Empirical likelihood methods based on influence functions*

Ming Zheng, Ziqiang Zhao, and Wen Yu†

Empirical likelihood methods based on estimating equations have been widely explored in existing literatures. When there exist unknown nuisance parameters in estimating functions, proper methods are required to deal with the nuisance parameters. In this paper, a new empirical likelihood approach is developed. The empirical likelihood functions are constructed based on the influence functions of the parameters of interest. The new method retains the nonparametric Wilks property of empirical likelihood, that is, the resulting log-empirical likelihood ratio statistics converge in distribution to chi-squared random variables. Several examples are discussed to illustrate the effectiveness of the new method. Simulation studies are conducted to assess the finite sample performances and a real example is provided.

AMS 2000 subject classifications: Primary 62G10; secondary 60G20.
Keywords and phrases: Chi-squared distribution, Empirical likelihood, Influence function, Nuisance parameter, Profile likelihood.

1. INTRODUCTION

Empirical likelihood was introduced by [7, 8] for constructing generalized likelihood ratio test statistics and corresponding confidence regions without specifying parametric models for the data. It was motivated by an earlier work of [15], which provided a way to construct confidence intervals for survival probabilities through constrained likelihood ratios. [7, 8] showed that the Wilks theorem of chi-squared limiting distribution for the log-parametric likelihood ratio statistic [20] continues to hold for the log-empirical likelihood ratio. The property is known as the nonparametric Wilks property for empirical likelihood. The empirical likelihood method is well recognized to possess several additional interesting properties including range preserving, transformation respecting, data decided shape for confidence regions and implicit studentizing carried out internally without the need to estimate variance explicitly. Moreover, [13] showed

that when the number of estimating functions is larger than that of unknown parameters, the empirical likelihood function is able to make the optimal linear combination of the original estimation functions automatically. It implies that the empirical likelihood approach provides an efficient way to extract information from data through statistical models. Because of all these features, the method has been extended to many areas in statistical analysis. [9] provides a detailed discussion on the method itself as well as the related applications.

The construction of empirical likelihood ratio includes maximizing a nonparametric likelihood function subject to appropriate constraints. In many cases, the constraints can be defined based on estimating functions for unknown parameters and this makes the empirical likelihood method applicable in various inference problems. Sometimes people are only interested in part of the unknown parameters and treat the others as nuisance parameters. Since nuisance parameters are necessary to describe the underlying generating schemes for observed data, usually the estimating functions derived for parameters of interest depend on unknown nuisance parameters as well. As a result, the empirical likelihood ratio for a certain parameter of interest may also depend on unknown nuisance parameters through the constraints. When the nuisance parameters involved are of finite dimension, [9] proposed a profile empirical likelihood approach, mimicking the idea of profile likelihoods in parametric models with nuisance parameters. Under suitable conditions, the profile log-empirical likelihood ratios still possess chi-squared limiting distributions with appropriate degrees of freedom.

In semiparametric models, the parameters of interest are still of finite dimension, but there exist infinite dimensional nuisance parameters. Thus, the estimation functions for parameters of interest usually involve infinite dimensional nuisance parameters. This brings difficulty to the profile empirical likelihood approach since the approach does not give out meaningful results when profiling out infinite dimensional nuisance parameters. One alternative approach is to replace the nuisance parameter in the estimating function by its consistent estimator and then construct the constraint in empirical likelihood function according to the form of the estimating function. This so-called plugged-in empirical likelihood method has been widely applied under various semiparametric models by much literature. For example, [14, 17]

on partially linear models, [11] on Cox models, [10, 5] on censored linear models, [12] on censored quantile models, [21, 26] on single-index models, [6, 24] on linear transformation models, [18, 19] on linear models with missing responses, among many others.

Although the plugged-in empirical likelihood method has an intuitive idea, the asymptotic property is changed because of the estimation of nuisance parameters. In many cases, instead of converging to a chi-squared distribution, the log-empirical likelihood ratios with estimators being plugged in converge to a weighted sum of several independent chi-squared distributions, c.f. [10, 18, 19, 5, 12, 21, 26, 6, 24], etc. The weights, depending on the limiting variances of the estimators for parameters of interest, are unknown. It means that the plugged-in empirical likelihood method loses the nonparametric Wilks property as well as the optimal linear combination feature. This makes the plugged-in empirical likelihood approach less attractive when competing with the normal approximation based inferential procedures.

According to the semiparametric theory, the asymptotic distribution of a regular and asymptotic linear (RAL) estimator for the parameter of interest is decided by its influence function (see, for example, [16]). Moreover, when the RAL estimator is solved from an estimating equation with the unknown nuisance parameter being replaced by its estimator, the corresponding influence function is generally different from the estimating function. Based on these facts, it seems to be reasonable if we construct the constraint in empirical likelihood function according to the form of influence function instead of estimating function. This idea comes from the fact that the influence functions play a vital role in semiparametric inferential procedures. In this paper, a new empirical likelihood method based on influence functions is developed and its large-sample properties are explored. We find out that in many cases, the new influence function based empirical likelihood method may retain the nonparametric Wilks property, even when the involved possibly infinite dimensional nuisance parameters are replaced by their estimators. Hence, by applying the new empirical likelihood method, there is no need to estimate variances or unknown weights explicitly when making inferences about the parameters of interest.

The rest of the paper is organized as follows. In Section 2, we propose new empirical likelihood methods based on influence functions when the nuisance parameters involved in estimating functions are of finite dimension. The resulting log-empirical likelihood ratio statistics are shown to follow the nonparametric Wilks theorem, which improves the plugged-in empirical likelihood method based on estimating functions. Moreover, we find that our proposed empirical likelihood ratio is asymptotically equivalent to the profile empirical likelihood ratio. The proposed methods are more valuable when the nuisance parameters involved are of infinite dimension, because under these circumstances the profile empirical likelihood is usually not feasible. In Section 3, several examples including two commonly used semipara-

metric models in survival analysis are discussed. We show that the proposed empirical likelihood methods can be applied and improve the results in existing literatures. Some numerical studies are conducted in Section 4 to validate the theoretical findings. A real example is also provided to illustrate the new method. Section 5 concludes. All technical details are summarized in the Appendix.

2. MAIN RESULTS

Let data be envisioned as realizations of random vectors $Z_1, \ldots, Z_n$, where Z_i is of d-dimension corresponding to the data collected on the i-th subject in a sample of n subjects chosen from some population of interest denoted by Z. Suppose that there exist a r-dimensional known function $m(\cdot, \cdot)$ defined on $\mathbb{R}^d \times \mathbb{R}^r$ and a r-dimensional unknown parameter ν_0 satisfying

$$(1) \qquad \mathsf{E}\big[m(Z; \nu_0)\big] = 0.$$

Throughout the paper we will assume that $Z_1, \ldots, Z_n$ are identically and independently distributed (i.i.d.) with the probability law of the population Z. Moreover, suppose that the parameter ν_0 is partitioned into two parts, denoted by θ_0 and γ_0 with $\theta_0 \in \mathbb{R}^p$ ($p < r$) being the parameter of interest while $\gamma_0 \in \mathbb{R}^{r-p}$ being the nuisance parameter.

Based on the random sample and (1), one can establish the following estimating equation

$$(2) \qquad \sum_{i=1}^{n} m\big(Z_i; \nu\big) = 0.$$

Under mild conditions, the equation (2) has a consistent solution known as the m-estimator for ν_0. Denote the solution by $\hat{\nu} = (\hat{\theta}^T, \hat{\gamma}^T)^T$. Then the m-estimator for the parameter of interest is given by $\hat{\theta}$. As we mentioned, the asymptotic property of $\hat{\theta}$ is determined by its corresponding influence function. Now suppose that the regularity conditions C.1 to C.4 listed in the Appendix hold. By the consistency of $\hat{\nu}$ and Taylor series expansion, one may show that

$$\sqrt{n}\big(\hat{\nu} - \nu_0\big) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A^{-1} m\big(Z_i; \nu_0\big) + o_p(1)$$

where $A = \mathsf{E}(\partial m(Z; \nu_0)/\partial \nu^T)$ is a $r \times r$ square matrix. Let $m_1(Z; \theta_0, \gamma_0)$ be the first p elements of $m(Z; \theta_0, \gamma_0) = m(Z; \nu_0)$, $m_2(Z; \theta_0, \gamma_0)$ be the remaining $r - p$ elements, $A_{11} = \mathsf{E}(\partial m_1(Z; \theta_0, \gamma_0)/\partial \theta^T)$, $A_{12} = \mathsf{E}(\partial m_1(Z; \theta_0, \gamma_0)/\partial \gamma^T)$, $A_{21} = \mathsf{E}(\partial m_2(Z; \theta_0, \gamma_0)/\partial \theta^T)$ and $A_{22} = \mathsf{E}(\partial m_2(Z; \theta_0, \gamma_0)/\partial \gamma^T)$. Simple algebra yields that

$$(3) \quad \sqrt{n}\big(\hat{\theta} - \theta_0\big)$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_{11.1}^{-1} \Big[m_1\big(Z_i; \theta_0, \gamma_0\big) - A_{12} A_{22}^{-1} m_2\big(Z_i; \theta_0, \gamma_0\big) \Big]$$
$$+ o_p(1),$$

where $A_{11.1} = A_{11} - A_{12}A_{22}^{-1}A_{21}$. (3) implies that the i-th influence function of $\hat{\theta}$, denoted by $\varphi_\theta(Z_i)$, is given by $A_{11.1}^{-1}[m_1(Z_i; \theta_0, \gamma_0) - A_{12}A_{22}^{-1}m_2(Z_i; \theta_0, \gamma_0)]$. The asymptotic normality of $\hat{\theta}$ can be derived immediately from (3) and the limiting variance-covariance matrix is decided by the influence function φ_θ.

Next we discuss the empirical likelihood based inferential procedure for θ_0. Following [9], the empirical likelihood function can be written as $\prod_{i=1}^n p_i$ with suitable constraints since data are assumed to be i.i.d. Because of the unknown nuisance parameter γ_0, the estimating function $m(Z; \nu_0)$ can not be used directly. According to the idea of plugged-in empirical likelihood, in addition to the standard unit total probability constraint $\sum_{i=1}^n p_i = 1$, one could use

$$(4) \qquad \sum_{i=1}^n p_i m_1(Z_i; \theta, \tilde{\gamma}) = 0,$$

where $\tilde{\gamma}$ is a certain consistent estimator for γ_0. Note that the constraint (4) follows the form of the estimating function corresponding to θ_0, with the nuisance parameter being replaced by its consistent estimator. Although the idea is straightforward, the resulting log-empirical likelihood ratio statistic may lose the nonparametric Wilks property and converge in distribution to a weighted sum of several independent chi-squared random variables. Some literature argued that this is because the components in the constraint (4) are no longer independent of each other with $\tilde{\gamma}$ being plugged in. However, we think that the dependence may not be the key problem. Considering that it is the influence function, not the estimating function, that determines the asymptotic property of $\hat{\theta}$, we propose to construct the constraint according to the form of φ_θ instead of m_1. Define

$$g(Z; \theta) = m_1(Z; \theta, \tilde{\gamma}) - \hat{A}_{12}\hat{A}_{22}^{-1}m_2(Z; \theta, \tilde{\gamma}),$$

where $\hat{A}_{12}$ and $\hat{A}_{22}$ are consistent estimators for A_{12} and A_{22}, respectively. The proposed constraint is defined to be

$$(5) \qquad \sum_{i=1}^n p_i g(Z_i; \theta) = 0,$$

and then the resulting empirical likelihood ratio is given by

$$\mathcal{R}_{IF}(\theta) = \sup\left\{\prod_{i=1}^n np_i \ \middle| \ \sum_{i=1}^n p_i = 1, \right.$$
$$\left. \sum_{i=1}^n p_i g(Z_i; \theta) = 0, \ p_i \geq 0\right\}.$$

Omitting the constant matrix $A_{11.1}$, $g(Z; \theta)$ could be viewed as an estimated version of the influence function with the unknown quantities being replaced by corresponding estimators. Since constraints are invariant by multiplying constants, the constraint (5) can be viewed as being constructed according to the form of the influence function.

The main reason that we use (5) instead of (4) is that the resulting log-empirical likelihood ratio statistic possesses nonparametric Wilks property, which is formally presented by the following theorem.

Theorem 2.1. *Suppose that the conditions C1–C4 listed in the Appendix hold. For any $\tilde{\gamma}$ such that $\sqrt{n}\|\tilde{\gamma} - \gamma_0\| = O_p(1)$, where $\|\cdot\|$ represents the Euclidean norm, $-2\log\mathcal{R}_{IF}(\theta_0)$ converges in distribution to χ_p^2 as $n \to \infty$, where χ_l^2 is a chi-squared random variable with l degrees of freedom.*

The nonparametric Wilks theorem requires consistent estimators for A_{12}, A_{22} and $\sqrt{n}$-consistent estimator for the nuisance parameters. These are not crucial requirements and for the current situation, we can use

$$\hat{A}_{12} = \frac{1}{n}\sum_{i=1}^n \frac{\partial m_1(Z_i; \hat{\theta}, \hat{\gamma})}{\partial \gamma^T}, \qquad \hat{A}_{22} = \frac{1}{n}\sum_{i=1}^n \frac{\partial m_2(Z_i; \hat{\theta}, \hat{\gamma})}{\partial \gamma^T}$$

and $\hat{\gamma}$, respectively. Other choices are available and the specific forms of the estimators do not affect the conclusion of Theorem 2.1. It is interesting to see that the proposed $\mathcal{R}_{IF}(\theta)$ is also a plugged-in empirical likelihood ratio and the components in its constraint are also dependent with each other. However, the chi-squared limiting distribution is preserved since the constraint is constructed based on the influence function instead of the estimating function.

When the nuisance parameter space is of finite dimension, [9] proposed a profile empirical likelihood approach to deal with the nuisance parameters. Define the joint empirical likelihood ratio for θ_0 and γ_0

$$\mathcal{R}(\theta, \gamma) = \sup\left\{\prod_{i=1}^n np_i \ \middle| \ \sum_{i=1}^n p_i = 1, \right.$$
$$\left. \sum_{i=1}^n p_i m(Z_i; \theta, \gamma) = 0, \ p_i \geq 0\right\}.$$

The profile empirical likelihood ratio is then given by

$$\mathcal{R}_P(\theta) = \max_\gamma \mathcal{R}(\theta, \gamma).$$

Under the conditions in Theorem 2.1, [9] claimed that $-2\log\mathcal{R}_P(\theta)$ converges in distribution to χ_p^2 as $n \to \infty$. We show in the Appendix that the proposed influence function based empirical likelihood is asymptotically equivalent to the profile empirical likelihood in the sense that $-2\log\mathcal{R}_{IF}(\theta_0)$ has the same asymptotic power as that of $-2\log\mathcal{R}_P(\theta_0)$ under contiguous alternatives. This finding justifies the usage of the proposed empirical likelihood ratio based on influence functions.

When the nuisance parameters involved in estimating functions are of infinite dimension, the profile empirical likelihood approach is no longer feasible. However, the proposed method may still be useful if one can derive the influence function for the parameter of interest and find appropriate

estimators for the unknown quantities. In the following section, we illustrate the effectiveness of the proposed method through several examples. Two commonly used semiparametric models in survival analysis, Cox model and censored linear regression model, with infinite dimensional nuisance parameters in the estimating functions are included.

3. EXAMPLES

Example I. Variance

Let $Z_1, \ldots, Z_n$ be i.i.d. random variables (i.e., $d = 1$) drawn from the distribution Z with finite fourth moment. Suppose that the parameter of interest is the variance of the distribution, i.e., $\sigma_0^2 = \mathsf{E}[(Z - \mu_0)^2]$ where $\mu_0 = \mathsf{E}(Z)$. The usual estimator for σ_0^2 is given by the sample variance

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Z_i - \bar{Z})^2$$

where $\bar{Z} = n^{-1} \sum_{i=1}^{n} Z_i$ is the sample mean. It is not difficult to calculate that

$$\begin{aligned} \sqrt{n}(\hat{\sigma}^2 - \sigma_0^2) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[(Z_i - \mu_0)^2 - \sigma_0^2 \right] \\ &\quad + \sqrt{n}(\bar{Z} - \mu_0)^2 + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[(Z_i - \mu_0)^2 - \sigma_0^2 \right] + o_p(1). \end{aligned}$$

It implies that the i-th influence function of $\hat{\sigma}^2$ is $\varphi(Z_i) = (Z_i - \mu_0)^2 - \sigma_0^2$. Note that $\bar{Z}$ is a $\sqrt{n}$-consistent estimator for μ_0. The empirical likelihood ratio based on the estimated influence function is given by

$$\begin{aligned} \mathcal{R}_{IF}(\sigma^2) = \sup\Bigg\{ \prod_{i=1}^{n} np_i \ \Bigg| \ \sum_{i=1}^{n} p_i = 1, \\ \sum_{i=1}^{n} p_i \left[(Z_i - \bar{Z})^2 - \sigma^2 \right] = 0, \ p_i \geq 0 \Bigg\}. \end{aligned}$$

By Theorem 2.1, we have that $-2 \log \mathcal{R}_{IF}(\sigma_0^2)$ converges in distribution to χ_1^2 as $n \to \infty$.

The joint empirical likelihood ratio for μ_0 and σ_0^2 takes the form of

$$\begin{aligned} \mathcal{R}(\mu, \sigma^2) = \sup\Bigg\{ \prod_{i=1}^{n} np_i \ \Bigg| \ \sum_{i=1}^{n} p_i = 1, \ \sum_{i=1}^{n} p_i Z_i = \mu, \\ \sum_{i=1}^{n} p_i Z_i^2 = \mu^2 + \sigma^2, \ p_i \geq 0 \Bigg\}. \end{aligned}$$

Thus, the profile empirical likelihood ratio is $\mathcal{R}_P(\sigma^2) = \max_\mu \mathcal{R}(\mu, \sigma^2)$.

Example II. Slope parameter in simple regression model

Next we turn to the simple regression model. Consider the i.i.d. 2-dimensional random vectors $(Y_1, X_1), \ldots, (Y_n, X_n)$ satisfying the simple regression model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, \ldots, n,$$

where ε_i's are i.i.d. random errors with mean zero and finite variance. Suppose that we are only interested in making inference about the slope parameter β_1. The least squares estimators, denoted by $\hat{\beta}_0$ and $\hat{\beta}_1$, are solved from the normal equations

$$\sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i) = 0, \qquad \sum_{i=1}^{n} X_i (Y_i - \beta_0 - \beta_1 X_i) = 0.$$

The empirical likelihood ratio for β_1 based on the estimating function directly is given by

$$\begin{aligned} \mathcal{R}_{EF}(\beta_1) = \sup\Bigg\{ \prod_{i=1}^{n} np_i \ \Bigg| \ \sum_{i=1}^{n} p_i = 1, \\ \sum_{i=1}^{n} p_i X_i (Y_i - \hat{\beta}_0 - \beta_1 X_i) = 0, \ p_i \geq 0 \Bigg\}, \end{aligned}$$

which results in a log-empirical likelihood ratio without the chi-squared limiting distribution. Let $\mu_X = \mathsf{E}(X_1)$, $\sigma_X^2 = \mathsf{Var}(X_1)$. One can show that

$$\begin{aligned} \sqrt{n}(\hat{\beta}_1 - \beta_1) \\ = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sigma_X^{-2} (X_i - \mu_X)(Y_i - \beta_0 - \beta_1 X_i) + o_p(1). \end{aligned}$$

It reveals that the i-th influence function of $\hat{\beta}_1$ is

$$\varphi(Y_i, X_i) = \sigma_X^{-2} (X_i - \mu_X)(Y_i - \beta_0 - \beta_1 X_i).$$

Omitting the proportional constant, the empirical likelihood ratio based on the estimated influence function is given by

$$\begin{aligned} \mathcal{R}_{IF}(\beta_1) = \sup\Bigg\{ \prod_{i=1}^{n} np_i \ \Bigg| \ \sum_{i=1}^{n} p_i = 1, \\ \sum_{i=1}^{n} p_i (X_i - \bar{X})(Y_i - \hat{\beta}_0 - \beta_1 X_i) = 0, \ p_i \geq 0 \Bigg\}, \end{aligned}$$

where $\bar{X} = n^{-1} \sum_{i=1}^{n} X_i$. By Theorem 2.1, we have that $-2 \log \mathcal{R}_{IF}(\beta_1)$ converges in distribution to χ_1^2 when β_1 is the true value of the slope parameter.

Let $\mathbf{X}_i = (1, X_i)^T$. The joint empirical likelihood ratio

for β_0 and β_1 takes the form of

$$\mathcal{R}(\beta_0, \beta_1) = \sup\left\{\prod_{i=1}^{n} np_i \;\middle|\; \sum_{i=1}^{n} p_i = 1, \right.$$
$$\left. \sum_{i=1}^{n} p_i \mathbf{X}_i(Y_i - \beta_0 - \beta_1 X_i) = 0, \;\; p_i \geq 0 \right\}.$$

Thus, the profile empirical likelihood ratio is $\mathcal{R}_P(\beta_1) = \max_{\beta_0} \mathcal{R}(\beta_0, \beta_1)$.

Example III. Regression parameters in Cox model

In survival analysis, the default choice of regression models is Cox model. For a random sample of n independent individual, let T_i be failure time for the i-th individual which is right censored by a censoring time C_i. The observation for each individual includes $\tilde{T}_i = \min(T_i, C_i)$, $\Delta_i = I\{T_i \leq C_i\}$ as well as a p-dimensional possibly time-dependent covariate process $X_i(t)$. Cox model assumes that the conditional hazard of T_i given the covariate process follows

$$\lambda\big(t|\overline{X}_i(t)\big) = \lambda_0(t) \exp\big(\beta_0^T X_i(t)\big),$$

where $\overline{X}_i(t) = \{X_i(u), 0 \leq u \leq t\}$, $\lambda_0(t)$ is an unknown baseline hazard and β_0 is p-dimensional regression parameter of interest. The maximum partial likelihood estimator, denoted by $\hat{\beta}_{PL}$, solves the following estimating equation

$$(6) \qquad \sum_{i=1}^{n} \int_0^{\infty} \big[X_i(t) - E(\beta, t)\big] dN_i(t) = 0,$$

where $N_i(t) = \Delta_i I\{\tilde{T}_i \leq t\}$,

$$E(\beta, t) = \frac{\sum_{i=1}^{n} X_i(t) Y_i(t) \exp(\beta^T X_i(t))}{\sum_{i=1}^{n} Y_i(t) \exp(\beta^T X_i(t))}$$

and $Y_i(t) = I\{\tilde{T}_i \geq t\}$. [11] proposed an empirical likelihood method based on (6) to make inference about β_0. However, their approach requires a known baseline hazard, which is not a common assumption for Cox model.

Under suitable regularity conditions, [1] showed that

$$\sqrt{n}\big(\hat{\beta}_{PL} - \beta_0\big)$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A^{-1} \int_0^{\infty} \big[X_i(t) - e(t)\big] dM_i(t) + o_p(1),$$

where

$$M_i(t) = N_i(t) - \int_0^t Y_i(u) \exp\big(\beta_0^T X_i(u)\big) \lambda_0(u) du,$$

$$e(t) = \frac{\mathsf{E}[X_1(t) Y_1(t) \exp(\beta_0^T X_1(t))]}{\mathsf{E}[Y_1(t) \exp(\beta_0^T X_1(t))]}$$

and A is a $p \times p$ - dimensional invertible matrix. Then the i-th influence function of the maximum partial likelihood estimator is given by

$$\varphi(\tilde{T}_i, \Delta_i, X_i) = A^{-1} \int_0^{\infty} \big[X_i(t) - e(t)\big] dM_i(t).$$

Note that the influence function here involves an infinite dimensional nuisance parameter $\lambda_0(t)$. Define

$$\hat{\Lambda}(\beta, t) = \int_0^t \frac{d \sum_{i=1}^{n} N_i(u)}{\sum_{i=1}^{n} Y_i(u) \exp(\beta^T X_i(u))}.$$

It is easy to see that $\hat{\Lambda}(\beta_0, t)$ is a $\sqrt{n}$-consistent estimator for $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ at each fixed t. Omitting the proportional constant matrix A^{-1}, we defined the following estimated version of the i-th influence function

$$g_i(\beta) = \int_0^{\infty} \big[X_i(t) - E(\beta, t)\big] d\hat{M}_i(\beta, t),$$

where

$$\hat{M}_i(\beta, t) = N_i(t) - \int_0^t Y_i(u) \exp\big(\beta^T X_i(u)\big) d\hat{\Lambda}(\beta, u).$$

The empirical likelihood ratio based on the estimated influence function is given by

$$\mathcal{R}_{IF}(\beta) = \sup\left\{\prod_{i=1}^{n} np_i \;\middle|\; \sum_{i=1}^{n} p_i = 1, \; \sum_{i=1}^{n} p_i g_i(\beta) = 0, \; p_i \geq 0 \right\}.$$

Note that the constraints used in $\mathcal{R}_{IF}(\beta)$ still reflects our basic idea of constructing constraints based on the influence functions. Under similar conditions assumed in [1], one can show that $-2 \log \mathcal{R}_{IF}(\beta_0)$ converges in distribution to χ_p^2. Moreover, the profile empirical likelihood approach can be applied to $\mathcal{R}_{IF}(\beta)$ when interest only lies in certain part of β_0. Chi-squared limiting distributions with appropriate degrees of freedom are expected. The proposed method can be extended to many semiparametric survival models with martingale-type estimating functions, c.f., [22, 25, 23].

Example IV. Regression parameters in censored regression model

An alternative model to Cox model is the linear regression model which assumes that the logarithm or certain transformation of the failure time T_i, denoted by Y_i, and its corresponding p-dimensional covariates vector X_i satisfy

$$\mathsf{E}\big(Y_i|X_i\big) = \beta_0^T \mathbf{X}_i,$$

where $\mathbf{X}_i = (1, X_i^T)^T$ and β_0 is $(p+1)$-dimensional regression parameter of interest. Y_i is right censored by a censoring time C_i. Let $\tilde{Y}_i = \min(Y_i, C_i)$ and $\Delta_i = I(Y_i \leq C_i)$. The observations are $(\tilde{Y}_i, \Delta_i, X_i)$ for $i = 1, \ldots, n$.

Assume that C_i is independent of (Y_i, X_i), and $C_1, \ldots, C_n$ are i.i.d. random variables with a common distribution func-

tion $G(t)$. When $G(t)$ is assumed to be continuous and known, in order to apply the least squares approach, [2] proposed an unbiased modification

$$Y_{i,G} = \frac{\Delta_i \tilde{Y}_i}{1 - G(\tilde{Y}_i)}.$$

Under some assumptions, it can be shown that $E(Y_{i,G}|X_i) = E(Y_i|X_i) = \beta_0^T \mathbf{X}_i$. However, in most cases, $G(t)$ is unknown. [2] proposed replacing $G(t)$ by its Kaplan-Meier estimator $\hat{G}(t)$ based on $(\tilde{Y}_i, 1 - \Delta_i)$ $i = 1, \ldots, n$. This leads to the modified least squares estimator

$$\hat{\beta}_{KSV} = \left(\sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i^T \right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i Y_{i,\hat{G}},$$

where

$$Y_{i,\hat{G}} = \frac{\Delta_i \tilde{Y}_i}{1 - \hat{G}(\tilde{Y}_i)}.$$

We may call $Y_{i,\hat{G}_n}$ "synthetic data" since such data are synthesized from the raw data $(\tilde{Y}_i, \Delta_i)$ to fit the regression model.

$\hat{\beta}_{KSV}$ is solved from the following modified normal equation

$$(7) \qquad \sum_{i=1}^{n} \mathbf{X}_i \left(\tilde{Y}_{i,\hat{G}} - \beta^T \mathbf{X}_i \right) = 0.$$

Based on the estimating function in (7), [10, 5] proposed empirical likelihood approaches. They both defined the following empirical likelihood ratio

$$\mathcal{R}_{EF}(\beta) = \sup \left\{ \prod_{i=1}^{n} np_i \; \middle| \; \sum_{i=1}^{n} p_i = 1, \right.$$
$$\left. \sum_{i=1}^{n} p_i \mathbf{X}_i \left(Y_{i,\hat{G}} - \beta^T \mathbf{X}_i \right) = 0, \; p_i \geq 0 \right\},$$

and showed that under some regularity conditions, $-2 \log \mathcal{R}_{EF}(\beta_0)$ converges in distribution to a weighted sum of $p+1$ independent chi-squared random variables. They argued that the terms in the constraint are no longer independent of each other because the infinite dimensional nuisance parameter $G(t)$ has to be estimated from the data. Thus, the log-empirical likelihood ratio loses the non-parametric Wilks property. However, in our understanding, the chi-squared limiting distribution can be preserved if one constructs the constraint according to the influence function of $\hat{\beta}_{KSV}$. It was proved by [4] that under suitable regularity conditions,

$$\sqrt{n} \left(\hat{\beta}_{KSV} - \beta_0 \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A^{-1} \Big[\mathbf{X}_i \left(Y_{i,G} - \beta_0^T \mathbf{X}_i \right)$$
$$+ \int_{-\infty}^{\tau} H(t) dM_i(t) \Big] + o_p(1),$$

where $A = \mathsf{E}(\mathbf{X}_1 \mathbf{X}_1^T)$, τ is a predetermined constant, $H(t)$ is a $p+1$-dimensional function depending on the joint distribution of Y_i, X_i and C_i,

$$M_i(t) = \left(1 - \Delta_i \right) I \left(\tilde{Y}_i \leq t \right) - \int_0^{t \wedge \tilde{Y}_i} \lambda_G(u) du,$$

and $\lambda_G(t)$ is the hazard function corresponding to $G(t)$. This implies that the i-th influence function of $\hat{\beta}_{KSV}$ is given by

$$\varphi \left(\tilde{Y}_i, \Delta_i, X_i \right) = A^{-1} \Big[\mathbf{X}_i \left(Y_{i,G} - \beta_0^T \mathbf{X}_i \right) + \int_{-\infty}^{\tau} H(t) dM_i(t) \Big].$$

$H(t)$ can be consistently estimated from the data and denote the estimator as $\hat{H}(t)$. Let $\hat{\Lambda}_G(t)$ be the Nelson-Aalen estimator for $\Lambda_G(t) = \int_{-\infty}^{t} \lambda_G(u) du$. Omitting the proportional constant matrix A^{-1}, we defined an estimated version of the i-th influence function

$$g_i(\beta) = \mathbf{X}_i \left(\tilde{Y}_{i,\hat{G}} - \beta^T \mathbf{X}_i \right) + \int_{-\infty}^{\tau} \hat{H}(t) d\hat{M}_i(t),$$

where

$$\hat{M}_i(t) = \left(1 - \Delta_i \right) I \left(\tilde{Y}_i \leq t \right) - \hat{\Lambda}_G \left(t \wedge \tilde{Y}_i \right).$$

According to our basic idea, we propose the following empirical likelihood ratio based on the estimated influence function

$$\mathcal{R}_{IF}(\beta) = \sup \left\{ \prod_{i=1}^{n} np_i \; \middle| \; \sum_{i=1}^{n} p_i = 1, \; \sum_{i=1}^{n} p_i g_i(\beta) = 0, \; p_i \geq 0 \right\}.$$

Different from the empirical likelihood ratios of [10, 5], the proposed empirical likelihood ratio follows the nonparametric Wilks theorem, i.e., $-2 \log \mathcal{R}_{IF}(\beta_0)$ converges in distribution to χ_{p+1}^2.

Note that in this and the previous example, the infinite dimensional nuisance parameters involved in the estimating functions can not be dealt with by profile empirical likelihood. However, the proposed influence function based empirical likelihood methods provide an easy way to retain the nonparametric Wilks property. Moreover, when there exist over-identified estimating functions with infinite dimensional nuisance parameters, our empirical likelihood ratios also keep the optimal linear combination property.

4. NUMERICAL STUDIES

4.1 Simulation studies

In this subsection we carry out some simulation studies to assess the finite sample performances of the proposed methods based on the four examples discussed in the previous section. For Example I, we generate data from normal distribution with mean 1 and variance σ_0^2. The sample size n is set

Table 1. Summarized simulation results for Example I

σ_0^2	n	$-2\log\mathcal{R}_{IF}$		$-2\log\mathcal{R}_P$	
		0.05	0.1	0.05	0.1
1	100	0.0620	0.1086	0.0634	0.1092
	200	0.0538	0.1038	0.0538	0.1038
2	100	0.0620	0.1086	0.0634	0.1092
	200	0.0538	0.1038	0.0538	0.1038
$1+\frac{3}{\sqrt{n}}$	100	0.5016	0.6300	0.5054	0.6324
	200	0.5098	0.6424	0.5116	0.6438
$2+\frac{3}{\sqrt{n}}$	100	0.2156	0.3010	0.2198	0.3030
	200	0.1902	0.2906	0.1908	0.2916

Table 2. Summarized simulation results for Example II

β_1	ε	n	$-2\log\mathcal{R}_{IF}$		$-2\log\mathcal{R}_P$	
			0.05	0.1	0.05	0.1
	Norm	100	0.0486	0.1048	0.0486	0.1056
1		200	0.0546	0.1038	0.0548	0.1040
	Exp	100	0.0624	0.1178	0.0634	0.1182
		200	0.0552	0.1060	0.0548	0.1058
	Norm	100	0.4000	0.5228	0.4012	0.5232
$1+\frac{3}{\sqrt{n}}$		200	0.4046	0.5360	0.4050	0.5362
	Exp	100	0.4284	0.5440	0.4308	0.5440
		200	0.4138	0.5428	0.4138	0.5442

to be 100 and 200. We use both the proposed log-empirical likelihood ratio $-2\log\mathcal{R}_{IF}(\sigma^2)$ and the profile log-empirical likelihood ratio $-2\log\mathcal{R}_P(\sigma^2)$ to test for $\mathsf{H}_0 : \sigma_0^2 = \sigma^2$ against $\mathsf{H}_1 : \sigma_0^2 \neq \sigma^2$. The testing value σ^2 is set to be 1 and 2. For the empirical sizes, the true variance σ_0^2 is chosen to equal to 1 and 2, respectively. For the estimated powers under contiguous alternatives, the true variance σ_0^2 is chosen to equal to $1+3/\sqrt{n}$ and $2+3/\sqrt{n}$, respectively. The nominal levels are chosen to be 0.05 and 0.1. 5,000 replicates are generated. The simulation results are summarized in Table 1.

From the results, we can see that the proposed influence function based empirical likelihood test gives out reasonable estimated sizes when H_0 holds. The test statistic seems to be more liberal for smaller sample sizes, which is noticed by much literature. [9] proposed several methods to decrease the coverage errors under small sample sizes. All the methods can be applied to the proposed test without much difficulty. The profile empirical likelihood test has almost the same asymptotic powers as the proposed one under contiguous alternatives. This validates our theoretical findings of the asymptotic equivalence between the two approaches. Note that the calculation of the proposed log-empirical likelihood ratio is quite easy since it has the same structure as the classic one. The standard algorithm discussed in [9] can be applied directly.

For Example II, we set $\beta_0 = 1$. The random error ε_i's follow either standard normal distribution (Norm) or exponential distribution (Exp). For exponential distribution, we first generate exponential random variables with mean 1 and then subtract the mean to obtain zero-mean random errors. The covariate X_i's are generated from uniform distribution on $[0,2]$. The sample size n is set to be 100 and 200. We use both the proposed log-empirical likelihood ratio $-2\log\mathcal{R}_{IF}(\beta_1)$ and the profile log-empirical likelihood ratio $-2\log\mathcal{R}_P(\beta_1)$ to test for $\mathsf{H}_0 : \beta_1 = 1$ against $\mathsf{H}_1 : \beta_1 \neq 1$. For the empirical sizes, the true value of β_1 is set to be 1, and for the estimated powers under contiguous alternatives, the true value of β_1 is set to be $1+3/\sqrt{n}$. The nominal levels are chosen to be 0.05 and 0.1. 5,000 replicates are generated. The simulation results are summarized in Table 2.

Similar findings can be found from Table 2. The proposed empirical likelihood test has appropriate estimated sizes in almost all scenarios. It also possesses comparable finite-sample performances with the profile empirical likelihood ratio test under both the null hypotheses and the contiguous alternatives.

For Example III, we set the baseline hazard $\lambda_0(t) \equiv 1$. Two covariates are independently drawn from Bernoulli distribution with success probability 0.5 and uniform distribution on $[0,1]$. The corresponding regression parameters, denoted by β_1 and β_2, are set to be -1 and 0.5, respectively. The censoring times are generated from uniform distribution on $[0, c]$, where c is chosen to yield censoring percentages of 30% and 50%. The sample size n is set to be 100 and 200. We first use the proposed $-2\log\mathcal{R}_{IF}(\beta_1, \beta_2)$ to test for $\mathsf{H}_0 : (\beta_1, \beta_2) = (-1, 0.5)$ against $\mathsf{H}_0 : (\beta_1, \beta_2) \neq (-1, 0.5)$. The nominal levels are chosen to be 0.05 and 0.1. According to the nonparametric Wilks theorem, the critical values are decided by the quantiles of a chi-squared distribution with 2 degrees of freedom. Besides the global inference, we also use the profile empirical likelihood ratio $-2\log\max_{\beta_2}\mathcal{R}_{IF}(\beta_1, \beta_2)$ to test for $\mathsf{H}_0^* : \beta_1 = -1$ against $\mathsf{H}_1^* : \beta_1 \neq -1$. The critical values are decided by the quantiles of a chi-squared distribution with 1 degree of freedom. 5,000 replicates are generated. The simulation results are summarized in Table 3.

Again, the proposed empirical likelihood ratio tests give out appropriate estimated sizes which are quite close to the corresponding nominal levels. The censoring percentage does not affect the performances of the proposed tests. With sample size increasing, the coverage errors decrease as expected.

For Example IV, we consider a linear model with a single covariate. The covariate is generated from uniform distribution $[0,2]$. The intercept parameter $\beta_0 = 1$ and the slop parameter $\beta_1 = 1$. The random error ε_i's follow normal distribution with mean 0 and variance 0.5. The censoring time C_i's are independently drawn from normal distribution with mean μ and variance 5, where μ is chosen to yield censoring percentages of 30% and 50%. The sample size n is set to be 100 and 200. We first use the proposed $-2\log\mathcal{R}_{IF}(\beta_0, \beta_1)$ to test for $\mathsf{H}_0 : (\beta_0, \beta_1) = (1, 1)$ against $\mathsf{H}_0 : (\beta_0, \beta_1) \neq (1, 1)$. The nominal levels are chosen to be 0.05 and 0.1. Similar
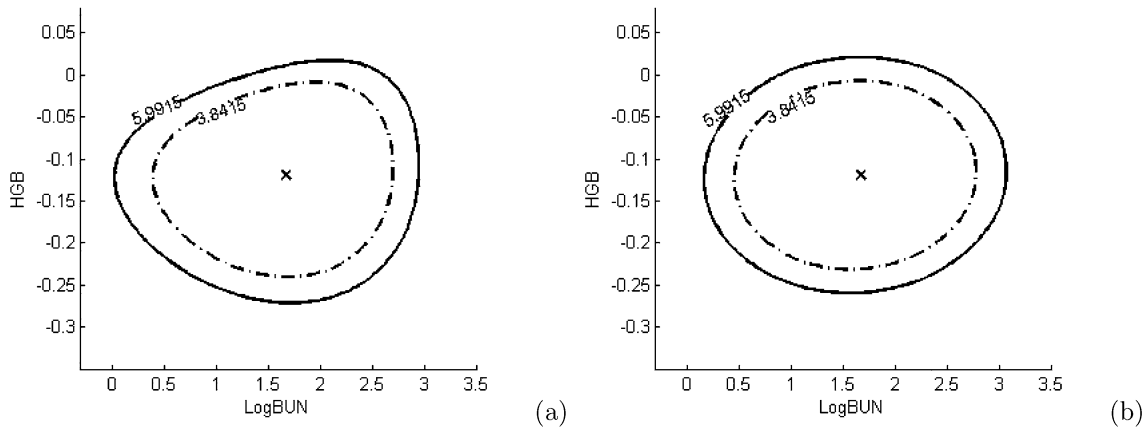
Figure 1. Confidence regions for the multiple myeloma study.

Table 3. Summarized simulation results for Example III

censoring percentage	n	H_0 v.s. H_1		H_0^* v.s. H_1^*	
		0.05	0.1	0.05	0.1
30%	100	0.0640	0.1182	0.0528	0.1052
	200	0.0598	0.1110	0.0538	0.1042
50%	100	0.0540	0.1064	0.0524	0.1006
	200	0.0500	0.1068	0.0484	0.0986

Table 4. Summarized simulation results for Example IV

censoring percentage	n	H_0 v.s. H_1		H_0^* v.s. H_1^*	
		0.05	0.1	0.05	0.1
30%	100	0.0588	0.1110	0.0560	0.1082
	200	0.0514	0.1052	0.0520	0.1000
50%	100	0.0628	0.1184	0.0530	0.1088
	200	0.0556	0.1070	0.0510	0.1062

to the previous simulation study, the critical values here are decided by the quantiles of a chi-squared distribution with 2 degrees of freedom. We also use the profile empirical likelihood ratio $-2 \log \max_{\beta_0} \mathcal{R}_{IF}(\beta_0, \beta_1)$ to test for $\mathsf{H}_0^* : \beta_1 = 1$ against $\mathsf{H}_1^* : \beta_1 \neq 1$. The critical values are decided by the quantiles of a chi-squared distribution with 1 degree of freedom. 5,000 replicates are generated. The simulation results are summarized in Table 4.

In general, we obtain similar observations from Table 4 to those from Table 3. It is interesting to point out that with different choices of the unbiased modification, the proposed empirical likelihood approach may provide an optimal linear combination asymptotically and thus result in more efficient estimators for the regression parameters.

4.2 A real example

Here we apply the proposed empirical likelihood method for Cox model to a study on multiple myeloma reported by [3]. In the study, 65 patients were treated with alkylating agents. 48 of them died during the study and 17 sur-

vived. This data is regarded as a main example in the PROC PHREG of the SAS/STAT User's Guide (1999, pp. 2608–2617, 2536–2641) to focus on Cox model with the logarithm of blood urea nitrogen, LogBUN and haemoglobin, HGB, as the covariates. Here we still used these two covariates and made inferences about the two corresponding regression coefficients. Based on the proposed log-empirical likelihood ratio, a confidence region for the two coefficients is obtained. The contour of the region is shown in Figure 1 (a). The outer loop corresponds to 95% confidence level. The center of the confidence region (marked by X), at which the empirical likelihood ratio attains 1, is just the maximum partial likelihood estimates $(1.6744, -0.1190)$ for the two coefficients. We also obtain the Wald-type confidence region which is shown in Figure 1 (b) with outer loop representing 95% level. Note that different from the Wald-type confidence region, the empirical likelihood based confidence region does not possess elliptical shape for the contour. The shape is decided by the data themselves, which is one of the features of empirical likelihood methods.

The empirical likelihood based confidence intervals for each coefficient could be obtained as the left (right, upper or lower) most points of the contours decided by quantiles of χ_1^2 distribution (the inner loops correspond to 0.95 quantile). The confidence intervals corresponding to LogBUN and HGB are $[0.3868, 2.7038]$ and $[-0.2404, -0.0082]$, respectively. The Wald-type counterparts are $[0.4747, 2.8741]$ and $[-0.2317, -0.0063]$. The two methods provide similar inference conclusions, except that the empirical likelihood based intervals are no longer symmetric about the maximum partial likelihood estimates.

5. CONCLUSIONS

We propose a new idea to develop empirical likelihood ratios for parameters of interest when there exist nuisance parameters in estimating functions. The constraints in empirical likelihood functions are constructed based on the influence functions for parameters of interest, instead of the

estimating functions. The resulting log-empirical likelihood ratio statistics are shown to possess chi-squared limiting distributions under suitable conditions. The property enables people to make inferences without the need to estimate variances explicitly. When the nuisance parameters involved are of finite dimension, the proposed approach is asymptotic equivalent to the profile empirical likelihood method. Moreover, the new approach can be applied to deal with infinite dimensional nuisance parameters under some semiparametric models. The standard algorithm developed for original empirical likelihood ratios can be used here without any difficulty.

Although the proposed methods show potential usage under some semiparametric models, no unified theory has been established for this influence function based empirical likelihood approach. We realize that there are some drawbacks of the proposed method. For example, sometimes it is very difficult to obtain an explicit form of the influence functions, and in other cases, the unknown quantities in the influence functions are difficult to estimate. Moreover, it is also of interest that what is the relationship between the convergence speed of the estimators for nuisance parameters and the nonparametric Wilks property. All these may become topics of our future communications.

APPENDIX

To prove the asymptotic properties in Section 2, we assume that the following regularity conditions hold.

$C1.$ When $\nu \neq \nu_0$, $\mathsf{E}[m(Z; \nu)] \neq 0$, i.e., ν_0 is identifiable.

$C2.$ $m(Z; \nu)$ has continuous first-order derivative with respect to ν.

$C3.$ There exist a neighborhood V of ν_0 and an integrable function $M(Z)$ with $\mathsf{E}[M(Z)] < \infty$ such that

$$\sup_{\nu \in V} \|m(Z; \nu)\|^3 \leq M(Z), \ \sup_{\nu \in V} \left\| \frac{\partial m(Z; \nu)}{\partial \nu^T} \right\| \leq M(Z).$$

$C4.$ A is non-degenerate and $\mathsf{E}[m(Z; \nu_0)m(Z; \nu_0)^T]$ is positively definite.

A.1 Proof of Theorem 2.1

We first prove the following lemmas.

Lemma A.1. *Under the conditions C1–C4, we have that* $\max_{1 \leq i \leq n} \|g(Z_i; \theta_0)\| = o_p(\sqrt{n})$.

Proof. Under the conditions C2 and C3, we have that $\sup_{\nu \in V} \|m(Z; \nu)\|^2$ is integrable. By arguments similar to those of the proof of Lemma 11.2 in [9], one can show that $\sup_{\nu \in V} \max_{1 \leq i \leq n} \|m(Z_i; \nu)\| = o_p(\sqrt{n})$. Define

$$g_0(Z; \theta) = m_1(Z; \theta, \tilde{\gamma}) - A_{12}A_{22}^{-1}m_2(Z; \theta, \tilde{\gamma}).$$

Since $\tilde{\gamma}$ is consistent, $(\theta_0^T, \tilde{\gamma}^T)^T$ belongs to V with probability tending to 1 as $n \to \infty$. By the condition C3,

$\|A_{12}A_{22}^{-1}\| < \infty$. Combing the above facts, we have that $\max_{1 \leq i \leq n} \|g_0(Z_i; \theta_0)\| = o_p(\sqrt{n})$. For $i = 1, \ldots, n$, define

$$r_i = g_0(Z_i; \theta_0) - g(Z_i; \theta_0)$$
$$= (\hat{A}_{12}\hat{A}_{22}^{-1} - A_{12}A_{22}^{-1})m_2(Z_i; \theta_0, \tilde{\gamma}).$$

By the consistency of $\hat{A}_{12}$ and $\hat{A}_{22}$, we have that

$$\max_{1 \leq i \leq n} \|r_i\| \leq \|\hat{A}_{12}\hat{A}_{22}^{-1} - A_{12}A_{22}^{-1}\| \max_{1 \leq i \leq n} \|m_2(Z_i; \theta_0, \tilde{\gamma})\|$$
$$= o_p(\sqrt{n}).$$

Then the conclusion follows immediately. □

Lemma A.2. *Under the conditions C1–C4, we have that* $n^{-1} \sum_{i=1}^{n} g(Z_i; \theta_0)g(Z_i; \theta_0)^T$ *converges in probability to* Σ *as* $n \to \infty$, *where*

$$\Sigma = M_{11} - M_{12}A_{22}^{-1}A_{12}^T - A_{12}A_{22}^{-1}M_{21}$$
$$+ A_{12}A_{22}^{-1}M_{22}A_{22}^{-1}A_{12}^T$$

with $M_{ij} = \mathsf{E}[m_i(Z; \theta_0, \gamma_0)m_j(Z; \theta_0, \gamma_0)^T]$, $i, j = 1, 2$.

Proof. By the condition C3 and the uniform law of large number, we have that $n^{-1} \sum_{i=1}^{n} m(Z_i; \nu)m(Z_i; \nu)^T$ converges in probability to $\mathsf{E}[m(Z; \nu)m(Z; \nu)^T]$ uniformly about $\nu \in V$ as $n \to \infty$. Thus, it can be shown that $n^{-1} \sum_{i=1}^{n} g_0(Z_i; \theta_0)g_0(Z_i; \theta_0)^T$ converges in probability to Σ. Simple algebra gives out

$$(8) \quad \frac{1}{n} \sum_{i=1}^{n} g(Z_i; \theta_0)g(Z_i; \theta_0)^T$$
$$= \frac{1}{n} \sum_{i=1}^{n} g_0(Z_i; \theta_0)g_0(Z_i; \theta_0)^T + \frac{1}{n} \sum_{i=1}^{n} g_0(Z_i; \theta_0)r_i^T$$
$$+ \frac{1}{n} \sum_{i=1}^{n} r_i g_0(Z_i; \theta_0)^T - \frac{1}{n} \sum_{i=1}^{n} r_i r_i^T.$$

By the consistency of $\hat{A}_{12}$ and $\hat{A}_{22}$, the last three terms on the right-hand-side of (8) are all of order $o_p(1)$. Thus, the conclusion follows immediately. □

Lemma A.3. *Under the conditions C1–C4, we have that* $n^{-1/2} \sum_{i=1}^{n} g(Z_i; \theta_0)$ *converges in distribution to* $N(0, \Sigma)$ *as* $n \to \infty$.

Proof. By Taylor expansion, we have that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_0(Z_i; \theta_0)$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[m_1(Z_i; \theta_0, \gamma_0) - A_{12}A_{22}^{-1}m_2(Z_i; \theta_0, \gamma_0) \right]$$
$$+ \frac{1}{n} \sum_{i=1}^{n} \left[\frac{\partial m_1(Z_i; \theta_0, \gamma^*)}{\partial \gamma^T} - A_{12}A_{22}^{-1} \frac{\partial m_2(Z_i; \theta_0, \gamma^*)}{\partial \gamma^T} \right]$$
$$\times \sqrt{n}(\tilde{\gamma} - \gamma_0),$$

where $\|\gamma^* - \gamma_0\| \leq \|\tilde{\gamma} - \gamma_0\|$. By the condition C3 and the law of large number, it can be shown that

$$\frac{1}{n} \sum_{i=1}^{n} \left[\frac{\partial m_1(Z_i; \theta_0, \gamma^*)}{\partial \gamma^T} - A_{12}A_{22}^{-1} \frac{\partial m_2(Z_i; \theta_0, \gamma^*)}{\partial \gamma^T} \right]$$

converges in probability to $A_{12} - A_{12}A_{22}^{-1}A_{22} = 0$. Meanwhile, by the central limit theorem, it can be shown that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[m_1(Z_i; \theta_0, \gamma_0) - A_{12}A_{22}^{-1} m_2(Z_i; \theta_0, \gamma_0) \right]$$

converges in distribution to $N(0, \Sigma)$ as $n \to \infty$. Thus, we have that $n^{-1/2} \sum_{i=1}^{n} g_0(Z_i; \theta_0)$ converges in distribution to $N(0, \Sigma)$ as $n \to \infty$.

On the other hand, simple algebra yields

$$(9) \quad \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(Z_i; \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_0(Z_i; \theta_0)$$
$$- \left(\hat{A}_{12}\hat{A}_{22}^{-1} - A_{12}A_{22}^{-1} \right) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m_2(Z_i; \theta_0, \tilde{\gamma}).$$

Again, by Taylor expansion, we have that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} m_2(Z_i; \theta_0, \tilde{\gamma})$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m_2(Z_i; \theta_0, \gamma_0) + \frac{1}{n} \sum_{i=1}^{n} \frac{\partial m_2(Z_i; \theta_0, \gamma^*)}{\partial \gamma^T}$$
$$\times \sqrt{n}(\tilde{\gamma} - \gamma_0) = O_p(1).$$

Thus, the second term on the right-hand-side of (9) is of order $o_p(1)$, which implies that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(Z_i; \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_0(Z_i; \theta_0) + o_p(1).$$

Then the conclusion follows immediately. □

Now we are ready to prove Theorem 2.1.

Proof. Since $\mathsf{E}[m_1(Z; \theta_0, \gamma_0) - A_{12}A_{22}^{-1}m_2(Z; \theta_0, \gamma_0)] = 0$, by arguments similar to those of the proof of Lemma 11.2 in [9], one can show that 0 has probability tending to 1 to fall into the convex hull of $(g(Z_1; \theta_0), \ldots, g(Z_n; \theta_0))^T$. When 0 is in the convex hull, by Lagrange multipliers it can be obtained that

$$(10) \quad \mathcal{R}_{IF}(\theta_0) = \prod_{i=1}^{n} \frac{1}{1 + \eta(\theta_0)^T g(Z_i; \theta_0)},$$

where $\eta(\theta_0)$ satisfies the following equation

$$(11) \quad \frac{1}{n} \sum_{i=1}^{n} \frac{g(Z_i; \theta_0)}{1 + \eta(\theta_0)^T g(Z_i; \theta_0)} = 0.$$

Based on (11), by arguments similar to those of the proof of Theorem 3.2 in [9], one can show that $\|\eta(\theta_0)\| = O_p(n^{-1/2})$. By this order, we derive from (11) that

$$(12) \quad \eta(\theta_0) = \Sigma_n^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} g(Z_i; \theta_0) \right) + o_p(1),$$

where $\Sigma_n = n^{-1} \sum_{i=1}^{n} g(Z_i; \theta_0)g(Z_i; \theta_0)^T$. Now take the logarithm of the right-hand-side of (10) and do Taylor expansion at $\eta = 0$, we get that

$$-2 \log \mathcal{R}_{IF}(\theta_0) = \sum_{i=1}^{n} \eta(\theta_0)^T g(Z_i; \theta_0)$$
$$- \frac{1}{2} \sum_{i=1}^{n} \eta(\theta_0)^T g(Z_i; \theta_0)g(Z_i; \theta_0)^T \eta(\theta_0)$$
$$+ O_p \left(\|\eta(\theta_0)\|^3 \sum_{i=1}^{n} \frac{\|g(Z_i; \theta_0)\|^3}{(1 + \tilde{\eta}^T g(Z_i; \theta_0))^3} \right),$$

where $\|\tilde{\eta}\| \leq \|\eta(\theta_0)\|$. By the condition C3 and Lemma A.1, we can conclude that

$$n\|\eta(\theta_0)\|^3 \times \frac{1}{n} \sum_{i=1}^{n} \frac{\|g(Z_i; \theta_0)\|^3}{(1 + \tilde{\eta}^T g(Z_i; \theta_0))^3} = o_p(1).$$

Now replacing $\eta(\theta_0)$ by the right-hand-side of (12), we have that

$$(13) \quad -2 \log \mathcal{R}_{IF}(\theta_0)$$
$$= \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(Z_i; \theta_0) \right)^T \Sigma_n^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(Z_i; \theta_0) \right)$$
$$+ o_p(1).$$

Combining (13), Lemmas A.2 and A.3, the desired conclusion follows immediately. □

A.2 The asymptotic equivalence of $\mathcal{R}_{IF}$ and $\mathcal{R}_P$

We show the asymptotic equivalence of the proposed method and profile empirical likelihood method under the following situation. Consider the hypothesis testing problem $\mathsf{H}_0 : \theta_0 = \theta$ v.s. $\mathsf{H}_1 : \theta_0 \neq \theta$, where θ is a specific value in $\mathbb{R}^p$. Let $f_0(z)$ be a density (or probability mass) function for the population Z satisfying (1), H_0 and the conditions C1–C4. Let $f_1(z)$ be a density (or probability mass) function satisfying (1) and H_1. Define the likelihood ratio $L_n = \prod_{i=1}^{n} f_1(Z_i) / \prod_{i=1}^{n} f_0(Z_i)$ and $S_n(\nu_0) = n^{-1/2} \sum_{i=1}^{n} m(Z_i; \nu_0)$, where $\nu_0 = \nu(f_0(z))$.

Suppose that when Z_i's follow $f_0(z)$, for any constant vector $c \in \mathbb{R}^r$, $c^T S_n(\nu_0)$ and $\log L_n$ are jointly asymptoti-

cally normal with proper parameters as $n \to \infty$. Then by LeCam's first lemma, $f_1(z)$ is contiguous to $f_0(z)$. From (13), we can show that under $f_0(z)$, i.e., $\theta_0 = \theta$,

$$-2 \log \mathcal{R}_{IF}(\theta)$$
$$= \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n g(Z_i; \theta) \right)^T \Sigma^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n g(Z_i; \theta) \right) + o_p(1).$$

By definition,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n g(Z_i; \theta) = A_1^T \frac{1}{\sqrt{n}} \sum_{i=1}^n m(Z_i; \theta, \gamma_0) + o_p(1),$$

where

$$A_1 = \begin{pmatrix} I \\ -A_{12} A_{22}^{-1} \end{pmatrix},$$

with I being a $p \times p$ identity matrix. Combining the last two equations yields

$$(14) \quad -2 \log \mathcal{R}_{IF}(\theta)$$
$$= \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n m(Z_i; \theta, \gamma_0) \right)^T U_1 \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n m(Z_i; \theta, \gamma_0) \right)$$
$$+ o_p(1),$$

where $U_1 = A_1^T \Sigma^{-1} A_1$. On the other hand, by the results from [13], we can show that under $f_0(z)$,

$$(15) \quad -2 \log \mathcal{R}_P(\theta)$$
$$= \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n m(Z_i; \theta, \gamma_0) \right)^T U_2 \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n m(Z_i; \theta, \gamma_0) \right)$$
$$+ o_p(1),$$

where $U_2 = M^{-1} - M^{-1} A_2 [A_2^T M^{-1} A_2]^{-1} A_2^T M^{-1}$ with

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}, \qquad A_2 = \begin{pmatrix} A_{12} \\ A_{22} \end{pmatrix}.$$

Careful calculation gives out that $U_1 = U_2$. Since $f_1(z)$ is contiguous to $f_0(z)$, (14) and (15) still hold when Z_i's follow $f_1(z)$. Moreover, by LeCam's third lemma, $n^{-1/2} \sum_{i=1}^n m(Z_i; \theta, \gamma_0)$ is asymptotically normally distributed when Z_i's follow $f_1(z)$. Thus, under $f_1(z)$, i.e., under contiguous alternative, $-2 \log \mathcal{R}_{IF}(\theta)$ and $-2 \log \mathcal{R}_P(\theta)$ have the same asymptotic distribution which is given by a non-central chi-squared distribution with p degrees of freedom and a certain non-central parameter. It implies that the two log-empirical likelihood ratio test statistics possess the same asymptotic power.

REFERENCES

[1] ANDERSON, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10** 1100–1120. MR0673646

[2] KOUL, H., SUSARLA, V. and VAN RYZIN, J. (1981). Regression analysis with randomly right-censord data. *Ann. Statist.* **9** 1276–1288. MR0630110

[3] KRALL, J. M., UTHOFF, V. A. and HARLEY, J. B. (1975). A step-up procedure for selecting variables associated with survival. *Biometrics* **31** 49–57.

[4] LAI, T. L., YING, Z. and ZHENG, Z. K. (1995). Asymptotic normality of a class of adaptive statistics with applications to synthetic data methods for censored regression. *J. Multivariate Anal.* **52** 259–279. MR1323333

[5] LI, G. and WANG, Q.-H. (2003). Empirical likelihood regression analysis for right censored data. *Statist. Sinica* **13** 51–68. MR1963919

[6] LU, W. and LIANG, Y. (2006). Empirical likelihood inference for linear transformation models. *J. Multivariate Anal.* **97** 1586–1599. MR2256232

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

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

[9] OWEN, A. B. (2001). *Empirical Likelihood.* Chapman and Hall, London.

[10] QIN, G. and JING, B.-Y. (2001a). Empirical likelihood for censored linear regression. *Scand. J. Stat.* **28** 661–673. MR1876506

[11] QIN, G. and JING, B.-Y. (2001b). Empirical likelihood for Cox regression model under random censorship. *Comm. Statist. Simulation Comput.* **30** 79–90. MR1862090

[12] QIN, G. and TSAO, M. (2003). Empirical likelihod inference for median regression models for censored survival data. *J. Multivariate Anal.* **85** 416–430. MR1983805

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

[14] SHI, J. and LAU, T. S. (2000). Empirical likelihood for partially linear models. *J. Multivariate Anal.* **72** 132–148. MR1747427

[15] THOMAS, D. R. and GRUNKEMEIER, G. L. (1975). Confidence interval estimation of survival probabilities for censored data. *J. Amer. Statist. Assoc.* **70** 865–871. MR0405766

[16] TSIATIS, A. A. (2006). *Semiparametric Theory and Missing Data.* Springer, New York. MR2233926

[17] WANG, Q.-H. and JING, B.-Y. (2003). Empirical likelihood for partial linear models. *Ann. Inst. Statist. Math.* **55** 585–595. MR2007800

[18] WANG, Q.-H. and RAO, J. N. K. (2002a). Empirical likelihood-based inference in linear models with missing data. *Scand. J. Stat.* **29** 563–576. MR1925575

[19] WANG, Q.-H. and RAO, J. N. K. (2002b). Empirical likelihood-based inference under imputation for missing response data. *Ann. Statist.* **30** 896–924. MR1922545

[20] WILKS, S. S. (1938). The large sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* **9** 60.

[21] XUE, L. and ZHU, L. (2006). Empirical likelihood for single-index models. *J. Multivariate Anal.* **97** 1295–1312. MR2279674

[22] YU, W. (2010). Empirical likelihood method for general additive-multiplicative hazard models. *Comm. Statist. Theory Methods* **39** 2977–2990. MR2755538

[23] YU, W., SUN, Y. and ZHENG, M. (2011). Empirical likelihood method for linear transformation models. *Ann. Inst. Statist. Math.* **63** 331–346.

[24] ZHAO, Y. (2010). Semiparametric inference for transformation models via empirical likelihood, *J. Multivariate Anal.* **101** 1846–1858. MR2651960

[25] ZHENG, M. and YU, W. (2011). Empirical likelihood method for the multivariate accelerated failure time models. *J. Statist. Plann. Inference.* **141** 972–983. MR2732965

[26] ZHU, L. and XUE, L. (2006). Empirical likelihood confidence regions in a partially linear single-index model. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 549–570. MR2278341

Ming Zheng
Department of Statistics, School of Management
Fudan University
China
E-mail address: mingzheng@fudan.edu.cn

Ziqiang Zhao
Department of Statistics, School of Management
Fudan University
China
E-mail address: johnzhao2835@gmail.com

Wen Yu
Department of Statistics, School of Management
Fudan University
China
E-mail address: wenyu@fudan.edu.cn