Empirical likelihood in some nonparametric and semiparametric models

Liugen Xue* and Lixing Zhu[†,‡]

In this very selective overview, we summarise the recent developments by our own and other, on the empirical likelihood in some nonparametric and semiparametric regression models. The models include the partially linear model, the single-index model, the partially linear single-index model, the varying coefficient model, and so on. The focus of this overview is to expatiate the adjustment and "bias correction" methodologies when Wilks' phenomenon does not hold. The adjustment or bias correction can make the limiting distributions tractable such that they can be directly used to construct the confidence regions of parameters of interest without the assistance of Monte Carlo approximation.

AMS 2000 subject classifications: Primary 62G08; secondary 62G05, 62H05.
Keywords and phrases: Adjustment, Bias correction, Confidence region, Empirical likelihood, Semiparametric regression models, Wilks' phenomenon.

1. INTRODUCTION

Since the empirical likelihood was proposed by Art Owen in 1988, there has been much research in literature ranging from parametric to nonparametric statistical inference. Hall and La Scale (1990) gave the first review to summarize the main properties of the empirical likelihood in mainly parametric models. Owen (2001) published the first book in this area to summarize the results up to that time. Chen and Keilegon's (2009) review focused on the recent development of the empirical likelihood. Xue and Zhu (2010) published a book that mainly summarizes the their recent results about nonparametric and semiparametric models when the empirical likelihood is applied.

Recent years, the empirical likelihood method has received great attention when we deal with statistical inference for nonparametric and semiparametric regression models. These models include fully nonparametric regression, single-index, partially linear single-index, varying coefficient models, so on and so forth. However, how to efficiently apply the empirical likelihood to these models is of particular interest and challenging. This is because for such models, the classical empirical likelihood ratios are often not asymptotically distribution-free. The main reason causing this difficulty is that in such models, there are two unknowns: the parameters of interest and some nonparametric link functions or additive functions, of which we need to regard them as infinite-dimensional nuisance parameters. When we consider constructing confidence regions for the parameters of interest in these models, plug-in estimators are needed to replace these unknown infinite-dimensional nuisance parameters. This is a commonly used method in literature that causes the classical empirical likelihood to often not have Wilks' phenomenon. A commonly used method relies on the assistance of the Monte Carlo approximation. A very useful alternative is to correct the bias such that limiting distributions are tractable. We in recent years have been studying this problem and proposed several adjustment and bias correction methods for the aforementioned models.

As there are a great number of works about the empirical likelihood in literature, it is impossible to give a complete overview. Thus, we only select six nonparametric and semiparametric models in this overview. The materials are organized in the following way. Section 2 describes the results on the single-index model. The classical bias corrections in literature are used. Section 3 expatiates the results on the partially linear single-index model to describe the new bias correction proposed in Zhu and Xue (2006). Section 4 summarizes the results on the partial linear model with longitudinal data, and Section 5 presents the bias correction method for the varying coefficient model with longitudinal data. Section 6 discusses a nonlinear errors-in-variables model with validation data. Section 7 introduces some other works. Most materials come from our research articles.

2. SINGLE-INDEX MODEL

2.1 Estimated empirical likelihood

Suppose that Y is a scalar response variable and X is p-dimension explanatory variable. The single-index model is

written as

(1) $$Y = g(\beta^{\mathrm{T}} X) + \varepsilon,$$

where $g(\cdot)$ is an unknown univariate link function, β is an unknown vector in R^p, and ε is a random error with $E(\varepsilon|X) = 0$ almost surely. For identifiability purposes, we typically assume that $\|\beta\| = 1$ with its first nonzero element being positive, where $\|\cdot\|$ denotes the Euclidean norm.

Xue and Zhu (2006) considered the following approach to construct confidence region of β by empirical likelihood. Suppose that the recorded data $\{(X_i, Y_i), 1 \le i \le n\}$ are generated by the model (1), this is

(2) $$Y_i = g(\beta^{\mathrm{T}} X_i) + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are independent and identically distributed (i.i.d.) random errors with $E(\varepsilon_i|X_i) = 0$, $1 \le i \le n$, and $X_i = (X_{i1}, \ldots, X_{ip})^{\mathrm{T}} \in R^p$.

Note that $\|\beta\| = 1$ is actually a constraint and then β has only $p-1$ free components. This constraint will be used to construct a $p-1$ dimensional confidence region of β, and the confidence region of the remaining component can be determined by the others automatically. Specifically, let $\beta = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$, and $\beta^{(r)} = (\beta_1, \ldots, \beta_{r-1}, \beta_{r+1}, \ldots, \beta_p)^{\mathrm{T}}$ be a $p-1$ dimensional parameter vector after deleting the rth component β_r of β. Without loss of generality, we assume that the β_r is a positive component (otherwise, consider $\beta_r = -(1 - \|\beta^{(r)}\|^2)^{1/2}$). Then, we write

$$\beta \equiv \beta(\beta^{(r)}) = (\beta_1, \ldots, \beta_{r-1}, (1 - \|\beta^{(r)}\|^2)^{1/2}, \beta_{r+1}, \ldots, \beta_p)^{\mathrm{T}}.$$

The true parameter $\beta^{(r)}$ satisfies the constraint $\|\beta^{(r)}\| < 1$. Thus, β is infinitely differentiable in a neighborhood of the true parameter $\beta^{(r)}$, the Jacobian matrix is

$$J_{\beta^{(r)}} = \frac{\partial \beta}{\partial \beta^{(r)}} = (\gamma_1, \ldots, \gamma_p)^{\mathrm{T}},$$

where γ_s ($1 \le s \le p, s \ne r$) is a $p-1$ dimensional unit vector with sth component 1, and $\gamma_r = -(1 - \|\beta^{(r)}\|^2)^{-1/2} \beta^{(r)}$.

Denote $X_i^{(r)} = (X_{i1}, \ldots, X_{i,r-1}, X_{ir}, \ldots, X_{ip})^{\mathrm{T}}$. Since $\beta^{\mathrm{T}} X_i = \beta^{(r)\mathrm{T}} X_i^{(r)} + (1 - \|\beta^{(r)}\|^2)^{1/2} X_{ir}$, both $g(\beta^{\mathrm{T}} X_i)$ and $g'(\beta^{\mathrm{T}} X_i)$ are the function of $\beta^{(r)}$. A direct way to construct the empirical likelihood ratio by the following auxiliary random vector

$$Z_i(\beta^{(r)}) = [Y_i - g(\beta^{\mathrm{T}} X_i)] g'(\beta^{\mathrm{T}} X_i) J_{\beta^{(r)}}^{\mathrm{T}} X_i.$$

Note that $E[Z_i(\beta^{(r)})] = 0$ if β is the true parameter. Whereas, when $E[Z_i(\beta^{(r)})] = 0$, an estimating equation $\sum_{i=1}^{n} Z_i(\beta^{(r)}) = 0$ can be constructed. Define the profile empirical log-likelihood ratio function

$$l_n(\beta^{(r)}) = -2 \max \left\{ \sum_{i=1}^{n} \log(np_i) \Big| p_i \ge 0, \right.$$
$$\left. \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i Z_i(\beta^{(r)}) = 0 \right\}.$$

To obtain the plug-in estimation of the unknown $g(\beta^{\mathrm{T}} X_i)$ and $g'(\beta^{\mathrm{T}} X_i)$ in $l_n(\beta^{(r)})$, they used local linear smoother (Fan and Gijbels 1996). The estimators of $g(t)$ and $g'(t)$ are of the following formulas

$$\hat{g}(t; \beta) = \sum_{i=1}^{n} W_{ni}(t; \beta) Y_i$$

and

$$\hat{g}'(t; \beta) = \sum_{i=1}^{n} \widetilde{W}_{ni}(t; \beta) Y_i,$$

for some weight functions specified in Xue and Zhu (2006).

Let $\hat{Z}_i(\beta^{(r)})$ be an estimator of $Z_i(\beta^{(r)})$ with $g(\beta^{\mathrm{T}} X_i)$ and $g'(\beta^{\mathrm{T}} X_i)$ being replaced by $\hat{g}(\beta^{\mathrm{T}} X_i; \beta)$ and $\hat{g}'(\beta^{\mathrm{T}} X_i; \beta)$ respectively for $i = 1, \ldots, n$. Then an estimated empirical log-likelihood is defined as

$$\hat{l}(\beta^{(r)}) = -2 \max \left\{ \sum_{i=1}^{n} \log(np_i) \Big| p_i \ge 0, \sum_{i=1}^{n} p_i = 1, \right.$$
$$\left. \sum_{i=1}^{n} p_i \hat{Z}_i(\beta^{(r)}) = 0 \right\}.$$

By the Lagrange multiplier method, $\hat{l}(\beta^{(r)})$ can be represented as

(3) $$\hat{l}(\beta^{(r)}) = 2 \sum_{i=1}^{n} \log \left(1 + \lambda^{\mathrm{T}} \hat{Z}_i(\beta^{(r)}) \right),$$

where λ is determined

(4) $$\frac{1}{n} \sum_{i=1}^{n} \frac{\hat{Z}_i(\beta^{(r)})}{1 + \lambda^{\mathrm{T}} \hat{Z}_i(\beta^{(r)})} = 0.$$

Under certain regularity conditions,

$$\hat{l}(\beta^{(r)}) \xrightarrow{D} w_1 \chi_{1,1}^2 + \cdots + w_{p-1} \chi_{1,p-1}^2,$$

where $\xrightarrow{D}$ denotes convergence in distribution, the weights w_i, for $1 \le i \le p-1$, are eigenvalues of $D(\beta^{(r)}) = V_0^{-1}(\beta^{(r)}) V(\beta^{(r)})$, and $\chi_{1,i}^2 (1 \le i \le p-1)$ are independent χ_1^2 variables.

To apply this result to construct a confidence region (interval) of $\beta^{(r)}$, we need to estimate the unknown weights w_i consistently and approximate the conditional distribution of the weighted sum $\hat{s} = \hat{w}_1 \chi_{1,1}^2 + \cdots + \hat{w}_{p-1} \chi_{1,p-1}^2$

given the data $\{(X_i, Y_i), 1 \leq i \leq n\}$. It is a computationally intensive method. Two adjustments of the empirical likelihood ratio are suggested to make the asymptotic distribution tractable.

2.2 Two adjusted empirical likelihood

The first adjustment is based on Rao and Scott (1981)'s method. Let $\rho(\beta^{(r)}) = (p-1)/\mathrm{tr}\{D(\beta^{(r)})\}$ with $\mathrm{tr}(\cdot)$ being the trace operator. Then following Rao and Scott (1981), the distribution of $\rho(\beta^{(r)})\sum_{i=1}^{p-1} w_i \chi_{1,i}^2$ can be approximated by a standard chi-square distribution with $p-1$ degrees of freedom, χ_{p-1}^2. This implies that the asymptotic distribution of the Rao-Scott adjusted empirical log-likelihood ratio $\hat{\rho}(\hat{\beta}^{(r)})\hat{l}(\beta^{(r)})$ can be approximated by χ_{p-1}^2. This can be achieved by using the above result and the consistency of $\hat{V}(\hat{\beta}^{(r)})$ and $\hat{V}_0(\hat{\beta}^{(r)})$, where $\hat{\rho}(\hat{\beta}^{(r)}) = (p-1)/\mathrm{tr}\{\hat{D}(\hat{\beta}^{(r)})\}$. The Rao-Scott adjusted empirical log-likelihood can be improved by replacing $\hat{\beta}^{(r)}$ in $\hat{\rho}(\hat{\beta}^{(r)})$ by $\beta^{(r)}$. An improved Rao-Scott adjusted empirical log-likelihood is defined as

$$\tilde{l}(\beta^{(r)}) = \hat{\rho}(\beta^{(r)})\hat{l}(\beta^{(r)}).$$

However, the accuracy of this approximation depends on the values of the w_i's.

The second adjustment is to make the empirical log-likelihood ratio asymptotically follow a standard chi-square distribution with $p-1$ degrees of freedom. Note that

$$\hat{\rho}(\hat{\beta}^{(r)}) = \frac{\mathrm{tr}\{\hat{V}^{-1}(\hat{\beta}^{(r)})\hat{V}(\hat{\beta}^{(r)})\}}{\mathrm{tr}\{\hat{V}_0^{-1}(\hat{\beta}^{(r)})\hat{V}(\hat{\beta}^{(r)})\}}.$$

By examining the asymptotic expansion of $\hat{l}(\beta^{(r)})$, we replace $\hat{V}(\hat{\beta}^{(r)})$ in $\hat{\rho}(\hat{\beta}^{(r)})$ by $\hat{B}(\hat{\beta}^{(r)}) = \{\sum_{i=1}^n \hat{Z}_i(\hat{\beta}^{(r)})\}\{\sum_{i=1}^n \hat{Z}_i(\hat{\beta}^{(r)})\}^{\mathrm{T}}$ and obtain a different adjustment factor

$$\hat{r}(\hat{\beta}^{(r)}) = \frac{\mathrm{tr}\{\hat{V}^{-1}(\hat{\beta}^{(r)})\hat{B}(\hat{\beta}^{(r)})\}}{\mathrm{tr}\{\hat{V}_0^{-1}(\hat{\beta}^{(r)})\hat{B}(\hat{\beta}^{(r)})\}}.$$

It can be shown that $\hat{r}(\hat{\beta}^{(r)})\hat{l}(\beta^{(r)})$ is an asymptotically χ_{p-1}^2 variable. To increase the accuracy of approximation, we replace $\hat{\beta}^{(r)}$ in $\hat{r}(\hat{\beta}^{(r)})$ by $\beta^{(r)}$, and define an adjusted empirical log-likelihood by

$$(5) \qquad \hat{l}_{\mathrm{ad}}(\beta^{(r)}) = \hat{r}(\beta^{(r)})\hat{l}(\beta^{(r)}).$$

In the following section, partially linear single-index model is investigated and a more sophisticated technique is suggested.

Under certain regularity conditions,

$$\hat{l}_{\mathrm{ad}}(\beta^{(r)}) \xrightarrow{D} \chi_{p-1}^2.$$

3. PARTIALLY LINEAR SINGLE-INDEX MODEL

3.1 Methodology

A partially linear single-index model for the dependence of a scalar response variable Y and two covariables X and Z has the form

$$(6) \qquad Y = g(\beta^{\mathrm{T}} X) + \theta^{\mathrm{T}} Z + \varepsilon,$$

where $(X, Z) \in R^p \times R^q$, $g(\cdot)$ is an unknown univariate link function, (β, θ) is an unknown vector in $R^p \times R^q$ with $\|\beta\| = 1$ and first nonzero component positive. Assume that the sample $\{(X_i, Z_i, Y_i), 1 \leq i \leq n\}$ are independent and identically distributed (i.i.d.) and

$$(7) \qquad Y_i = g(\beta^{\mathrm{T}} X_i) + \theta^{\mathrm{T}} Z_i + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $\{\varepsilon_i, 1 \leq i \leq n\}$ are i.i.d. random errors with $E(\varepsilon_i | X_i, Z_i) = 0$, $1 \leq i \leq n$; $X_i = (X_{i1}, \ldots, X_{ip})^{\mathrm{T}} \in R^p$ and $Z_i = (Z_{i1}, \ldots, Z_{iq})^{\mathrm{T}} \in R^q$.

Introduce an auxiliary random vector $\xi_i(\beta, \theta) = [Y_i - g(\beta^{\mathrm{T}} X_i) - \theta^{\mathrm{T}} Z_i]\Lambda_i$, where $\Lambda_i = (g'(\beta^{\mathrm{T}} X_i)X_i^{\mathrm{T}} J_{\beta^{(r)}}, Z_i^{\mathrm{T}})^{\mathrm{T}}$ where g' is the derivative of g with respect to $\beta^{(r)}$. Note that $E[\xi_i(\beta, \theta)] = 0$ if (β, θ) is the true parameter. Using this, an empirical log-likelihood ratio function is defined as

$$l_n(\beta, \theta) = -2 \max \left\{ \sum_{i=1}^n \log(np_i) \Big| p_i \geq 0, \sum_{i=1}^n p_i = 1, \right.$$
$$\left. \sum_{i=1}^n p_i \xi_i(\beta, \theta) = 0 \right\}.$$

Similar to that in Section 2, the two unknown functions $g(\beta^{\mathrm{T}} X_i)$ and $g'(\beta^{\mathrm{T}} X_i)$ in $l_n(\beta, \theta)$ can be estimated by local linear smoothing method as

$$(8) \qquad \hat{g}(t; \beta, \theta) = \sum_{i=1}^n W_{ni}(t; \beta)(Y_i - \theta^{\mathrm{T}} Z_i)$$

and

$$(9) \qquad \hat{g}'(t; \beta, \theta) = \sum_{i=1}^n \widetilde{W}_{ni}(t; \beta)(Y_i - \theta^{\mathrm{T}} Z_i),$$

When replacing $\xi_i(\beta, \theta)$ in $l_n(\beta, \theta)$ by $\hat{\xi}_i(\beta, \theta)$ in terms of using, for $i = 1, \ldots, n$, $\hat{g}(\beta^{\mathrm{T}} X_i; \beta, \theta)$ and $\hat{g}'(\beta^{\mathrm{T}} X_i; \beta, \theta)$ in place of $g(\beta^{\mathrm{T}} X_i)$ and $g'(\beta^{\mathrm{T}} X_i)$, an estimated empirical log-likelihood, say $\tilde{l}_n(\beta, \theta)$, can be defined. Such plug-in estimators make that $\tilde{l}_n(\beta, \theta)$ is asymptotically a weighted sum of independent standard chi-squared variables, each with one degree of freedom and an unknown weight. Although the adjustments in Section 2 can be used, we will propose a more sophisticated technique here.

Note that $\hat{\xi}_i$ are weighted residuals. To eliminate a remainder which has a slower convergence rate than $\sqrt{n}$ when

an optimal bandwidth is used, we consider a conditional centering to correct bias. The bias-corrected empirical log-likelihood (BCEL) ratio is defined as

$$(10) \quad \hat{l}(\beta, \theta) = -2 \max \left\{ \sum_{i=1}^{n} \log(np_i) \middle| p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \right.$$
$$\left. \sum_{i=1}^{n} p_i \hat{\eta}_i(\beta, \theta) = 0 \right\},$$

where

$$(11) \quad \hat{\eta}_i(\beta, \theta) = [Y_i - \hat{g}(\beta^{\mathrm{T}} X_i; \beta, \theta) - \theta^{\mathrm{T}} Z_i] \big[\hat{g}'(\beta^{\mathrm{T}} X_i; \beta, \theta)$$
$$\times \big(X_i - \hat{\mu}_1(\beta^{\mathrm{T}} X_i; \beta) \big)^{\mathrm{T}} J_{\beta^{(r)}},$$
$$(Z_i - \hat{\mu}_2(\beta^{\mathrm{T}} X_i; \beta))^{\mathrm{T}} \big]^{\mathrm{T}},$$

$\hat{\mu}_1(t; \beta)$ and $\hat{\mu}_2(t; \beta)$ are the estimators of $\mu_1(t) = E[X|\beta^{\mathrm{T}} X = t]$ and $\mu_2(t) = E[Z|\beta^{\mathrm{T}} X = t]$ respectively in the following.

$$\hat{\mu}_1(t; \beta) = \sum_{i=1}^{n} W_{ni}(t; \beta) X_i$$

and

$$\hat{\mu}_2(t; \beta) = \sum_{i=1}^{n} W_{ni}(t; \beta) Z_i.$$

Under certain regularity conditions, $\hat{l}^*(\beta^{(r)}, \theta) \xrightarrow{D} \chi^2_{p+q-1}$, where $\hat{l}^*(\beta^{(r)}, \theta)$ is the same as $\hat{l}(\beta, \theta)$ replacing the Jacobian matrix $J_{\beta^{(r)}}$ by $J^*_{\beta^{(r)}} = (\gamma_1^*, \ldots, \gamma_p^*)^{\mathrm{T}}$, $\gamma_s^*(s \neq r)$ is a $(p-1)$-dimensional unit vector with sth component 1 and $\gamma_r^* = -(1 - \|\beta^{(r)}\|^2)^{-1/2} \beta^{(r)}$.

This result shows that once we obtain the confidence regions of $(\beta^{(r)}, \theta)$, the confidence region of (β, θ) can be immediately obtained through the relation $\beta_r = (1 - \|\beta^{(r)}\|^2)^{1/2}$.

We now consider the pure single-index model studied in Section 2. Rewrite (11) as

$$\hat{\eta}_i(\beta) = [Y_i - \hat{g}(\beta^{\mathrm{T}} X_i; \beta)] \hat{g}'(\beta^{\mathrm{T}} X_i; \beta) J^{\mathrm{T}}_{\beta^{(r)}} \big(X_i - \hat{\mu}_1(\beta^{\mathrm{T}} X_i) \big).$$

Let $\hat{l}(\beta)$ denote $\hat{l}(\beta, \theta)$ with $\hat{\eta}_i(\beta, \theta)$ replaced by $\hat{\eta}_i(\beta)$ in (10). We have

$$\hat{l}^*(\beta^{(r)}) \xrightarrow{D} \chi^2_{p-1},$$

where $\hat{l}^*(\beta^{(r)})$ is $\hat{l}(\beta)$ with $J_{\beta^{(r)}}$ replaced by $J^*_{\beta^{(r)}}$.

When $\beta = 1$, model (6) is reduced to a partially linear model. Then estimation and statistical inference for β makes no sense. In this case, we introduce a random vector $\hat{\eta}_i(\theta) = [Y_i - \hat{g}(X_i) - \theta^{\mathrm{T}} Z_i] [Z_i - \hat{\mu}_2(X_i)]$. Let $\hat{l}(\theta)$ denote $\hat{l}(\beta, \theta)$ with $\hat{\eta}_i(\beta, \theta)$ being replaced by $\hat{\eta}_i(\theta)$. All the above results can be directly used to construct confidence regions of the parameters of interest.

We have that under some regularity conditions, $\hat{l}(\theta) \xrightarrow{D} \chi^2_q$.

4. A PARTIAL LINEAR MODEL WITH CENSORED AND LONGITUDINAL DATA

For longitudinal data, Zeger and Diggle (1994) considered the partial linear model of the form

$$(12) \qquad Y(t) = X^{\mathrm{T}}(t)\beta + \theta(t) + \varepsilon(t),$$

where $Y(t)$ is the response variable and $X(t)$ is the $p \times 1$ covariate vector at time t, β is a $p \times 1$ vector of unknown regression coefficients, $\theta(t)$ is an unspecified baseline function of t, and $\varepsilon(t)$ is a zero-mean stochastic process. Here t ranges over a nondegenerate compact interval $\mathcal{T}$, without loss of generality assumed to be the unit interval $[0, 1]$. Furthermore, the observation times are censored at the end of follow-up. As such, this model is with both censored and longitudinal data.

4.1 Naive empirical likelihood

Suppose that we have a random sample of n subjects. For the ith subject, the response variable $Y_i(t)$ and the covariate vectors $X_i(t)$ are collected at time points $t = t_{i1}, \ldots, t_{in_i}$, $i = 1, \ldots, n$, where n_i is the total number of observations on the ith subject. Thus,

$$(13) \qquad Y_i(t_{ij}) = X_i^{\mathrm{T}}(t_{ij})\beta + \theta(t_{ij}) + \varepsilon_i(t_{ij})$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$. Assume that both $(X_i(t_{ij}), Y_i(t_{ij}))$ and $\varepsilon_i(t_{ij})$ from different subjects are independent and $E\{\varepsilon_i(t_{ij})|X_i(t_{ij})\} = 0$. We also assume in our asymptotic study that n_i is bounded but the number of subjects n goes to infinity.

To deal with censoring of the observation times, a counting process approach is applied, which is similar to Lin and Ying (2001) and Fan and Li (2004). The time points at which the observations on the ith subject are made are characterized by the counting process $N_i(t) \equiv \sum_{j=1}^{n_i} I(t_{ij} \leq t)$, where $I(\cdot)$ is the indicator function. Let $N_i(t) = N_i^*\{\min(t, C_i)\}$, where $N_i^*(t)$ is a counting process in continuous time $t \in [0, 1]$ and C_i is the follow-up or censoring time. Following Lin and Ying (2001), the counting process $N_i(t)$ is a random sample from a certain population, and both $X_i(t)$ and $Y_i(t)$ were observed at the jump points $N_i(t)$. The censoring time C_i is allowed to depend on the vector of covariates $X_i(t)$ in an arbitrary manner. In this section the censoring mechanism is assumed to be noninformative in the sense that

$$E\{Y_i(t)|X_i(t), C_i \geq t\} = E\{Y_i(t)|X_i(t)\}.$$

Lin and Ying (2001) considered two situations, depending on whether or not the observation times are independent of the covariates $X(t)$. When the observation times depend on the covariates, following Lin and Ying (2001), assume that

$$E\{dN_i^*(t)|X_i(t), Y_i(t), C_i \geq t\} = \exp\{\gamma^{\mathrm{T}} X_i(t)\} d\Lambda(t),$$
$$i = 1, \ldots, n,$$

where γ is a vector of unknown parameters and $\Lambda(\cdot)$ is an arbitrary nondecreasing function. When $\gamma = 0$, the observation times are independent of the covariates.

Note that

$$E\{Y_i(t_{ij})\} = E\{X_i^{\mathrm{T}}(t_{ij})\}\beta + \theta(t_{ij}).$$

This together with (13) yields

(14)
$$Y_i(t_{ij}) - E\{Y_i(t_{ij})\} = [X_i(t_{ij}) - E\{X_i(t_{ij})\}]^{\mathrm{T}}\beta + \varepsilon_i(t_{ij}).$$

Consider constructing the empirical likelihood ratio for β.

Let $w(\cdot)$ be a bounded nonnegative weight function with a compact support, $[a, b]$, where $0 < a < b < 1$. We assume that $a \geq h$ and $1 - b \geq h$, where h is a bandwidth defined in (15) below. Then the auxiliary random vectors are introduced as

$$Z_i(\beta) = \int_0^1 w(t)\breve{X}_i(t)\{\breve{Y}_i(t) - \breve{X}_i^{\mathrm{T}}(t)\beta\}dN_i(t), \quad i = 1, \ldots, n,$$

where $\breve{X}_i(t) = X_i(t) - E\{X_i(t)\}$ and $\breve{Y}_i(t) = Y_i(t) - E\{Y_i(t)\}$. Note that $E\{Z_i(\beta)\} = 0$ when β is the true parameter. When two estimators of $E\{X_i(t)\}$ and $E\{Y_i(t)\}$ are used as plug-in estimators in $E\{Z_i(\beta)\}$, an empirical likelihood ratio can then constructed as was done in the previous sections.

Consider two mean function models,

$$X(t) = m_X(t) + \epsilon(t), \qquad Y(t) = m_Y(t) + \delta(t),$$

where $m_X(t) = E\{X(t)\}$, $m_Y(t) = E\{Y(t)\}$, and $\epsilon(t)$ and $\delta(t)$ are the zero-mean error terms. This is a nonparametric regression problem, and $m_X(t)$ and $m_Y(t)$ can be estimated by kernel smoother as follows and the estimators are respectively defined as

$$\hat{m}_X(t) = \sum_{i=1}^n \sum_{j=1}^{n_i} W_{ij}(t) X_i(t_{ij}),$$

$$\hat{m}_Y(t) = \sum_{i=1}^n \sum_{j=1}^{n_i} W_{ij}(t) Y_i(t_{ij}),$$

where

(15) $$W_{ij}(t) = K_h(t_{ij} - t) \bigg/ \sum_{k=1}^n \sum_{l=1}^{n_k} K_h(t_{kl} - t),$$

h is a bandwidth, $K_h(\cdot) = K(\cdot/h)$ and K is a kernel function.

Therefore, an estimator $\hat{Z}_i(\beta)$ of $Z_i(\beta)$ can be obtained by substituting $m_X(t)$ and $m_Y(t)$ of $Z_i(\beta)$ with $\hat{m}_X(t)$ and $\hat{m}_Y(t)$; that is,

$$\hat{Z}_i(\beta) = \int_0^1 w(t)\tilde{X}_i(t)\left\{\tilde{Y}_i(t) - \tilde{X}_i^{\mathrm{T}}(t)\beta\right\}dN_i(t),$$
$$i = 1, \ldots, n,$$

where $\tilde{X}_i(t) = X_i(t) - \hat{m}_X(t)$ and $\tilde{Y}_i(t) = Y_i(t) - \hat{m}_Y(t)$. An estimated empirical likelihood ratio for β is defined as

$$\hat{L}(\beta) = \max\left\{\prod_{i=1}^n (np_i)\bigg| p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i\hat{Z}_i(\beta) = 0\right\}.$$

For $\theta(t)$, a profile empirical log-likelihood ratio is defined by

$$\tilde{l}\{\theta(t)\} = -2\max\left\{\sum_{i=1}^n \log(np_i)\bigg| p_i \geq 0, \sum_{i=1}^n p_i = 1, \right.$$
$$\left. \sum_{i=1}^n p_i\hat{\eta}_i\{\theta(t)\} = 0\right\},$$

where

$$\hat{\eta}_i\{\theta(t)\} = \sum_{j=1}^{n_i}\left\{Y_i(t_{ij}) - X_i^{\mathrm{T}}(t_{ij})\hat{\beta} - \theta(t)\right\}K_h(t_{ij} - t).$$

In other words, although the plug-in estimators of these nonparametric functions $E\{X(t)\}$ and $E\{Y(t)\}$ have slower rates of convergence than the estimator of β, this does not affect the first-order behaviour of the empirical likelihood in the proposed empirical likelihood ratio. The main reason for this is the use of the centered $X(t)$ and $Y(t)$ which is a kind of bias correction for the empirical likelihood function $\hat{L}(\beta)$. For the classical partly linear model, Wang and Jing (2003) derived a similar result.

However, for the baseline function $\theta(\cdot)$, the situation is not so good, and then a particular bias correction designed for $\tilde{l}\{\theta(t)\}$ is necessary.

4.2 Bias-corrected empirical likelihood

4.2.1 Mean-corrected empirical likelihood

Xue and Zhu (2007b) derived that an asymptotic presentation of the profile empirical loglikelihood ratio $\tilde{l}\{\theta(t_0)\}$ as

(16) $$\tilde{l}\{\theta(t_0)\} = \left(\sum_{i=1}^n \hat{\eta}_i\{\theta(t_0)\}\right)^2 \bigg/ \sum_{i=1}^n \hat{\eta}_i^2\{\theta(t_0)\} + o_P(1)$$

which converges in distribution to χ_1^2 under some regularity conditions and the bandwidth satisfying that $nh^2/\log n \to \infty$ and $nh^5 \to 0$. That is, we have to undersmooth the estimator of $\theta(t_0)$. When the optimal bandwidth is used and the empirical likelihood confidence intervals for $\theta(t)$ is constructed, we must make a correction for $\tilde{l}\{\theta(t)\}$. Let

$$\hat{\xi}\{\theta(t)\} = \left(2\hat{b}(t) \cdot \frac{1}{\sqrt{nh}}\sum_{i=1}^n \hat{\eta}_i\{\theta(t)\} - \hat{b}^2(t)\right)$$
$$\times \left(\frac{1}{nh}\sum_{i=1}^n \hat{\eta}_i^2\{\theta(t)\}\right)^{-1}.$$

A mean-corrected empirical loglikelihood ratio is defined as

(17) $$\tilde{l}^*\{\theta(t)\} = \tilde{l}\{\theta(t)\} - \hat{\xi}\{\theta(t)\}.$$

Then $\tilde{l}^*\{\theta(t_0)\} \to \chi_1^2$.

4.2.2 Residual-adjusted empirical likelihood

A more sophisticated correction can be carried out through the asymptotic expansion of the empirical likelihood ratio. We can adjust the weighted residuals $\hat{\eta}_i(\theta(t))$ and then obtain an adjusted empirical likelihood ratio. Introduce the auxiliary random variables

$$\hat{\eta}_i^*\{\theta(t)\} = \sum_{j=1}^{n_i} \big[Y_i(t_{ij}) - X_i^T(t_{ij})\hat{\beta} - \theta(t)$$
$$- \{\hat{\theta}(t_{ij}) - \hat{\theta}(t)\}\big]K_h(t_{ij} - t).$$

A residual-adjusted empirical loglikelihood ratio can be defined as

$$\hat{l}^*\{\theta(t)\} = -2\max\left\{\sum_{i=1}^n \log(np_i)\Big| p_i \geq 0, \sum_{i=1}^n p_i = 1,\right.$$
$$\left.\sum_{i=1}^n p_i\hat{\eta}_i^*\{\theta(t)\} = 0\right\}.$$

This correction is better in form than the mean-corrected empirical likelihood ratio because it guarantees the nonnegativity of the ratio. In addition, by the residual-adjustment in $\hat{\eta}_i^*\{\theta(t)\}$, we not only correct the bias, but also avoid undersmoothing the baseline function $\theta(t)$. $\hat{l}^*\{\theta(t_0)\}$ is asymptotically chi-squared χ_1^2.

5. VARYING COEFFICIENT MODEL

5.1 Classical empirical likelihood

For longitudinal data, another useful model is the varying coefficient model:

(18) $$Y_{ij} = X_i^T(t_{ij})\beta(t_{ij}) + \varepsilon_i(t_{ij}),$$

where, for all $t_{ij} \in R$, $X_i(t_{ij}) = (1, X_{i1}(t_{ij}), \ldots, X_{ik}(t_{ij}))^T$ are real-valued covariates at time t_{ij}, $\beta(\cdot) = (\beta_0(\cdot), \ldots, \beta_k(\cdot))^T$, $\beta_r(\cdot)$ are smooth functions for all $r = 0, \ldots, k$, $\varepsilon_i(\cdot)$ are mean 0 stochastic processes, and $\varepsilon_i(\cdot)$ and $X_i(\cdot)$ are independent.

For a given time $t \in R$, we can define a least squares estimator of $\beta(t)$ by a minimizer of the sample version of the conditional mean squared error $E\{[Y(t) - X^T(t)\beta(t)]^2|t\}$, or the solution to $E\{[Y(t) - X^T(t)\beta(t)]X(t)|t\} = 0$. This is equivalent to the minimizer of $E\{[Y(t) - X^T(t)\beta(t)]^2|t\}f(t)$, or the solution of $E\{[Y(t) - X^T(t)\beta(t)]X(t)|t\}f(t) = 0$. As nonparametric conditional expectation given t is involved, a local smoothing method is needed to obtain the sample ver-

sion. To define the empirical likelihood estimator, we employ the constraint $E\{[Y(t) - X^T(t)\beta(t)]X(t)|t\}f(t) = 0$. With this, the auxiliary random vectors are introduced as follows:

(19) $$Z_i(\beta(t)) = \sum_{j=1}^{n_i} \big[Y_{ij} - X_i^T(t_{ij})\beta(t)\big]X_i(t_{ij})K_h(t - t_{ij}),$$

where h is a bandwidth, $K_h(\cdot) = K(\cdot/h)$, and K is a kernel function.

Note that the $\{Z_i(\beta(t)); 1 \leq i \leq n\}$ are independent, and that $E\big[Z_i(\beta(t))\big] = 0$. A classical empirical log-likelihood ratio function for $\beta(t)$ can be defined by

$$\mathcal{R}(\beta(t)) = -2\max\left\{\sum_{i=1}^n \log(np_i)\Big| p_i \geq 0, \sum_{i=1}^n p_i = 1,\right.$$
$$\left.\sum_{i=1}^n p_iZ_i(\beta(t)) = 0\right\},$$

where $p_i = p_i(t), i = 1, \ldots, n$. It has been showed that under some regularity conditions $\mathcal{R}(\beta(t))$ converges in distribution to, when $Nh \to \infty$ and $Nh^5 \to 0$, χ_{k+1}^2. This result means that even nonparametric smoother is involved in defining $Z_i(\beta(t))$ in (19), the Wilks' phenomenon still holds when undersmoothing is used. But undersmoothing makes a difficulty to appropriately select the bandwidth in the kernel estimation.

5.2 Residual-adjusted empirical likelihood

A more sophisticated correction can be carried out through the asymptotic expansion of the empirical likelihood ratio without undersmoothing.

Introduce the following auxiliary random vectors

$$\hat{Z}_i(\beta(t)) = \sum_{j=1}^{n_i} \big\{Y_{ij} - X_i^T(t_{ij})\beta(t)$$
$$- X_i^T(t_{ij})[\hat{\beta}(t_{ij}) - \hat{\beta}(t)]\big\}X_i(t_{ij})K_h(t - t_{ij}).$$

Clearly, $\hat{Z}_i$ are adjustments of Z_i of (19). An adjusted empirical log-likelihood ratio function for $\beta(t)$ can be defined as

$$\hat{\mathcal{R}}(\beta(t)) = -2\max\left\{\sum_{i=1}^n \log(np_i)\Big| p_i \geq 0, \sum_{i=1}^n p_i = 1,\right.$$
$$\left.\sum_{i=1}^n p_i\hat{Z}_i(\beta(t)) = 0\right\}.$$

It can be shown that $\mathcal{R}(\beta(t))$ is asymptotically chi-squared under mild conditions with the optimal bandwidth $h = O(N^{-1/5})$. That is, $\hat{\mathcal{R}}(\beta(t_0))$ can also be asymptotically χ_{k+1}^2.

6. ERRORS-IN-VARIABLES MODEL

6.1 Estimated empirical likelihood

Consider the parametric regression model

(20) $$Y = g(X, \beta) + \varepsilon$$

Here X is a d-variate explanatory variable, Y is a scalar response and the noise variable ε satisfies $E(\varepsilon|X) = 0$. In particular, heteroscedasticity is allowed and no independence between X and ε is required. Also Y and/or some of the X-coordinates may be discrete so existence of joint or marginal densities may and will not be imposed. The vector $\beta = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ of regression parameters is the unknown target, since it fully specifies the regression function g.

However, in many cases, the explanatory variable X is available only subject to noise. The observed variable is its surrogate $\tilde{X}$. In other words, what one observes are independent replicates $(\tilde{X}_i, Y_i)$, $1 \le i \le N$, of $(\tilde{X}, Y)$ rather than (X, Y), where the relationship between $\tilde{X}_i$ and X_i may be specified or not. If not, the missing information for inference has to be taken from a sample $(\tilde{X}_i, X_i)$, $N+1 \le i \le N+n$, of so-called validation data being independent of the primary sample.

Put

$$G(\tilde{X}, \beta) = E[g(X, \beta)|\tilde{X}],$$

where $\tilde{X}$ is a surrogate variable. The dimension of $\tilde{X}$, say d_1, may or may not differ from d. Throughout

(21) $$E[\varepsilon|\tilde{X}] = 0$$

is assumed. Conclude that

(22) $$Y = G(\tilde{X}, \beta) + \eta$$

with $E[\eta|\tilde{X}] = 0$. Recall that $(\tilde{X}_i, Y_i)$, $1 \le i \le N$, is our primary sample. When $G(\cdot)$ is unknown, estimating β is not possible. To circumvent this problem, a sample $(\tilde{X}_i, X_i)$, $N+1 \le i \le N+n$, will be needed to estimate G. Let, for any $\tilde{\beta}$,

$$G(\tilde{X}, \tilde{\beta}) = E[g(X, \tilde{\beta})|\tilde{X}]$$

and put

$$Z_i(\tilde{\beta}) = G^{(1)}(\tilde{X}_i, \tilde{\beta}) \left\{ Y_i - G(\tilde{X}_i, \tilde{\beta}) \right\}, \quad 1 \le i \le N,$$

where

(23) $$G^{(1)}(\tilde{X}_i, \tilde{\beta}) = \frac{\partial}{\partial \beta} G(\tilde{X}_i, \beta)|_{\beta = \tilde{\beta}} = E\left[g^{(1)}(X_i, \beta)|\tilde{X}_i \right]\Big|_{\beta = \tilde{\beta}}$$

and

(24) $$g^{(1)}(X_i, \tilde{\beta}) = \left(\frac{\partial}{\partial \beta_1} g(X_i, \beta), \ldots, \frac{\partial}{\partial \beta_p} g(X_i, \beta) \right)^{\mathrm{T}}\Big|_{\beta = \tilde{\beta}}.$$

Clearly, when $\tilde{\beta} = \beta$ is the true parameter, $E[Z_i(\beta)] = 0$.

When G is unknown, we have to consider the estimated empirical likelihood ratio

$$\hat{l}_N(\beta) = -2 \max \sum_{i=1}^{N} \ln(N p_i),$$

where now the maximum has taken over the set of $p = (p_1, \ldots, p_N)$ satisfying $\sum_{i=1}^{n} p_i \hat{Z}_i(\beta) = 0$, $p_i \ge 0$ and $\sum_{i=1}^{N} p_i = 1$. Here

$$\hat{Z}_i(\beta) = \hat{G}^{(1)}(\tilde{X}_i, \beta)\{Y_i - \hat{G}(\tilde{X}_i, \beta)\}.$$

The functions $\hat{G}$ and $\hat{G}^{(1)}$ are nonparametric plug-in estimators of G and $G^{(1)}$, respectively, obtained from the validation data. The following multivariate Nadaraya-Watson estimators are used here

$$\hat{G}(\tilde{x}, \beta) = \frac{\sum_{k=N+1}^{N+n} g(X_k, \beta) K\left(\frac{\|\tilde{X}_k - \tilde{x}\|}{h} \right)}{\max\left(1, \sum_{k=N+1}^{N+n} K\left(\frac{\|\tilde{X}_k - \tilde{x}\|}{h} \right) \right)},$$

$$\hat{G}^{(1)}(\tilde{x}, \beta) = \frac{\partial}{\partial \beta} \hat{G}(\tilde{x}, \beta)$$

$$= \frac{\sum_{k=N+1}^{N+n} g^{(1)}(X_k, \beta) K\left(\frac{\|\tilde{X}_k - \tilde{x}\|}{h} \right)}{\max\left(1, \sum_{k=N+1}^{N+n} K\left(\frac{\|\tilde{X}_k - \tilde{x}\|}{h} \right) \right)}.$$

Here K is a nonnegative kernel function defined on the real line and $h = h_n$ is a positive bandwidth tending to zero as $n \to \infty$.

As we know, $\hat{G}$ and $\hat{G}^{(1)}$ are no longer $\sqrt{n}$-consistent, which will create a bias for the estimated empirical likelihood ratio. Actually, since smoothing only incorporates

To circumvent this, for large parts of the paper, we first study the likelihood ratio, for a given bandwidth $h = h_n > 0$, at β_h, where β_n minimizes the L^2-distance between Y and the smoothed $G(\tilde{x}, \tilde{\beta})$, namely

$$\bar{G}_n(\tilde{x}, \tilde{\beta}) = \frac{\int G(y, \tilde{\beta}) K\left(\frac{\|y - \tilde{x}\|}{h} \right) \tilde{\mu}(dy)}{\int K\left(\frac{\|y - \tilde{x}\|}{h} \right) \tilde{\mu}(dy)}.$$

Here, $\tilde{\mu}$ is the distribution of $\tilde{X}$. In other words,

$$\beta_h = \arg \min_{\tilde{\beta}} E[Y - \bar{G}_n(\tilde{X}, \tilde{\beta})]^2.$$

Since $\bar{G}_n(\tilde{X}, \tilde{\beta}) \to G(\tilde{X}, \tilde{\beta})$ as $h \to 0$ for each $\tilde{\beta}$ we may expect that $\beta_h \to \beta$. A large part of our work is devoted to the likelihood ratio $\hat{l}_N$ evaluated at β_h. For statistical inference about β it remains to study $\beta - \beta_h$, which is purely analytic and nonstochastic. As it will turn out, $\hat{l}_N(\beta_h)$ will not be asymptotically chi-squared anymore. Rather it will be a sum of weighted independent χ_1^2-variables. The asymptotic

distribution will also depend on the ratio n/N. In particular, we have to distinguish the two cases $\gamma < \infty$ and $\gamma = \infty$, where

$$\gamma = \lim(N/n).$$

For $l_N(\beta)$, a similar representation holds true, with $Z_i(\beta)$ instead of $\hat{Z}_i(\beta_h)$. From this we see that $l_N(\beta)$ is asymptotically chi-squared. For $\hat{l}_N(\beta_h)$, however, the limit will be different. It turns out that a crucial role will be played by the size of γ. the asymptotic behaviours with the case $\gamma < \infty$ and $\gamma = \infty$ are investigated as follows.

1. Under certain regularity conditions and $\gamma < \infty$ as $n, N \to \infty$,

$$\hat{l}_N(\beta_h) \xrightarrow{D} w_1 \chi^2_{1,1} + \cdots + w_p \chi^2_{1,p},$$

where $\chi^2_{1,i}$, $1 \le i \le p$, are independent χ^2_1-variables and the weights w_i, $1 \le i \le p$, are the eigenvalues of the matrix $D(\beta) = (\Sigma^{\mathrm{T}}(\beta))^{1/2} \Sigma_0(\beta)^{-1} (\Sigma(\beta))^{1/2}$. Here

$$\Sigma(\beta) = \Sigma_0(\beta) + \gamma E\big[G^{(1)}(\tilde{X}, \beta) G^{(1)}(\tilde{X}, \beta)^{\mathrm{T}} \big(g(X, \beta) - G(\tilde{X}, \beta)\big)^2\big].$$

2. When $n, N \to \infty$ in such a way that $\gamma = \infty$, then

$$\frac{n}{N} \hat{l}_N(\beta_h) \xrightarrow{D} w_1^* \chi^2_{1,1} + \cdots + w_p^* \chi^2_{1,p},$$

where the weights w_i^*, $1 \le i \le p$, are the eigenvalues of the matrix

$$D^*(\beta) = (\Sigma_1^{\mathrm{T}}(\beta))^{1/2} \Sigma_0(\beta)^{-1} (\Sigma_1(\beta))^{1/2}$$

and

$$\Sigma_1(\beta) = E\big[G^{(1)}(\tilde{X}, \beta) G^{(1)}(\tilde{X}, \beta)^{\mathrm{T}} (g(X, \beta) - G(\tilde{X}, \beta))^2\big].$$

6.2 Adjusted empirical likelihood

Apart from the above direct way to approximate the asymptotic distributions, Rao and Scott's (1981) approximation can be an alternative as used in Section 2.

However, the accuracy of this approximation still depends on the values of the w_i's. Another adjusted empirical log-likelihood (AEL) was then suggested, see Stute, Xue and Zhu (2007) for details. AEL is asymptotically chi-squared with p degrees of freedom. The adjustment follows an idea of Rao and Scott (1981), and is similar to that proposed by Wang and Rao (2002a). However, with a different ratio of n and N, we have to select different adjustment factors. We first consider the case $\gamma < \infty$. Note that $\hat{\rho}(\beta)$ can be written as

$$\hat{\rho}(\beta) = \frac{tr\{\hat{\Sigma}(\beta)^{-1} \hat{\Sigma}(\beta)\}}{tr\{\hat{\Sigma}_0(\beta)^{-1} \hat{\Sigma}(\beta)\}}.$$

By examining the asymptotic expansion of $\hat{l}(\beta)$, an adjustment factor can be defined as

$$\hat{r}(\beta) = \frac{tr\{\hat{\Sigma}(\beta)^{-1} \hat{B}(\beta)\}}{tr\{\hat{\Sigma}_0(\beta)^{-1} \hat{B}(\beta)\}},$$

where $\hat{B}(\beta) = \big\{\sum_{i=1}^N \hat{Z}_i(\beta)\big\} \big\{\sum_{i=1}^N \hat{Z}_i(\beta)\big\}^{\mathrm{T}}$. The adjusted empirical log-likelihood is defined as

$$(25) \qquad \hat{l}_{ad}(\beta) = \hat{r}(\beta) \hat{l}(\beta).$$

We can show that asymptotically $\hat{r}(\beta_h) \hat{l}(\beta_h)$ is a χ^2_p-variable. From the following two results we can see that with such an adjustment, we can get rid of the unknown weights, and, unlike the simple extension of Rao and Scott (1981), we do not use the estimator $\hat{\beta}$ in the adjustment factor. The result is as follows.

1. Under certain regularity conditions, when $\lim_{N \to \infty}(N/n) = \gamma < \infty$, $\hat{l}_{ad}(\beta_h)$ is asymptotically a χ^2_p-variable:

$$\hat{l}_{ad}(\beta_h) \xrightarrow{D} \chi^2_p.$$

From the above idea, we can similarly select an adjustment factor when $\gamma = \infty$. Let $\hat{B}_1(\beta) = \big\{\sum_{i=N+1}^{N+n} \hat{Z}_i^*(\beta)\big\} \big\{\sum_{i=N+1}^{N+n} \hat{Z}_i^*(\beta)\big\}^{\mathrm{T}}$, where, for $N + 1 \le i \le N + n$,

$$\hat{Z}_i^*(\hat{\beta}) = \hat{G}^{(1)}(\tilde{X}_i, \hat{\beta}) \big\{g(X_i, \hat{\beta}) - \hat{G}(\tilde{X}_i, \hat{\beta})\big\}$$

and $\hat{\beta}$ is the least squares estimation of β, that is,

$$(26) \quad \hat{\beta} = \arg\min_\beta \left\{ N^{-1} \sum_{i=1}^N \left(Y_i - \hat{G}(\tilde{X}_i, \beta)\right)^2 \right\}.$$

The adjustment factor is defined as

$$\hat{r}^*(\beta) = \frac{tr\{\hat{\Sigma}_1(\beta)^{-1} \hat{B}_1(\beta)\}}{tr\{\hat{\Sigma}_0(\beta)^{-1} \hat{B}_1(\beta)\}}.$$

The adjusted empirical likelihood is defined as

$$(27) \qquad \hat{l}_{ad}^*(\beta) = \hat{r}^*(\beta)(n/N) \hat{l}(\beta).$$

2. Under some regularity conditions, when $\lim_{N \to \infty} n = \infty$, $\lim_{N \to \infty}(N/n) = \infty$. $\hat{l}_{ad}^*(\beta_h)$ is a

$$\hat{l}_{ad}^*(\beta_h) \xrightarrow{D} \chi^2_p.$$

7. OTHER WORKS

7.1 Empirical likelihood and estimating equations

Estimating equations provide an extremely flexible way to describe parameters and the corresponding statistics. For

a random variable $X \in R^d$, a parameter $\theta \in R^p$, and a vector-valued function $m(X,\theta) \in R^s$ suppose that

$$(28) \qquad E[m(X,\theta)] = 0.$$

The usual setting has $p = s$ and then under conditions on $m(X,\theta)$ and possibly on F, there is a unique solution θ. In this just determined case, the true value θ_0 may be estimated by solving

$$(29) \qquad \frac{1}{n}\sum_{i=1}^{n} m(X,\hat{\theta}) = 0$$

for $\hat{\theta}$. To write a vector mean by equation (28), we take $m(X,\theta) = X - \theta$, and then equation (29) gives $\hat{\theta} = \bar{X}$. For $P(X \in A)$ take $m(X,\theta) = I\{X \in A\} - \theta$. For a continuously distributed scalar X and $\theta \in R$, the function $m(X,\theta) = I\{X \leq \theta\} - \tau$ defines θ as the τ quantile of X. Owen (2001, Section 3.6) described tail probabilities and quantiles in more detail.

Equation (29) is known as an estimating equation, and $m(X,\theta)$ is called estimating function. Most maximum likelihood estimators are defined through estimating equations.

The underdetermined case $s < p$ can also be useful. Then (28) and (29) might each have an $s - p$ dimensional solution set of θ values. Some functions of θ may be precisely determined from the data, while the others will not.

In econometrics, considerable interest attaches to the overdetermined case with $s > p$. In problems with $s > p$ the fact that (28) holds is a special feature of F and constitutes important side information. Even when (28) holds for the true F_0, it will not ordinarily hold for the nonparametric maximum likelihood estimate $\hat{F}$, in which case (29) has no solution. The generalized method of moments looks for a value $\hat{\theta}$ that comes close to solving (28). An empirical likelihood approach to this problem was described in Owen (2001, Section 3.10). Also see Qin and Lawless (1994).

The empirical likelihood and estimating equations are well suited to each other. The empirical likelihood ratio function for θ is defined by

$$\mathcal{R}(\theta) = \max\left\{ \prod_{i=1}^{n} np_i \Big| p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \right.$$
$$\left. \sum_{i=1}^{n} p_i m(X_i,\theta) = 0 \right\}.$$

Owen (2001) showed that $-2\log\mathcal{R}(\theta_0) \to \chi_p^2$ in distribution as $n \to \infty$, where θ_0 satisfies $E[m(X,\theta)] = 0$.

7.2 Empirical likelihood in missing data problems

In the section, we introduce two missing data problems. In each problem, we explore the use of empirical likelihood to effectively combine unbiased estimating equations when the number of estimating equations is greater than the number of unknown parameters. Most of material on this section comes from Qin et al. (2009). Related works can be found in Xue et al. (2011), Xue (2009a,b), Xue and Xue (2011), and Yang et al. (2009).

7.2.1 Covariates missing at random in regression model

Let Y be a response variable and (X, Z) be a vector of random covariates. Assume that Y and (X, Z) are related by a regression model

$$(30) \qquad Y = \mu(X, Z, \beta) + \varepsilon,$$

where $\mu(X, Z, \beta)$ is a possibly nonlinear link function indexed by an unknown $p \times 1$ vector parameter β and ε is a random error that satisfies $E(\varepsilon|X, Z) = 0$ so that $E(Y|X, Z) = \mu(X, Z, \beta)$. Suppose that a sample of n observations is collected where (Y, X, Z) are completely observed on a subset of the sample but only (Y, X) are observed on the remaining sample. Let δ be an indicator variable, which equals 1 if Z is observed and 0 if Z is missing. The missing data mechanism associated with the missingness of Z is characterized by the conditional distribution of δ given (Y, X, Z), which is assumed to satisfy

$$(31) \qquad P(\delta = 1|Y, X, Z) = P(\delta = 1|Y, X) = w(Y, X, \eta),$$

where w is a specified probability distribution function for given η, a $q \times 1$ unknown vector parameter. Under (31), the data is MAR. Let $(\delta_i, Y_i, X_i, Z_i)$ be a generic symbol for an observation on (δ, Y, X, Z). We are interested in estimating the regression parameter β. Let $U(Y, X, Z, \beta)$ be a set of unbiased estimating functions for β, that is, $E[U(Y, X, Z, \beta)] = 0$. In the absence of missing data, so that $\delta_i = 1, i = 1, \ldots, n$, we may estimate β by solving $\sum_{i=1}^{n} U(Y_i, X_i, Z_i, \beta) = 0$. A common choice for $U(Y, X, Z, \beta)$ is $a(X, Z, \beta)\{Y - \mu(X, Z, \beta)\}$ where $a(X, Z, \beta)$ is a vector of known functions of (X, Z), up to the unknown parameter β. Write

$$(32)$$
$$\begin{cases} g_1(\delta, Y, X, Z, \beta, \eta) = \dfrac{\delta \times U(Y, X, Z, \beta)}{w(Y, X, \eta)}, \\[2mm] g_2(\delta, Y, X, Z, \beta, \eta, \gamma) = \dfrac{\delta - w(Y, X, \eta)}{w(Y, X, \eta)}\varphi(Y, X, \beta, \gamma), \end{cases}$$

where φ is a $p \times 1$ vector of known functions of (Y, X), up to the unknown parameter β and another unknown (vector) parameter γ. The optimal choice for φ is given by $E\{U(Y, X, Z, \beta)|Y, X\}$. Since this optimal function is unknown, it needs to be estimated using the observed data. One possible approach is to fit a flexible regression model for $\varphi(Y, X, \beta, \gamma)$ by regressing $U(Y, X, Z, \beta)$ on (Y, X) using the data with complete information on (Y, X, Z).

Since $E\{g_1(\delta, Y, X, Z, \beta, \eta)\} = 0$ and $E\{g_2(\delta, Y, X, Z, \beta, \eta, \gamma)\} = 0$, we can estimate β based on the following system of estimating equations:

$$(33) \qquad \begin{cases} \sum_{i=1}^{n} g_1(\delta_i, Y_i, X_i, Z_i, \beta, \hat{\eta}) = 0, \\ \sum_{i=1}^{n} g_2(\delta_i, Y_i, X_i, Z_i, \beta, \hat{\eta}, \gamma) = 0, \end{cases}$$

where $\hat{\eta}$ is a consistent estimate of η based on the data $\{(\delta_i, Y_i, X_i), 1 \leq i \leq n\}$ under the missing probability model $w(Y, X, \eta)$. Since $E\{g_2(\delta, Y, X, Z, \beta, \eta, \gamma)\} = 0$ for any choice of $\varphi(Y, X, \beta, \gamma)$, therefore, g_2 is not sensitive to the underlying parameter β and by itself, g_2 cannot be used to estimate β. However, g_2 can be used for improving upon the Horvitz-Thompson (HT) estimating function, g_1, as follows. Since the number of estimating functions in (33) is $2p$, which is greater than the dimension, p, of β, the question arises as how to combine the two sets of estimating functions in (33). In (33), g_1 and $g_1 - g_2$ yield, respectively, the system of estimating equations, but other linear combinations are also possible. The empirical likelihood method of Owen (1988, 1990) and Qin and Lawless (1994) provides an effective method for combining estimating equations when the number of estimating equations is greater than the number of parameters. See Section 7.1. The idea is as follows. Let (β_0, η_0) denote the true values of (β, η), and let

$$(34) \qquad g(\delta, Y, X, Z, \beta, \eta) = \begin{pmatrix} g_1(\delta, Y, X, Z, \beta, \eta) \\ g_2(\delta, Y, X, \beta, \eta) \end{pmatrix},$$

then $E[g(\delta, Y, X, Z, \beta_0, \eta_0)] = 0$. Since the dimension of $g(\delta, Y, X, Z, \beta, \eta)$ is higher than the dimension of β, which is p, we employ the empirical likelihood method (Qin and Lawless, 1994) to seek an optimal combination of the estimating functions $g_1(\delta, Y, X, Z, \beta, \eta)$ and $g_2(\delta, Y, X, \beta, \eta)$. To this end, suppose $(\delta, Y, X, Z) \sim F$ and let $L = \prod_{i=1}^{n} p_i$, where $p_i = dF(\delta_1, Y_i, X_i, Z_i), i = 1, \ldots, n$, are nonnegative jump sizes with total mass that sums to 1. For fixed β, we need to maximize L subject to the constraints.

$$p_i \geq 0, \qquad \sum_{i=1}^{n} p_i = 1, \qquad \sum_{i=1}^{n} p_i g(\delta_i, Y_i, X_i, Z_i, \beta, \hat{\eta}_{\mathrm{ML}}) = 0.$$

After profiling the p_i's, the profile empirical log-likelihood of β is given by

$$l_1(\beta) = -\sum_{i=1}^{n} \log\{1 + \lambda^T g(\delta_i, Y_i, X_i, Z_i, \beta, \hat{\eta}_{\mathrm{ML}})\} - n \log n,$$

where $\lambda = \lambda(\beta)$ is determined by

$$\frac{1}{n} \sum_{i=1}^{n} \frac{g(\delta_i, Y_i, X_i, Z_i, \beta, \hat{\eta}_{\mathrm{ML}})}{1 + \lambda^T g(\delta_i, Y_i, X_i, Z_i, \beta, \hat{\eta}_{\mathrm{ML}})} = 0,$$

where $\hat{\eta}_{\mathrm{ML}}$ is the maximum binomial likelihood estimator of η, based on $\{(\delta_i, Y_i, X_i), 1 \leq i \leq n\}$, which maximizes the binomial likelihood

$$L_{\mathrm{B}}(\eta) = \prod_{i=1}^{n} \{w(Y_i, X_i, \eta)\}^{\delta_i} \{1 - w(Y_i, X_i, \eta)\}^{\delta_i}.$$

Let $\hat{\beta}_{\mathrm{EL1}}$ denote the value of β that maximizes $l_1(\beta)$. We call $\hat{\beta}_{\mathrm{EL1}}$ the pseudo maximum empirical likelihood estimator of β. Qin et al. (2009) proved the asymptotic normality of $\hat{\beta}_{\mathrm{EL1}}$. The empirical likelihood ratio statistic for testing $H_0 : \beta = \beta_0$ is given by $R_1(\beta_0) = 2\{l_1(\hat{\beta}_{\mathrm{EL1}}) - l_1(\beta_0)\}$. When H_0 is true, $R_1(\beta_0)$ is asymptotically a weighted sum of independent χ_1^2 variables with unknown weights.

Since the distribution of $R_1(\beta_0)$ has the unknown weights needed to be estimated, it can be not used directly to construct the confidence region. Therefore, we must to consider the another method. We also construct the empirical likelihood ratio function of (β, η), and obtain the maximum empirical likelihood estimation of (β, η) when η in unknown. For this, let

$$g_3(\delta, Y, X, \eta) = \frac{\{\delta - w(Y, X, \eta)\} w_1(Y, X, \eta)}{w(Y, X, \eta)\{1 - w(Y, X, \eta)\}},$$

$$G(\delta, Y, X, Z, \beta, \eta) = \begin{pmatrix} g(\delta, Y, X, Z, \beta, \eta) \\ g_3(\delta, Y, X, \eta) \end{pmatrix},$$

where g is defined in (34), and $w_1(Y, X, \eta) = \partial w(Y, X, \eta)/\partial \eta$. Then $E[G(\delta, Y, X, Z, \beta_0, \eta_0)] = 0$. The dimension of $G(\delta, Y, X, Z, \beta, \eta)$ is $2p + q$, which is higher than $p + q$, the dimension of (β, η). For fixed (β, η), we need to maximize $L = \prod_{i=1}^{n} p_i$ subject to the constraints.

$$p_i \geq 0, \qquad \sum_{i=1}^{n} p_i = 1, \qquad \sum_{i=1}^{n} p_i G(\delta_i, Y_i, X_i, Z_i, \beta, \eta) = 0.$$

After profiling the p_i's, the profile empirical log-likelihood ratio function of (β, η) is given by

$$l_2(\beta, \eta) = -\sum_{i=1}^{n} \log\{1 + \lambda^T G(\delta_i, Y_i, X_i, Z_i, \beta, \eta)\} - n \log n,$$

where $\lambda = \lambda(\beta, \eta)$ is determined by

$$\frac{1}{n} \sum_{i=1}^{n} \frac{g(\delta_i, Y_i, X_i, Z_i, \beta, \eta)}{1 + \lambda^T g(\delta_i, Y_i, X_i, Z_i, \beta, \eta)} = 0,$$

where $\hat{\eta}_{\mathrm{ML}}$ is the maximum binomial likelihood estimator of η, based on $\{(\delta_i, Y_i, X_i), 1 \leq i \leq n\}$, which maximizes the binomial likelihood

$$L_{\mathrm{B}}(\eta) = \prod_{i=1}^{n} \{w(Y_i, X_i, \eta)\}^{\delta_i} \{1 - w(Y_i, X_i, \eta)\}^{\delta_i}.$$

Let $(\hat{\beta}_{\mathrm{EL2}}, \hat{\eta}_{\mathrm{EL}})$ denote the maximum empirical likelihood estimator of β that maximizes $l_2(\beta, \eta)$. Qin et al. (2009)

proved the asymptotic normality of $(\hat{\beta}_{\text{EL2}}, \hat{\eta}_{\text{EL}})$, and $\hat{\beta}_{\text{EL2}}$ and $\hat{\eta}_{\text{EL}}$ are asymptotically independent. In addition, Qin et al. (2009) point out that $\hat{\eta}_{\text{EL}}$ and $\hat{\eta}_{\text{ML}}$ are asymptotically equivalent, and that are asymptotically efficient for estimating η in parametric models. The empirical likelihood ratio statistic for testing $H_0 : \beta = \beta_0$ is given by $R_2(\beta_0) = 2\{l_2(\hat{\beta}_{\text{EL2}}, \hat{\eta}_{\text{EL}}(\beta_0)) - l_2(\beta_0, \hat{\eta}_{\text{EL}}(\beta_0))\}$, where $\hat{\eta}_{\text{EL}}(\beta_0)$ is the maximum empirical likelihood estimator of η subject to $\beta = \beta_0$. When H_0 is true, $R_2(\beta_0) \to \chi_p^2$ in distribution as $n \to \infty$.

7.2.2 Surrogate response

Surrogate response data arise frequently in medical and social science research. Utilizing surrogate response to improve estimation efficiency is a challenging statistical problem.

Let Y be a response variable, X be a covariate vector, and S be a surrogate for Y. Suppose S and X can always be observed, but Y may be missing in some of the observations. Let $\delta = 1$ if Y is observed and $\delta = 0$ if Y is missing. We assume a parametric model $f(Y|X, \beta)$ for the conditional density of Y given X, where β is an unknown $p \times 1$ vector parameter. Furthermore, we assume $P(\delta = 1|Y, S, X) = w(S, X, \eta)$, where w is a known function and η is a $q \times 1$ vector of unknown parameters. If Y is continuous, a practical example is when $S = I(Y < c)$, for a known constant c. In this case, the missingness probability of Y depends on the value of S. Based on the data $\{(\delta_i, S_i, X_i), 1 \le i \le n\}$, the parameter η can be estimated using the binomial log-likelihood

$$l_{\text{B}}(\eta) = \sum_{i=1}^{n} \{\delta_i \log w(S_i, X_i, \eta) + (1-\delta_i) \log(1 - w(S_i, X_i, \eta))\}.$$

Consider the following unbiased estimating equations:

$$\sum_{i=1}^{n} g(\delta_i, Y_i, S_i, X_i, \beta, \eta, \gamma) = 0,$$

where

$$g(\delta, Y, S, X, \beta, \eta, \gamma) = \begin{pmatrix} g_1(\delta, Y, S, X, \beta, \eta) \\ g_2(\delta, S, X, \beta, \eta, \gamma) \\ g_3(\delta, S, X, \eta), \end{pmatrix},$$

$$g_1(\delta, Y, S, X, \beta, \eta) = \frac{\delta}{w(S, X, \eta)} \frac{\partial \log f(Y|X, \beta)}{\partial \beta},$$

$$g_2(\delta, S, X, \beta, \eta, \gamma) = \frac{\delta - w(S, X, \eta)}{w(S, X, \eta)} \frac{\partial \log f(S, X, \beta, \gamma)}{\partial \beta},$$

$$g_3(\delta, S, X, \eta) = \frac{\delta - w(S, X, \eta)}{w(S, X, \eta)\{1 - w(S, X, \eta)\}} \frac{\partial w(S, X, \eta)}{\partial \eta},$$

$$f(S, X, \beta, \gamma) = \int f(y|X, \beta) f(S|y, X, \gamma) dy,$$

γ is an unknown (vector of) parameter(s) to be determined using the data. It is easy to verify that $E\{g(\delta, Y, S, X, \beta, \eta, \gamma)\} = 0$. If η is known, the last estimating function $g_3(\delta, S, X, \eta)$ is not needed. Again, the number of estimating equations is greater than the number of parameters. Similar to the method of 7.2.1, we can construct the empirical likelihood ratio statistics of (β, η, γ). Here omit the detail.

8. PLUG-IN EMPIRICAL LIKELIHOOD

We describe the general framework. The basic idea of empirical likelihood is to regard the observations $X_1, \ldots, X_n$ as if they are i.i.d. from a fixed and unknown d-dimensional distribution F of X, and to model F by a multinomial distribution concentrated on the observations. Inference for the parameter of interest, $\theta_0 = \theta(F) \in \Theta$, is then carried out using a p-dimensional estimating function of the form $m_n(X, \theta, h)$, where h is a (possibly infinite-dimensional) 'nuisance' parameter with unknown true value $h_0 = h(F) \in \mathcal{H}$. When h_0 is known, it can replace h in the empirical likelihood ratio function

$$L_n(\theta, h) = \max \left\{ \prod_{i=1}^{n} (np_i) \middle| p_i \ge 0, \sum_{i=1}^{n} p_i = 1, \right.$$
$$\left. \sum_{i=1}^{n} p_i m_n(\theta, h) = 0 \right\}.$$

leading to a confidence region $\{\theta | L_n(\theta, h) > c\}$ for θ_0, where c is a suitable positive constant, and the maximum of the empty set is defined to be zero. The constant c can be calibrated using Owen's (1990) empirical likelihood theorem, provided $m_n = m$ does not depend on n: if the observations are i.i.d. and $m(X, \theta_0, h_0)$ has zero mean and a positive definite covariance matrix, then $-2 \log L_n(\theta_0, h_0) \xrightarrow{D} \chi_p^2$, where χ_p^2 has a chi-squared distribution with p degrees of freedom.

Hjort et al. (2009) establish a generalization of Owen's result in which the unknown h_0 is replaced by an estimator $\hat{h}$, and the estimating function is allowed to depend on n. Let $\{a_n\}$ be a sequence of positive constants bounded away from zero, and U a non-degenerate p-dimensional random vector. In most of the applications we consider, $a_n = 1$ and $U \sim N_p(0, V_1)$, where the covariance matrix V_1 is positive definite, but the extra generality can be useful in some applications. Let V_2 denote a $p \times p$ positive definite covariance matrix. The following conditions are needed.

(A0) $P\{L_n(\theta_0, \hat{h}) = 0\} \to 0$.

(A1) $\sum_{i=1}^{n} m_n(X_i, \theta_0, \hat{h}) \xrightarrow{D} U$.

(A2) $a_n \sum_{i=1}^{n} m_n(X_i, \theta_0, \hat{h}) m_n^T(X_i, \theta_0, \hat{h}) \xrightarrow{P} V_2$.

(A3) $a_n \max_{1 \le i \le n} \|m_n(X_i, \theta_0, \hat{h})\| \xrightarrow{P} 0$.

Hjort et al. (2009) proved the following result: If (A0)–(A3) hold, then

$$-2a_n^{-1} \log L_n(\theta_0, \hat{h}) \xrightarrow{D} U^T V_2^{-1} U.$$

This result will provide a way of calibrating $\{\theta \mid L_n(\theta, \hat{h}) > c\}$ as a confidence region for θ_0. When $U \sim N_p(0, V_1)$ with V_1 positive definite, the limit distribution above may be expressed as $w_1 \chi_{1,1}^2 + \ldots + w_p \chi_{1,p}^2$, where the $\chi_{1,j}^2$ are independent chi-squared random variables with one degree of freedom and the weights $w_1, \ldots, w_p$ are the eigenvalues of $V_2^{-1} V_1$, cf. Lemma 3 of Qin and Jing (2001). If, in addition, V_1 and V_2 coincide, we have the standard χ_p^2 limit distribution. When V_1 and V_2 are not identical, the weights $w_1, \ldots, w_p$ may need to be estimated, for example via consistent estimators $\hat{V}_1$, $\hat{V}_2$ and computing the eigenvalues of $\hat{V}_2^{-1} \hat{V}_1$. It is not possible to say anything in general about estimation of V_1 and V_2, which will depend on the structure of the specific application. More remarks can be found in Hjort et al. (2009), here omits the details.

REFERENCES

CHEN, S. X. and VAN KEILEGOM, I. (2009). A review on empirical likelihood methods for regression. *Test* **18** 415–447. MR2566404

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modeling and its Applications.* London: Chapman and Hall. MR1383587

FAN, J. and LI. R. Z. (2004). New estimation and model selection procedures for semiparametric modeling in longgitudinal data analysis. *J. Amer. Statist. Assoc.* **99** 710–723. MR2090905

HALL, P. and LA SCALA, B. (1990). Methodology and algorithms of empirical likelihood. *Internat. Statist. Rev.* **58** 109–127.

HJORT, N. L., MCKEAGUE, N. L. and VAN KEILEGOM, I. (2009). Extending the scope of empirical likelihood. *Ann. Statist.* **37** 1079–1111. MR2509068

LIN, D. Y. and YING, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data (with Discussion). *J. Amer. Statist. Assoc.* **96** 103–126. MR1952726

OWEN, A. B. (1988). Empirical likelihood ratio confidence intervals for a single function. *Biometrika* **75** 237–249. MR0946049

OWEN, A. B. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18** 90–120. MR1041387

OWEN, A. B. (1991). Empirical likelihood for linear models. *Ann. Statist.* **19** 1725–1747. MR1135146

OWEN, A. B. (2001). *Empirical Likelihood.* Chapman & Hall, New York.

QIN, J. and LAWLESS, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22** 300–325. MR1272085

QIN, G. and JING, B. Y. (2001). Empirical likelihood for censored linear regression. *Scandinavian J. Statist.* **28** 661–673. MR1876506

QIN, J., ZHANG, B. and LEUNG, D. H. Y. (2009). Empirical likelihood in missing data problems. *J. Amer. Statist. Assoc.* **104** 1492–1503. MR2750574

RAO, J. N. K. and SCOTT, A. J. (1981). The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables. *J. Amer. Statist. Assoc.* **76** 221–230. MR0624328

STUTE, W., XUE, L. G. and ZHU, L. X. (2007). Empirical likelihood inference in nonlinear error in covariables models with validation data. *J. Amer. Statist. Assoc.* **102** 332–346. MR2345546

WANG, Q. H. and JING, B.Y. (2001). Empirical likelihood for a class of functionals of survival distribution with censored data, *Ann. Inst. Statist. Math.* **53** 517–527. MR1868888

WANG, Q. H. and RAO, J. N. K. (2002a). Empirical likelihood-based inference in linear errors-in-covariables models with validation data. *Biometrika* **89** 345–358. MR1913963

XUE, D., XUE, L. G. and CHENG, W. H. (2011). Empirical likelihood for generalized linear models with missing responses. *J. Statist. Plann. Inference* **141** 2007–2020. MR2772207

XUE, L. G. (2009a). Empirical likelihood for linear models with missing responses. *J. Multivariate Anal.* **100** 1353–1366. MR2514134

XUE, L. G. (2009b). Empirical likelihood confidence intervals for response mean with data missing at random. *Scand. J. Statist.* **36** 671–685. MR2573302

XUE L. G. and XUE D. (2011). Empirical likelihood for semiparametric regression model with missing response data. *J. Mult. Anal.* **102** 723–740. MR2772332

XUE, L. G. and ZHU, L. X. (2006). Empirical likelihood for single-index model. *J. Multivariate Anal.* **97** 1295–1312. MR2279674

XUE, L. G. and ZHU, L. X. (2007a). Empirical likelihood for a varying coefficient model with longitudinal data. *J. Amer. Statist. Assoc.* **102** 642–654. MR2370858

XUE, L. G. and ZHU, L. X. (2007b). Empirical likelihood semiparametric regression analysis for longitudinal data. *Biometrika* **94** 921–937. MR2416799

XUE, L. G. and ZHU, L. X. (2010). *Empirical Likelihood in Nonparametric and Memiparametric Model.* Science Press, Beijing.

YANG, Y. P., XUE, L. G. and CHENG, W. H. (2009). Empirical likelihood for a partially linear model with covariate data missing at random. *J. Statist. Plann. Inference* **139** 4143–4153. MR2558357

ZEGER, S. L. and DIGGLE, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* **50** 689–699.

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* **68** 549–570. MR2278341

Liugen Xue
Beijing University of Technology
Beijing, China
E-mail address: lgxue@bjut.edu.cn

Lixing Zhu
Hong Kong Baptist University
Hong Kong, China
E-mail address: lzhu@hkbu@edu.hk