $\epsilon > 0$, there exists a large enough C (in the definition of γ_n) such that

$$P\{\sup_{\|u\|=1} l((a_0, \beta_0) + \gamma_n u) < l(a_0, \beta_0)\} \geq 1 - \epsilon, \tag{3}$$

when n is big enough.

We have

$$l(b_0 + \gamma_n u) - l(b_0) = \gamma_n U(b_0)^T u + \frac{1}{2}\gamma_n^2 u^T \partial U(b_0) u + r_n, \tag{4}$$

where r_n is equal to

$$\frac{1}{6}\sum_{j,k,l}(b_j - b_{0j})(b_k - b_{0k})(b_l - b_{0l})\frac{\partial^2 U_l(\tilde{b})}{\partial b_j \partial b_k}$$

where U_l is the l-th component of U, and $\tilde{b}$ is a value between b_0 and $b = b_0 + \gamma_n u$.

We first consider

$$U(b_0) = \sum_i \int_0^\tau Q_i - \frac{S_n^{(1)}(m_{0n}, t)[Q]}{S_n^{(0)}(m_{0n}, t)} dN_i(t),$$

where $m_{0n}(R) = Za_0 + X^T \beta_0$.

Similar to Lemma 5.3 of Huang [7], we have

$$P_{\Delta n}\frac{S_n^{(1)}(m_{0n},t)[Q]}{S_n^{(0)}(m_{0n},t)}-\frac{S_n^{(1)}(m_0,t)[Q]}{S_n^{(0)}(m_0,t)}=P_{\Delta}\frac{S^{(1)}(m_{0n},t)[Q]}{S^{(0)}(m_{0n},t)}-\frac{S^{(1)}(m_0,t)[Q]}{S^{(0)}(m_0,t)}+o_p(n^{-1/2}) \tag{5}$$

where $m_0(R)=\phi_0(W)+X^T\beta_0$. Using Lemma A.4 of Huang [7] and that $\|m_{0n}-m_0\|=O(K^{-d})$, we have

$$P_{\Delta}\frac{S^{(1)}(m_{0n},t)[Q]}{S^{(0)}(m_{0n},t)}-\frac{S^{(1)}(m_0,t)[Q]}{S^{(0)}(m_0,t)}=O(K^{-d}). \tag{6}$$

Thus

$$U(b_0)=\sum_i\int_0^\tau Q_i-\frac{S_n^{(1)}(m_0,t)[Q]}{S_n^{(0)}(m_0,t)}dN_i(t)+O_p(\sqrt{n}+nK^{-d}).$$

Let ξ_n denote the first term on the right hand side above, direct algebraic calculations show that

$$\begin{aligned}E(\xi_n^T\xi_n)&=\operatorname{tr}(E[\xi_n\xi_n^T])\\&=n\operatorname{tr}\left(E\int_0^\tau V_n(m_0,t)[Q]S_n^{(0)}(m_0,t)\lambda_0(t)dt\right).\end{aligned}$$

Since

$$\begin{aligned}\operatorname{tr}(E[V_n(m_0,t)[Q]S_n^{(0)}(m_0,t)])&=\operatorname{tr}\left(E\left[\sum_i(Q_i-G_n(m_0,t)[Q])^{\otimes 2}Y_i(t)\exp\{m_0(R_i)\}\right]\right)\\&\le E[\operatorname{tr}(Q_i^{\otimes 2}Y_i\exp\{m_0(R_i)\})]=O(K+p),\end{aligned}$$

we have $\|\xi_n\|_2=O_p(\sqrt{n(K+p)})$ and thus

$$\|U(b_0)\|=O_p(\sqrt{n(K+p)}+nK^{-d}). \tag{7}$$

Next, we have

$$\begin{aligned}-\partial U(b_0)&=\sum_i\int_0^\tau\frac{S_n^{(2)}(m_{0n},t)[Q]S_n^{(0)}(m_{0n},t)-(S_n^{(1)}(m_{0n},t)[Q])^{\otimes 2}}{(S_n^{(0)}(m_{0n},t))^2}dN_i(t)\\&=\sum_i\int_0^\tau\frac{S_n^{(2)}(m_0,t)[Q]S_n^{(0)}(m_0,t)-(S_n^{(1)}(m_0,t)[Q])^{\otimes 2}}{(S_n^{(0)}(m_0,t))^2}dN_i(t)+O(nK^{-d})\end{aligned} \tag{8}$$

where again we used Lemma A.4 of Huang [7]. Similar to Lemmas A.2 and A.4 of Bradic et al. [1], we can show that

$$\begin{aligned}&\sum_i\int_0^\tau\frac{S_n^{(2)}(m_0,t)[Q]S_n^{(0)}(m_0,t)-(S_n^{(1)}(m_0,t)[Q])^{\otimes 2}}{(S_n^{(0)}(m_0,t))^2}dN_i(t)\\&=n\int_0^\tau V(m_0,t)[Q]S^{(0)}(m_0,t)\lambda_0(t)dt+o_p(n),\end{aligned} \tag{9}$$

and thus the minimum eigenvalue of $-\partial U(b_0)/n$ is bounded away from zero.

Then, as in the proof of Theorem 4.2 in [1],

$$r_n=O_p(n\gamma_n^3). \tag{10}$$

Combining (4)–(10), we get $\|\hat{b}-b_0\|=O_p(\gamma_n)$, which implies the statement of the theorem. □

Proof of Theorem 2. Let $h_n^*=(h_{n1}^*,\dots,h_{np}^*)^T$ be the spline functions that approximate $h^*=(h_1^*,\dots,h_p^*)^T$ with $\|h_{nj}^*-h_j^*\|_\infty=O(K^{-d})$. Since $\hat{b}=(\hat{a},\hat{\beta})$ maximizes the partial likelihood (2), it is easy to see that $v=0$ maximizes

$$\sum_{i=1}^n\int_0^\tau\hat{m}(R_i)+(X_i-h_n^*(W_i))^Tv-\log\left(\sum_kY_k(t)\exp\{\hat{m}(R_i)+(X_i-h_n^*(W_i))^Tv\}\right)dN_i(t),$$

where $\hat{m}(R)=Z^T\hat{a}+X^T\hat{\beta}$. The first order condition gives

$$\sum_i\int_0^\tau(X_i-h_n^*(W_i))-\frac{S_n^{(1)}(\hat{m},t)[X-h_n^*(W)]}{S_n^{(0)}(\hat{m},t)}dN_i(t)=0. \tag{11}$$

Consider the difference

$$
\begin{aligned}
& n^{-1}\sum_i \int_0^\tau (X_i - h^*(W_i)) - \frac{S_n^{(1)}(\hat{m},t)[X-h^*(W)]}{S_n^{(0)}(\hat{m},t)} dN_i(t) \\
& \quad - n^{-1}\sum_i \int_0^\tau (X_i - h_n^*(W_i)) - \frac{S_n^{(1)}(\hat{m},t)[X-h_n^*(W)]}{S_n^{(0)}(\hat{m},t)} dN_i(t) \\
& \quad = n^{-1}\sum_i \int_0^\tau (h_n^*(W_i) - h^*(W_i)) - \frac{S_n^{(1)}(\hat{m},t)[h_n^*(W)-h^*(W)]}{S_n^{(0)}(\hat{m},t)} dN_i(t) \\
& \quad \equiv A_{1n} + A_{2n} + A_{3n} + A_{4n}
\end{aligned}
$$

where

$$
\begin{aligned}
A_{1n} &= (P_{\Delta n} - P_\Delta)\left\{ h_n^*(W_i) - h^*(W_i) - \frac{S^{(1)}(\hat{m},t)[h_n^*(W)-h^*(W)]}{S^{(0)}(\hat{m},t)} \right\}, \\
A_{2n} &= P_{\Delta n}\left\{ \frac{S^{(1)}(\hat{m},t)[h_n^*(W)-h^*(W)]}{S^{(0)}(\hat{m},t)} - \frac{S_n^{(1)}(\hat{m},t)[h_n^*(W)-h^*(W)]}{S_n^{(0)}(\hat{m},t)} \right\}, \\
A_{3n} &= P_\Delta\left\{ \frac{S^{(1)}(m_0,t)[h_n^*(W)-h^*(W)]}{S^{(0)}(m_0,t)} - \frac{S^{(1)}(\hat{m},t)[h_n^*(W)-h^*(W)]}{S^{(0)}(\hat{m},t)} \right\}, \\
A_{4n} &= P_\Delta\left(h_n^*(W_i) - h^*(W_i) - \frac{S^{(1)}(m_0,t)[h_n^*(W)-h^*(W)]}{S^{(0)}(m_0,t)} \right).
\end{aligned}
$$

By the maximal inequality and the entropy calculations in Lemma A.1 and Corollary A.1 of Huang [7], we have $A_{1n} = o_p(n^{-1/2})$. Similar to Lemma A.3 of Huang [7], $A_{2n} = o_p(n^{-1/2})$. Similar to (6) and using that $\|h_{nj}^* - h_j^*\| = O(K^{-d})$, $A_{3n} = o_p(n^{-1/2})$ and finally, we note that since for any function h, $S^{(1)}(m_0,t)[h]/S^{(0)}(m_0,t) = E[h(R)|T=t,\Delta=1]$ [20, Lemma 2], $A_{4n} = 0$.

Thus we have that

$$
n^{-1}\sum_i \int_0^\tau (X_i - h^*(W_i)) - \frac{S_n^{(1)}(\hat{m},t)[X-h^*(W)]}{S_n^{(0)}(\hat{m},t)} dN_i(t) = o_p(n^{-1/2}). \tag{12}
$$

Similar to (5),

$$
\begin{aligned}
& n^{-1}\sum_i \int_0^\tau (X_i - h^*(W_i)) - \frac{S_n^{(1)}(\hat{m},t)[X-h^*(W)]}{S_n^{(0)}(\hat{m},t)} dN_i(t) \\
& \quad = n^{-1}\sum_i \int_0^\tau (X_i - h^*(W_i)) - \frac{S_n^{(1)}(m_0,t)[X-h^*(W)]}{S_n^{(0)}(m_0,t)} dN_i(t) \\
& \quad - P_\Delta\left(\frac{S^{(1)}(\hat{m},t)[X-h^*(W)]}{S^{(0)}(\hat{m},t)} - \frac{S^{(1)}(m_0,t)[X-h^*(W)]}{S^{(0)}(m_0,t)} \right) + o_p(n^{-1/2}).
\end{aligned} \tag{13}
$$

Direct Taylor expansion shows that

$$
\begin{aligned}
& P_\Delta\left(\frac{S^{(1)}(\hat{m},t)[X-h^*(W)]}{S^{(0)}(\hat{m},t)} - \frac{S^{(1)}(m_0,t)[X-h^*(W)]}{S^{(0)}(m_0,t)} \right) \\
& \quad = P_\Delta\left(X - h^*(W) - \frac{S^{(1)}(m_0,t)[X-h^*(W)]}{S^{(0)}(m_0,t)} \right)\left(X - \frac{S^{(1)}(m_0,t)[X]}{S^{(0)}(m_0,t)} \right)(\hat{\beta}-\beta_0) \\
& \quad + P_\Delta\left(X - h^*(W) - \frac{S^{(1)}(m_0,t)[X-h^*(W)]}{S^{(0)}(m_0,t)} \right)\left(\hat{\phi}(W) - \phi_0(W) - \frac{S^{(1)}(m_0,t)[\hat{\phi}-\phi_0]}{S^{(0)}(m_0,t)} \right) + O_p(\|\hat{m}-m_0\|^2) \\
& \quad = P_\Delta\left(X - h^*(W) - \frac{S^{(1)}(m_0,t)[X-h^*(W)]}{S^{(0)}(m_0,t)} \right)^{\otimes 2}(\hat{\beta}-\beta_0) + o_p(n^{-1/2}),
\end{aligned} \tag{14}
$$

where the last step used that

$$P_\Delta\left(X - h^*(W) - \frac{S^{(1)}(m_0,t)[X-h^*(W)]}{S^{(0)}(m_0,t)}\right)\left(\hat\phi(W) - \phi_0(W) - \frac{S^{(1)}(m_0,t)[\hat\phi-\phi_0]}{S^{(0)}(m_0,t)}\right) = 0,$$

by the definition of h^*.

Furthermore, similar to A_{2n} above,

$$\begin{aligned}
&n^{-1}\sum_i \int_0^\tau (X_i - h^*(W_i)) - \frac{S_n^{(1)}(m_0,t)[X-h^*(W)]}{S_n^{(0)}(m_0,t)} dN_i(t)\\
&\quad = n^{-1}\sum_i \int_0^\tau (X_i - h^*(W_i)) - \frac{S_n^{(1)}(m_0,t)[X-h^*(W)]}{S_n^{(0)}(m_0,t)} dM_i(t)\\
&\quad = n^{-1}\sum_i \int_0^\tau (X_i - h^*(W_i)) - \frac{S^{(1)}(m_0,t)[X-h^*(W)]}{S^{(0)}(m_0,t)} dM_i(t) + o_p(n^{-1/2}) \qquad (15)
\end{aligned}$$

and can be seen to be asymptotically normal, where $M_i(t) = N_i(t) - \int_0^t Y_i(t)\exp[m_0(R_i)]\lambda_0(t)dt$. Combining (12)–(15), the asymptotic normality of $\hat\beta$ follows. We also note that

$$\begin{aligned}
&P_\Delta\left(X - h^*(W) - \frac{S^{(1)}(m_0,t)[X-h^*(W)]}{S^{(0)}(m_0,t)}\right)^{\otimes 2}\\
&\quad = P_\Delta(X - h^*(W) - E[X-h^*(W)]|T=t, \Delta=1)^{\otimes 2}\\
&\quad = P_\Delta(X - h^*(W) - a^*(T))^{\otimes 2}. \quad \square
\end{aligned}$$

Proof of Theorems 3 and 4. Using the results similar to those proved in the previous theorems, the asymptotic study for empirical likelihood ratio becomes relatively simple.

We write $\lambda = \|\lambda\|\theta$ where θ is a unit vector. We have

$$\frac{1}{n}\sum_{i=1}^n \frac{\theta^T\hat\eta_i}{1+\lambda^T\hat\eta_i} = 0.$$

Let $Y_i = \lambda^T\hat\eta_i$. Substituting $1/(1+Y_i) = 1 - Y_i/(1+Y_i)$ into the above displayed equation, we have

$$\theta^T\bar{\hat\eta} = \|\lambda\|\theta^T\hat S\theta,$$

where

$$\bar{\hat\eta} = \frac{1}{n}\sum_{i=1}^n \hat\eta_i, \qquad \hat S = \frac{1}{n}\sum_{i=1}^n \frac{\hat\eta_i\hat\eta_i^T}{1+\lambda^T\hat\eta_i}.$$

Define $S = \frac{1}{n}\sum_{i=1}^n \hat\eta_i\hat\eta_i^T$. Then we have

$$\begin{aligned}
\|\lambda\|\theta^T S\theta &\le \|\lambda\|\theta^T\hat S\theta(1 + \|\lambda\|\max_i\|\hat\eta_i\|)\\
&= \theta^T\bar{\hat\eta}(1+\|\lambda\|\max_i\|\hat\eta_i\|). \qquad (16)
\end{aligned}$$

We now use that $\max_i\|\hat\eta_i\| \le \max_i\|\eta_i\| + \max_i\|\hat\eta_i - \eta_i\| = O_p(\sqrt{p}) + \max_i\|\hat\eta_i - \eta_i\|$ ($\max_i\|\eta_i\| = O_p(\sqrt{p})$ since we used the simplifying assumption X is bounded and the score function for the definition of η_i is basically a projection) and that

$$\begin{aligned}
\hat\eta_i - \eta_i = &\int_0^\tau h^*(W_i) - \hat h^*(W_i) - \frac{S^{(1)}(\phi_0(W)+X^T\beta_0,t)[h^*-\hat h^*]}{S^{(0)}(\phi_0(W)+X^T\beta_0,t)} dN_i(t)\\
&+ \int_0^\tau \frac{S_n^{(1)}(\phi_0(W)+X^T\beta_0,t)[X-h^*]}{S_n^{(0)}(\phi_0(W)+X^T\beta_0,t)} - \frac{S_n^{(1)}(\hat\phi(W;\beta_0)+X^T\beta_0,t)[X-h^*]}{S_n^{(0)}(\hat\phi(W;\beta_0)+X^T\beta_0,t)} dN_i(t)\\
&+ \int_0^\tau \frac{S_n^{(1)}(\phi_0(W)+X^T\beta_0,t)[h^*-\hat h^*]}{S_n^{(0)}(\phi_0(W)+X^T\beta_0,t)} - \frac{S_n^{(1)}(\hat\phi(W;\beta_0)+X^T\beta_0,t)[h^*-\hat h^*]}{S_n^{(0)}(\hat\phi(W;\beta_0)+X^T\beta_0,t)} dN_i(t)\\
&+ \int_0^\tau \frac{S^{(1)}(\phi_0(W)+X^T\beta_0,t)[h^*-\hat h^*]}{S^{(0)}(\phi_0(W)+X^T\beta_0,t)} - \frac{S_n^{(1)}(\phi_0(W)+X^T\beta_0,t)[h^*-\hat h^*]}{S_n^{(0)}(\phi_0(W)+X^T\beta_0,t)} dN_i(t)\\
=:&\ A_1 + A_2 + A_3 + A_4. \qquad (17)
\end{aligned}$$

From the results of Theorem 1, $\|\hat{\phi}-\phi_0\| = O_p(\delta_n)$ where $\delta_n = \sqrt{(K+p)/n} + K^{-d}$. The first term in (17) has mean zero and by calculating its second moment, it is straightforward to show $A_1 = O_p(\sqrt{p}\delta_n)$. Using Lemmas A.3 and A.6 in [7], we can similarly show $A_2 = O_p(\sqrt{p}\delta_n)$ (note this convergence is uniform in i due to these lemmas). Again using the same lemmas in that paper, $A_3 = O_p(\sqrt{p}\delta_n^2)$, $A_4 = O_p(\sqrt{p}\delta_n^2)$. These taken together imply that $\max_i \|\hat{\eta}_i\| = O_p(\sqrt{p})$.

Using that $\theta^T\bar{\hat{\eta}} = O_p(\sqrt{p/n})$, (16) implies

$$\|\lambda\|(\theta^T S\theta - o_p(1)) = O_p(\sqrt{p/n}).$$

Using (17), it can be shown that the eigenvalues of S is bounded away from zero by assumption (C5), which then gives $\|\lambda\| = \sqrt{p/n}$.

Now, we have

$$\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_i\left\{1 - Y_i + \frac{Y_i^2}{1-Y_i}\right\} \\
&= \bar{\hat{\eta}} - S\lambda + \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\eta}_i Y_i^2}{1-Y_i}.
\end{aligned}$$

The last term above is of order $O_p(\max_i \|\hat{\eta}_i\|\lambda^T S\lambda) = O_p(p^{3/2}/n)$.

Now we may write

$$\begin{aligned}
-2\log R(\beta_0) &= 2\log(1+Y_i) \\
&= 2\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} Y_i^2 + o_p(1) \\
&= 2n\lambda^T\bar{\hat{\eta}} - n\lambda^T S\lambda + o_p(1) \\
&= n\bar{\hat{\eta}}^T S^{-1}\bar{\hat{\eta}} + o_p(1).
\end{aligned}$$

Using (17), it can be shown exactly as in Lemma B.5 of Li et al. [12] that

$$\frac{-2\log R(\beta_0) - p}{\sqrt{2p}} \xrightarrow{d} N(0,1).$$

Note the proof of Theorem 3 is based on profiling out the nonparametric part (the score function is only for the parametric part). To show Theorem 4 for inferences on $\beta^{(1)}$, we need to actually profile out $\beta^{(2)}$ together with the nonparametric part. Conceptually we just combine the nonparametric part ϕ together with $\beta^{(2)}$ which is the part that needs to be profiled out. Then the proof follows line by line as for Theorem 3 without change (with the slight simplification due to that the dimension is fixed now for the score function of $\beta^{(1)}$). $\square$

References

[1] J. Bradic, J. Fan, J. Jiang, Regularization for Cox's proportional hazards model With NP-dimensionality, Annals of Statistics 39 (6) (2011) 3092–3120.
[2] S. Chen, L. Peng, Y. Qin, Effects of data dimension on empirical likelihood, Biometrika 96 (3) (2009) 711–722.
[3] P. Du, S. Ma, H. Liang, Penalized variable selection procedure for Cox models with semiparametric relative risk, Annals of Statistics 38 (4) (2010) 2092–2117.
[4] J. Fan, I. Gijbels, M. King, Local likelihood and local partial likelihood in hazard regression, The Annals of Statistics 25 (4) (1997) 1661–1690.
[5] G. Heller, The cox proportional hazards model with a partly linear relative risk function, Lifetime Data Analysis 7 (3) (2001) 255–277.
[6] N.L. Hjort, I.W. McKeague, I.V. Keilegom, Extending the scope of empirical likelihood, Annals of Statistics 37 (2009) 1079–1111.
[7] J. Huang, Efficient estimation of the partly linear additive Cox model, Annals of Statistics 27 (5) (1999) 1536–1563.
[8] J. Huang, J.L. Horowitz, F. Wei, Variable selection in nonparametric additive models, Annals of Statistics 38 (4) (2010) 2282–2313.
[9] Z. Huang, R. Zhang, Efficient empirical-likelihood-based inferences for the single-index model, Journal of Multivariate Analysis 102 (5) (2011) 937–947.
[10] Z. Huang, R. Zhang, Profile empirical-likelihood inferences for the single-index-coefficient regression model, Statistics and Computing 23 (4) (2013) 455–465.
[11] C. Leng, C. Tang, Penalized empirical likelihood and growing dimensional general estimating equations, Biometrika 99 (3) (2012) 703–716.
[12] G. Li, L. Lin, L. Zhu, Empirical likelihood for a varying coefficient partially linear model with diverging number of parameters, Journal of Multivariate Analysis 105 (2012) 85–111.
[13] G. Li, L. Zhu, L. Xue, S. Feng, Empirical likelihood inference in partially linear single-index models for longitudinal data, Journal of Multivariate Analysis 101 (3) (2010) 718–732.
[14] J. Nielsen, O. Linton, P. Bickel, On a semiparametric survival model with flexible covariate effect, Annals of Statistics 26 (1998) 215–241.
[15] F. O'Sullivan, Nonparametric estimation in the Cox model, The Annals of Statistics 21 (1993) 124–145.
[16] A.B. Owen, Empirical likelihood ratio confidence intervals for a single functional, Biometrika 75 (2) (1988) 237–249.
[17] A. Owen, Empirical likelihood ratio confidence regions, Annals of Statistics 18 (1) (1990) 90–120.
[18] J. Qin, J. Lawless, Empirical likelihood and general estimating equations, Annals of Statistics 22 (1) (1994) 300–325.
[19] P. Sasieni, Information bounds for the conditional hazard ratio in a nested family of regression models, Journal of the Royal Statistical Society: Series B (Statistical Methodology) (1992) 617–635.
[20] P. Sasieni, Non-orthogonal projections and their application to calculating the information in a partly linear Cox model, Scandinavian Journal of Statistics 19 (1992) 215–233.
[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 (3) (2009) 444–462.

[22] C. Tang, C. Leng, Penalized high-dimensional empirical likelihood, Biometrika 97 (4) (2010) 905–920.
[23] L. Wang, X. Liu, H. Liang, R. Carroll, Estimation and variable selection for generalized additive partially linear models, Annals of Statistics 39 (4) (2011) 1827–1851.
[24] L. Xue, L. Zhu, Empirical likelihood for single-index models, Journal of Multivariate Analysis 97 (6) (2006) 1295–1312.
[25] L. Xue, L. Zhu, Empirical likelihood for a varying coefficient model with longitudinal data, Journal of the American Statistical Association 102 (478) (2007) 642–654.
[26] L. Xue, L. Zhu, Empirical likelihood semiparametric regression analysis for longitudinal data, Biometrika 94 (4) (2007) 921–937.
[27] J. You, Y. Zhou, Empirical likelihood for semiparametric varying-coefficient partially linear regression models, Statistics & Probability Letters 76 (4) (2006) 412–422.
[28] L. Zhu, L. Lin, X. Cui, G. Li, Bias-corrected empirical likelihood in a multi-link semiparametric model, Journal of Multivariate Analysis 101 (4) (2010) 850–868.
[29] L. Zhu, L. Xue, Empirical likelihood confidence regions in a partially linear single-index model, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68 (3) (2006) 549–570.