Wang, Qihua; Rao, J. N. K.

Working Paper

Empirical likelihood-based inference in linear errors-in-covariables models with validation data

SFB 373 Discussion Paper, No. 2001,64

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Empirical Likelihood-based Inference in Linear Errors-in-Covariables Models with Validation Data

By QIHUA WANG

Institute of Applied Mathematics, Academy of Mathematics and System Science
Chinese Academy of Science, Beijing 100080, P. R. China
qhwang@amath4.amt.ac.cn

AND J. N. K. RAO

School of Mathematics and Statistics, Carleton University
Ottawa, K1S 5B6, Canada
jrao@math.carleton.ca

SUMMARY

Linear errors-in-covariables models are considered, assuming the availability of independent validation data on the covariables in addition to primary data on the response variable and surrogate covariables. We first develop an estimated empirical log-likelihood with the help of validation data and prove that its asymptotic distribution is that of a weighted sum of independent standard χ_1^2 random variables with unknown weights. By estimating the unknown weights consistently, an estimated empirical likelihood confidence region on the regression parameter vector is constructed. We also suggest an adjusted empirical log-likelihood and prove that its asymptotic distribution is a standard χ^2. To avoid estimating the unknown weights or the adjustment factor, we propose a partially smoothed bootstrap empirical log-likelihood to construct a confidence region which has asymptotically correct coverage probability. A simulation study is conducted to compare the proposed methods with a normal approximation based method in term of coverage accuracy and average length of the confidence interval.

Some key words: Bootstrap empirical likelihood; Confidence region; Measurement error; Surrogate variables.

1. INTRODUCTION

Let Y be a response variable and X be a d-vector of explanatory variables. We consider the linear regression model

$$Y = X^T\beta + e, \tag{1.1}$$

where $\beta \in \mathbf{R}^d$ is a d-vector of regression coefficients, e is a random error, and given X, the errors e are independent and identically distributed.

In many applications, the exact measurement of X may be difficult, time consuming or expensive, so a surrogate $\widetilde{X}$ is observed instead of X. Generally, the relationship between the surrogate variables $\widetilde{X}$ and the true variables X can be rather complicated compared to the classical additive error structure usually assumed (Fuller, 1987). For example, $\widetilde{X}$ may follow the model $\widetilde{X} = \mathcal{F}(X, \epsilon)$, where ϵ is independent of (Y, X) and $\mathcal{F}$ is an arbitrary function. This situation presents serious difficulties in making valid statistical inferences on the regression parameter vector β. One solution is to postulate distributional assumptions on the conditional distribution of X given $\widetilde{X}$, but the resulting inferences could be sensitive to the assumed distribution.

In this paper, we consider settings where some validation data are available to relate X and $\widetilde{X}$. In particular, we assume that independent validation data $\{(X_i, \widetilde{X}_i)_{i=N+1}^{N+n}\}$ are available in addition to the primary data $\{(Y_i, \widetilde{X}_i)_{i=1}^{N}\}$.Pepe & Fleming (1991), Wittes, Lakatos & Probstfield (1989), and Pepe (1992) gave some examples where such validation data are available. In one example of Pepe (1992) related to school-based smoking prevention projects, current smoking behavior was generally assessed through self-reporting by using questionnaires. But, self-reporting data may be subject to error. Chemical analysis of saliva samples for the presence of cotinine yields a more accurate measure of current smoking behavior. However, due to higher costs relative to self-reporting, only a small subset of subjects enrolled in the study can be subjected to cotinine analysis.

With the help of validation data, Stefanski & Carroll (1985), Carroll & Wand (1991), Sepanski & Lee (1995), Wang (1999) and the above referenced authors developed suitable methods, but the existing literature does not deal with empirical likelihood method to handle error-in-covariables models. It is well known that the empirical likelihood method, introduced by Owen (1988), is especially useful for constructing confidence regions (intervals). It has many advantages over some classical and modern methods, such as the normal approximation-based method and the bootstrap method. In particular, it does not impose prior constraints on region shape, does not require the construction of a pivotal quantity, and the region is range preserving and transformation respecting (Hall & La Scala, 1990). Owen (1991) and Chen (1993, 1994) developed empirical likelihood method for the linear regression model (1.1) when X is measured exactly. In this case, the empirical log-likelihood is asymptotically distributed as a standard χ^2.

To develop empirical likelihood method in the presence of errors in covariables, we first calibrate the model (1.1) and obtain an equivalent calibration regression model with $E[X|\widetilde{X}]$ as the covariate vector. Clearly, the empirical log-likelihood function based on the calibration model contains unknown covariate vector $E[X|\widetilde{X}]$ in addition to the unknown parameter vector β. To make β identifiable, we replace $E[X|\widetilde{X}]$ by an estimator obtained from the validation data, leading to an estimated empirical log-likelihood. However, the distribution of the estimated empirical log-likelihood is asymptotically a weighted sum of independent χ_1^2 variables with unknown weights. As a result, it cannot be applied directly to construct confidence regions for β. To overcome this difficulty, several different methods may be used. In the first method, the unknown weights are estimated consistently so that the distribution of the estimated weighted sum of chi-square variables can be calculated from the data. In the second method, the estimated empirical log-likelihood is adjusted so that the resulting adjusted empirical log-likelihood is asymptotically distributed as a standard χ^2. In the third method, a bootstrap empirical log-likelihood is defined so that the distribution of the bootstrap empirical log-likelihood approximates that of the estimated empirical log-likelihood.

The rest of this paper is organized as follows. In §2, we define an estimated empirical log-likelihood and then show that it is asymptotically distributed as a weighted sum of independent χ^2 variables. In §3, we define an adjusted empirical log-likelihood and obtain its asymptotic distribution as a standard χ^2, and then use it to define a confidence region for β. In §4, we develop a partially smoothed bootstrap empirical likelihood method to construct a confidence region for β which has asymptotically correct coverage probability. In §5, we conduct a simulation study to compare the proposed methods and some classical methods. Proofs of theorems in §2-4 are delegated to the Appendix.

2. ESTIMATED EMPIRICAL LOG-LIKELIHOOD

In what follows, we assume that in addition to the primary data consisting of N independent and identically distributed observations $\{(Y_j, \widetilde{X}_j)\}_{j=1}^N$, validation data consisting of n independent and identically distributed observations $\{(\widetilde{X}_j, X_j)_{j=N+1}^{N+n}\}$ are available. Furthermore, we assume that $E[e|\widetilde{X}] = 0$. This assumption may be reasonable in some cases. A special case for the assumption to be true is that e is independent of $\widetilde{X}$ and $Ee = 0$. A more general case is that the error $e = \sigma(X)\epsilon'$ with $E\epsilon' = 0$ and ϵ' is assumed to be independent of both X and $\widetilde{X}$. This is the case when $\widetilde{X} = \mathcal{F}(X, \epsilon)$ and ϵ' is assumed to be independent of both X and ϵ, where ϵ is the measurement error associated with X and $\mathcal{F}$ is an arbitrary function.

Let $u(\widetilde{X}) = E[X|\widetilde{X}]$, then (1.1) may be rewritten as

$$Y = u^T(\widetilde{X})\beta + \eta \tag{2.1}$$

with $\eta = e + X^T\beta - u^T(\widetilde{X})\beta$. This motivates us to introduce the auxiliary random vector $Z_i(\beta) = u(\widetilde{X}_i)\{Y_i - u^T(\widetilde{X}_i)\beta\}$. Now noting that $EZ_i(\beta) = 0$ if β is the true parameter, the problem of testing whether β is the true parameter is equivalent to testing whether $EZ_i(\beta) = 0$ for $i = 1, 2, \cdots, N$. By Owen (1991), this may be done using the empirical likelihood. Let $p_1, p_2, \cdots, p_N$ be the nonnegative numbers summing to unity. Then, the empirical log-likelihood, evaluated at β, is defined as

$$l_N(\beta) = -2 \max_{\sum_{i=1}^N p_i Z_i = 0, \sum_{i=1}^N p_i = 1} \sum_{i=1}^N \log(N p_i).$$

If β is the true parameter, then $l_N(\beta)$ can be shown to be asymptotically distributed as a standard χ^2 with d degrees of freedom, χ^2_d. However, this result cannot be used to make inference about β because $l_N(\beta)$ contains the unknown terms $u(\widetilde{X}_i)$, and hence β is not identifiable. A natural way to solve this problem is to replace $u(\cdot)$ in $l_N(\beta)$ by an estimator. With the validation data, $u(\widetilde{x})$ can be estimated by

$$\widehat{u}_n(\widetilde{x}) = \frac{\sum_{j=N+1}^{N+n} X_j K\left(\frac{\widetilde{x} - \widetilde{X}_j}{h_n}\right)}{\sum_{j=N+1}^{N+n} K\left(\frac{\widetilde{x} - \widetilde{X}_j}{h_n}\right)}, \tag{2.2}$$

where $K(\cdot)$ is a kernel density function and h_n is a bandwidth that decreases to zero as n increase to infinity. To avoid technical difficulties due to small values in the denominator of $\widehat{u}_n(\widetilde{x})$ given by (2.2), we define an estimated empirical log-likelihood function by replacing $u(\cdot)$ in $l_N(\beta)$ with a truncated version of $\widehat{u}_n(\cdot)$. Let

$$\widehat{f}_n(\widetilde{x}) = (n h_n^d)^{-1} \sum_{j=N+1}^{N+n} K\left(\frac{\widetilde{x} - \widetilde{X}_j}{h_n}\right)$$

and $\widetilde{f}_n(\widetilde{x}) = \max\{\widehat{f}_n(\widetilde{x}), b_n\}$ for some positive constant sequence b_n. The truncated version of $\widehat{u}_n(\widetilde{x})$ is then given by

$$\widetilde{u}_n(\widetilde{x}) = \frac{\widehat{u}_n(\widetilde{x}) \widehat{f}_n(\widetilde{x})}{\widetilde{f}_n(\widetilde{x})}. \tag{2.3}$$

Let $Z_{in}(\beta)$ denote $Z_i(\beta)$ with $u(\widetilde{X}_i)$ replaced by $\widetilde{u}_n(\widetilde{X}_i)$ for $i = 1, 2, \cdots, N$. Then, we define an estimated empirical log-likelihood as

$$\widehat{l}(\beta) = -2 \max_{\sum_{i=1}^N p_i Z_{in}(\beta) = 0, \sum_{i=1}^N p_i = 1} \sum_{i=1}^N \log(N p_i),$$

suppressing the subscripts n and N on $\widehat{l}(\beta)$ for simplicity. By the Lagrange multiplier method, $\widehat{l}(\beta)$ can be represented as

$$\widehat{l}(\beta) = 2\sum_{i=1}^{N} \log\{1 + \lambda^T Z_{in}(\beta)\}, \tag{2.4}$$

where λ satisfies

$$\frac{1}{N}\sum_{i=1}^{N} \frac{Z_{in}(\beta)}{1 + \lambda^T Z_{in}(\beta)} = 0. \tag{2.5}$$

By Taylor's expansion and some mild regularity conditions given in Appendix 1, it can be shown that

$$\widehat{l}(\beta) = \left\{\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_{in}(\beta)\right\}^T \left\{\frac{1}{N}\sum_{i=1}^{N} Z_{in}(\beta)Z_{in}^T(\beta)\right\}^{-1} \left\{\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_{in}(\beta)\right\} + o_p(1).$$

Further, $\sum_{i=1}^{N} Z_{in}(\beta)/\sqrt{N}$ can be shown to be asymptotically normal with zero mean vector and covariance matrix

$$\begin{aligned} V(\beta) &= E[u(\widetilde{X})u^T(\widetilde{X})\{Y - u^T(\widetilde{X})\beta\}^2] \\ &\quad + \gamma E\left[u(\widetilde{X})u^T(\widetilde{X})\{(X - u(\widetilde{X}))^T\beta\}^2\right], \end{aligned}$$

where $\gamma = \lim(N/n)$. Also, $\sum_{i=1}^{N} Z_{in}(\beta)Z_{in}^T(\beta)/N \xrightarrow{p} V_0(\beta)$, where

$$V_0(\beta) = E\{u(\widetilde{X})u^T(\widetilde{X})(Y - u^T(\widetilde{X})\beta)^2\}.$$

This implies that $\widehat{l}(\beta)$ is not asymptotically χ_d^2 since $V(\beta) \neq V_0(\beta)$ when $\gamma \neq 0$. Actually, it is asymptotically a weighted sum of independent χ_1^2 variables, as stated in Theorem 1.

THEOREM 1 *Under the regularity conditions listed in Appendix 1, if β is the true value of the parameter, then*

$$\widehat{l}(\beta) \xrightarrow{\mathcal{L}} w_1\chi_{1,1}^2 + \cdots + w_d\chi_{1,d}^2,$$

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

To apply Theorem 1 to construct a confidence region (interval) for β, we must estimate the unknown weights w_i consistently. By the "plug in" method, $V(\beta)$ and $V_0(\beta)$ can be estimated consistently by

$$\begin{aligned}\widehat{V}(\widehat{\beta}) &= N^{-1}\sum_{i=1}^{N}\{\widetilde{u}_n(\widetilde{X}_i)\widetilde{u}_n^T(\widetilde{X}_i)(Y_i-\widetilde{u}_n^T(\widetilde{X}_i)\widehat{\beta})^2\} \\ &\quad +\frac{N}{n^2}\sum_{i=N+1}^{N+n}\{\widetilde{u}_n(\widetilde{X}_i)\widetilde{u}_n^T(\widetilde{X}_i)\left[(X_i-\widetilde{u}_n(\widetilde{X}_i))^T\widehat{\beta}\right]^2\}\end{aligned}$$

and

$$\widehat{V}_0(\widehat{\beta}) = N^{-1}\sum_{i=1}^{N}\{\widetilde{u}_n(\widetilde{X}_i)\widetilde{u}_n^T(\widetilde{X}_i)(Y_i-\widetilde{u}_n^T(\widetilde{X}_i)\widehat{\beta})^2\},$$

respectively. Here $\widehat{\beta}$ is the least squares estimator of β based on (2.1) and the "plug in " method:

$$\widehat{\beta} = \widehat{\Sigma}^{-1}\widehat{A} \tag{2.6}$$

with

$$\begin{aligned}\widehat{\Sigma} &= \tfrac{1}{N}\sum_{i=1}^{N}\widetilde{u}_n(\widetilde{X}_i)\widetilde{u}_n^T(\widetilde{X}_i), \\ \widehat{A} &= \tfrac{1}{N}\sum_{i=1}^{N}\widetilde{u}_n^T(\widetilde{X}_i)Y_i.\end{aligned}$$

This implies that the eigenvalues of $\widehat{D}(\widehat{\beta}) = \widehat{V}_0^{-1}(\widehat{\beta})\widehat{V}(\widehat{\beta})$, say $\widehat{w}_i$, estimate w_i consistently for $i = 1, 2, \cdots, d$. Let $\widehat{c}_\alpha$ be the $1-\alpha$ quantile of the conditional distribution of the weighted sum $\widehat{S} = \widehat{w}_1\chi^2_{1,1} + \cdot + \widehat{w}_d\chi^2_{1,d}$ given the data. Then, the confidence region for β with asymptotically correct coverage probability $1-\alpha$ can be defined as

$$\widehat{I}_\alpha(\widetilde{\beta}) = \{\widetilde{\beta} : \widehat{l}(\widetilde{\beta}) \le \widehat{c}_\alpha\}.$$

In practice, the conditional distribution of the weighted sum $\widehat{S}$ given the data $\{(X_i, \widetilde{X}_i)_{i=N+1}^{N+n}\}$ and $\{(Y_i, \widetilde{X}_i)_{i=1}^{N}\}$ can be obtained using Monte Carlo simulations by repeatedly generating independent samples $\chi^2_{1,1}, \cdots, \chi^2_{1,d}$ from χ^2_1 distribution.

In the absence of measurement error, $D(\beta)$ reduces to an identity matrix so that $w_i = 1$ for $1 \leq i \leq d$, and Theorem 1 reduces to Owen's (1991) result that the empirical log-likelihood is asymptotically χ^2_d.

3. ADJUSTED EMPIRICAL LOG-LIKELIHOOD

Following Rao & Scott (1981), the distribution of $\rho(\beta)(\sum_{i=1}^d w_i \chi^2_{1,i})$ can be approximated by χ^2_d, where $\rho(\beta) = d/\mathrm{tr}\{D(\beta)\}$ and $\mathrm{tr}(A)$ denotes the trace of a matrix A. This implies that the asymptotic distribution of $\tilde{l}(\beta) = \widehat{\rho}(\widehat{\beta})\widehat{l}(\beta)$ can be approximated by χ^2_d by Theorem 2.1 and the consistency of $\widehat{V}(\widehat{\beta})$ and $\widehat{V}_0(\widehat{\beta})$, where $\widehat{\rho}(\widehat{\beta}) = d/\mathrm{tr}\{\widehat{D}(\widehat{\beta})\}$. However, the accuracy of this approximation depends on the values of $w_i's$. Next, we give an adjusted empirical log-likelihood whose asymptotic distribution is exactly a standard chi-square. Note that

$$\widehat{\rho}(\widehat{\beta}) = \frac{\mathrm{tr}\{\widehat{V}^{-1}(\widehat{\beta})\widehat{V}(\widehat{\beta})\}}{\mathrm{tr}\{\widehat{V}_0^{-1}(\widehat{\beta})\widehat{V}(\widehat{\beta})\}}. \tag{3.1}$$

By examining the asymptotic expansion of $\widehat{l}(\beta)$, we replace $\widehat{V}(\widehat{\beta})$ in (3.1) by $\widehat{H}(\widehat{\beta}) = \{\sum_{i=1}^n Z_{in}(\widehat{\beta})\}\{\sum_{i=1}^n Z_{in}(\widehat{\beta})\}^T$ and get a different adjustment factor

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

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

$$\widehat{l}_{ad}(\beta) = \widehat{r}(\beta)\widehat{l}(\beta). \tag{3.2}$$

THEOREM 2. *Under the regularity conditions listed in the Appendix, if β is the true value of parameter, then $\widehat{l}_{ad}(\beta)$ is asymptotically χ^2_d. That is,*

$$\widehat{l}_{ad}(\beta) \xrightarrow{\mathcal{L}} \chi^2_d.$$

Based on Theorem 2, $\widehat{l}_{ad}(\beta)$ can be used to construct a confidence region for β. Let

$$\widehat{I}_{ad,\alpha}(\widetilde{\beta}) = \{\widetilde{\beta} : \widehat{l}_{ad}(\widetilde{\beta}) \le c_\alpha\}$$

with $P(\chi^2_d \le c_\alpha) = 1 - \alpha$. Then, by Theorem 2, $\widehat{I}_{ad,\alpha}(\widetilde{\beta})$ gives a confidence region for β with asymptotically correct coverage probability $1 - \alpha$.

Clearly, the Rao-Scott adjusted empirical log-likelihood $\widetilde{l}(\beta)$ can also be improved by replacing $\widehat{\rho}(\widehat{\beta})$ in $\widetilde{l}(\beta)$ by $\widehat{\rho}(\beta)$. The improved Rao-Scott adjusted empirical log-likelihood reduces to the adjusted empirical log-likelihood given by (3.2) when $d = 1$.

4. PARTIALLY SMOOTHED BOOTSTRAP EMPIRICAL LOG-LIKELIHOOD

In this section, we develop a partially smoothed bootstrap empirical log-likelihood method to construct a confidence region for β with asymptotically correct coverage probability $1 - \alpha$. Let μ_N be the empirical distribution of the primary data $\{(\widetilde{X}_i, Y_i), i = 1, 2, \cdots, N\}$. Given $(\widetilde{X}_1, Y_1), \cdots, (\widetilde{X}_N, Y_N)$, let $(\widetilde{X}_1^*, Y_1^*), \cdots, (\widetilde{X}_M^*, Y_M^*)$ be independent, with common distribution μ_N. Let ν_n be the empirical distribution of the validation data $\{(\widetilde{X}_i, X_i), i = N+1, \cdots, N+n\}$. Given $(\widetilde{X}_{N+1}, X_{N+1}), \cdots,$ $(\widetilde{X}_{N+n}, X_{N+n}))$, let $(\widetilde{X}_{M+1}^*, X_{M+1}^*), \cdots, (\widetilde{X}_{M+m}^*, X_{M+m}^*)$ be independent, with common distribution ν_n. It should be pointed that the primary and validation data are bootstrapped independently.

Let $\widetilde{u}_m^*(\cdot)$ be the bootstrap analog of $\widetilde{u}_n(\cdot)$ given by (2.3), based on the bootstrap primary data $\{(\widetilde{X}_i^*, Y_i^*)_{i=1}^M\}$ and bootstrap validation data $\{(\widetilde{X}_i^*, X_i^*)_{i=M+1}^{M+m}\}$. Denote the bootstrap auxiliary random vector by $Z_{im}^*(\widehat{\beta}) = \widetilde{u}_m^*(\widetilde{X}_i^*)(Y_i^* - \widetilde{u}_m^{*T}(\widetilde{X}_i^*)\widehat{\beta})$, where $\widehat{\beta}$ is given by (2.6). The bootstrap empirical log-likelihood function is defined as

$$\widehat{l}^*(\widehat{\beta}) = 2\sum_{i=1}^{M} \log(1 + \lambda^* Z_{im}^*(\widehat{\beta})), \tag{4.1}$$

where λ^* satisfies (2.5) with $Z_{in}(\beta)$ and N replaced by $Z_{im}^*(\widehat{\beta})$ and M, respectively.

To prove that the asymptotic distribution of $\widehat{l}^*(\widehat{\beta})$ approximates that of $\widehat{l}(\beta)$, we need to ensure that $\widetilde{X}_{M+1}^*, \cdots, \widetilde{X}_{M+m}^*$ have a probability density. This motivates us to use a smoothed bootstrap obtained as follows. Let $\widetilde{X}^{**}_{M+i} = \widetilde{X}^{*}_{M+i} + h_n \xi_{M+i}$ for $i = 1, 2, \cdots, m$, where h_n is the bandwidth sequence used in §2 and $\xi_{M+i}, i = 1, 2, \cdots, m$ are independent and identically distributed random variables with the common probability density $K(\cdot)$, the kernel function used in §2. We define $\widehat{l}^{**}(\widehat{\beta})$ to be $\widehat{l}^{*}(\widehat{\beta})$ with $\widetilde{X}^{*}_{M+1}, \cdots, \widetilde{X}^{*}_{M+m}$ replaced by $\widetilde{X}^{**}_{M+1}, \cdots, \widetilde{X}^{**}_{M+m}$. Because this method uses smoothed bootstrap sample only partially, we call this method as partially smoothed bootstrap.

In what follows, we denote P^* as the partially smoothed bootstrap probability.

THEOREM 3 *Assume the regularity conditions of Theorem 1. Then, along almost all sample sequences,* $(\widetilde{X}_i, Y_i)$ *for* $1 \le i \le N$ *and* $(\widetilde{X}_j, X_j)$ *for* $N + 1 \le j \le N + n$,

$$\sup_x |P(\widehat{l}(\beta) \le x) - P^*(\widehat{l}^{**}(\widehat{\beta}) \le x)| \longrightarrow 0.$$

as n, N, m *and* M *tend to infinity*

As pointed out by a referee M and m only need to go to infinity and they can but need not match N and n respectively. The distribution of $\widehat{l}^{**}(\widehat{\beta})$ can be calculated by Monte Carlo simulation. Hence, the bootstrap approximation result in Theorem 3 can be used to construct a confidence region for β. Let $\widehat{c}^{**}_{\alpha}$ be the $1 - \alpha$-quantile of the distribution of $\widehat{l}^{**}(\widehat{\beta})$ and

$$\widehat{I}^{**}_{\alpha} = \{\tilde{\beta} : \widehat{l}(\tilde{\beta}) \le \widehat{c}^{**}_{\alpha}\}.$$

Then, by Theorem 3, $\widehat{I}^{**}_{\alpha}$ gives a bootstrap confidence regions for β with asymptotically correct coverage probability $1 - \alpha$.

Compared to the estimated empirical log-likelihood and the adjusted empirical log-likelihood, an advantage of the bootstrap empirical log-likelihood is that it avoids estimating the unknown weights and the adjusting factor. This is especially attractive in some cases when the unknown weights or the adjustment factor are difficult to estimate efficiently. Hall & La Scala(1990) pointed out that a construction of any multivariate bootstrap region requires a decision on how the region should be shaped, and it can be rather difficult to make the decision in a manner which depends only on objective, data-driven criteria. On the other hand, the shape of the confidence region of the bootstrap empirical likelihood is determined automatically by the data configuration.

5. SIMULATION RESULTS

To evaluate the performance of the proposed methods, we conducted a simulation study for the two cases, $d = 1$ and $d = 2$. We compared the four empirical likelihood methods, namely the estimated empirical likelihood, the adjusted empirical likelihood , the improved Rao-Scott adjusted empirical likelihood, and the partly smoothed bootstrap empirical likelihood, with a normal approximation-based method and the standard bootstrap empirical likelihood in terms of coverage accuracies of confidence regions (intervals) and average lengths of confidence intervals. The normal approximation confidence region (interval) was calculated based on the asymptotic normality of $\widehat{\beta}$, defined in (2.6), with asymptotic covariance matrix estimated by

$$\widehat{\Omega} = \widehat{\Sigma}^{-1}\widehat{V}(\widehat{\beta})\widehat{\Sigma}^{-1},$$

where $\widehat{\Sigma}$ and $\widehat{V}(\widehat{\beta})$ are defined in §2.

5.1. *Case 1*: $d = 1$

We considered the linear model (1.1) with $\beta = 1.5$ and e distributed as $N(0, 1)$. The surrogate variable $\widetilde{X}$ was generated such that $\widetilde{X} = 1.25X + \delta\epsilon$, where X and ϵ have the standard exponential distribution and $N(0, 1)$ distribution respectively, and δ is the standard deviation of the measurement error. Results for $\delta = 0.4, 0.6$ are reported. Simulations were run with validation and primary data sizes $(n, N) =$ $(10, 30), (30, 90),$ $(60, 180),$ $(10, 50), (30, 150)$ and $(60, 300)$; bootstrap sample sizes were taken as $m = n$ and $M = N$. The bandwidth h_n was taken to be $n^{-\frac{2}{5}}$ and $b_n = n^{-\frac{1}{24}}$, and the kernel $K(x)$ to be $\frac{1}{2}$ if $|x| \leq 1$, and zero otherwise. The coverage probabilities and the average lengths of the confidence intervals, with nominal level 0.95, were computed by using 5000 simulation runs. The simulation results are presented in Tables 1 and 2.

Insert Tables 1 and 2 here

Since the improved Rao-Scott adjusted empirical likelihood reduces to the adjusted empirical likelihood when $d = 1$, we do not consider the improved Rao-Scott adjusted empirical likelihood in Tables 1 and 2. Table 1 shows that the standard bootstrap empirical likelihood method behaves quite badly in terms of coverage accuracy. It appears that the confidence interval based on this method does not have asymptotically correct coverage probability because the coverage accuracy changes unstably with the sample size. It is clear from Tables 1 and 2 that the normal approximation method is consistently and sometimes grossly over covering and it does this by using long intervals. The adjusted empirical likelihood method consistently achieves higher coverage levels, close to 0.95 as n and N increase, and shorter intervals than the estimated empirical likelihood method, although the differences are not substantial. Also, the adjusted method generally leads to shorter intervals than the smoothed bootstrap empirical likelihood method. This might explain why the smoothed method is generally over covering while the adjusted method leads to coverage levels lower than 0.95. However, the over covering of the smoothed method is smaller than the over covering of the normal approximation method. Comparing the values in Tables 1 and 2 for the estimated, adjusted and smoothed methods as n increases with N/n constant, we see that the coverage levels, for a given δ, rapidly approach the nominal level, 0.95, and the interval lengths rapidly decrease, especially as n increases from 10 to 30. Also, the methods, for a given δ, perform better in terms of coverage levels and interval lengths as N increases with n constant: $N/n = 5$ versus $N/n = 3$. But, the improvement is small compared to increasing n with N/n constant. We also note that the performance of the methods, for a given (n, N), improves as the standard deviation of the measurement error, δ, decreases.

All in all, our limited simulation study suggests that the adjusted and smoothed bootstrap empirical likelihood methods perform better than the remaining methods for $d = 1$. Also, the smoothed method might be preferred over the adjusted method if n is small, $n = 10$, because the adjusted method leads to significant under coverage while the smoothed method is conservative. For larger $n(\geq 30)$, the adjusted method is better.

5.2. *Case 2*: $d = 2$

We considered the linear model (1.1) with $d = 2$, $\beta = (1.5, 2.2)^T$ and X generated from the standard bivariate normal distribution. The surrogate variable $\widetilde{X}$ was generated such that $\widetilde{X} = 1.25X + \delta\epsilon$, where ϵ is distributed as a bivariate normal with zero mean vector and covariance matrix

$$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

and δ is the standard deviation of the measurement error. Results for $\delta = 0.4, 0.6$ are reported. Simulations were run with validation and primary data sizes $(n, N) = (20, 60), (60, 120)$ and $(100, 300)$. The bandwidth h_n, was taken to be $n^{-\frac{1}{5}}$ and $b_n = n^{-\frac{1}{24}}$, and the kernel density function was taken to be the product kernel $K(x_1, x_2) = K_0(x_1)K_0(x_2)$, where

$$K_0(x) = \begin{cases} \frac{15}{16}(1 - x^2)^2, & -1 \leq x \leq 1, \\ 0, & \text{otherwise} \end{cases}$$

The coverage probabilities, computed from 5000 simulation runs, are presented in Table 3. We have not calculated the volumes of the confidence regions because it could take a lot of computational work. As a result, our simulation study for $d = 2$ is limited in scope.

Insert Tables 3 here

In Table 3, the empirical coverage level of the normal approximation confidence region was calculated from the Wald statistic based on $\widehat{\beta}$, where $\widehat{\beta}$ is given by (2.6). It is clear that the standard bootstrap empirical likelihood performs poorly with coverage probabilities substantially lower than the nominal level of 0.95. This result again shows that the distribution of the empirical likelihood might not be well-approximated by that of its standard bootstrap empirical likelihood. Secondly, the normal approximation method leads to coverage probabilities significantly lower than the nominal level 0.95 when n is not large. Among the remaining methods, the adjusted empirical likelihood, the smoothed bootstrap empirical likelihood and the improved Rao-Scott adjusted empirical likelihood seem to perform better than the estimated empirical likelihood.

6. CONCLUDING REMARKS

When the dimension of X and hence of $\widetilde{X}$ is large, the curse of dimensionality may occur because of the estimation of $E[X|\widetilde{X}]$. In this case, a more appealing approach is to consider a dimension reduction model by assuming that $E[X|\widetilde{X}] = (g_1(\widetilde{X}^T\gamma_1), \cdots, g_d(\widetilde{X}^T\gamma_d))$. The parameters $\gamma_1, \cdots, \gamma_d$ can first be estimated by sliced inverse regression techniques due to Li (1991). Then, we can estimate $g_1(\cdot), \cdots, g_d(\cdot)$, respectively, by the univariate kernel regression method. After obtaining the estimator of $E[X|\widetilde{X}]$, we can develop empirical likelihood methods and obtain results similar to those in the previous sections. Asymptotic inference theory in this setting is worth further investigation.

ACKNOWLEDGEMENTS

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada, National Natural Science Foundation of China and a grant for excellent Ph.D. dissertation work in China awarded to the first author.

APPENDIX 1

Regularity conditions

For any vector a, denote by $\|a\|$ the Euclidean norm. Let $\mathcal{D}^m$ be the class of all continuous function $q(\cdot)$ on R^d, $m > d$, such that the derivatives

$$\frac{\partial^{i_1}}{\partial x_1^{i_1}} \frac{\partial^{i_2}}{\partial x_2^{i_2}} \cdots \frac{\partial^{i_d}}{\partial x_d^{i_d}} q(x_1, x_2, \cdots, x_d)$$

are uniformly bounded for $0 \le i_1 + i_2 + \cdots + i_d \le m$. The following assumptions are needed for the proofs of Theorems 1-3.

C1: $E\|X\|^2 < \infty$,

C2: $EY^2 < \infty$,

C3: $f_{\widetilde{X}}(\widetilde{x}) \in \mathcal{D}^m$,

C4: $u(\cdot) \in \mathcal{D}^m$,

C5: $\sqrt{n} b_n \int_{\widetilde{\mathcal{X}}} (\|u(\widetilde{x})\|^2 I[f_{\widetilde{X}}(\widetilde{x}) \le b_n])\, d\widetilde{x} \to 0$,

C6: $K(\cdot)$ is a nonnegative and bounded kernel function of order m with bounded support,

C7: $b_n^4 n h_n^{2d} \to \infty$,

C8: $\frac{n h_n^{2m}}{b_n^4} \to 0$,

C9: $\frac{h_n^{m-\frac{d}{2}}}{b_n^2} \to 0$,

C10: $\frac{nh_n^{3d}}{b_n^4} \to 0$.

C11: $\frac{N}{n} \to \gamma$.

APPENDIX 2

Proofs

Sketch of Proof of Theorem 1. Applying Taylor's expansion to (2.4), we get

$$\hat{l}(\beta) = 2\sum_{i=1}^{N}\{\lambda^T Z_{in}(\beta) - \frac{1}{2}(\lambda^T Z_{in}(\beta))^2\} + o_p(1). \qquad (A.1)$$

By (2.5), it follows that

$$0 = \frac{1}{N}\sum_{i=1}^{N}\frac{Z_{in}(\beta)}{1+\lambda^T Z_{in}(\beta)} = \frac{1}{N}\sum_{i=1}^{N} Z_{in}(\beta) - \frac{1}{N}\sum_{i=1}^{N} Z_{in}(\beta)Z_{in}^T(\beta)\lambda + \frac{1}{N}\sum_{i=1}^{N}\frac{Z_{in}(\beta)\{\lambda^T Z_{in}(\beta)\}^2}{1+\lambda^T Z_{in}(\beta)}. \qquad (A.2)$$

This implies that

$$\lambda = \Big\{\sum_{i=1}^{N} Z_{in}(\beta)Z_{in}^T(\beta)\Big\}^{-1}\sum_{i=1}^{N} Z_{in}(\beta) + o_p(N^{-\frac{1}{2}}) \qquad (A.3)$$

and

$$\sum_{i=1}^{N}\lambda Z_{in}(\beta) = \sum_{i=1}^{N}\{\lambda^T Z_{in}(\beta)\}^2 + o_p(1) \qquad (A.4)$$

by using the results $\max_{1\le i\le n}\|Z_{in}(\beta)\| = o_p(n^{\frac{1}{2}})$, $\frac{1}{n}\sum_{i=1}^{n} Z_{in}(\beta)Z_{in}^T(\beta) = O_p(1)$ and $\lambda = O_p(n^{-\frac{1}{2}})$.

Let

$$\widehat{W}_1(\beta) = \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_{in}(\beta)\right)^T \hat{V}^{-1}(\beta)\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_{in}(\beta)\right)$$

Then, (A.1), (A.2), (A.3) and (A.4) together yield

$$\hat{l}(\beta) = \widehat{W}_1(\beta) + o_p(1). \qquad (A.5)$$

Note that $\hat{V}_0(\beta) \xrightarrow{p} V_0(\beta)$ and $\sum_{i=1}^{n} Z_{in}(\beta)/\sqrt{n} = O_p(1)$. By (A.5), we have

$$\hat{l}(\beta) = \left(\frac{1}{\sqrt{n}}V^{-\frac{1}{2}}(\beta)\sum_{i=1}^{n} Z_{in}(\beta)\right)^T D(\beta)\left(\frac{1}{\sqrt{n}}V^{-\frac{1}{2}}(\beta)\sum_{i=1}^{n} Z_{in}(\beta)\right) + o_p(1) \qquad (A.6)$$

where $D(\beta)$ is defined in Theorem 1.

Let $\widetilde{D} = \text{diag}(w_1, \cdots, w_p)$, where w_i for $1 \leq i \leq p$ are defined in Theorem 1. Then, an orthonormal matrix Q exists such that $Q^T \widetilde{D} Q = D(\beta)$. By (A.6), we have

$$\hat{l}(\beta) = \left(\frac{1}{\sqrt{n}} Q V^{-\frac{1}{2}}(\beta) \sum_{i=1}^{n} Z_{in}(\beta) \right)^T \widetilde{D} \left(\frac{1}{\sqrt{n}} Q V^{-\frac{1}{2}}(\beta) \sum_{i=1}^{n} Z_{in}(\beta) \right) + o_p(1) \qquad (A.7)$$

Let $U_i = (\widetilde{X}_i, \epsilon_i)$ and $V_j = (\widetilde{X}_j, X_j)$. First, it can be proved that

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z_{in}(\beta) = \frac{1}{n\sqrt{N}} \sum_{i=1}^{N} \Psi_n(U_i, V_j; h_n) + o_p(1), \qquad (A.8)$$

where

$$\begin{aligned} \Psi_n(U_i, V_j; h_n) &= u(\widetilde{X}_i)\eta_i + \frac{K\left(\frac{\widetilde{X}_i - \widetilde{X}_j}{h_n}\right)(X_j - u(\widetilde{X}_i))\eta_i}{h_n^d f_{\widetilde{X}}(\widetilde{X}_i)} \\ &+ \frac{K\left(\frac{\widetilde{X}_i - \widetilde{X}_j}{h_n}\right)(X_j - u(\widetilde{X}_i))^T \beta u(\widetilde{X}_i)}{h_n^d f_{\widetilde{X}}(\widetilde{X}_i)}. \end{aligned}$$

Note that Q is an orthonormal matrix. Then, by (A.8), using arguments similar to those used in Wang (1999), we can prove that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Q V^{-\frac{1}{2}}(\beta) Z_{in}(\beta) \xrightarrow{\mathcal{L}} N(0, I_p)), \qquad (A.9)$$

where I_p is the $p \times p$ identity matrix. (A.7) and (A.9) together prove Theorem 2.1.

Sketch of Proof of Theorem 2 Let

$$\widehat{W}_2(\beta) = \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z_{in}(\beta) \right)^T \hat{V}^{-1}(\beta) \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z_{in}(\beta) \right).$$

Recalling the definition of $\hat{l}_{ad}(\beta)$, by (A.5) it follows that

$$\hat{l}_{ad}(\beta) = \widehat{W}_2(\beta) + o_p(1). \qquad (A.10)$$

Notice that $\hat{V}(\beta) \xrightarrow{p} V(\beta)$. This together with (A.9) proves that $\widehat{W}_2(\beta)$, and hence $\hat{l}_{ad}(\beta)$ by (A.10) is asymptotically distributed as χ^2_d.

Sketch of Proof of Theorem 3 Let E^* denote the bootstrap expectation. Let $f^*_{\widetilde{X}^{**}}(\cdot)$ be the probability density of $\widetilde{X}^{**}$ and $u^*(\cdot) = E^*(X^* | \widetilde{X}^{**} = \cdot)$. Standard arguments can be used to prove that

$$E^*\|X^*\|^2 = \frac{1}{n}\sum_{i=N+1}^{n+N} \|X_i\| \xrightarrow{a.s.} E\|X\|^2 < \infty, \quad E^*Y^{*2} = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 \xrightarrow{a.s.} EY^2 < \infty \tag{A.11}$$

and

$$f^*_{\widetilde{X}^{**}}(\cdot) \in \mathcal{D}^m \ \text{ and } \ u^*(\cdot) \in \mathcal{D}^m \ \text{ with probability 1.} \tag{A.12}$$

Next, we prove that

$$\sqrt{n}b_n \int_{\widetilde{\mathcal{X}}} \|\widetilde{u}_n(\widetilde{x})\|^2 I[\widehat{f}_n(\widetilde{x}) \le b_n]\, d\widetilde{x} \xrightarrow{a.s.} 0. \tag{A.13}$$

Observe that

$$\begin{aligned}
& \sqrt{n}b_n \textstyle\int_{\widetilde{\mathcal{X}}} \|\widetilde{u}_n(\widetilde{x})\|^2 I[\widehat{f}_n(\widetilde{x}) \le b_n]\, d\widetilde{x} \\
= \ & \sqrt{n}b_n \textstyle\int_{\widetilde{\mathcal{X}}} \|\widetilde{u}_n(\widetilde{x})\|^2 I[\widehat{f}_n(\widetilde{x}) \le b_n, f_{\widetilde{X}}(\widetilde{x}) \le b_n]\, d\widetilde{x} \\
& + \sqrt{n}b_n \textstyle\int_{\widetilde{\mathcal{X}}} \|\widetilde{u}_n(\widetilde{x})\|^2 I[\widehat{f}_n(\widetilde{x}) \le b_n, b_n < f_{\widetilde{X}}(\widetilde{x}) \le 2b_n]\, d\widetilde{x} \qquad (A.14) \\
& + \sqrt{n}b_n \textstyle\int_{\widetilde{\mathcal{X}}} \|\widetilde{u}_n(\widetilde{x})\|^2 I[\widehat{f}_n(\widetilde{x}) \le b_n, f_{\widetilde{X}}(\widetilde{x}) > 2b_n]\, d\widetilde{x} \\
:= \ & \triangle_{n1} + \triangle_{n2} + \triangle_{n3}.
\end{aligned}$$

By assumption C5, C3, C4, C6, C7 and C8, we have

$$\begin{aligned}
\triangle_{n1} & \le 2\sqrt{n}b_n \textstyle\int_{\widetilde{\mathcal{X}}} \|\widetilde{u}_n(\widetilde{x}) - u(\widetilde{x})\|^2 I[f_{\mathcal{X}}(\widetilde{x}) \le b_n]\, d\widetilde{x} + o_p(1) \\
& \le 2\sqrt{n}b_n \left(\sup_{\widetilde{x}} \|\widetilde{u}_n(\widetilde{x}) - u(\widetilde{x})\|)^2 \textstyle\int_{\widetilde{\mathcal{X}}} I[f_{\mathcal{X}}(\widetilde{x}) \le b_n]\, d\widetilde{x}\right) + o_p(1) \\
& = 2\sqrt{n}b_n(O_p((nh_n^d b_n^2)^{-1}) + O_p(h_n^{2m})) \textstyle\int_{\widetilde{\mathcal{X}}} I[f_{\widetilde{X}}(\widetilde{x}) \le b_n]\, d\widetilde{x} + o_p(1) = o_p(1).
\end{aligned} \tag{A.15}$$

Similarly, we can prove that

$$\triangle_{n2} = o_p(1). \tag{A.16}$$

It is easy to see that

$$\triangle_{n3} \le \sqrt{n}b_n \int_{\widetilde{\mathcal{X}}} \|\widetilde{u}_n(\widetilde{x})\|^2 I[|\widehat{f}_n(\widetilde{x}) - f(\widetilde{x})| > b_n]\, d\widetilde{x}.$$

Hence, for any $\epsilon > 0$ we have

$$P(|\triangle_{n3}| > \epsilon) \le P(\sup_{\widetilde{x}} |\widehat{f}_n(\widetilde{x}) - f(\widetilde{x})| > b_n). \tag{A.17}$$

Standard arguments can be used to prove that

$$\sup_{\widetilde{x}} |\widehat{f}_n(\widetilde{x}) - f(\widetilde{x})| = O_p((nh_n^d)^{-\frac{1}{2}}) + O_p(h_n^m)$$

by C3 and C4. This proves that

$$\frac{\sup_{\widetilde{x}} |\widehat{f}_n(\widetilde{x}) - f(\widetilde{x})|}{b_n} = O_p((nh_n^d b_n^2)^{-\frac{1}{2}}) + O_p\left(\frac{h_n^m}{b_n}\right) = o_p(1), \tag{A.18}$$

as $nh_n^d b_n^2 \to \infty$ and $\frac{h_n^m}{b_n} \to 0$, which are implied by C7 and C9, respectively. (A.17) and (A.18) together prove $\triangle_{n3} = o_p(1)$. This together with (A.14), (A.15) and (A.16) proves (A.13)

By (A.11)–(A.13), conditions C6–C10 and arguments similar to those used in the proof of Theorem 1, we can prove that along almost all sample sequences, given $(\widetilde{X}_i, Y_i)$ for $1 \le i \le N$ and $(\widetilde{X}_j, X_j)$ for $N+1 \le j \le N+n$, as m and M, and n and N go to infinity $\widehat{l}^{**}(\widehat{\beta})$ has the same asymptotic weighted chi-square distribution as $\widehat{l}(\beta)$. This together with Theorem 1 proves Theorem 3.

REFERENCES

CARROLL, R.J. & WAND, M.P.(1991). Semiparametric estimation in logistic measurement error models, *J. Roy. Statis. Assoc.,* Ser. B, **53**, 652-63.

CHEN, S.X. (1993). On the accuracy of empirical likelihood confidence regions for linear regression model. *Ann. Inst. Statist. Math.* **45**, 621-37.

CHEN, S.X. (1994). Empirical likelihood confidence intervals for linear regression coefficients. *J. Multi. Anal.* **49**, 24-40.

FULLER, W.A. (1987). *Measurement Error Models*, New York: Wiley.

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

LI, K.C.(1991). Sliced inverse regression for dimension reduction (with discussion), *J. Amer. Statist. Assoc.*. **86**, 337-42.

OWEN, A. (1988). Empirical likelihood ratio confidence intervals for single functional. *Biometrika* **75**, 237-49.

OWEN, A.(1991). Empirical likelihood for linear models. *Ann. Statist.* **19**, 1725-47.

PEPE, M.S.(1992), Inference using surrogate outcome data and a validation sample, *Biometrika*, **79**, 355-65.

PEPE, M.S. & Fleming, T.R.(1991). A general nonparametric method for dealing with errors in missing or surrogate covariate data, *J. Amer. Statist. Assoc.,* **86**, 108-13.

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

SEPANSKI, J.H. & LEE, L.F.(1995), Semiparametric estimation of nonlinear error-in-variables models with validation study, *J. Nonparam. Statist.* **4**, 365-394.

STEFANSKI, L.A. & CARROLL,R.J. (1985), Covariate measurement error in generalized linear models. *Biometrika* **72**, 583-592.

WANG, Q.H. (1999). Estimation of partial linear error-in-variables models with validation data, *J. Multi. Anal.* **69**, 30-64.

WITTES, J., LAKATOS, E. & PROBSTFIELD, J.(1989). Surrogate endpoints in clinical trials: Cardiovascular diseases., *Statist. Med.* **8**, 415-25

Table 1. Coverage probabilities of confidence intervals on β for selected values of δ and sample sizes (n, N) when nominal level is 0.95

$\frac{N}{n}$	δ	(n, N)	EEL	AEL	SBEL	BEL	NA
3	0.4	(10,30)	.9124	.9190	.9754	.9780	.9938
		(30,90)	.9340	.9373	.9612	.8574	.9831
		(60,180)	.9427	.9489	.9577	.7682	.9630
	0.6	(10,30)	.9052	.9094	.9839	.9732	.9958
		(30,90)	.9210	.9238	.9737	.8321	.9863
		(60,180)	.9414	.9428	.9594	.9911	.9682
5	0.4	(10,50)	.9180	.9220	.9742	.6729	.9905
		(30,150)	.9389	.9391	.9583	.8925	.9800
		(60,300)	.9486	.9494	.9546	.9876	.9607
	0.6	(10,50)	.9096	.9136	.9804	.9758	.9935
		(30,150)	.9253	.9284	.9699	.9820	.9827
		(60,300)	.9456	.9460	.9587	.8954	.9632

EEL: estimated empirical likelihood; AEL: adjusted empirical likelihood; SBEL: smoothed bootstrap empirical likelihood; BEL: standard bootstrap empirical likelihood; NA: normal approximation.

Table 2. Average lengths of confidence intervals on β for selected values of δ and sample sizes (n, N) when nominal level is 0.95

$\frac{N}{n}$	δ	(n, N)	EEL	AEL	SBEL	NA
3	0.4	(10,30)	1.5600	1.5500	1.8900	2.0147
		(30,90)	0.4900	0.4300	0.5700	1.0267
		(60,180)	0.3300	0.3200	0.3800	0.5748
	0.6	(10,30)	1.8200	1.7800	2.0200	2.9758
		(30,90)	0.6900	0.6600	0.7300	1.3526
		(60,180)	0.3600	0.3400	0.4500	0.7480
5	0.4	(10,50)	1.5400	1.5100	1.8100	1.9403
		(30,150)	0.4500	0.3900	0.5200	0.9241
		(60,300)	0.3100	0.3000	0.3300	0.3801
	0.6	(10,50)	1.7900	1.7300	1.8800	2.0221
		(30,150)	0.6100	0.5800	0.6200	0.9542
		(60,300)	0.3300	0.3300	0.3500	0.4149

EEL: estimated empirical likelihood; AEL: adjusted empirical likelihood; SBEL: smoothed bootstrap empirical likelihood; NA: normal approximation.

Table 3. Coverage probabilities of confidence regions on β for selected values of δ and sample sizes (n, N) when nominal level is 0.95

δ	(n, N)	EEL	AEL	IRSAEL	SBEL	BEL	NA
	(20,60)	.8970	.9060	.9208	.9814	.8670	.8824
0.4	(60,120)	.9133	.9290	.9257	.9692	.9072	.9062
	(100,300)	.9432	.9429	.9601	.9575	.8902	.9328
	(20,60)	.8820	.9001	.9134	.9950	.8924	.8674
0.6	(60,120)	.9133	.9192	.9247	.9820	.9011	.9028
	(100,300)	.9422	.9427	.9616	.9579	.8728	.9319

EEL: estimated empirical likelihood; AEL: adjusted empirical likelihood; IRSAEL: improved Rao-Scott adjusted empirical likelihood; SBEL: smoothed bootstrap empirical likelihood; BEL: bootstrap empirical likelihood; NA: normal approximation.