Biometrika (2017), **104**, 1, *pp.* 165–179
Printed in Great Britain

doi: 10.1093/biomet/asw072
Advance Access publication 18 February 2017

On pseudolikelihood inference for semiparametric models with boundary problems

By Y. Chen

Department of Biostatistics and Epidemiology, University of Pennsylvania, 210 Blockley Hall, 423 Guardian Drive, Philadelphia, Pennsylvania 19104, U.S.A.

ychen123@upenn.edu

J. Ning

Department of Biostatistics, The University of Texas MD Anderson Cancer Center, 1400 Pressler St, Unit 1411, FCT4.6006, Houston, Texas 77030, U.S.A.

jning@mdanderson.org

Y. Ning

Department of Statistical Science, Cornell University, Comstock Hall 1188, Ithaca, New York 14853, U.S.A.

yn265@cornell.edu

K.-Y. Liang

Department of Life Sciences, National Yang-Ming University, Taipei City 112, Taiwan

kyliang@ym.edu.tw

and K. Bandeen-Roche

Department of Biostatistics, Johns Hopkins University, 615 N. Wolfe St., Baltimore, Maryland 21205, U.S.A.

kbandeen@jhsph.edu

Summary

Consider a semiparametric model indexed by a Euclidean parameter of interest and an infinite-dimensional nuisance parameter. In many applications, pseudolikelihood provides a convenient way to infer the parameter of interest, where the nuisance parameter is replaced by a consistent estimator. The purpose of this paper is to establish the asymptotic behaviour of the pseudolikelihood ratio statistic under semiparametric models. In particular, we consider testing the hypothesis that the parameter of interest lies on the boundary of its parameter space. Under regularity conditions, we establish the equivalence between the asymptotic distributions of the pseudolikelihood ratio statistic and a likelihood ratio statistic for a normal mean problem with a misspecified covariance matrix. This result holds when the nuisance parameter is estimated at a rate slower than the usual rate in parametric models. We study three examples in which the asymptotic distributions are shown to be mixtures of

© 2017 Biometrika Trust

chi-squared variables. We conduct simulation studies to examine the finite-sample performance of the pseudolikelihood ratio test.

Some key words: Likelihood ratio test; Multivariate survival model; Pseudolikelihood; Semiparametric model.

1. INTRODUCTION

Consider a semiparametric model indexed by two parameters: a parameter of interest $\theta \in \Omega \subset \mathbb{R}^d$ and a nuisance parameter ϕ lying in a Banach space $\mathcal{H}$ with a norm $\|\cdot\|$. Semiparametric models have been widely used in a variety of settings; see for example Gu & Zhang (1993), Murphy (1995), Huang (1996) and Cheng (2009). Asymptotic theory for semiparametric maximum likelihood estimation can be found in Bickel et al. (1993), van der Vaart (2000) and Kosorok (2008). The consistency of the bootstrap for M-estimators was established by Dixon et al. (2005) and Cheng & Huang (2010). Semiparametric likelihood ratio inference based on the profile likelihood has been developed by Murphy & van der Vaart (1997, 2000) and Banerjee (2005). Chen et al. (2014) studied the local identification of nonparametric and semiparametric models.

Pseudolikelihood provides an approach to inference on θ in the presence of the nuisance parameter ϕ (Gong & Samaniego, 1981). The key idea is that the inference for θ can be based on $L^*(\theta) = L(\theta, \hat{\phi})$, where $\hat{\phi}$ is a consistent estimator of ϕ and $L(\theta, \phi)$ is the loglikelihood. Unlike the profile likelihood, which estimates the nuisance parameter ϕ by $\hat{\phi}(\theta) = \arg\max_{\phi \in \mathcal{H}} L(\theta, \phi)$, the pseudolikelihood is constructed by substituting a consistent estimator $\hat{\phi}$ that is free of θ, so the information equality does not hold (Gong & Samaniego, 1981). When the nuisance parameter ϕ is of finite dimension, under certain regularity conditions Liang & Self (1996) derived the asymptotic distribution of the pseudolikelihood ratio test for $\theta = \theta_0$. One of the regularity conditions is that θ_0 must be an interior point of its parameter space, but in many applications θ_0 lies on the boundary of the parameter space. For parametric models, this boundary problem has been studied by Chernoff (1954), Kudo (1963), Chant (1974), Shapiro (1985), Self & Liang (1987) and Chen & Liang (2010). In particular, Chen & Liang (2010) derived the asymptotic distribution of the pseudolikelihood ratio test statistic for boundary problems when the nuisance parameter is of finite dimension. However, to the best of our knowledge, there is no systematic theoretical study of the boundary problem under semiparametric models.

The primary purpose of this paper is to develop a general theory on pseudolikelihood ratio inference for semiparametric models in cases where the parameter of interest may lie on the boundary of the parameter space. In a similar spirit to that of the profile likelihood (Murphy & van der Vaart, 2000), the theoretical justification for the pseudolikelihood in semiparametric models is more difficult than in Gong & Samaniego (1981), Liang & Self (1996) and Chen & Liang (2010), for the following reasons. First, the estimator of the nuisance parameter ϕ may converge at a rate slower than the usual rate in parametric models. Second, unlike in parametric models, standard Taylor expansions cannot be used to deal with the remainder terms in likelihood expansions. To overcome these challenges, we establish our main results using empirical processes. Under certain regularity conditions, our main results cover cases in which the nuisance parameter is estimated with a rate slower than $n^{1/2}$. In addition, the sensitivity of the likelihood to the nuisance parameter is characterized by the Fréchet derivative (Bickel et al., 1993). We establish a general theorem on the asymptotic distribution of the pseudolikelihood ratio test for $\theta = \theta_0$, which allows θ_0 to lie on the boundary of its parameter space. The general theory is verified and illustrated by copula and nested copula models for survival data and by weighted likelihoods for missing data. We show that the pseudolikelihood ratio test performs well in simulation studies, while the naive test that ignores the boundary problem has a conservative Type I error rate and much lower power.

2. Examples and main results

2·1. *Examples*

Example 1. Suppose that C_θ is a distribution function with density c_θ on $[0, 1]^2$ for some θ in $\mathbb{R}$. Let $(\tilde{Y}_1, \tilde{Y}_2)$ denote the paired failure times, and let (S_1, S_2) and (f_1, f_2) denote the corresponding marginal survival functions and density functions, respectively. Assuming that $(\tilde{Y}_1, \tilde{Y}_2)$ comes from the C_θ copula, the joint survival function and density function of $(\tilde{Y}_1, \tilde{Y}_2)$ are

$$S_\theta(y_1, y_2) = C_\theta\{S_1(y_1), S_2(y_2)\},$$
$$f_\theta(y_1, y_2) = c_\theta\{S_1(y_1), S_2(y_2)\} f_1(y_1) f_2(y_2) \quad (y_1, y_2 \geqslant 0).$$

Let (C_1, C_2) denote paired censoring times. For $i = 1, \ldots, n$, assume that $(\tilde{Y}_{1i}, \tilde{Y}_{2i})$ and (C_{1i}, C_{2i}) are independent data. For each i, we observe $Y_{ji} = \tilde{Y}_{ji} \wedge C_{ji}$ and $\delta_{ji} = I(\tilde{Y}_{ji} \leqslant C_{ji})$ for $j = 1, 2$.

For concreteness, we consider the Clayton copula model (Clayton, 1978), defined by

$$C_\theta(u, v) = (u^{1-\theta} + v^{1-\theta} - 1)^{1/(1-\theta)}$$

for $\theta > 1$ and $C_\theta(u, v) = uv$ for $\theta = 1$. Given n paired data $(Y_{1i}, Y_{2i}), \ldots, (Y_{ni}, Y_{ni})$, write $\{S_1(Y_{1i}), S_2(Y_{2i})\}$ as (u_i, v_i) for notational simplicity. The loglikelihood function can be written as

$$L(\theta, S_1, S_2) = \sum_{i=1}^{n} \Bigg[\delta_{1i}\delta_{2i} \log c_\theta(u_i, v_i) + \delta_{1i}(1 - \delta_{2i}) \log \left\{ \frac{\partial}{\partial u} C_\theta(u_i, v_i) \right\}$$
$$+ (1 - \delta_{1i})\delta_{2i} \log \left\{ \frac{\partial}{\partial v} C_\theta(u_i, v_i) \right\} + (1 - \delta_{1i})(1 - \delta_{2i}) \log C_\theta(u_i, v_i) \Bigg],$$

where

$$\frac{\partial C_\theta(u, v)}{\partial u} = u^{-\theta}(u^{1-\theta} + v^{1-\theta} - 1)^{1/(1-\theta)-1},$$
$$\frac{\partial C_\theta(u, v)}{\partial v} = v^{-\theta}(u^{1-\theta} + v^{1-\theta} - 1)^{1/(1-\theta)-1},$$

$c_\theta(u, v) = \theta u^{-\theta} v^{-\theta} (u^{1-\theta} + v^{1-\theta} - 1)^{1/(1-\theta)-2}$, and θ characterizes the association between the paired failure times $(\tilde{Y}_1, \tilde{Y}_2)$.

In practice, the marginal survival functions S_1 and S_2 are unknown nuisance parameters. To make inference on θ, Shih & Louis (1995) proposed a log-pseudolikelihood function $L^*(\theta) = L(\theta, \hat{S}_1, \hat{S}_2)$, where $\hat{S}_1$ and $\hat{S}_2$ are Kaplan–Meier estimators of S_1 and S_2. In association analysis for bivariate survival times, a typical hypothesis of interest is $H_0 : \theta = 1$, i.e., no association between two failure times. The null hypothesis $\theta = 1$ lies on the boundary of the parameter space $\Omega = [1, \infty)$. However, this hypothesis testing problem with the boundary constraint is not covered by the theory of Shih & Louis (1995).

Example 2. Bandeen-Roche & Liang (1996) proposed a class of models for failure time data that accounts for multiple levels of clustering. For simplicity of notation, we consider a cluster of two levels, such as households and villages. Here, for illustration, we assume that there are three individual members and two households that are clustered as $\{1\}$ and $\{2, 3\}$. When the Clayton copula is used to model the failure times $(\tilde{Y}_1, \tilde{Y}_2, \tilde{Y}_3)$, its joint survival function is

$$S_\theta(y_1,y_2,y_3) = \left[S_1(y_1)^{1-\theta_2} + \left\{S_2(y_2)^{1-\theta_1} + S_3(y_3)^{1-\theta_1} - 1\right\}^{(\theta_2-1)/(\theta_1-1)} - 1\right]^{1/(1-\theta_2)},$$

where $(y_1,y_2,y_3) \in \mathbb{R}_+^3$ with $\mathbb{R}_+ = (0,\infty)$, $S_1(y_1)$, $S_2(y_2)$ and $S_3(y_3)$ are the marginal survival functions, θ_1 characterizes the association within the same household, and θ_2 characterizes the association between two individuals from different households in the same village. To ensure that $S_\theta(y_1,y_2,y_3)$ is nonnegative, Bandeen-Roche & Liang (1996) required that $\theta_1 \geqslant \theta_2 \geqslant 1$.

Let n denote the number of villages. For $i = 1,\ldots,n$, we observe $Y_{ji} = \tilde{Y}_{ji} \wedge C_{ji}$ and $\delta_{ji} = I(\tilde{Y}_{ji} \leqslant C_{ji})$ $(j = 1,2,3)$, where (C_1, C_2, C_3) denote censoring times. Based on the observed data, we can specify the likelihood function, the form of which is provided in the Supplementary Material. Similar to Example 1, Bandeen-Roche & Liang (1996) proposed a pseudolikelihood approach for inference on θ. One hypothesis of interest is $H_0 : \theta_1 = \theta_2 = 1$, i.e., no association among all failure times within the same village. In this example, H_0 is on the boundary of the parameter space $\Omega = \{(\theta_1, \theta_2) \in \mathbb{R}^2 : \theta_1 \geqslant \theta_2 \geqslant 1\}$.

Example 3. Assume that a study involves the collection of independent and identically distributed observations (Y_i, X_i) $(i = 1,\ldots,n)$, where Y_i is the outcome and X_i is a d-dimensional auxiliary covariate. Let $f(y;\theta)$ denote the density function of Y_i, which is indexed by a finite-dimensional parameter θ. Let V_i denote a binary missing data indicator, with $V_i = 1$ if Y_i is observable and $V_i = 0$ if Y_i is missing. Our goal is to estimate θ based on the observed data (V_iY_i, V_i, X_i) $(i = 1,\ldots,n)$. Assume that Y_i is independent of V_i given X_i, known in the literature as missingness at random. Under this assumption, an inverse probability weighting estimator for θ is derived by solving a set of weighted estimating equations (Robins et al., 1994; Scharfstein et al., 1999). Under the likelihood framework, an equivalent estimator can be obtained by maximizing the weighted loglikelihood function

$$L(\theta,\pi) = \sum_{i=1}^{n} \frac{V_i}{\pi(X_i)} \log f(Y_i;\theta),$$

where $\pi(X_i) = \mathrm{pr}(V_i = 1 \mid X_i)$ is the probability of observing Y_i given covariates. The function $\pi(\cdot)$ is often an unknown infinite-dimensional nuisance parameter and must be estimated. When X_i is low-dimensional, one can estimate π by the Nadaraya–Watson estimator $\hat{\pi}$ (Nadaraya, 1964; Watson, 1964). The inference on θ can be based on the weighted pseudolikelihood function $L^*(\theta) = L(\theta, \hat{\pi})$. Suppose that θ is univariate and we are interested in the null hypothesis $H_0 : \theta = 0$. If there is prior knowledge that $\theta \geqslant 0$, then $\theta = 0$ is on the boundary of the parameter space $\Omega = [0, \infty)$.

2·2. *Main results*

In the following theoretical development, we consider a general setting. Given independent and identically distributed observations $(O_1,\ldots,O_n)$, let $L(\theta,\phi) = \sum_{i=1}^n m(\theta,\phi)(O_i)$ denote a generic objective function. For instance, $L(\theta, \phi)$ reduces to the loglikelihood function if $m(\theta, \phi)(O_i)$ is the log probability density function of the data O_i; see Examples 1 and 2. The term $m(\theta, \phi)(O_i)$ can also be a weighted loglikelihood function; see Example 3. Let $\hat{\phi}$ denote an estimator of the nuisance parameter ϕ in the Banach space $\mathcal{H}$. Define $L^*(\theta) = L(\theta, \hat{\phi})$. We refer to $L^*(\theta)$ as the pseudolikelihood, even though $L(\theta, \phi)$ can be an objective function other than the loglikelihood. The parameter space of θ is denoted by $\Omega \subset \mathbb{R}^d$. Let $\hat{\theta} = \arg\max_{\theta \in \Omega} L^*(\theta)$ denote the maximum pseudolikelihood estimator. The pseudolikelihood ratio statistic T for the null hypothesis $H_0 : \theta = \theta_0$ versus the alternative hypothesis $H_1 : \theta \in \Omega - \{\theta_0\}$ is defined as

$$T = 2\left\{\sup_{\theta \in \Omega} L^*(\theta) - L^*(\theta_0)\right\}.$$

To handle the nonparametric component ϕ, we introduce the following submodel notation (Kosorok, 2008). For any fixed ϕ, let $\phi_t \in \mathcal{H}$ be a smooth curve running through ϕ at $t = 0$. The loglikelihood for the parametric submodel indexed by (θ, t) is $L(\theta, \phi_t)$. Let $a = (\partial/\partial t)\phi_t|_{t=0}$ be the direction vector in a tangent set for the nuisance parameter. Write

$$m_1(\theta, \phi) = \frac{\partial}{\partial \theta} m(\theta, \phi), \qquad m_2(\theta, \phi)[a] = \frac{\partial}{\partial t} m(\theta, \phi_t)\Big|_{t=0},$$
$$m_{11}(\theta, \phi) = \frac{\partial^2}{\partial \theta \partial \theta^{\mathrm{T}}} m(\theta, \phi), \qquad m_{12}(\theta, \phi)[a] = \frac{\partial^2}{\partial \theta \partial t} m(\theta, \phi_t)\Big|_{t=0}.$$

Given a measurable function g, we write $\mathbb{P}_n g = n^{-1} \sum_{i=1}^n g(O_i)$ and write $\mathbb{P} g = \int g \, \mathrm{d}\mathbb{P}$ for the expectation of g. We use $|\cdot|$ to denote the L_2-norm in the Euclidean space, and $\|\cdot\|$ to denote the norm in the Banach space $\mathcal{H}$. Let ϕ_0 denote the true value of ϕ. We assume the following regularity conditions.

Condition 1. There exists some $c_1 > 0$ such that

$$|\hat{\theta} - \theta_0| = o_{\mathrm{p}}(1), \quad \|\hat{\phi} - \phi_0\| = O_{\mathrm{p}}(n^{-c_1}).$$

Condition 2. For any $\delta_n \to 0$, any $\theta \in \Omega$ and some $D > 0$,

$$\sup_{|\theta - \theta_0| \leqslant \delta_n,\, \|\phi - \phi_0\| \leqslant D n^{-c_1}} \left| n^{1/2}(\mathbb{P}_n - \mathbb{P})\{m_1(\theta, \phi) - m_1(\theta_0, \phi_0)\} \right| = o_{\mathrm{p}}(1).$$

Condition 3. For some $c_2 > 1$ satisfying $c_1 c_2 > 1/2$ and any $\|\phi - \phi_0\| \leqslant D n^{-c_1}$,

$$\left| \mathbb{P}\{m_1(\theta_0, \phi) - m_1(\theta_0, \phi_0) - m_{12}(\theta_0, \phi_0)(\phi - \phi_0)\} \right| = O(\|\phi - \phi_0\|^{c_2}).$$

Condition 4. As $n \to \infty$, $n^{1/2}\mathbb{P}_n m_1(\theta_0, \phi_0)$ and $\mathbb{P} m_{12}(\theta_0, \phi_0)\{n^{1/2}(\hat{\phi} - \phi_0)\}$ jointly converge in distribution to $N(0, \Sigma)$, where Σ is a positive-definite matrix and can be partitioned as Σ_{11}, Σ_{12} and Σ_{22} accordingly.

Condition 5. The information and covariance matrices

$$I_{11} = -\mathbb{P} m_{11}(\theta_0, \phi_0), \quad I_{11}^* = \Sigma_{11} + 2\Sigma_{12} + \Sigma_{22}$$

are positive definite.

Condition 6. For any $\delta_n \to 0$, any $\theta \in \Omega$ and some $D > 0$,

$$\sup_{|\theta-\theta_0|\leqslant\delta_n,\, \|\phi-\phi_0\|\leqslant Dn^{-c_1}} \left|\mathbb{P}_n m_{11}(\theta,\phi) - \mathbb{P} m_{11}(\theta_0,\phi_0)\right| = o_{\mathrm{p}}(1).$$

A major difference between the above regularity conditions and those in the semiparametric literature is that we do not require θ_0 to be interior to its parameter space. Moreover, in contrast to the efficient Fisher information matrix in profile likelihood estimation (Murphy & van der Vaart, 2000), for pseudolikelihood the information matrix I_{11} may not be identical to I_{11}^*, where I_{11}^* is the covariance of the pseudo-score function $n^{1/2}\mathbb{P}_n m_1(\theta_0,\phi_0) + \mathbb{P} m_{12}(\theta_0,\phi_0)\{n^{1/2}(\hat\phi-\phi_0)\}$. This is also one of the major differences between the profile likelihood and the pseudolikelihood.

Our Conditions 1–6 are imposed on the model and the estimators instead of on the parameter space. These conditions are similar to the regularity conditions in the literature; see, for example, Cheng & Huang (2010). Specifically, Conditions 1–3 are similar to Conditions S1–S3 in Cheng & Huang (2010) for studying M-estimators. The consistency of $\hat\theta$ in Condition 1 can be established by the M-estimator theory in Kosorok (2008, § 14.2). The rate of convergence of $\hat\phi$ must be established case by case. A sufficient condition for Condition 2 is that the class of functions $\{m_1(\theta,\phi) : |\theta-\theta_0| \leqslant \delta_1, \|\phi-\phi_0\| \leqslant \delta_2\}$ is Donsker for some $\delta_1,\delta_2 > 0$ with a square-integrable envelope function. Condition 3 essentially requires that $m_1(\theta_0,\phi)$ be Fréchet differentiable at ϕ_0 (Bickel et al., 1993, p. 455), and it holds with $c_2 = 2$ if the second derivatives of $m_1(\theta_0,\phi)$ with respect to ϕ are bounded in a neighbourhood of ϕ_0. In this case, our pseudolikelihood estimation can accommodate a convergence rate of the nonparametric component that is slower than $n^{1/2}$, provided that $c_1 > 1/4$. This extends the scope of the pseudolikelihood theory in Gong & Samaniego (1981). Condition 4 usually holds automatically if $\hat\phi$ is $n^{1/2}$-consistent, but may need additional work for estimators with convergence rates slower than $n^{1/2}$ (Wong & Severini, 1991). A general strategy for verifying Condition 4 is to exploit the concrete form of $\hat\phi$ and find the influence function of $\mathbb{P} m_{12}(\theta_0,\phi_0)\{n^{1/2}(\hat\phi-\phi_0)\}$. We illustrate such derivations in Example 3; see the Supplementary Material for technical details. Condition 5 requires I_{11} and I_{11}^* to be positive definite. A similar sufficient condition for Condition 6 to hold is that the class of functions $\{m_{11}(\theta,\phi) : |\theta-\theta_0| \leqslant \delta_1, \|\phi-\phi_0\| \leqslant \delta_2\}$ should be Glivenko–Cantelli with an integrable envelope function.

Assume that Ω is a convex set, as in our examples. The following lemma provides a general quadratic approximation of $2n\mathbb{P}_n\{m(\theta,\hat\phi) - m(\theta_0,\hat\phi)\}$.

LEMMA 1. *Let* $W(\theta) = -n^{1/2}(\theta-\theta_0) + I_{11}^{-1}U(\theta_0)$, *where*

$$U(\theta_0) = n^{1/2}\mathbb{P}_n m_1(\theta_0,\phi_0) + \mathbb{P} m_{12}(\theta_0,\phi_0)\{n^{1/2}(\hat\phi-\phi_0)\}.$$

If Conditions 1–6 *hold, then for any* $\theta \in \Omega$ *such that* $\theta-\theta_0 = o(1)$, *as* $n \to \infty$ *we have*

$$2n\,\mathbb{P}_n\{m(\theta,\hat\phi) - m(\theta_0,\hat\phi)\} = -W(\theta)^{\mathrm{T}} I_{11} W(\theta) + U(\theta_0)^{\mathrm{T}} I_{11}^{-1} U(\theta_0) + o_{\mathrm{p}}(1 + n^{1/2}|\theta-\theta_0|)^2.$$

A proof is provided in the Appendix. To use this lemma to derive the asymptotic distribution of the pseudolikelihood ratio test T, we need to characterize the geometry of Ω. To this end, we introduce the definition of an approximating cone (Chernoff, 1954; Shapiro, 1985; Self & Liang, 1987). Recall that a cone $C(\theta_0)$ with vertex at θ_0 is a set of points such that if $x \in C(\theta_0)$, then $a(x-\theta_0)+\theta_0 \in C(\theta_0)$ for any $a \geqslant 0$.

DEFINITION 1. *The set Ω is approximated at θ_0 by a cone with vertex at θ_0, referred to as the approximating cone $C_\Omega(\theta_0)$, if*

$$\inf_{x \in C_\Omega(\theta_0)} |x - y| = o(|y - \theta_0|), \quad y \in \Omega, \qquad \inf_{y \in \Omega} |x - y| = o(|x - \theta_0|), \quad x \in C_\Omega(\theta_0).$$

For instance, if $\Omega = [1, \infty)$ and $\theta_0 = 1$, then $C_\Omega(\theta_0) = [1, \infty)$. Similarly, a sphere in $\mathbb{R}^2$ can be approximated at a boundary point by a half-plane tangent to the sphere at that point. In addition, if θ_0 is an interior point of $\Omega \subset \mathbb{R}^d$, then $C_\Omega(\theta_0) = \mathbb{R}^d$. See Chernoff (1954), Shapiro (1985) and Self & Liang (1987) for more examples. Using Lemma 1, we derive the limiting distribution of the pseudolikelihood ratio statistic T under the null hypothesis.

THEOREM 1. *Assume that Conditions 1–6 hold and the parameter space Ω is approximated at θ_0 by a cone $C_\Omega(\theta_0)$. Then the pseudolikelihood ratio statistic T converges weakly to*

$$T(Z) = Z^{\mathrm{T}} I_{11} Z - \inf_{h \in C_\Omega(0)} \{(Z - h)^{\mathrm{T}} I_{11} (Z - h)\} \tag{1}$$

as $n \to \infty$, where $Z \sim N(0, I_{11}^{-1} I_{11}^* I_{11}^{-1})$ and $C_\Omega(0)$ is a cone obtained by translating $C_\Omega(\theta_0)$ so that its vertex lies at the origin. There are two scenarios as follows.

(i) *If θ_0 is an interior point of Ω, then $T(Z)$ reduces to $T(Z) = Z^{\mathrm{T}} I_{11} Z$ for $Z \sim N(0, I_{11}^{-1} I_{11}^* I_{11}^{-1})$, and $T(Z)$ is distributed as a weighted sum of d independent χ_1^2 variables with weights the eigenvalues of $I_{11}^* I_{11}^{-1}$.*
(ii) *If θ_0 is a boundary point of Ω, then the distribution of $T(Z)$ depends on the shape of $C_\Omega(0)$, and is generally a mixture of distributions as described by Chen & Liang (2010, Lemma 2).*

A proof is given in the Appendix. This theorem shows that the asymptotic distribution of T is the same as that of the likelihood ratio statistic for a normal mean problem with a misspecified covariance. To see why, consider a random variable Z from the distribution $N(0, I_{11}^{-1} I_{11}^* I_{11}^{-1})$. If we incorrectly assume that Z follows $N(h, I_{11}^{-1})$ with known covariance matrix I_{11}^{-1} but unknown mean vector $h \in C_\Omega(0)$, the corresponding loglikelihood ratio test for testing $h = 0$ based on only one observation of Z follows the same distribution as $T(Z)$.

The main difficulties when applying the general results in Theorem 1 are two-fold. First, for boundary problems the approximating cone $C_\Omega(\theta_0)$ has to be determined case by case. In § 2·3, we give its form and the calculation of $T(Z)$ for the examples of § 2·1. Further examples can be found in Self & Liang (1987) and Chen & Liang (2010). Second, the calculation of I_{11}^* is more challenging than in Chen & Liang (2010) due to the presence of a nuisance parameter of infinite dimension. To address this issue, the following corollary provides an explicit formula for calculating I_{11}^*. In addition, this corollary deals with important special cases where the asymptotic distribution of the pseudolikelihood ratio statistic T is simplified.

COROLLARY 1. *Assume that the conditions of Theorem 1 hold and there exists a zero-mean function $\alpha(\theta_0, \phi_0)(O_i)$ such that*

$$\mathbb{P} m_{12}(\theta_0, \phi_0)\{n^{1/2}(\hat{\phi} - \phi_0)\} = n^{1/2} \mathbb{P}_n \alpha(\theta_0, \phi_0) + o_{\mathrm{p}}(1).$$

Then, as $n \to \infty$, the asymptotic distribution of the pseudolikelihood ratio statistic T is the same as the distribution of $T(Z)$ defined in (1), where $Z \sim N(0, I_{11}^{-1} I_{11}^* I_{11}^{-1})$ with

$$I_{11}^* = \operatorname{cov}\{m_1(\theta_0, \phi_0)(O_i)\} + 2 \operatorname{cov}\{m_1(\theta_0, \phi_0)(O_i), \alpha(\theta_0, \phi_0)(O_i)\} + \operatorname{cov}\{\alpha(\theta_0, \phi_0)(O_i)\}.$$

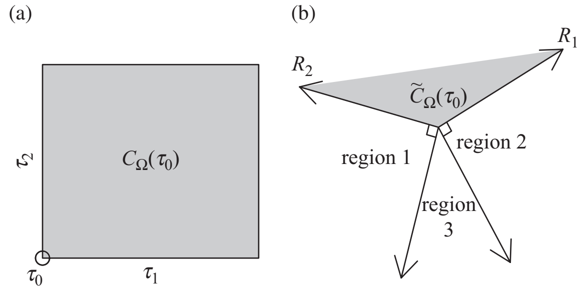

Fig. 1. The partitions of parameter space considered in Example 2: (a) the parameter space for $\tau = (\tau_1, \tau_2)$, with the shaded region representing admissible parameter values; (b) the transformed parameter space, where the shaded region $\tilde{C}_\Omega(\tau_0)$ represents admissible parameter values. The asymptotic distribution of T is a mixture of χ^2_2, χ^2_1 and χ^2_0 distributions, with mixing probabilities depending on the angles in $\tilde{C}_\Omega(\tau_0)$.

In addition, if $\mathbb{P}[\{2m_1(\theta_0, \phi_0) + \alpha(\theta_0, \phi_0)\}\alpha^{\mathrm{T}}(\theta_0, \phi_0)] = 0$ *and* $\mathbb{P}\{m_1(\theta_0, \phi_0)m_1^{\mathrm{T}}(\theta_0, \phi_0)\} = I_{11}$*, we have* $I^*_{11} = I_{11}$ *and therefore* $Z \sim N(0, I_{11}^{-1})$.

The proof follows directly from Theorem 1. When $I^*_{11} = I_{11}$, T has the same limiting distribution of the likelihood ratio statistic as it would have if the nuisance parameter were known, and the naive test that ignores the boundary constraints by comparing the pseudolikelihood ratio test statistic with χ^2_d always leads to conservative Type I error rates and a loss of power.

2·3. *Examples revisited*

Example 1. Recall that the null hypothesis $\theta = 1$ is on the boundary of the parameter space $\Omega = [1, \infty)$. Then $C_\Omega(0) = [0, \infty)$. Following the calculation in Shih & Louis (1995, p. 1389), the conditions in Corollary 1 hold and we obtain $I^*_{11} = I_{11}$. Therefore, equation (1) reduces to

$$T(Z) = Z^2 I_{11} - Z^2 I(Z < 0) I_{11} = Z^2 I(Z > 0) I_{11},$$

where $Z \sim N(0, I_{11}^{-1})$. Thus, the asymptotic distribution of T is a mixture of χ^2_0 and χ^2_1 with mixing probabilities 0·5 and 0·5, where χ^2_0 is a point mass at 0. Additional details on the verification of Conditions 1–6 are given in the Supplementary Material.

Example 2. Denote the parameter value under the null hypothesis by $\theta_0 = (1, 1)^{\mathrm{T}}$. The approximating cone in this case is $C_\Omega(\theta_0) = \{(t_1, t_2) : t_1 \geqslant t_2 \geqslant 0\}$. For ease of derivation, a simple reparameterization from $\theta = (\theta_1, \theta_2)^{\mathrm{T}}$ to $\tau = (\tau_1, \tau_2)^{\mathrm{T}}$, where $\tau_1 = \theta_2 - 1$ and $\tau_2 = \theta_1 - \theta_2$, yields the approximating cone $C_\Omega(\tau_0) = [0, \infty) \times [0, \infty)$ with $\tau_0 = (0, 0)^{\mathrm{T}}$, which is illustrated in Fig. 1(a).

By arguments similar to those in Shih & Louis (1995), equation (1) reduces to

$$T(Z_\tau) = Z_\tau^{\mathrm{T}} I_{\tau\tau} Z_\tau - \inf_{\tau \in [0,\infty)\times[0,\infty)} (Z_\tau - \tau)^{\mathrm{T}} I_{\tau\tau} (Z_\tau - \tau),$$

where $Z_\tau \sim N(0, I_{\tau\tau}^{-1})$ and $I_{\tau\tau} = E\{-\partial^2 \log f(O_i; \tau_0, \phi_0)/\partial\tau^2\}$. Here $\phi_0 = (S_{10}, S_{20}, S_{30})$ are the true values of (S_1, S_2, S_3), $O_i = (Y_{i1}, Y_{i2}, Y_{i3})$, and $f(\cdot)$ is the density function of O_i. Let

$I_{\tau\tau} = R^{\mathrm{T}}R$, where R is a 2×2 nonsingular matrix, and write $\tilde{C}_\Omega(\tau_0) = \{\tilde{\tau} : \tilde{\tau} = R\tau$ for any $\tau \in C_\Omega(\tau_0)\}$ and $\tilde{Z}_\tau = RZ_\tau$. Then $T(Z_\tau)$ can be rewritten as

$$T(\tilde{Z}_\tau) = |\tilde{Z}_\tau|^2 - \inf_{\tilde{\tau} \in \tilde{C}_\Omega(\tau_0)} \left|\tilde{Z}_\tau - \tilde{\tau}\right|^2. \tag{2}$$

The calculation of the second term in (2) depends on the location of $\tilde{Z}_\tau$ relative to the boundary of $\tilde{C}_\Omega(\tau_0)$. Four different regions must be considered separately, as illustrated in Fig. 1(b): the shaded region represents $\tilde{C}_\Omega(\tau_0)$; the angle in the shaded area is less than 180°, since the convexity of $C_\Omega(\tau_0)$ is preserved under the linear transformation $\tau \to R\tau$.

Denote the columns of R by R_1 and R_2, and denote the inner product of vectors a and b by $\langle a, b\rangle = a^{\mathrm{T}}b$. Then (2), namely the asymptotic distribution of T, can be written as

$$T(\tilde{Z}_\tau) = \begin{cases} |\tilde{Z}_\tau|^2 \sim \chi_2^2, & \tilde{Z}_\tau \in \tilde{C}_\Omega(\tau_0), \\ \left(\dfrac{\langle \tilde{Z}_\tau, R_2\rangle}{|R_2|}\right)^2 \sim \chi_1^2, & \tilde{Z}_\tau \in \text{region } 1, \\ \left(\dfrac{\langle \tilde{Z}_\tau, R_1\rangle}{|R_1|}\right)^2 \sim \chi_1^2, & \tilde{Z}_\tau \in \text{region } 2, \\ 0, & \tilde{Z}_\tau \in \text{region } 3. \end{cases}$$

Since the distribution of $\tilde{Z}_\tau$ is symmetric about the origin, the probabilities of $\tilde{Z}_\tau$ being from certain regions are completely determined by the angles of these regions (Chernoff, 1954). The mixing probability for the shaded region is

$$p_s = \cos^{-1}\left[\left\{I_{\tau\tau}^{(1,1)} I_{\tau\tau}^{(2,2)}\right\}^{-1/2} I_{\tau\tau}^{(1,2)}\right] \Big/ (2\pi), \tag{3}$$

where $I_{\tau\tau}^{(i,j)}$ is the (i,j) element of the 2×2 matrix $I_{\tau\tau}$. Therefore, under $H_0 : \theta_1 = \theta_2 = 1$, the asymptotic distribution of the pseudolikelihood ratio test T is a mixture of χ_2^2, χ_1^2 and χ_0^2 distributions with mixing probabilities p_s, 0·5 and $0{\cdot}5 - p_s$, respectively.

Example 3. We consider the cases $\Omega = [0, \infty)$ and $\Omega = \mathbb{R}$. In the former, the approximating cone is $C_\Omega(0) = [0, \infty)$. So (1) reduces to $T(Z) = Z^2 I_{11} - Z^2 I(Z < 0) I_{11} = Z^2 I(Z > 0) I_{11}$, where $Z \sim N(0, I_{11}^{-1} I_{11}^{*} I_{11}^{-1})$. The asymptotic distribution of T is a mixture of χ_0^2 and $I_{11}^{*} I_{11}^{-1} \chi_1^2$ distributions with mixing probabilities 0·5 and 0·5. In the latter case, $\theta = 0$ is an interior point and $C_\Omega(0) = \mathbb{R}$. Hence, the asymptotic distribution of T is weighted chi-squared, $I_{11}^{*} I_{11}^{-1} \chi_1^2$. Unlike the previous two examples, in which the nuisance parameter is estimated at an $n^{1/2}$ rate, the Nadaraya–Watson estimator $\hat{\pi}$ in this example has a slower rate. The explicit forms of I_{11}^{*} and I_{11} and details on verifying Conditions 1–6 are presented in the Supplementary Material.

3. Simulations

We conducted simulation studies in the settings of Examples 1 and 2. We first applied the pseudolikelihood ratio test for the dependence between bivariate survival times in Example 1. To generate the paired failure times, we used the rmvdc function in the R package copula (Yan, 2007; R Development Core Team, 2017). The marginal distributions of the failure times were

Table 1. *Empirical rejection rates (%) of the pseudolikelihood ratio test for testing association between bivariate survival times in Example 1 over 5000 replications*

		Censoring % = 0%		Censoring % = 15%		Censoring % = 30%	
n	θ	$\alpha = 5\%$ Rejection	$\alpha = 1\%$ Rejection	$\alpha = 5\%$ Rejection	$\alpha = 1\%$ Rejection	$\alpha = 5\%$ Rejection	$\alpha = 1\%$ Rejection
100	1·0	5·5	1·3	5·4	1·1	5·4	1·1
	1·3	36·8	16·6	34·6	14·6	33·5	14·1
	1·5	65·2	39·3	62·9	36·5	60·2	34·2
	1·7	84·4	65·0	82·3	61·2	80·2	58·4
	2·0	97·1	88·6	95·9	86·1	94·7	83·7
200	1·0	5·4	1·0	5·5	1·0	5·2	1·1
	1·3	55·8	29·6	55·5	29·5	55·5	28·6
	1·5	87·0	68·6	86·3	67·4	85·5	66·7
	1·7	97·7	91·6	97·4	90·1	96·8	89·3
	2·0	99·9	99·5	99·9	99·1	99·8	98·7
400	1·0	5·2	1·1	5·1	1·0	5·4	1·3
	1·3	77·2	53·7	76·0	52·8	76·1	52·2
	1·5	98·6	93·3	98·1	91·6	97·8	91·6
	1·7	100·0	99·8	99·9	99·6	99·9	99·4
	2·0	100·0	100·0	100·0	100·0	100·0	100·0

Weibull with shape parameter 2 and unit scale parameter. Independent censoring times were generated from uniform distributions on $(0, 5{\cdot}4)$ and $(0, 2{\cdot}7)$, corresponding to 15% and 30% censoring. To evaluate the size of the test, we drew bivariate failure times from a Clayton copula with $\theta = 1$, corresponding to an independence scenario. To evaluate the power of the tests, we implemented a similar procedure but set θ to 1·3, 1·5, 1·7 and 2·0. We set the number of pairs at 100, 200 or 400. For each generated dataset, we compared the pseudolikelihood ratio statistic with the $0{\cdot}5\chi_0^2 + 0{\cdot}5\chi_1^2$ distribution. Table 1 shows the estimated levels of Type I error and power from 5000 replications of the test. When the null hypothesis is true, the rejection rates of the pseudolikelihood ratio test were all within 95% confidence intervals for the nominal levels, i.e., 0·7–1·3% for nominal level 1% and 4·4–5·6% for nominal level 5%. A plot of the quantiles of the test statistics against those of the asymptotic distribution indicates that the latter works well at levels other than 5% and 1%; see the Supplementary Material. We also compared the test statistics with a χ_1^2 distribution as if the boundary constraint were ignored. This naive test was too conservative under all scenarios considered; see the Supplementary Material. The power of the pseudolikelihood ratio test increased with increasing values of the association parameter θ. At the nominal level of 5%, the pseudolikelihood ratio test had about 80% power when θ was 1·7 at a sample size of 100, when θ was 1·5 at a sample size of 200, and when θ was 1·3 at a sample size of 400. The power slightly decreased with increased censoring; thus censoring has a relatively small effect on the power.

In the second simulation study, we tested for associations among all failure times within the same village in the model of Example 2. To generate the multivariate failure times with the hierarchical structure, we used the R package nacopula (Hofert & Mächler, 2011). The marginal distributions of the failure times were chosen to be standard exponential. Independent censoring times were generated from uniform distributions to have censoring percentages of 15% and 30%. For each generated dataset, we compared the pseudolikelihood ratio test statistic with the $(0{\cdot}5 - \hat{p}_s)\chi_0^2 + 0{\cdot}5\chi_1^2 + \hat{p}_s\chi_2^2$ distribution, where $\hat{p}_s$ can be estimated empirically by (3), as described in § 2·3. Table 2 shows the estimated Type I error rates and power from 5000

Table 2. *Empirical rejection rates* (%) *of the pseudolikelihood ratio test for testing associations in the model of Example* 2 *over* 5000 *replications*

n	(θ_1, θ_2)	Censoring % = 0% $\alpha = 5\%$ Rejection	Censoring % = 0% $\alpha = 1\%$ Rejection	Censoring % = 15% $\alpha = 5\%$ Rejection	Censoring % = 15% $\alpha = 1\%$ Rejection	Censoring % = 30% $\alpha = 5\%$ Rejection	Censoring % = 30% $\alpha = 1\%$ Rejection
100	(1·0,1·0)	5·4	1·5	5·0	1·0	5·0	1·0
	(1·1, 1·1)	32·5	15·0	27·1	12·3	25·1	10·1
	(1·2, 1·1)	55·7	32·8	48·9	25·9	40·4	18·6
	(1·3, 1·1)	74·1	52·6	67·8	44·7	57·1	32·4
	(1·2, 1·2)	71·4	51·5	66·1	44·4	57·0	33·5
	(1·3, 1·2)	84·4	68·1	78·1	59·7	70·1	45·3
200	(1·0, 1·0)	5·0	1·5	4·9	0·9	4·6	1·0
	(1·1, 1·1)	53·7	32·3	45·7	25·2	37·7	18·5
	(1·2, 1·1)	81·8	62·6	74·7	51·7	62·5	37·8
	(1·3, 1·1)	95·1	86·1	90·6	76·0	82·4	61·6
	(1·2, 1·2)	92·5	85·1	88·1	75·6	80·0	61·6
	(1·3, 1·2)	98·5	95·4	96·5	90·3	91·8	79·5
400	(1·0, 1·0)	4·9	1·0	4·9	1·1	4·8	0·8
	(1·1, 1·1)	78·5	61·1	69·7	49·4	57·4	35·9
	(1·2, 1·1)	97·8	92·7	94·0	84·6	87·1	69·7
	(1·3, 1·1)	100·0	99·4	99·5	97·6	97·6	91·3
	(1·2, 1·2)	99·5	98·8	98·6	96·7	96·0	90·5
	(1·3, 1·2)	100·0	99·9	100·0	99·8	99·5	98·2

simulations. Similar to our findings in Example 1, the proposed test had sizes close to nominal, suggesting that the asymptotic approximation performs well. The naive test, which ignores the boundary problem, led to conservative Type I error rates and substantial loss of power; see the Supplementary Material.

4. Discussion

If the maximizer of the log-pseudolikelihood $L^*(\theta)$ does not have a closed-form solution, iterative algorithms are needed to maximize $L^*(\theta)$. Cheng (2013) provided a general algorithm for maximizing the log profile likelihood and established its rate of convergence. Unlike the algorithm in Cheng (2013), the pseudolikelihood approach does not require iterative updating of the nuisance parameter estimate, since $\hat{\phi}$ is free of θ by definition. It is of interest to extend the theoretical results on the algorithm in Cheng (2013) to pseudolikelihood estimation.

In regular statistical models, the likelihood ratio test is known to be asymptotically optimal, whereas the pseudolikelihood ratio test may lose efficiency due to the use of a generic estimator of the nuisance parameter. For a class of nonregular models in which the parameters are not identifiable under the null hypothesis, Song et al. (2009) proposed optimal tests based on the integrated profile likelihood. One future direction of research is to study the optimality of these tests when the parameter of interest lies on the boundary of the parameter space.

This paper focuses on the pseudolikelihood, which relies on the availability of a consistent estimator for the nuisance parameter that is free of the parameter of interest. In some situations, such a consistent estimator may not be available. In such cases, likelihood ratio inference can be considered. The theoretical results for the semiparametric likelihood ratio test

(Murphy & van der Vaart, 1997) with a boundary problem can be developed following an argument similar to that given in the present paper. We leave this work for future investigation.

ACKNOWLEDGEMENT

We thank the editor, the associate editor and three referees for their helpful comments, which have significantly improved the paper. The authors are grateful to Drs John Breitner and Peter Zandi for providing the Cache County data. This research was supported in part by the U.S. National Institutes of Health.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes more technical details of Examples 1–3, further results from the simulation studies, and a real-data example.

APPENDIX

Proof of Lemma 1

We apply a second-order Taylor expansion to the pseudolikelihood ratio:

$$
\begin{aligned}
&2n\mathbb{P}_n\{m(\theta,\hat{\phi}) - m(\theta_0,\hat{\phi})\} \\
&\quad = 2n^{1/2}\{n^{1/2}(\theta-\theta_0)\}^{\mathrm{T}}\mathbb{P}_n m_1(\theta_0,\hat{\phi}) + n^{1/2}(\theta-\theta_0)^{\mathrm{T}}\mathbb{P}_n m_{11}(\tilde{\theta},\hat{\phi})n^{1/2}(\theta-\theta_0),
\end{aligned} \tag{A1}
$$

where $\tilde{\theta} = \theta_0 + t(\theta - \theta_0)$ for some $t \in [0,1]$. By the convexity of Ω, we have $\tilde{\theta} \in \Omega$. By Condition 2,

$$n^{1/2}(\mathbb{P}_n - \mathbb{P})m_1(\theta_0,\hat{\phi}) = n^{1/2}(\mathbb{P}_n - \mathbb{P})m_1(\theta_0,\phi_0) + o_{\mathrm{p}}(1).$$

Using $\mathbb{P}m_1(\theta_0,\phi_0) = 0$ together with Condition 3, we obtain

$$
\begin{aligned}
n^{1/2}\mathbb{P}_n m_1(\theta_0,\hat{\phi}) &= n^{1/2}\mathbb{P}m_1(\theta_0,\hat{\phi}) + n^{1/2}\mathbb{P}_n m_1(\theta_0,\phi_0) + o_{\mathrm{p}}(1) \\
&= n^{1/2}\mathbb{P}_n m_1(\theta_0,\phi_0) + \mathbb{P}m_{12}(\theta_0,\phi_0)\{n^{1/2}(\hat{\phi}-\phi_0)\} \\
&\quad + O_{\mathrm{p}}(n^{1/2}\|\hat{\phi}-\phi_0\|^{c_2}) + o_{\mathrm{p}}(1).
\end{aligned} \tag{A2}
$$

Since $\|\hat{\phi}-\phi_0\| = O_{\mathrm{p}}(n^{-c_1})$ and $c_1c_2 > 1/2$, we have $n^{1/2}\|\hat{\phi}-\phi_0\|^{c_2} = o_{\mathrm{p}}(1)$. By Conditions 5 and 6,

$$\mathbb{P}_n m_{11}(\tilde{\theta},\hat{\phi}) = -I_{11} + o_{\mathrm{p}}(1). \tag{A3}$$

Combining (A1), (A2) and (A3), we obtain the following quadratic expansion of the likelihood ratio statistic:

$$
\begin{aligned}
2n\mathbb{P}_n\{m(\theta,\hat{\phi}) - m(\theta_0,\hat{\phi})\} &= 2\{n^{1/2}(\theta-\theta_0)\}^{\mathrm{T}}n^{1/2}\mathbb{P}_n m_1(\theta_0,\phi_0) \\
&\quad + 2\{n^{1/2}(\theta-\theta_0)\}^{\mathrm{T}}\mathbb{P}m_{12}(\theta_0,\phi_0)\{n^{1/2}(\hat{\phi}-\phi_0)\} \\
&\quad - n^{1/2}(\theta-\theta_0)^{\mathrm{T}}I_{11}n^{1/2}(\theta-\theta_0) + o_{\mathrm{p}}(1+n^{1/2}|\theta-\theta_0|)^2,
\end{aligned} \tag{A4}
$$

which is equivalent to

$$2n\mathbb{P}_n\{m(\theta,\hat{\phi}) - m(\theta_0,\hat{\phi})\} = -W(\theta)^{\mathrm{T}} I_{11} W(\theta) + U(\theta_0)^{\mathrm{T}} I_{11}^{-1} U(\theta_0) + o_{\mathrm{p}}(1 + n^{1/2}|\theta - \theta_0|)^2 \tag{A5}$$

where $W(\theta) = -n^{1/2}(\theta - \theta_0) + I_{11}^{-1} U(\theta_0)$ and $U(\theta_0) = n^{1/2}\mathbb{P}_n m_1(\theta_0,\phi_0) + \mathbb{P} m_{12}(\theta_0,\phi_0)\{n^{1/2}(\hat{\phi} - \phi_0)\}$. This completes the proof.

Proof of Theorem 1

For notational simplicity, let $\hat{h} = n^{1/2}(\hat{\theta} - \theta_0)$ and $U = n^{1/2}\mathbb{P}_n m_1(\theta_0,\phi_0) + \mathbb{P} m_{12}(\theta_0,\phi_0)\{n^{1/2}(\hat{\phi} - \phi_0)\}$. Replacing θ in (A4) by $\hat{\theta}$, we have

$$2n\mathbb{P}_n\{m(\hat{\theta},\hat{\phi}) - m(\theta_0,\hat{\phi})\} = 2\hat{h}^{\mathrm{T}} U - \hat{h}^{\mathrm{T}} I_{11} \hat{h} + o_{\mathrm{p}}(1 + |\hat{h}|)^2.$$

Since $\mathbb{P}_n\{m(\hat{\theta},\hat{\phi}) - m(\theta_0,\hat{\phi})\} \geqslant 0$, U is bounded in probability by Condition 4 and I_{11} is positive definite by Condition 5, we have

$$0 \leqslant K|\hat{h}| - K'|\hat{h}|^2 + o_{\mathrm{p}}(1 + |\hat{h}|)^2$$

for some positive constants K and K'. If $|\hat{h}| = o_{\mathrm{p}}(1)$, then $|\hat{h}| = O_{\mathrm{p}}(1)$. Otherwise, $1 + |\hat{h}|$ is of the order of $|\hat{h}|$, and we have $K'|\hat{h}|^2 \leqslant K|\hat{h}| + o_{\mathrm{p}}(|\hat{h}|^2)$. This implies that $|\hat{h}| = O_{\mathrm{p}}(1)$ and thus $\hat{h}$ is uniformly tight. Write $W_h = -h + I_{11}^{-1} U$. In the following, we shall prove that

$$2n\mathbb{P}_n\{m(\hat{\theta},\hat{\phi}) - m(\theta_0,\hat{\phi})\} - \sup_{h \in C_\Omega(0)} (-W_h^{\mathrm{T}} I_{11} W_h + U^{\mathrm{T}} I_{11}^{-1} U) = o_{\mathrm{p}}(1). \tag{A6}$$

By Condition 5, we have that U converges weakly to $N(0, I_{11}^*)$. By (A6) and the continuous mapping theorem,

$$\sup_{h \in C_\Omega(0)} (-W_h^{\mathrm{T}} I_{11} W_h + U^{\mathrm{T}} I_{11}^{-1} U) \to \sup_{h \in C_\Omega(0)} \left\{-(Z - h)^{\mathrm{T}} I_{11}(Z - h) + Z^{\mathrm{T}} I_{11} Z\right\}$$

as $n \to \infty$, where $Z \sim N(0, I_{11}^{-1} I_{11}^* I_{11}^{-1})$. Hence, Slutsky's theorem implies that the pseudolikelihood ratio test $2n\mathbb{P}_n\{m(\hat{\theta},\hat{\phi}) - m(\theta_0,\hat{\phi})\}$ converges weakly to $\sup_{h \in C_\Omega(0)}\{-(Z - h)^{\mathrm{T}} I_{11}(Z - h) + Z^{\mathrm{T}} I_{11} Z\}$.

It remains to show that (A6) holds. Since $\hat{\theta}$ is root-n consistent, (A5) gives

$$2n\mathbb{P}_n\{m(\hat{\theta},\hat{\phi}) - m(\theta_0,\hat{\phi})\} - \sup_{h \in \Omega_n} (-W_h^{\mathrm{T}} I_{11} W_h + U^{\mathrm{T}} I_{11}^{-1} U) = o_{\mathrm{p}}(1),$$

where $\Omega_n = \{n^{1/2}(\theta - \theta_0) : \theta \in \Omega\}$. Comparing with (A6), we only need to show that $\inf_{h \in \Omega_n} W_h^{\mathrm{T}} I_{11} W_h = \inf_{h \in C_\Omega(0)} W_h^{\mathrm{T}} I_{11} W_h + o_{\mathrm{p}}(1)$. Similar to the proof of root-n convergence of $\hat{\theta}$, we can show that the minimizer of $W_h^{\mathrm{T}} I_{11} W_h$ in Ω_n is bounded in probability. By the definition of Ω_n, for any $h \in \Omega_n$ with $|h| = O(1)$, there exists $\theta \in \Omega$ such that $h = n^{1/2}(\theta - \theta_0)$. By the definition of the approximating cone, there exists a sequence $\bar{\theta} \in C_\Omega(\theta_0)$ such that $|\bar{\theta} - \theta| = o(|\theta - \theta_0|) = o(n^{-1/2})$. Let $\bar{h} = n^{1/2}(\bar{\theta} - \theta_0)$. We have that $\bar{h}$ belongs to the cone $C_\Omega(0)$ and $|\bar{h} - h| = o(1)$. Then

$$\begin{aligned}(I_{11}^{-1}U - h) I_{11} (I_{11}^{-1}U - h) &= (I_{11}^{-1}U - \bar{h} + \bar{h} - h) I_{11} (I_{11}^{-1}U - \bar{h} + \bar{h} - h) \\ &\geqslant (I_{11}^{-1}U - \bar{h}) I_{11} (I_{11}^{-1}U - \bar{h}) - O_{\mathrm{p}}(|\bar{h} - h|) - O_{\mathrm{p}}(|\bar{h} - h|^2) \\ &= (I_{11}^{-1}U - \bar{h}) I_{11} (I_{11}^{-1}U - \bar{h}) + o_{\mathrm{p}}(1).\end{aligned}$$

Hence

$$\inf_{h\in\Omega_n}(I_{11}^{-1}U-h)I_{11}(I_{11}^{-1}U-h) \geqslant \inf_{h\in\Omega_n}(I_{11}^{-1}U-\bar{h})I_{11}(I_{11}^{-1}U-\bar{h})+o_{\mathrm{p}}(1)$$
$$=(I_{11}^{-1}U-\bar{h})I_{11}(I_{11}^{-1}U-\bar{h})+o_{\mathrm{p}}(1)$$
$$\geqslant \inf_{\bar{h}\in C_{\Omega}(0)}(I_{11}^{-1}U-\bar{h})I_{11}(I_{11}^{-1}U-\bar{h})+o_{\mathrm{p}}(1). \quad \text{(A7)}$$

Similarly, we can show that the minimizer of $W_h^{\mathrm{T}}I_{11}W_h$ in $C_\Omega(0)$ is bounded in probability. For any $\bar{h}\in C_\Omega(0)$ with $|\bar{h}|=O(1)$, there exists $\bar{\theta}\in C_\Omega(\theta_0)$ satisfying $\bar{h}=n^{1/2}(\bar{\theta}-\theta_0)$. By the definition of the approximating cone, there exists a sequence $\theta\in\Omega$ such that $|\bar{\theta}-\theta|=o(|\bar{\theta}-\theta_0|)=o(n^{-1/2})$. Let $h=n^{1/2}(\theta-\theta_0)$. It can be seen that $h\in\Omega_n$ and $|\bar{h}-h|=o(1)$. Following arguments similar to those leading to (A7), we can show that

$$\inf_{\bar{h}\in C_{\Omega}(0)}(I_{11}^{-1}U-\bar{h})I_{11}(I_{11}^{-1}U-\bar{h}) \geqslant \inf_{h\in\Omega_n}(I_{11}^{-1}U-h)I_{11}(I_{11}^{-1}U-h)+o_{\mathrm{p}}(1).$$

Together these results imply that

$$\inf_{\bar{h}\in C_{\Omega}(0)}(I_{11}^{-1}U-\bar{h})I_{11}(I_{11}^{-1}U-\bar{h}) = \inf_{h\in\Omega_n}(I_{11}^{-1}U-h)I_{11}(I_{11}^{-1}U-h)+o_{\mathrm{p}}(1).$$

This completes the proof of (1). When θ_0 is an interior point, (1) reduces to $T(Z)=Z^{\mathrm{T}}I_{11}Z$ upon taking $h=Z$, which is a weighted sum of d independent χ_1^2 variables with the weights being the eigenvalues of $I_{11}^{*}I_{11}^{-1}$ by Theorem 4.4.4 of Graybill (1976). When θ_0 is a boundary point of Ω, (1) has the same form as equation (2) in Chen & Liang (2010). Thus, the distribution of $T(Z)$ follows from Lemma 2 of Chen & Liang (2010).

REFERENCES

BANDEEN-ROCHE, K. & LIANG, K.-Y. (1996). Modelling failure-time associations in data with multiple levels of clustering. *Biometrika* **83**, 29–39.

BANERJEE, M. (2005). Likelihood ratio tests under local alternatives in regular semiparametric models. *Statist. Sinica* **15**, 635–44.

BICKEL, P. J., KLAASSEN, C., RITOV, Y. & WELLNER, J. (1993). *Efficient and Adaptive Inference in Semiparametric Models*. Baltimore, Maryland: Johns Hopkins University Press.

CHANT, D. (1974). On asymptotic tests of composite hypotheses in nonstandard conditions. *Biometrika* **61**, 291–8.

CHEN, X., CHERNOZHUKOV, V., LEE, S. & NEWEY, W. K. (2014). Local identification of nonparametric and semiparametric models. *Econometrica* **82**, 785–809.

CHEN, Y. & LIANG, K.-Y. (2010). On the asymptotic behaviour of the pseudolikelihood ratio test statistic with boundary problems. *Biometrika* **97**, 603–20.

CHENG, G. (2009). Semiparametric additive isotonic regression. *J. Statist. Plan. Infer.* **139**, 1980–91.

CHENG, G. (2013). How many iterations are sufficient for efficient semiparametric estimation? *Scand. J. Statist.* **40**, 592–618.

CHENG, G. & HUANG, J. Z. (2010). Bootstrap consistency for general semiparametric M-estimation. *Ann. Statist.* **38**, 2884–915.

CHERNOFF, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Statist.* **25**, 573–8.

CLAYTON, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141–51.

DIXON, J. R., KOSOROK, M. R. & LEE, B. L. (2005). Functional inference in semiparametric models using the piggyback bootstrap. *Ann. Inst. Statist. Math.* **57**, 255–77.

GONG, G. & SAMANIEGO, F. J. (1981). Pseudo maximum likelihood estimation: Theory and applications. *Ann. Statist.* **9**, 861–9.

GRAYBILL, F. (1976). *Theory and Application of the Linear Model*. North Scituate, Massachusetts: Duxbury Press.

GU, M. & ZHANG, C.-H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *Ann. Statist.* **21**, 611–24.

HOFERT, M. & MÄCHLER, M. (2011). Nested Archimedean copulas meet R: The nacopula package. *J. Statist. Software* **39**, 1–20.
HUANG, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.* **24**, 540–68.
KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer.
KUDO, A. (1963). A multivariate analogue of the one-sided test. *Biometrika* **50**, 403–18.
LIANG, K.-Y. & SELF, S. G. (1996). On the asymptotic behaviour of the pseudolikelihood ratio test statistic. *J. R. Statist. Soc.* B **58**, 785–96.
MURPHY, S. A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23**, 182–98.
MURPHY, S. A. & VAN DER VAART, A. W. (1997). Semiparametric likelihood ratio inference. *Ann. Statist.* **25**, 1471–509.
MURPHY, S. A. & VAN DER VAART, A. W. (2000). On profile likelihood. *J. Am. Statist. Assoc.* **95**, 449–65.
NADARAYA, E. A. (1964). On estimating regression. *Theory Prob. Appl.* **9**, 141–2.
R DEVELOPMENT CORE TEAM (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org.
ROBINS, J. M., ROTNITZKY, A. & ZHAO, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Am. Statist. Assoc.* **89**, 846–66.
SCHARFSTEIN, D. O., ROTNITZKY, A. & ROBINS, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J. Am. Statist. Assoc.* **94**, 1096–120.
SELF, S. G. & LIANG, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Statist. Assoc.* **82**, 605–10.
SHAPIRO, A. (1985). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika* **72**, 133–44.
SHIH, J. H. & LOUIS, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics* **51**, 1384–99.
SONG, R., KOSOROK, M. R. & FINE, J. P. (2009). On asymptotically optimal tests under loss of identifiability in semiparametric models. *Ann. Statist.* **37**, 2409–44.
VAN DER VAART, A. W. (2000). *Asymptotic Statistics*. New York: Cambridge University Press.
WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā* A **26**, 359–72.
WONG, W. H. & SEVERINI, T. A. (1991). On maximum likelihood estimation in infinite dimensional parameter spaces. *Ann. Statist.* **19**, 603–32.
YAN, J. (2007). Enjoy the joy of copulas: With a package copula. *J. Statist. Software* **21**, 1–21.

[*Received on* 6 *May* 2014. *Editorial decision on* 15 *October* 2016]

Biometrika (2012), **99**, 1, *pp*. 1–12
© 2012 Biometrika Trust
Printed in Great Britain
Advance Access publication on 31 July 2012

Web-based Supplementary Materials for "On pseudolikelihood inference for semiparametric models with boundary problems"

BY Y. CHEN

Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

ychen123@upenn.edu

J. NING

Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, U.S.A.

Y. NING

Department of Statistical Science, Cornell University, Ithaca, New York 14853, U.S.A.

K.-Y. LIANG

Department of Life Sciences, National Yang-Ming University, Taipei City 112, Taiwan

AND K. BANDEEN-ROCHE

Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland 21205, U.S.A.

SUMMARY

The Supplementary Materials contain more technical details on Examples 1, 2 and 3, more details on simulation studies, and a data example. It is organized as follows. The technical details on Examples 1, 2 and 3 are shown in Sections A, B, C. In particular, the regularity conditions C1–C6 for Examples 1 and 3 are verified in Sections A and C. Simulation results for the naive test are shown in Section D, and some additional simulation results comparing the empirical quantiles of the pseudolikelihood ratio test statistics and the theoretical quantiles are shown in Section E. Section F contains the analysis of a real data set.

A. TECHNICAL DETAILS ON EXAMPLE 1

Recall that in Example 1, we have the log likelihood

$$
\begin{aligned}
L(\theta, S_1, S_2) \ &= \sum_{i=1}^{n} \delta_{1i}\delta_{2i} \log c_\theta(u_i, v_i) + \delta_{1i}(1-\delta_{2i}) \log\left\{\frac{\partial}{\partial u} C_\theta(u_i, v_i)\right\} \\
&\quad + (1-\delta_{1i})\delta_{2i} \log\left\{\frac{\partial}{\partial v} C_\theta(u_i, v_i)\right\} + (1-\delta_{1i})(1-\delta_{2i}) \log C_\theta(u_i, v_i),
\end{aligned}
$$

where $\partial C_\theta(u,v)/\partial u = u^{-\theta}(u^{1-\theta} + v^{1-\theta} - 1)^{1/(1-\theta)-1}$, $\partial C_\theta(u,v)/\partial v = v^{-\theta}(u^{1-\theta} + v^{1-\theta} - 1)^{1/(1-\theta)-1}$, $c_\theta(u,v) = \theta u^{-\theta} v^{-\theta}(u^{1-\theta} + v^{1-\theta} - 1)^{1/(1-\theta)-2}$. After some tedious algebra, we can show that

$$\frac{\partial}{\partial\theta}\log c_\theta(u,v)=\frac{1}{\theta}-\log u-\log v+\frac{1}{(1-\theta)^2}\log(u^{1-\theta}+v^{1-\theta}-1)+\Big(2-\frac{1}{1-\theta}\Big)\frac{u^{1-\theta}\log u+v^{1-\theta}\log v}{u^{1-\theta}+v^{1-\theta}-1},$$

for $\theta>1$ and $1+\log u+\log v+\log u\log v$ for $\theta=1$,

$$\frac{\partial}{\partial\theta}\log\left\{\frac{\partial}{\partial u}C_\theta(u,v)\right\}=-\log u+\frac{1}{(1-\theta)^2}\log(u^{1-\theta}+v^{1-\theta}-1)+\Big(1-\frac{1}{1-\theta}\Big)\frac{u^{1-\theta}\log u+v^{1-\theta}\log v}{u^{1-\theta}+v^{1-\theta}-1},$$

for $\theta>1$ and $\log u\log v$ for $\theta=1$,

$$\frac{\partial}{\partial\theta}\log\left\{\frac{\partial}{\partial v}C_\theta(u,v)\right\}=-\log v+\frac{1}{(1-\theta)^2}\log(u^{1-\theta}+v^{1-\theta}-1)+\Big(1-\frac{1}{1-\theta}\Big)\frac{u^{1-\theta}\log u+v^{1-\theta}\log v}{u^{1-\theta}+v^{1-\theta}-1},$$

for $\theta>1$ and $\log u\log v$ for $\theta=1$, and finally

$$\frac{\partial}{\partial\theta}\log C_\theta(u,v)=\frac{1}{(1-\theta)^2}\log(u^{1-\theta}+v^{1-\theta}-1)-\frac{1}{1-\theta}\frac{u^{1-\theta}\log u+v^{1-\theta}\log v}{u^{1-\theta}+v^{1-\theta}-1},$$

for $\theta>1$ and $\log u\log v$ for $\theta=1$. Putting all pieces together, we can find the explicit form of $m(\theta,S_1,S_2)$ and $m_1(\theta,S_1,S_2)$, which are defined in Section 2.2 with $\phi=(S_1,S_2)$. In the following, we now check conditions C1–C6 under some mild regularity conditions. Assume that θ_0 is the unique maximizer of $\mathbb{P}L(\theta,S_{10},S_{20})$, where S_{10},S_{20} are the true values of S_1 and S_2. This is a standard identifiability condition in the literature. We also assume that the end of study time is τ and there exists a small constant $\delta>0$ such that $S_{10}(\tau)\geq\delta$ and $S_{20}(\tau)\geq\delta$. This is mainly for some technical reason, because in the log likelihood function we have $\log S_1(y)$ and $\log S_2(y)$, which diverge to infinity if $S_1(y)$ and $S_2(y)$ are close to 0. The norm for the nuisance function is taken as $\|S_1-S_1'\|=\sup_{t\in[0,\tau]}|S_1(t)-S_1'(t)|$. In addition, to apply the technique in empirical processes, we assume Ω is compact. For simplicity, we consider $\Omega=[1,D]$, for some large constant D. Assume $G_{10}(\tau)>0$ and $G_{20}(\tau)>0$, where $G_{10}(\cdot)$ and $G_{20}(\cdot)$ are the true survival functions of the censoring times C_1 and C_2. This assumption is used to show the consistency of the Kaplan–Meier estimator.

To check the consistency of $\hat\theta$, Theorem 2.12 in Kosorok (2008) can be applied if

$$\sup_{\theta\in\Omega,\|S_1-S_{10}\|\leq\eta_n,\|S_1-S_{10}\|\leq\eta_n}|\mathbb{P}_n m(\theta,S_1,S_2)-\mathbb{P}m(\theta,S_{10},S_{20})|=o_p(1),\tag{A1}$$

for some η_n converging to 0. This is because the Kaplan–Meier estimator is consistent for S_{10} and S_{20}. Then (A1) holds, if

$$\sup_{\theta\in\Omega,\|S_1-S_{10}\|\leq\eta_n,\|S_1-S_{10}\|\leq\eta_n}|\mathbb{P}m(\theta,S_1,S_2)-\mathbb{P}m(\theta,S_{10},S_{20})|=o_p(1),$$

and $\mathcal{F}_m=\{m(\theta,S_1,S_2):\theta\in\Omega,\|S_1-S_{10}\|\leq\eta,\|S_1-S_{10}\|\leq\eta\}$ for some $\eta>0$ is Glivenko–Cantelli. The first condition holds by applying the dominated convergence theorem, because we can easily check that $|\log c_\theta(u,v)|$, $|\log\partial C_\theta(u,v)/\partial u|$, $|\log\partial C_\theta(u,v)/\partial v|$ and $|\log C_\theta(u,v)|$ for any $\theta\in\Omega$ and $\delta/2\leq u,v\leq 1$ are bounded above by a constant. Define $\mathcal{F}_1=\{\log S_1:\|S_1-S_{10}\|\leq\eta\}$. We know that the set of all survival functions is Vapnik–Chervonenkis-major (van der Vaart & Wellner, 1996). Since S_1 belongs to a subset of all survival functions, $\{S_1:\|S_1-S_{10}\|\leq\eta\}$ is also Vapnik–Chervonenkis-major and so is $\mathcal{F}_1$ by the definition of Vapnik–Chervonenkis-major. Then, we can apply Theorem 2.6.14 in van der Vaart & Wellner (1996) to conclude that $\mathcal{F}_1$ is a Donsker class. It remains to check that their integrability condition in Theorem 2.6.14 holds, that is $\int\{\mathrm{pr(F>x)}\}^{1/2}\mathrm{dx}<\infty$, where F is the envelop function. This is true, because for any $t\in[0,\tau]$, $|\log S_1(t)|\leq|\log(\delta-\eta)|$, provided η is small enough. Then, we can take the envelope function to be $F=|\log(\delta-\eta)|$. The integrability condition holds. Thus, $\mathcal{F}_1$ is a Donsker class. In the following, define

$$\mathcal{F}_2^*=\{\log(S_1^{1-\theta}+S_2^{1-\theta}-1)/(1-\theta):\|S_1-S_{10}\|\leq\eta,\|S_2-S_{20}\|\leq\eta,1<c\leq\theta\leq D\},\tag{A2}$$

and we check $\mathcal{F}_2^*$ is also a Donsker class. This is because $\{S_1 : \|S_1 - S_{10}\| \leq \eta\}$, $\{S_2 : \|S_2 - S_{20}\| \leq \eta\}$ and $\{1 < c \leq \theta \leq D\}$ are all Donsker. If the function $g(x_1, x_2, x_3) = \log(x_2^{1-x_1} + x_3^{1-x_1} - 1)/(1 - x_1)$ is Lipschitz, then Theorem 2.10.6 in van der Vaart & Wellner (1996) implies $\mathcal{F}_2^*$ is a Donsker class. It is easily seen that

$$\frac{\partial}{\partial x_1} g(x_1, x_2, x_3) = \frac{1}{(1-x_1)^2} \log(x_2^{1-x_1} + x_3^{1-x_1} - 1) - \frac{1}{1-x_1} \frac{x_2^{1-x_1} \log x_2 + x_3^{1-x_1} \log x_3}{x_2^{1-x_1} + x_3^{1-x_1} - 1}.$$

Since $x_1 \in [c, D]$, $x_2, x_3 \in [\delta - \eta, 1]$, this partial derivative is bounded above by a constant uniformly over x_1, x_2, x_3. Similarly, the partial derivatives of $g(x_1, x_2, x_3)$ with respect to x_2 and x_3 are also bounded. This verifies the Lipschitz condition in Theorem 2.10.6 of van der Vaart & Wellner (1996). Consider $\mathcal{F}_2 := \{\log(S_1^{1-\theta} + S_2^{1-\theta} - 1)/(1-\theta) : \|S_1 - S_{10}\| \leq \eta, \|S_2 - S_{20}\| \leq \eta, \theta \in \Omega\}$. Following the proof of Theorem 2.10.2 in van der Vaart & Wellner (1996), $\mathcal{F}_2$ is also a Donsker class. Similarly, define $\mathcal{F}_3 = \{\log(S_1^{1-\theta} + S_2^{1-\theta} - 1) : \|S_1 - S_{10}\| \leq \eta, \|S_2 - S_{20}\| \leq \eta, \theta \in \Omega\}$, and we can verify that $\mathcal{F}_3$ is also a Donsker class. It is easy to check that $m(\theta, S_1, S_2)$ is obtained by adding or multiplying functions in $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ and $\{\theta : \theta \in \Omega\}$. By the permanence of Donsker property, $\mathcal{F}_m$ is Donsker and therefore Glivenko–Cantelli. This verifies the consistency of $\hat{\theta}$. In addition, we have $\|\hat{S}_1 - S_{10}\| = O_p(n^{-1/2})$ and $\|\hat{S}_2 - S_{20}\| = O_p(n^{-1/2})$ for the Kaplan–Meier estimator; see Section 2.2.5 of Kosorok (2008). This justifies Condition C1.

To check Condition C2, we only need to prove that $\mathcal{F}_{m_1} = \{m_1(\theta, S_1, S_2) : \theta \in \Omega, \|S_1 - S_{10}\| \leq \eta, \|S_1 - S_{10}\| \leq \eta\}$ for some $\eta > 0$ is a Donsker class with a square integrable envelope function. The proof is very similar to the previous one. We omit the details.

In the following, we verify Condition (C3). In fact, because $\hat{S}_1$ and $\hat{S}_2$ have the root-n rate, it suffices to show the following weaker version of Condition (C3): for any $\|S_1 - S_{10}\| = O(n^{-1/2})$ and $\|S_2 - S_{20}\| = O(n^{-1/2})$, it remains to show that

$$\Big| \mathbb{P}\left\{m_1(\theta_0, S_1, S_2) - m_1(\theta_0, S_{10}, S_{20}) - m_{12}(\theta_0, S_{10}, S_{20})[S - S_0]\right\} \Big| = o(n^{-1/2}), \qquad \text{(A3)}$$

where $S = (S_1, S_2)$ and $S_0 = (S_{10}, S_{20})$. By Proposition 1 of Bickel et al. (1993), (A3) holds if for some $\epsilon > 0$,

$$\sup \left\{ \Big| \frac{\partial}{\partial t} \mathbb{P} m_1(\theta_0, S_{10} + ts_1, S_{20} + ts_2) \Big| : \|s_1\| + \|s_2\| \leq 1, |t| < \epsilon \right\} < \infty. \qquad \text{(A4)}$$

It is easy to see that

$$\begin{aligned} &\frac{\partial}{\partial t} \mathbb{P} m_1(\theta_0, S_{10} + ts_1, S_2 + ts_2) \\ &= \frac{\partial}{\partial t} \int\int \Big[\delta_1\delta_2\{1 + \log(S_{10} + ts_1) + \log(S_{20} + ts_2)\} + \log(S_{10} + ts_1)\log(S_{20} + ts_2)\Big] d\mathbb{P}. \end{aligned}$$

By the assumption that $S_{10}(\tau) \geq \delta$ and $S_{20}(\tau) \geq \delta$ for some constant $\delta > 0$, $|\partial \log\{S_{10}(y) + ts_1(t)\}/\partial t| = |s_1(y)/\{S_{10}(y) + ts_1(y)\}| \leq 1/\{S_{10}(y) - t\} \leq 2/\delta$ for ϵ small enough. This implies that the derivative of $\log\{S_{10}(y) + ts_1(y)\}$ is uniformly bounded by a constant. Similarly, we can check that the derivative of the above integrand is uniformly bounded and therefore integrable, because $\delta_1\delta_2$ has finite expectation. By the dominated convergence theorem, we can interchange the order of derivative and integral. This leads to

$$\begin{aligned} &\Big| \frac{\partial}{\partial t} \mathbb{P} m_1(\theta_0, S_{10} + ts_1, S_2 + ts_2) \Big| \\ &= \Big| \int\int \Big\{ \delta_1\delta_2 \Big(1 + \frac{s_1}{S_{10} + ts_1} + \frac{s_2}{S_{20} + ts_2}\Big) + \frac{s_1}{S_{10} + ts_1} \log(S_{20} + ts_2) + \frac{s_2}{S_{20} + ts_2} \log(S_{10} + ts_1) \Big\} d\mathbb{P} \Big| \\ &\leq \int\int \Big\{ \delta_1\delta_2 \Big(1 + \Big|\frac{s_1}{S_{10} + ts_1}\Big| + \Big|\frac{s_2}{S_{20} + ts_2}\Big|\Big) + \Big|\frac{s_1}{S_{10} + ts_1} \log(S_{20} + ts_2)\Big| + \Big|\frac{s_2}{S_{20} + ts_2} \log(S_{10} + ts_1)\Big| \Big\} d\mathbb{P}. \end{aligned}$$

As seen above, $|s_1(y)/\{S_{10}(y)+ts_1(y)\}| \le 2/\delta$, and similarly $|\log\{S_{10}(y)+ts_1(y)\}| \le -\log(\delta/2)$ for ϵ small enough. Then there exists a constant $C>0$ such that

$$\Big|\frac{\partial}{\partial t}\mathbb{P}m_1(\theta_0, S_{10}+ts_1, S_2+ts_2)\Big| \le C\int\int \delta_1\delta_2 d\mathbb{P} \le C\int \delta_1 d\mathbb{P} < \infty.$$

This verifies (A4) and the smoothness condition of Condition C3.

The joint asymptotic normality in Condition C4 and the existence of information and variance matrices in Condition C5 can be verified following the proof of Theorem 2 in Shih & Louis (1995).

Finally, we shall verify Condition C6. As an illustration, we focus on $\partial^2 \log c_\theta(u,v)/\partial\theta^2$. That is

$$\begin{aligned}\frac{\partial^2}{\partial\theta^2}\log c_\theta(u,v) &= -\frac{1}{\theta^2} - \frac{2}{(\theta-1)^3}\log(u^{1-\theta}+v^{1-\theta}-1) - \frac{2}{(\theta-1)^2}\frac{u^{1-\theta}\log u + v^{1-\theta}\log v}{u^{1-\theta}+v^{1-\theta}-1}\\ &\quad -\Big(2-\frac{1}{1-\theta}\Big)\frac{\{u^{1-\theta}(\log u)^2+v^{1-\theta}(\log v)^2\}(u^{1-\theta}+v^{1-\theta}-1)}{(u^{1-\theta}+v^{1-\theta}-1)^2}\\ &\quad +\Big(2-\frac{1}{1-\theta}\Big)\frac{\{u^{1-\theta}\log u+v^{1-\theta}\log v\}^2}{(u^{1-\theta}+v^{1-\theta}-1)^2}, \quad (\theta>1)\end{aligned}$$

and $-1+4\log u\log v$ for $\theta=1$. To prove $\mathcal{F}_{m_{11}} = \{m_{11}(\theta, S_1, S_2) : \theta\in\Omega, \|S_1-S_{10}\|\le\eta, \|S_1-S_{10}\|\le\eta\}$ for some $\eta>0$ is Glivenko–Cantelli, we basically follow the similar step to the proof of Condition (C1) by checking equation (A2) is a Donsker class. In fact, it suffices to show that each piece of $\partial^2\log c_\theta(u,v)/\partial\theta^2$ is a Donsker class and apply the permanence of Donsker property. The detailed verification is very tedious and is omitted.

B. LIKELIHOOD FUNCTION ON EXAMPLE 2

Let n denote the number of villages. For $i=1,\ldots,n$, write $u_i=(u_{1i},u_{2i},u_{3i})^{\mathrm{T}}$ for $(S_1(Y_{1i}), S_2(Y_{2i}), S_3(Y_{3i}))^{\mathrm{T}}$, and let $\delta_i=(\delta_{1i},\delta_{2i},\delta_{3i})$ be the censoring indicators. Then given u_i and δ_i, $i=1,\ldots,n$, the likelihood of $\theta=(\theta_1,\theta_2)$ and $\phi=(S_1,S_2,S_3)$ is proportional to

$$\begin{aligned}&\exp\{L(\theta,\phi)\}\\ &=\prod_{i=1}^{n}\Bigg(\Bigg[\theta_2 u_{1i}^{-\theta_2}(u_{2i}u_{3i})^{-\theta_1}\Big\{(\theta_1-\theta_2)A_i^{\frac{\theta_2-2\theta_1+1}{\theta_1-1}}B_i^{\frac{2\theta_2-1}{1-\theta_2}}+(2\theta_2-1)A_i^{\frac{2(\theta_2-\theta_1)}{\theta_1-1}}B_i^{\frac{3\theta_2-2}{1-\theta_2}}\Big\}\Bigg]^{\prod_{q=1}^{3}\delta_{qi}}\\ &\times\Big(u_{1i}^{-\theta_2}B_i^{\frac{\theta_2}{1-\theta_2}}\Big)^{\delta_{1i}(1-\delta_{2i})(1-\delta_{3i})}\prod_{j=2}^{3}\Big(u_{ji}^{-\theta_1}A_i^{\frac{\theta_2-\theta_1}{\theta_1-1}}B_i^{\frac{\theta_2}{1-\theta_2}}\Big)^{\frac{\prod_{q=1}^{3}(1-\delta_{qi})\delta_{ji}}{1-\delta_{ji}}}\\ &\times\prod_{l=2}^{3}\Big(\theta_2 u_{1i}^{-\theta_2}u_{li}^{-\theta_1}A_i^{\frac{\theta_2-\theta_1}{\theta_1-1}}B_i^{\frac{2\theta_2-1}{1-\theta_2}}\Big)^{\frac{\delta_{1i}\prod_{q=2}^{3}(1-\delta_{qi})\delta_{li}}{1-\delta_{li}}}\Bigg[(u_{2i}u_{3i})^{-\theta_1}\Big\{\theta_2 A_i^{\frac{2(\theta_2-\theta_1)}{\theta_1-1}}B_i^{\frac{2\theta_2-1}{1-\theta_2}}\\ &+(\theta_1-\theta_2)A_i^{\frac{\theta_2-2\theta_1+1}{\theta_1-1}}B_i^{\frac{\theta_2}{1-\theta_2}}\Big\}\Bigg]^{(1-\delta_{1i})\delta_{2i}\delta_{3i}}\Big(B_i^{\frac{1}{1-\theta_2}}\Big)^{\prod_{q=1}^{3}(1-\delta_{qi})}\Bigg)\end{aligned}$$

where $A_i=u_{2i}^{1-\theta_1}+u_{3i}^{1-\theta_1}-1$ and $B_i=u_{1i}^{1-\theta_2}+(u_{2i}^{1-\theta_1}+u_{3i}^{1-\theta_1}-1)^{(\theta_2-1)/(\theta_1-1)}-1$. In principle, one can follow the similar steps to verify Conditions C1–C6. However, due to the more complex structure of the log likelihood in this example, the score function and information matrix are much more involved. We leave the detailed verification of regularity conditions for future investigations.

C. Technical Details on Example 3

Recall that the weighted log likelihood function is

$$L(\theta, \pi) = \sum_{i=1}^{n} \frac{V_i}{\pi(X_i)} \log f(Y_i; \theta).$$

In this example, π is the nuisance parameter, which is comparable to ϕ as defined in Section 2.2. This yields

$$m(\theta, \pi) = \frac{V}{\pi(X)} \log f(Y; \theta), \qquad m_1(\theta, \pi) = \frac{V}{\pi(X)} \frac{\partial}{\partial \theta}\{\log f(Y; \theta)\}.$$

In the following, we check conditions C1–C6 under some mild regularity conditions. Assume that θ_0 is the unique maximizer of $\mathbb{P} \log f(Y; \theta)$. This is a standard identifiability condition in the literature. For simplicity, we assume X is a continuous univariate covariate within $[0, 1]$. If X is discrete, one can use the sample average estimator to estimate $\pi(\cdot)$, which has the parametric root-n rate. To illustrate our method for nonparametric estimators, we focus on the continuous case. We also assume that there exists a small constant $\delta > 0$ such that $\pi_0(x) \geq \delta$ for any $x \in [0, 1]$, where we let π_0 denote the true value of π. This assumption is standard in the missing data literature; see Robins et al. (1994) and Scharfstein et al. (1999). Let $\mathcal{F}$ denote the class of functions f on $[0, 1]$ such that $\sup_{x\in[0,1]} |f(x)| \leq 1$ and any rth derivative is uniformly bounded for $1 \leq r \leq k$ for some integer $k > 0$. Assume that $\pi_0 \in \mathcal{F}$ and $p \in \mathcal{F}$, where $p(x)$ is the density function of X. Also assume $0 < \delta \leq p(x) < 1/\delta$. The norm for the nuisance function is taken as $\|\pi - \pi'\| = \sup_{x\in[0,1]} |\pi(x) - \pi'(x)|$. We assume Ω is compact. For simplicity, we take $\Omega = [A_1, A_2]$ for some constants A_1 and A_2. Assume $\sup_{\theta\in\Omega} |\log f(Y; \theta)| \leq g(Y)$, $\sup_{\theta\in\Omega} |\partial \log f(Y; \theta)/\partial\theta| \leq g(Y)$, and $\sup_{\theta\in\Omega} |\partial^2 \log f(Y; \theta)/\partial\theta^2| \leq g(Y)$ for some integrable function $g(Y)$. In addition, suppose $\partial^2 \log f(Y; \theta)/\partial\theta^2$ is a continuous function of θ. Assume the information matrix and covariance matrix

$$I_{11} = -\mathbb{P}\Big[\frac{\partial^2}{\partial\theta^2}\{\log f(Y; \theta)\}\Big], \;\; I_{11}^* = \mathbb{P}\Big\{\frac{S^2(Y; \theta_0)}{\pi_0(X)}\Big\} - \mathbb{P}\Big[\frac{\{1 - \pi_0(X)\}E^2\{S(Y; \theta_0) \mid X\}}{\pi_0(X)}\Big] \tag{C1}$$

are positive definite, where $S(Y; \theta) = \partial \log f(Y; \theta)/\partial\theta$.

Recall that the Nadaraya–Watson estimator of $\pi(x)$ is defined as

$$\hat{\pi}(x) = \frac{\sum_{i=1}^{n} V_i K_h(x - X_i)}{\sum_{i=1}^{n} K_h(x - X_i)},$$

where $K_h(\cdot) = K(\cdot/h)$ and $K(\cdot)$ is a kernel function of order k and h is the bandwidth parameter. By van der Vaart (2000), the estimator $\hat{\pi}$ satisfies $\|\hat{\pi} - \pi_0\| = O_P\{(nh)^{-1/2} + h^k\}$. Assume that $h = n^{-a}$ for some $a > 0$. In the following, we can show that we need $1/(2k) < a < 1/2$ to check some of conditions. We also comment that the optimal bandwidth is taken as $a = 1/(2k + 1)$, which is smaller than $1/(2k)$. In other words, we need to choose a smaller h to under-smooth the curve.

To verify Condition C1, we first check the consistency of $\hat{\theta}$. In the following, we always assume $\pi \in \mathcal{F}$ without stated explicitly. Similar to Example 1, Theorem 2.12 in Kosorok (2008) can be applied if

$$\sup_{\theta\in\Omega, \|\pi-\pi_0\|\leq\eta_n} |\mathbb{P}_n m(\theta, \pi) - \mathbb{P} m(\theta, \pi_0)| = o_p(1), \tag{C2}$$

for some η_n converging to 0. To show (C2), we first consider

$$\begin{aligned}\lim_{\|\pi-\pi_0\|\to 0} \sup_{\theta\in\Omega} |\mathbb{P} m(\theta, \pi) - \mathbb{P} m(\theta, \pi_0)| &\leq \lim_{\|\pi-\pi_0\|\to 0} \int \sup_{\theta\in\Omega} |m(\theta, \pi) - m(\theta, \pi_0)| d\mathbb{P} \\ &\leq \lim_{\|\pi-\pi_0\|\to 0} \int g(Y) |\pi^{-1}(X) - \pi_0^{-1}(X)| d\mathbb{P} = o_p(1),\end{aligned}$$

where the last step follows from the dominated convergence theorem to interchange the limit with integral and the fact that $g(Y)$ is integrable. Next we need to show that $\mathcal{F}_m = \{m(\theta, \pi) : \theta \in \Omega, \|\pi - \pi_0\| \leq \eta\}$

for some $\eta > 0$ is Glivenko–Cantelli. By Example 19.10 in van der Vaart (2000), $\mathcal{F}$ is Donsker. Since $\pi(x) \geq \delta - \eta > 0$ for η sufficiently small, $\{1/\pi : \|\pi - \pi_0\| \leq \eta\}$ is also Donsker by Example 19.20 in van der Vaart (2000). In addition, our assumptions ensure that $\log f(Y;\theta)$ is Lipschitz in θ. Hence, the class of functions $\{\log f(Y;\theta) : \theta \in \Omega\}$ is Donsker by Example 19.7 in van der Vaart (2000). By the permanence of Donsker property, $\mathcal{F}_m$ is Donsker and therefore Glivenko–Cantelli. This verifies the consistency of $\hat{\theta}$. This justifies Condition C1.

To show Condition C2, we only need to prove that $\mathcal{F}_{m_1} = \{m_1(\theta,\pi) : \theta \in \Omega, \|\pi - \pi_0\| \leq \eta\}$ for some $\eta > 0$ is a Donsker class. The proof is the same as that of $\mathcal{F}_m$.

In the following, we verify Condition C3. In fact, we will show that Condition C3 holds with $c_2 = 2$. Since by Condition C1, we have $c_1 = \min\{(1-a)/2, ak\}$. Hence, $c_1 c_2 = \min\{(1-a), 2ak\} > 1/2$ holds as long as $a < 1/2$ and $ak > 1/4$. Consider the following parametric submodel $\pi_t = \pi_0 + (\pi - \pi_0)t$. It has $\partial \pi_t/\partial t|_{t=0} = \pi - \pi_0$. Then, by definition

$$m_{12}(\theta_0,\pi_0)[\pi - \pi_0] = \frac{\partial^2}{\partial\theta\partial t} m(\theta_0,\phi_t)\bigg|_{t=0} = -\frac{V(\pi-\pi_0)}{\pi_0^2}\frac{\partial}{\partial\theta}\{\log f(Y;\theta)\}.$$

Then, we have that

$$\begin{aligned}
\Big|\mathbb{P}\{m_1(\theta_0,\phi) - m_1(\theta_0,\phi_0) - m_{12}(\theta_0,\phi_0)[\phi-\phi_0]\}\Big| &= \Big|\int V\frac{\partial}{\partial\theta}\{\log f(Y;\theta)\}\Big(\frac{1}{\pi} - \frac{1}{\pi_0} + \frac{\pi-\pi_0}{\pi_0^2}\Big)d\mathbb{P}\Big| \\
&\leq \int \Big|\frac{\partial}{\partial\theta}\{\log f(Y;\theta)\}\Big|\frac{(\pi-\pi_0)^2}{\pi\pi_0^2}d\mathbb{P} \\
&\leq \|\pi-\pi_0\|^2 2/\delta^3 \int g(Y)d\mathbb{P} = O(\|\pi-\pi_0\|^2),
\end{aligned}$$

where π is sufficiently close to π_0. Thus, Condition C3 holds with $c_2 = 2$.

Next, we check Condition C4. The key part is to find the influence function of $\mathbb{P}m_{12}(\theta_0,\pi_0)[n^{1/2}(\hat{\pi} - \pi_0)]$. Denote the kernel estimator of density of X by

$$\hat{p}(x) = \frac{1}{nh}\sum_{i=1}^n K_h(x - X_i).$$

Let $S(Y;\theta) = \partial \log f(Y;\theta)/\partial\theta$ denote the score function of Y. Simple algebra shows that

$$\begin{aligned}
-\mathbb{P}m_{12}(\theta_0,\pi_0)[n^{1/2}(\hat{\pi}-\pi_0)] &= \int \frac{n^{1/2}\{\hat{\pi}(x) - \pi_0(x)\}}{\pi_0(x)}\frac{\partial}{\partial\theta}\{\log f(y;\theta_0)\}d\mathbb{P} \\
&= \frac{1}{n^{1/2}}\sum_{i=1}^n \int \frac{S(y;\theta_0)}{\pi_0(x)\hat{p}(x)}\{V_i - \pi_0(X_i)\}\frac{1}{h}K_h(x - X_i)d\mathbb{P} \\
&+ \frac{1}{n^{1/2}}\sum_{i=1}^n \int \frac{S(y;\theta_0)}{\pi_0(x)\hat{p}(x)}\{\pi_0(X_i) - \pi_0(x)\}\frac{1}{h}K_h(x - X_i)d\mathbb{P}. \quad \text{(C3)}
\end{aligned}$$

Denote I_1 and I_2 for the two terms of equation (C3), respectively. Since $0 < \delta \leq p(x) < 1/\delta$, we have $\sup_{x\in[0,1]}|p(x)/\hat{p}(x) - 1| = o_p(1)$, where $p(x)$ is the density of X. Thus,

$$\begin{aligned}
I_1 &= \frac{1}{n^{1/2}}\sum_{i=1}^n \int \frac{S(y;\theta_0)}{\pi_0(x)}\{V_i - \pi_0(X_i)\}\frac{1}{h}K_h(x - X_i)p(y \mid x)dydx + o_p(1) \\
&= \frac{1}{n^{1/2}}\sum_{i=1}^n \int \frac{S(y;\theta_0)}{\pi_0(X_i + uh)}\{V_i - \pi_0(X_i)\}K(u)p(y \mid X_i + uh)dydu + o_p(1) \\
&= \frac{1}{n^{1/2}}\sum_{i=1}^n \{V_i - \pi_0(X_i)\}\int \frac{E\{S(Y;\theta_0) \mid X_i + uh\}}{\pi_0(X_i + uh)}K(u)du + o_p(1),
\end{aligned}$$

where the second equation follows from the change of variable by introducing $u = (x - X_i)/h$. Following the arguments in page 2087 of Huang & Wang (2010), we can show that

$$I_1 = \frac{1}{n^{1/2}} \sum_{i=1}^{n} \{V_i - \pi_0(X_i)\} \frac{E\{S(Y;\theta_0) \mid X_i\}}{\pi_0(X_i)} + o_p(1).$$

Applying the Taylor expansion of $\pi_0(X_i) - \pi_0(x)$, we can show that

$$\begin{aligned} I_2 &= -\frac{1}{n^{1/2}} \sum_{i=1}^{n} \int \frac{S(y;\theta_0)}{\pi_0(x)} \{\pi_0(x) - \pi_0(X_i)\} \frac{1}{h} K_h(x - X_i) p(y \mid x) dy dx + o_p(1) \\ &= -\frac{1}{n^{1/2}} \sum_{i=1}^{n} \int \frac{S(y;\theta_0)}{\pi_0(x)} \Big\{ \sum_{j=1}^{k-1} A_{ij}(x - X_i)^j + R_i (x - X_i)^k \Big\} \frac{1}{h} K_h(x - X_i) p(y \mid x) dy dx + o_p(1), \end{aligned}$$

where $A_{ij} = \pi_0^{(j)}(X_i)/j!$, $R_i = \pi_0^{(k)}(\tilde{X}_i)/k!$ and $\tilde{X}_i$ lies between X_i and x. Similarly, letting $u = (x - X_i)/h$, we have

$$\begin{aligned} I_2 &= -\frac{1}{n^{1/2}} \sum_{i=1}^{n} \int \frac{S(y;\theta_0)}{\pi_0(X_i + uh)} \Big\{ \sum_{j=1}^{k-1} A_{ij}(uh)^j + R_i (uh)^k \Big\} K(u) p(y \mid X_i + uh) dy du + o_p(1) \\ &= -\frac{1}{n^{1/2}} \sum_{i=1}^{n} \int \frac{E\{S(Y;\theta_0) \mid X_i + uh\}}{\pi_0(X_i + uh)} \Big\{ \sum_{j=1}^{k-1} A_{ij}(uh)^j + R_i (uh)^k \Big\} K(u) du + o_p(1). \end{aligned}$$

By the definition of the kernel function, we have $\int u^j K(u) du = 0$ for $j = 1, ..., k-1$. Then, the arguments in Wang et al. (2002); Huang & Wang (2010) yield that

$$|I_2| \leq \frac{h^k}{n^{1/2}} \sum_{i=1}^{n} \frac{|E\{S(Y;\theta_0) \mid X_i\}|}{\pi_0(X_i)} \int |R_i| |u|^k K(u) du + o_p(1) = O_P(h^k n^{1/2}).$$

Since $h = n^{-a}$, we have $|I_2| = o_p(1)$ provided $a > 1/(2k)$. Combining the asymptotic expansions of I_1 and I_2, we prove that

$$\mathbb{P} m_{12}(\theta_0, \pi_0)[n^{1/2}(\hat{\pi} - \pi_0)] = -\frac{1}{n^{1/2}} \sum_{i=1}^{n} \{V_i - \pi_0(X_i)\} \frac{E\{S(Y;\theta_0) \mid X_i\}}{\pi_0(X_i)} + o_p(1).$$

Applying the multivariate central limit theorem and Slutsky's theorem, we show that $n^{1/2}\mathbb{P}_n m_1(\theta_0, \phi_0)$ and $\mathbb{P} m_{12}(\theta_0, \phi_0)[n^{1/2}(\hat{\phi} - \phi_0)]$ jointly converge in distribution to $N(0, \Sigma)$, where

$$\Sigma_{11} = \mathbb{P}\Big\{ \frac{S^2(Y;\theta_0)}{\pi_0(X)} \Big\}, \ \ \Sigma_{12} = -\mathbb{P}\Big[\frac{\{1 - \pi_0(X)\} E^2\{S(Y;\theta_0) \mid X\}}{\pi_0(X)} \Big],$$

and

$$\Sigma_{22} = \mathbb{P}\Big[\frac{\{1 - \pi_0(X)\} E^2\{S(Y;\theta_0) \mid X\}}{\pi_0(X)} \Big].$$

To check Condition C5, it is easily seen that the information matrix has the form of I_{11} in (C1) by applying the double expectation rule and the assumption that $E\{V w(Y) \mid X\} = E(V \mid X) E\{w(Y) \mid X\} = \pi_0(X) E\{w(Y) \mid X\}$ for any function w. By our assumption, I_{11} is positive definite. Similarly, I_{11}^* has the desired form in (C1) and is also positive definite by assumption.

To show Condition C6, we need to show the following two parts. First,

$$\lim_{\|\pi - \pi_0\| \to 0} \lim_{\theta \to \theta_0} |\mathbb{P} m_{11}(\theta, \pi) - \mathbb{P} m_{11}(\theta_0, \pi_0)| = o_p(1), \ \text{ where } \ m_{11}(\theta, \pi) = \frac{V}{\pi(X)} \frac{\partial^2}{\partial \theta^2} \log f(Y;\theta).$$

Similar to the proof of Condition C1, we can show that $|m_{11}(\theta, \pi) - m_{11}(\theta_0, \pi_0)|$ is upper bounded by an integrable function, such that the dominated convergence theorem can be applied to interchange the limit with integral. Second, we need to show $\mathcal{F}_{m_{11}} = \{m_{11}(\theta, \pi) : \theta \in \Omega, \|\pi - \pi_0\| \leq \eta\}$ for some $\eta > 0$ is Glivenko–Cantelli. The key is to apply Example 19.8 of van der Vaart (2000) to conclude that $\{\partial^2 \log f(Y;\theta)/\partial\theta^2 : \theta \in \Omega\}$ is Glivenko–Cantelli. By the Glivenko–Cantelli preservation theory in Theorem 9.26 of Kosorok (2008), $\mathcal{F}_{m_{11}}$ is Glivenko–Cantelli. Then the verification of Condition C6 is complete.

D. SIMULATION

We conducted three simulation studies under the settings of Examples 1 and 2. We compared the Type I errors and power of the proposed test and the naive test in which the boundary problem is ignored and the associated p-value is derived from chi-squared distributions. We first applied the pseudolikelihood ratio test to test for the dependence between bivariate survival times as we described in Example 1. The same data generating procedures are used. For each generated dataset, we compared the pseudolikelihood ratio test statistic with the distribution of $0.5\chi_0^2 + 0.5\chi_1^2$, denoted as the proposed test, and also with χ_1^2, denoted as the naive test. Table 1 shows the estimated Type I errors and power from 5000 replications of the tests. The nominal level for the type I error was set to 0.05 and 0.01. When the null hypothesis is true, the Type I errors of the proposed test were all within 95% confidence intervals for the nominal levels, i.e., 0.007–0.013 for the nominal level of 0.01 and 0.044–0.056 for the nominal level of 0.05 under all scenarios considered. This suggests that the asymptotic approximation for the pseudolikelihood ratio test is adequate for moderate sample sizes. In contrast, the naive test was too conservative under all scenarios considered: range of Type I errors is $[0.027, 0.030]$ at the nominal level of 0.05 and $[0.005, 0.008]$ at the nominal level of 0.01. As expected, the powers of the two tests increased with the increasing values of the association parameter θ. At the nominal level of 0.05, the proposed test had about 80% power when θ was 1.7 at a sample size of 100, and when θ was 1.5 at a sample size of 200, and when θ was 1.3 at a sample size of 400. The power slightly decreased with the increase in the degree of censoring, indicating that the censoring percentage has a relatively small impact on the power. In a power comparison, the proposed test exhibited a superior level of power compared to the naive test: at nominal levels of 0.05 and 0.01, the naive test resulted in power losses of up to 32.2% and 55.1%, respectively.

In the second simulation study, we applied the pseudolikelihood ratio test to test for associations between all failure times within the same village in the model described in Example 2. Again, the same data generating procedures are used. For each generated dataset, we compared the pseudolikelihood ratio test statistic with the distribution of $(0.5 - \hat{p}_s)\chi_0^2 + 0.5\chi_1^2 + \hat{p}_s\chi_2^2$, denoted as the proposed test, and also with the distribution of χ_2^2, denoted as the naive test. Table 2 shows the estimated Type I errors and powers from 5000 simulations. Similar to our findings in Example 1, the proposed test had sizes close to the nominal levels, suggesting the asymptotic approximation for the test performs well. In contrast, the naive test, which ignores the boundary problem, led to conservative Type I errors and substantial loss of power.

In summary, simulation studies under two different models suggest that the semiparametric pseudolikelihood ratio test performs well in moderate sample size settings, and that the naive test that ignores the boundary problem gives conservative Type I errors and much lower power.

E. ADDITIONAL SIMULATION RESULTS

We also created two quantile-quantile plots from Examples 1 and 2. Figures 1 and 2 show results from 5000 simulations under the null hypothesis with sample size of 200. The left panel of each figure shows the quantiles of the pseudolikelihood ratio test statistics from 5000 simulations plotted against the corresponding quantiles of chi-square distribution, i.e., the naive approach. The right panel of each Figure shows the quantiles of the pseudolikelihood ratio test statistics from 5000 simulations plotted against the quantile of the asymptotic distribution, i.e., the proposed approach. The quantile-quantile plots suggest that the theoretical approximation works well at levels in addition to 5% and 1%.

Table 1. *Empirical rejection rates (%) for the pseudolikelihood ratio test for testing association between bivariate survival time in Example* 1 *in* 5000 *simulations. Upper entry: naive method by comparing pseudolikelihood ratio test statistics with χ^2_1 distribution; lower entry: proposed test by comparing pseudolikelihood ratio test statistics with the correct limiting distribution.*

		censoring % = 0%		censoring % = 15%		censoring % = 30%	
n	θ	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$
		Rejection (%)	Rejection	Rejection	Rejection	Rejection	Rejection
100	1.0	3.0	0.7	2.9	0.5	2.9	0.6
		5.5	1.3	5.4	1.1	5.4	1.1
	1.3	26.3	10.7	24.5	10.0	22.7	9.8
		36.8	16.6	34.6	14.6	33.5	14.1
	1.5	55.4	30.4	51.4	27.4	48.0	25.8
		65.2	39.3	62.9	36.5	60.2	34.2
	1.7	76.3	55.4	73.6	51.6	70.9	48.4
		84.4	65.0	82.3	61.2	80.2	58.4
	2.0	94.3	88.3	92.5	80.0	90.5	76.4
		97.1	88.6	95.9	86.1	94.7	83.7
200	1.0	2.9	0.5	2.7	0.6	2.5	0.6
		5.4	1.0	5.5	1.0	5.2	1.1
	1.3	43.1	22.1	43.5	21.0	43.0	21.4
		55.8	29.6	55.5	29.5	55.5	28.6
	1.5	79.9	59.9	78.6	58.9	77.5	57.3
		87.0	68.6	86.3	67.4	85.5	66.7
	1.7	95.8	87.1	95.3	85.1	94.1	83.9
		97.7	91.6	97.4	90.1	96.8	89.3
	2.0	99.8	99.0	99.6	98.4	99.4	97.7
		99.9	99.5	99.9	99.1	99.8	98.7
400	1.0	2.7	0.5	2.7	0.5	2.9	0.8
		5.2	1.1	5.1	1.0	5.4	1.3
	1.3	67.6	44.0	65.2	42.6	66.3	42.8
		77.2	53.7	76.0	52.8	76.1	52.2
	1.5	97.0	89.6	95.9	87.6	95.7	87.1
		98.6	93.3	98.1	91.6	97.8	91.6
	1.7	99.9	99.4	99.8	99.1	99.7	98.8
		100.0	99.8	99.9	99.6	99.9	99.4
	2.0	100.0	100.0	100.0	100.0	100.0	100.0
		100.0	100.0	100.0	100.0	100.0	100.0

F. A DATA EXAMPLE

Dementia is a progressive degenerative medical condition and is one of the leading causes of death in the United States and Canada. There is evidence that dementia aggregates in families (Hendrie, 1998). We applied the proposed pseudolikelihood ratio test to check the aggregation of dementia in families who participated in the Cache County Study on Memory, Health, and Aging (Breitner et al., 1999). Information on the age at onset of dementia or censoring age was recorded for mothers and children of families who participated in the study.

Table 2. *Empirical rejection rates (%) for the pseudolikelihood ratio test for testing associations in model described in Example 2 in 5,000 simulations. Upper entry: naive method by comparing pseudolikelihood ratio test statistics with χ^2_2 distribution; lower entry: proposed test by comparing pseudolikelihood ratio test statistics with the correct limiting distribution.*

n	(θ_1, θ_2)	censoring % = 0%		censoring % = 15%		censoring % = 30%	
		$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$
		Rejection (%)	Rejection	Rejection	Rejection	Rejection	Rejection
100	(1.0,1.0)	2.3	0.8	1.5	0.3	1.5	0.3
		5.4	1.5	5.0	1.0	5.0	1.0
	(1.1,1.1)	18.1	7.4	14.4	5.8	12.3	4.0
		32.5	15.0	27.1	12.3	25.1	10.1
	(1.2,1.1)	35.6	18.7	29.4	13.3	22.1	9.2
		55.7	32.8	48.9	25.9	40.4	18.6
	(1.3,1.1)	56.1	34.9	48.8	25.7	36.7	18.4
		74.1	52.6	67.8	44.7	57.1	32.4
	(1.2,1.2)	56.7	34.9	49.3	28.6	37.4	18.7
		71.4	51.5	66.1	44.4	57.0	33.5
	(1.3,1.2)	72.0	51.1	64.3	42.9	50.0	28.7
		84.4	68.1	78.1	59.7	70.1	45.3
200	(1.0,1.0)	1.9	0.8	1.4	0.3	1.5	0.4
		5.0	1.5	4.9	0.9	4.6	1.0
	(1.1,1.1)	35.4	19.0	28.9	13.7	21.6	8.6
		53.7	32.3	45.7	25.2	37.7	18.5
	(1.2,1.1)	66.5	45.7	57.1	34.6	43.9	22.8
		81.8	62.6	74.7	51.7	62.5	37.8
	(1.3,1.1)	88.2	74.3	79.2	61.0	66.4	44.0
		95.1	86.1	90.6	76.0	82.4	61.6
	(1.2,1.2)	86.6	75.5	79.4	62.7	67.1	45.8
		92.5	85.1	88.1	75.6	80.0	61.6
	(1.3,1.2)	96.1	90.1	92.4	81.8	83.3	65.2
		98.5	95.4	96.5	90.3	91.8	79.5
400	(1.0,1.0)	1.3	0.3	1.2	0.1	1.6	0.4
		4.9	1.0	4.9	1.1	4.8	0.8
	(1.1,1.1)	65.2	47.3	54.4	35.0	40.7	22.2
		78.5	61.1	69.7	49.4	57.4	35.9
	(1.2,1.1)	94.4	84.9	87.0	72.4	74.3	53.7
		97.8	92.7	94.0	84.6	87.1	69.7
	(1.3,1.1)	99.6	98.4	98.1	93.9	93.3	82.4
		100.0	99.4	99.5	97.6	97.6	91.3
	(1.2,1.2)	99.1	97.8	97.3	94.3	92.0	83.8
		99.5	98.8	98.6	96.7	96.0	90.5
	(1.3,1.2)	99.9	99.7	99.9	99.1	98.4	95.4
		100.0	99.9	100.0	99.8	99.5	98.2

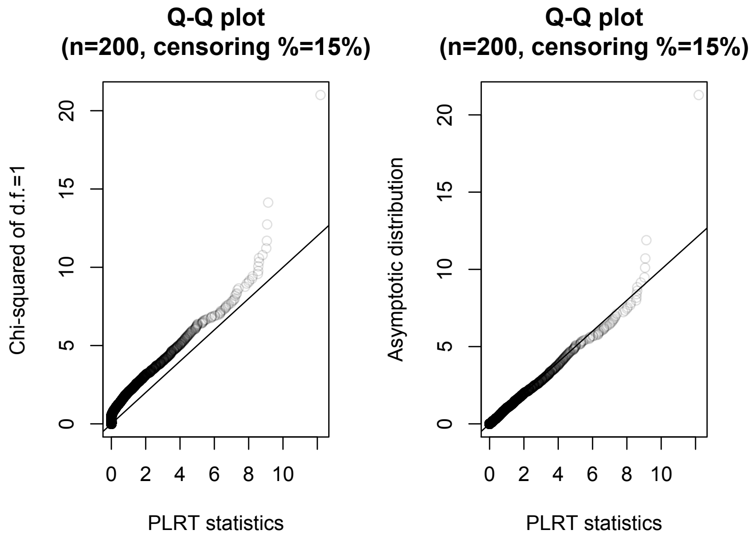

Fig. 1. Left panel: quantile-quantile plot of pseudolikelihood ratio test statistics in Example 1 against quantile of χ_1^2 distribution. Right panel: quantile-quantile plot of pseudolikelihood ratio test statistics against quantile of the distribution $0.5\chi_0^2 + 0.5\chi_1^2$.

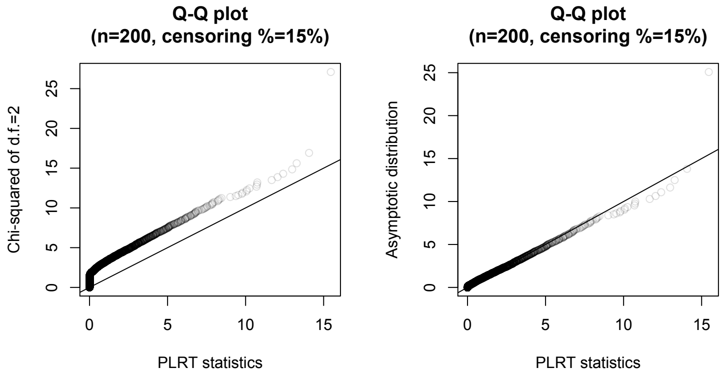

Fig. 2. Left panel: quantile-quantile plot of pseudolikelihood ratio test statistics in Example 2 against quantile of χ_2^2 distribution. Right panel: quantile-quantile plot of pseudolikelihood ratio test statistics against quantile of the distribution $(0.5 - p_s)\chi_0^2 + 0.5\chi_1^2 + p_s\chi_2^2$.

Following Bandeen-Roche & Liang (2002), we analyzed pairs that comprised the mother and the oldest child to study the association of dementia in families. We excluded pairs for which a member had been diagnosed with dementia or had died before age 55. After excluding pairs with missing data, we had 3,635 pairs available. Among the 3,635 pairs, there were 40 pairs in which both members had dementia, 158 pairs in which the child had dementia and the mother's outcome was censored, 419 pairs in which the child's outcome was censored and the mother had dementia, and 3,018 pairs in which both members' outcomes were censored. In the analysis, the observed data for the ith family are recorded as

$(y_{i1}, y_{i2}, \delta_{i1}, \delta_{i2})$, where y_{i1} is the observed time of the oldest child, y_{i2} is the observed time of the mother, and δ_{ij} is the censoring indicator corresponding to y_{ij}.

We assumed a Clayton copular model for the paired observations. The marginal survival functions of the children's dementia onsets and the mother's dementia onset were respectively estimated by using the Kaplan–Meier method. Applying the two-stage estimation procedure as we described in Example 1, we obtained the estimator of θ to be 2.44. This estimate implies that the cross ratio between dementia onsets of children and those of their mothers within families is 3.44 with standard error of 0.96, suggesting that the multiplicative increase in the risk of dementia onset for children whose mothers have dementia, versus those whose mothers are without dementia is 3.44. The small value of the p-value ($p < 0.001$) from the proposed pseudolikelihood ratio test strongly indicates that there is a high positive correlation between the onset of dementia in children and the onset of dementia in their mother.

REFERENCES

BANDEEN-ROCHE, K. & LIANG, K. Y. (2002). Modelling multivariate failure time associations in the presence of a competing risk. *Biometrika* **89**, 299–314.

BICKEL, P. J., KLAASSEN, C., RITOV, Y. & WELLNER, J. (1993). *Efficient and Adaptive Inference in Semiparametric Models*. Johns Hopkins University Press, Baltimore.

BREITNER, J. C., WYSE, B. W., ANTHONY, J. C., WELSH-BOHMER, K. A., STEFFENS, D. C.AND NORTON, M. C., TSCHANZ, J. T., PLASSMAN, B. L., MEYER, M. R., SKOOG, I. & KHACHATURIAN, A. (1999). APOE-epsilon4 count predicts age when prevalence of AD increases, then declines: The Cache County Study. *Neurology* **53**, 321–331.

HENDRIE, H. C. (1998). Epidemiology of dementia and Alzheimers disease. *American Journal of Geriatric Psychiatry* **6**, S3–18.

HUANG, B. & WANG, Q. (2010). Semiparametric analysis based on weighted estimating equations for transformation models with missing covariates. *Journal of Multivariate Analysis* **101**, 2078–2090.

KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York.

ROBINS, J. M., ROTNITZKY, A. & ZHAO, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89**, 846–866.

SCHARFSTEIN, D. O., ROTNITZKY, A. & ROBINS, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association* **94**, 1096–1120.

SHIH, J. H. & LOUIS, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics* **51**, 1384–1399.

VAN DER VAART, A. & WELLNER, J. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Science & Business Media.

VAN DER VAART, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press, New York.

WANG, Q., RAO, J. et al. (2002). Empirical likelihood-based inference under imputation for missing response data. *Annals of Statistics* **30**, 896–924.

[*Received April* 2012. *Revised September* 2012]