Biostatistics (2015), **0**, 0, *pp.* 1–9
doi:10.1093/biostatistics/BIOSTS-14203˙supp˙2

Supplementary Materials for "Semiparametric likelihood inference for left-truncated and right censored data"

CHIUNG-YU HUANG *

Sidney Kimmel Comprehensive Cancer Center and Department of Biostatistics,

Johns Hopkins University, Baltimore, Maryland 21205, U.S.A.

cyhuang@jhu.edu

JING NING

The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA

JING QIN

National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, U.S.A.

1. Asymptotic Behavior

Sketch proof of Theorem 1. For fixed $\boldsymbol{\theta}$, denote the maximizer of $\ell_n(\boldsymbol{\theta}, F)$ by $\widehat{F}_{\boldsymbol{\theta}}$. Obviously, $\widehat{\boldsymbol{\psi}}_n = (\widehat{\boldsymbol{\theta}}_n, \widehat{F}_n)$ is just the joint maximizer of $\ell_n(\boldsymbol{\psi})$. By a similar argument as in the proof of Property 1 in Vardi (1989), we can show that maximizing the log-likelihood function ℓ_n for a fixed $\boldsymbol{\theta}$ is equivalent to maximizing a strictly log-concave problem over a convex region, hence implying a unique maximizer $\widehat{F}_{\boldsymbol{\theta}}$ for each $\boldsymbol{\theta}$ in the compact set Θ. Then the compactness of Θ and the continuity of the profile likelihood $\ell_n(\boldsymbol{\theta}, \hat{F}_{\boldsymbol{\theta}})$ imply the existence of the maximum likelihood estimator. Furthermore, condition (A4) ensures the uniqueness of the maximum likelihood estimator.

Next, we show that both $\boldsymbol{\theta}_0$ and F_0 are identifiable. Suppose we have $P_{\boldsymbol{\theta},F} = P_{\boldsymbol{\theta}_0,F_0}$ almost every-

© The Author 2015. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

where under $P_{\boldsymbol{\theta}_0,F_0}$. Consider the densities on $\delta=0$, we see that $\bar{F}(y)h(a,\boldsymbol{\theta})/\mu(\boldsymbol{\theta},F)=\bar{F}_0(y)h(a,\boldsymbol{\theta}_0)/\mu(\boldsymbol{\theta}_0,F_0)$ for almost all $0\leqslant a\leqslant y\leqslant\tau$. Let $t^*=\sup\{t:F_0(t)<1\}$. Because F_0 is assumed to be continuous, there exists a sequence $\{y_m^*,m\geqslant 1\}$, $y_m^*<\tau$, converging to t^*. As $m\to\infty$, we have $h(a,\boldsymbol{\theta})/h(a,\boldsymbol{\theta}_0)=\mu(\boldsymbol{\theta},F)/\mu(\boldsymbol{\theta}_0,F_0)\times\bar{F}_0(y_m^*)/\bar{F}(y_m^*)\to\mu(\boldsymbol{\theta},F)/\mu(\boldsymbol{\theta}_0,F_0)$. The limit of the right side of the equation is independent of a. Hence $\boldsymbol{\theta}=\boldsymbol{\theta}_0$. Moreover, for $\delta=1$, we have $f(y)h(a,\boldsymbol{\theta})/\mu(\boldsymbol{\theta},F)=f_0(y)h(a,\boldsymbol{\theta}_0)/\mu(\boldsymbol{\theta}_0,F_0)$ for $0\leqslant a\leqslant y\leqslant\tau$. Combining with the previous result, we can show that that $f(y)/f_0(y)=1$ for all $y\in[0,\tau]$.

The proof of consistency for $\widehat{\boldsymbol{\psi}}_n$ is similar to those of Murphy (1995) and Parner (1998), thus we only state the main results for the proof. Since $\widehat{\boldsymbol{\psi}}_n$ is bounded, by Helly's selection theorem, there exists a convergent subsequence $\widehat{\boldsymbol{\psi}}_{n_k}=(\widehat{\boldsymbol{\theta}}_{n_k},\widehat{F}_{n_k})$, whose limit is denoted by $\boldsymbol{\psi}^*=(\boldsymbol{\theta}^*,F^*)\in\Theta\times\mathcal{F}$. It suffices to show that $\boldsymbol{\psi}^*=\boldsymbol{\psi}_0$ for any convergence subsequence. This can be accomplished by applying the classical Kullback-Leibler information approach. Specifically, we choose $\widetilde{\boldsymbol{\theta}}=\boldsymbol{\theta}_0$ and

$$\widetilde{F}_n(t)=\frac{n^{-1}\sum_{t_l\leqslant t}\sum_{i=1}^n H(a_i,\boldsymbol{\theta}_0)^{-1}\left[\delta_i I(y_i=t_l)+E_0\{(1-\delta_i)I(y_i=t_l)\}\right]}{n^{-1}\sum_{k=1}^L\sum_{j=1}^n H(a_j,\boldsymbol{\theta}_0)^{-1}\left[\delta_j I(y_j=t_k)+E_0\{(1-\delta_j)I(y_j=t_k)\}\right]},$$

where E_0 is the expectation under $\boldsymbol{\psi}_0$. In fact, if $\boldsymbol{\psi}_0$ was used as the initial value in the Expectation-Maximization algorithm in Section 2, $\widetilde{F}_n$ is simply the one-step estimator of F. Applying the Glivenko-Cantelli Theorem and a standard argument for Donsker class, we can show that $\widetilde{F}_n(t)$ converges to F_0 almost surely and uniformly in $[0,\tau]$. By the strong law of large numbers for empirical processes, we can show that $n_k^{-1}\{\ell_{n_k}(\widehat{\boldsymbol{\theta}}_{n_k},\widehat{F}_{n_k})-\ell_{n_k}(\widetilde{\boldsymbol{\theta}},\widetilde{F}_n)\}$ converges almost surely to the negative Kullback-Leibler distance between $P_{\boldsymbol{\psi}^*}$ and $P_{\boldsymbol{\psi}_0}$, where $P_{\boldsymbol{\psi}}$ is the probability measure under the parameter $\boldsymbol{\psi}=(\boldsymbol{\theta},F)$. Because ℓ_{n_k} is maximized at $\widehat{\boldsymbol{\psi}}_{n_k}$, we have $\ell_{n_k}(\widehat{\boldsymbol{\theta}}_{n_k},\widehat{F}_{n_k})-\ell_{n_k}(\widetilde{\boldsymbol{\theta}},\widetilde{F})\geqslant 0$. Hence $P_{\boldsymbol{\psi}^*}=P_{\boldsymbol{\psi}_0}$ almost surely. Thus by model identifiability, we have $\boldsymbol{\psi}^*=\boldsymbol{\psi}_0$. Because every convergent subsequence of $\widehat{\boldsymbol{\psi}}_n$ converges to the same limit $\boldsymbol{\psi}_0$, $\widehat{\boldsymbol{\psi}}_n$ must converge to $\boldsymbol{\psi}_0$ for any $t\in[0,\tau]$. The convergence is almost surely, since we only use the strong law of large numbers at most countably many times. The continuity and monotonicity of F_0 thus ensures the uniform convergence of $\widehat{F}_n$ in $[0,\tau]$.

We now prove the asymptotic normality of $n^{1/2}\{\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, \widehat{F}_n(t) - F_0(t)\}$ by applying the general Z-estimator convergence theorem (Theorem 3.3.1 in van der Vaart and Wellner, 1996). For theoretical development, we reparametrize the model and express the log-likelihood in terms of the hazard function:

$$\ell_n(\varphi) = \sum_{\mathbf{i=1}}^{\mathbf{n}} \left[\delta_{\mathbf{i}} \log\{\lambda(\mathbf{y_i})\} - \boldsymbol{\Lambda}(\mathbf{y_i}) + \log\left\{\mathbf{h}(\mathbf{a_i}, \boldsymbol{\theta})\right\} - \log\left\{\mu(\boldsymbol{\theta}, \boldsymbol{\Lambda})\right\}\right].$$

where $\boldsymbol{\varphi} = (\boldsymbol{\theta}, \Lambda)$, $\lambda(t) = d\Lambda(t)/dt$, and $\mu(\boldsymbol{\theta}, \Lambda) = \int_0^\tau h(u, \boldsymbol{\theta}) \exp\{-\Lambda(u)\} du$. Let $\boldsymbol{\psi}_0 = (\boldsymbol{\theta}_0, \Lambda_0)$ be the true parameter values, and let S_0 and λ_0 be the survival function and the hazard function that correspond to the true cumulative hazard function Λ_0.

Denote by $\widehat{\boldsymbol{\varphi}}_n = (\widehat{\boldsymbol{\theta}}_n, \widehat{\Lambda}_n)$ the maximum likelihood estimator that maximizes $\ell_n(\boldsymbol{\varphi})$. Taking the derivative of ℓ_n with respect to $\boldsymbol{\theta}$, we obtain the (normalised) score function for $\boldsymbol{\theta}$,

$$\boldsymbol{U}_{n1}(\boldsymbol{\theta}, \Lambda) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{h^{(1)}(a_i, \boldsymbol{\theta})}{h^{(0)}(a_i, \boldsymbol{\theta})} - \frac{\int_0^\tau h^{(1)}(u, \boldsymbol{\theta}) S(u) du}{\int_0^\tau h^{(0)}(u, \boldsymbol{\theta}) S(u) du} \right\},$$

where, for convenience, we define the functions $\boldsymbol{h}^{(k)}(t, \boldsymbol{\theta}) = \partial h^{(k-1)}(t, \boldsymbol{\theta})/\partial \boldsymbol{\theta}$, $k = 1, 2$, and $h^{(0)}(t, \boldsymbol{\theta}) = h(t, \boldsymbol{\theta})$. To derive the likelihood equation for the nonparametric component Λ, we consider a submodel defined by $\Lambda_\alpha(t) = \int_0^t \{1 + \alpha\eta(u)\} d\Lambda(u)$, where η is any bounded, integrable function on $[0, \tau]$. By taking the derivative of $\ell_n(\boldsymbol{\theta}, \Lambda_\alpha)$ with respect to α, evaluating it at $\alpha = 0$, and setting $\eta(\cdot) = 1(\cdot \leqslant t)$, the (normalised) likelihood equation for Λ is given by

$$U_{n2}(\boldsymbol{\theta}, \Lambda)(t) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \delta_i I(y_i \leqslant t) - \int_0^t I(y_i \geqslant u) d\Lambda(u) + \int_0^t \int_u^\tau \frac{S(w)}{\mu(\boldsymbol{\theta}, \Lambda)} h(w, \boldsymbol{\theta}) dw d\Lambda(u) \right\}.$$

Denote the vector of likelihood equations by $\boldsymbol{U}_n(\boldsymbol{\varphi})(t) = \{\boldsymbol{U}_{n1}(\boldsymbol{\varphi}), U_{n2}(\boldsymbol{\varphi})(t)\}$. The maximum likelihood estimator $\widehat{\boldsymbol{\varphi}}_n$ is the solution to the system $\boldsymbol{U}_{n1}(\boldsymbol{\varphi}) = 0$ and $U_{n2}(\boldsymbol{\varphi})(t) = 0$ for $t \in [0, \tau]$.

Let E_0 denote the expectation under the true value $\boldsymbol{\varphi}_0$. Define $\boldsymbol{\mathcal{U}}(\boldsymbol{\varphi})(t) = \{\boldsymbol{\mathcal{U}}_1(\boldsymbol{\varphi}), \mathcal{U}_2(\boldsymbol{\varphi})(t)\}$ with $\boldsymbol{\mathcal{U}}_1(\boldsymbol{\varphi}) = E_0\{\boldsymbol{U}_{n1}(\boldsymbol{\varphi})\}$ and $\mathcal{U}_2(\boldsymbol{\psi}) = E_0\{U_{n2}(\boldsymbol{\varphi})\}$. The asymptotic normality of $\widehat{\boldsymbol{\varphi}}_n$ can be established by verifying the three main conditions of the general Z-estimator convergence theorem (Theorem 3.3.1 in van der Vaart and Wellner, 1996): Fréchet differentiability, weak convergence of the likelihood equations $\sqrt{n}\boldsymbol{U}_n(\boldsymbol{\varphi}_0)$, and the stochastic approximation of the likelihood equations.

We first show that $\boldsymbol{\mathcal{U}}$ is Fréchet differentiable at $\boldsymbol{\varphi}_0$ and its Fréchet derivative is continuously invertible. The Gâteaux derivative of $\boldsymbol{\mathcal{U}}$ exists at any $(\boldsymbol{\theta}, \Lambda) \in \Theta \times L_2[0,\tau]$, where $L_2[0,\tau]$ is the space of functions with finite L_2 norm on $[0,\tau]$. Considering submodels $(\boldsymbol{\theta}_\alpha, \Lambda_\alpha) = (\boldsymbol{\theta}_0 + \alpha\boldsymbol{\theta}, \Lambda_0 + \alpha\Lambda)$, the Gâteaux variations of $\boldsymbol{\mathcal{U}}$ at $(\boldsymbol{\theta}_0, \Lambda_0)$ can be obtained by taking the derivative of $\boldsymbol{\mathcal{U}}(\boldsymbol{\theta}_\alpha, \Lambda_\alpha)$ with respect to α and evaluating at $\alpha = 0$. Specifically, let $\mu_0 = \int_0^\tau h^{(0)}(u, \boldsymbol{\theta}_0) S_0(u) du$ and $\boldsymbol{\mu}_1 = \int_0^\tau \boldsymbol{h}^{(1)}(u, \boldsymbol{\theta}_0) S_0(u) du$, then the Gâteaux derivative of $\boldsymbol{\mathcal{U}}$ at $(\boldsymbol{\theta}_0, \Lambda_0)$ is given by

$$\dot{\boldsymbol{\mathcal{U}}}_0(\boldsymbol{\psi}) = -\begin{pmatrix} \boldsymbol{\sigma}_{11} & \boldsymbol{\sigma}_{12} \\ \boldsymbol{\sigma}_{21} & \sigma_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{\theta} \\ \Lambda \end{pmatrix},$$

where

$$\begin{aligned}
\boldsymbol{\sigma}_{11}(\boldsymbol{\theta}) &= \left(E_0 \left[\left\{ \frac{\boldsymbol{h}^{(1)}(A, \boldsymbol{\theta})}{h^{(0)}(A, \boldsymbol{\theta})} \right\}^{\otimes 2} \right] - \left\{ \mu_0^{-1} \boldsymbol{\mu}_1 \right\}^{\otimes 2} \right)' \boldsymbol{\theta} = J_0 \boldsymbol{\theta}, \\
\boldsymbol{\sigma}_{12}(\Lambda) &= \int_0^\tau \int_u^\tau \frac{S_0(w)}{\mu_0} \left\{ \frac{\boldsymbol{\mu}_1}{\mu_0} h^{(0)}(w, \boldsymbol{\theta}_0) - \boldsymbol{h}^{(1)}(w, \boldsymbol{\theta}_0) \right\} dw \, d\Lambda(u), \\
\boldsymbol{\sigma}_{21}(\boldsymbol{\theta})(t) &= \left[\int_0^t \int_u^\tau \frac{S_0(w)}{\mu_0} \left\{ \frac{\boldsymbol{\mu}_1}{\mu_0} h^{(0)}(w, \boldsymbol{\theta}_0) - \boldsymbol{h}^{(1)}(w, \boldsymbol{\theta}_0) \right\} dw \, d\Lambda_0(u) \right]' \boldsymbol{\theta}, \\
\sigma_{22}(\Lambda) &= \int_0^t \frac{S_0(u)}{\mu_0} \left\{ \int_0^u h^{(0)}(v, \boldsymbol{\theta}_0) F_c(u - v) dv \right\} d\Lambda(u) \\
&\quad + \int_0^\tau \int_u^\tau \frac{S_0(w)}{\mu_0} h^{(0)}(w, \boldsymbol{\theta}_0) \left\{ \Lambda_0(t \wedge w) - K(t) \right\} dw \, d\Lambda(u),
\end{aligned}$$

with $u \wedge w = \min(u, w)$, S_c being the survival function of the censoring time, and

$$K(t) = \int_0^t \int_u^\tau S_0(w) {\mu_0}^{-1} h^{(0)}(w, \boldsymbol{\theta}_0) dw d\Lambda(u).$$

Thus it is easy to see that the mapping from $\boldsymbol{\psi} \in \Theta \times L_2[0,\tau]$ to the derivative of $\boldsymbol{\mathcal{U}}$ at $\boldsymbol{\psi}$ is continuous. Hence by a similar argument as the proof of Lemma 15.8 in Kosorok (2008), we can show that $\boldsymbol{\mathcal{U}}$ is Fréchet differentiable and its derivative at $\boldsymbol{\psi}_0$ is given by $\dot{\boldsymbol{\mathcal{U}}}_0$. Note that the operator $\boldsymbol{\mathcal{U}}_0$ is a linear continuous operator defined on $\mathbb{R}^p \times L_2[0,\tau]$, where $L_2[0,\tau]$ is a Banach space. If the inverse operator $\dot{\boldsymbol{\mathcal{U}}}_0$ exists, then it must be continuous by Banach's continuous inverse theorem (Zeidler, 1995, page 179). Hence, to prove the continuous invertibility of $\boldsymbol{\mathcal{U}}_0$, we only need to show that the inverse operator of $\dot{\boldsymbol{\mathcal{U}}}_0$ exists.

Straightforward algebra shows that if $\boldsymbol{\sigma}_{11}$ and $\Phi = \sigma_{22} - \boldsymbol{\sigma}_{21}\boldsymbol{\sigma}_{11}^{-1}\boldsymbol{\sigma}_{12}$ are invertible, then the inverse

of $\dot{\boldsymbol{\mathcal{U}}}_0$ is

$$\dot{\boldsymbol{\mathcal{U}}}_0^{-1} = -\begin{pmatrix} \boldsymbol{\sigma}_{11}^{-1} + \boldsymbol{\sigma}_{11}^{-1}\boldsymbol{\sigma}_{12}\Phi^{-1}\boldsymbol{\sigma}_{21}\boldsymbol{\sigma}_{11}^{-1} & -\boldsymbol{\sigma}_{11}^{-1}\boldsymbol{\sigma}_{12}\Phi^{-1} \\ -\Phi^{-1}\boldsymbol{\sigma}_{21}\boldsymbol{\sigma}_{11}^{-1} & \Phi^{-1} \end{pmatrix}.$$

The operator $\boldsymbol{\sigma}_{11}(\boldsymbol{\theta}) = \boldsymbol{J}_0\boldsymbol{\theta}$ is a linear operator, where the matrix J_0 is the Fisher information for $\boldsymbol{\theta}$ if Λ_0 is known. By Assumption (A4), the matrix J_0 is singular. Hence $\boldsymbol{\sigma}_{11}$ is invertible.

Define the functions

$$Q(t) = \frac{S_0(u)}{\mu_0}\left\{\int_0^u h^{(0)}(w,\boldsymbol{\theta}_0)F_c(u-w)dw\right\}$$

and

$$\begin{aligned} R(t,u) = &\int_u^\tau \frac{S_0(w)}{\mu_0}h^{(0)}(w,\boldsymbol{\theta}_0)\left\{\Lambda_0(t\wedge w) - K(t)\right\}dw \\ &- \left[\int_u^\tau \frac{S_0(w)}{\mu_0}\left\{\frac{\boldsymbol{\mu}_1}{\mu_0}h^{(0)}(w,\boldsymbol{\theta}_0) - \boldsymbol{h}^{(1)}(w,\boldsymbol{\theta}_0)\right\}dw\right]' J_0^{-1} \\ &\times \left[\int_0^t\int_u^\tau \frac{S_0(w)}{\mu_0}\left\{\frac{\boldsymbol{\mu}_1}{\mu_0}h^{(0)}(w,\boldsymbol{\theta}_0) - \boldsymbol{h}^{(1)}(w,\boldsymbol{\theta}_0)\right\}dw\, d\Lambda_0(u)\right], \end{aligned}$$

with F_c being the cumulative distribution of the censoring time. Then

$$\Phi(\Lambda) = \sigma_{22}(\Lambda) - \boldsymbol{\sigma}_{21}\boldsymbol{J}_0^{-1}\boldsymbol{\sigma}_{12}(\Lambda) = \int_0^t Q(u)d\Lambda(u) + \int_0^\tau R(t,u)d\Lambda(u). \tag{1.1}$$

The invertibility of Φ is equivalent to show that there exists a unique solution to the equation $\Phi(\Lambda) = \widetilde{\Lambda}$ for any function $\widetilde{\Lambda} \in L_2[0,\tau]$. Define $\dot{R}(t,u) = \partial R(t,u)/\partial t$. Taking the derivative with respect to t on both sides of (1.1), we have

$$d\widetilde{\Lambda}(t) = Q(t)d\Lambda(t) + \int_0^\tau \dot{R}(t,u)d\Lambda(u),$$

which is a Fredholm equation of the second type. By Assumptions (A1) $\sim$ (A3), the bivariate function $\dot{R}(t,u)$ is continuous on $[0,\tau]\times[0,\tau]$ and the function $Q(t)$ is continuous and bounded away from 0 for $t > 0$. Then it follows from the classical theory for integral equation (Tricomi, 1985, Chap 2) that there is a unique solution $d\Lambda(t)$ to the Fredholm integral equation, characterized by

$$d\Lambda(t) = \frac{d\widetilde{\Lambda}(t)}{Q(t)} - \int_0^\tau \frac{\eta(t,w)}{Q(t)}d\widetilde{\Lambda}(w)$$

where the function $\eta(u,v)$ satisfies

$$\eta(u,w) = -\frac{\dot{R}(u,w)}{Q(w)} - \int_0^\tau \eta(u,s)\frac{\dot{R}(s,w)}{Q(w)}ds.$$

Thus we show the invertibility of the functional Φ, where the inverse operator is

$$\Phi^{-1}(\Lambda)(t) = \int_0^t \frac{d\Lambda(u)}{Q(u)} - \int_0^\tau \int_0^t \frac{\eta(u,v)}{Q(t)}du\, d\Lambda(v).$$

It is easy to see that $\boldsymbol{U}_n(\boldsymbol{\varphi}) = \{\boldsymbol{U}_{n1}(\boldsymbol{\varphi}), U_{n2}(\boldsymbol{\varphi})\}$ is the sum of independently and identically distributed stochastic processes. The weak convergence of $\sqrt{n}\{\boldsymbol{U}_{n1}(\boldsymbol{\varphi}_0) - \boldsymbol{\mathcal{U}}_1(\boldsymbol{\varphi}_0)\}$ to a multivariate normal distribution $\boldsymbol{\mathcal{W}}_1$ follows from the multivariate central limit theorem. Moreover, by applying the central limit theorem for processes with bounded variation (see Example 2.11.16 in van der Vaart and Wellner, 1996), it can be shown that $\sqrt{n}\{U_{n2}(\boldsymbol{\varphi}_0) - \mathcal{U}_2(\boldsymbol{\varphi}_0)\}$ converges weakly to a tight Gaussian process $\mathcal{W}_2$. Thus the weak convergence of $\sqrt{n}\boldsymbol{U}_n(\boldsymbol{\varphi}_0)$ to $\mathbb{W} = (\mathbb{W}_1, \mathbb{W}_2)$ follows from the continuous mapping theorem.

Finally, to apply the Z-theorem for infinite dimensional estimating equations, we need to establish the stochastic approximation $\| \sqrt{n}\{\boldsymbol{U}_n(\widehat{\boldsymbol{\varphi}}_n) - \mathcal{U}(\widehat{\boldsymbol{\varphi}}_n)\} - \sqrt{n}\{\boldsymbol{U}_n(\boldsymbol{\varphi}_0) - \mathcal{U}(\boldsymbol{\varphi}_0)\} \| = o_p(1)$. Let $\boldsymbol{U}(\boldsymbol{\theta},\Lambda)(t) = \{\boldsymbol{U}_1(\boldsymbol{\theta},\Lambda), U_2(\boldsymbol{\theta},\Lambda)(t)\}$ the likelihood equations based on a single observation (a, y, δ), that is,

$$\boldsymbol{U}_1(\boldsymbol{\theta},\Lambda) = \left\{ \frac{\boldsymbol{h}^{(1)}(a,\boldsymbol{\theta})}{h^{(0)}(a,\boldsymbol{\theta})} - \frac{\int_0^\tau \boldsymbol{h}^{(1)}(u,\boldsymbol{\theta})S(u)du}{\int_0^\tau h^{(0)}(u,\boldsymbol{\theta})S(u)du} \right\}.$$

and

$$U_2(\boldsymbol{\theta},\Lambda)(t) = \delta I(y \leqslant t) - \int_0^t I(y \geqslant u)d\Lambda(u) + \int_0^t \int_u^\tau \frac{S(w)}{\mu(\boldsymbol{\theta},\Lambda)}h(w,\boldsymbol{\theta})dwd\Lambda(u).$$

The likelihood equations are defined on $\Theta \times \mathbb{H}$. Let $\bar{\mathbb{H}}$ be the closed linear subspace generated by $\mathbb{H}$. Thus $\bar{\mathbb{H}} \subset BV[0,\tau]$, where $BV[0,\tau]$ is the space of functions of bounded variation on $[0,\tau]$. Let the norm $\|\cdot\|_{\Theta\times\bar{\mathbb{H}}}$ on $\Theta \times \bar{\mathbb{H}}$ defined as $\| (\boldsymbol{\theta},\Lambda) \|_{\Theta\times\bar{\mathbb{H}}} = \| \boldsymbol{\theta} \| + \| \Lambda \|_v$, where $\|\cdot\|$ is the Euclidean norm and $\|\cdot\|_v$ is the total variation norm.

We now show that the class of functions $\{\boldsymbol{U}(\boldsymbol{\varphi})(t) - \boldsymbol{U}(\boldsymbol{\varphi}_0)(t) : \| \boldsymbol{\varphi} - \boldsymbol{\varphi}_0 \|_{\Theta\times\bar{\mathbb{H}}} < \epsilon,\ t \in [0,\tau]\}$ is P_0-Donsker. It follows (A2) that the classes of functions $\{\boldsymbol{h}^{(1)}(a,\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ and $\{h^{(0)}(a,\boldsymbol{\theta}) :$

$\boldsymbol{\theta} \in \Theta\}$ are bounded, and hence both are P_0-Donsker. Moreover, because the function S_0 is of bounded variation on $[0, \tau]$ and $\int_0^\tau h^{(0)}(u, \boldsymbol{\theta})S(u)du$ is bounded away from 0, the classes of bounded functions $\{\int_0^\tau \boldsymbol{h}^{(1)}(u, \boldsymbol{\theta})S(u)du :\| \boldsymbol{\varphi} - \boldsymbol{\varphi}_0 \|_{\Theta\times\bar{\mathbb{H}}} < \epsilon\}$ and $\{\{\int_0^\tau h^{(0)}(u, \boldsymbol{\theta})S(u)du\}^{-1} :\| \boldsymbol{\varphi} - \boldsymbol{\varphi}_0 \|_{\Theta\times\bar{\mathbb{H}}} < \epsilon\}$ are also P_0-Donsker. Thus the class of functions $\{\boldsymbol{U}_1(\boldsymbol{\theta}, \Lambda) :\| \boldsymbol{\varphi} - \boldsymbol{\varphi}_0 \|_{\Theta\times\bar{\mathbb{H}}} < \epsilon\}$ is P_0-Donsker, as the summation and production of Donsker classes are also Donsker. Similarly, we can show that the classes $\{\int_0^t \int_u^\tau S(w)h(w, \boldsymbol{\theta})dwd\Lambda(u) :\| \boldsymbol{\varphi} - \boldsymbol{\varphi}_0 \|_{\Theta\times\bar{\mathbb{H}}} < \epsilon, t \in [0, \tau]\}$ and $\{\mu(\boldsymbol{\theta}, \Lambda) :\| \boldsymbol{\varphi} - \boldsymbol{\varphi}_0 \|_{\Theta\times\bar{\mathbb{H}}} < \epsilon\}$ are P_0-Donsker, as they have uniformly bounded envelop functions. Again, by applying the fact that sums, productions and Lipschitz transformations of P_0-Donsker classes are still P_0-Donsker, we can show that $\{U_2(\boldsymbol{\theta}, \Lambda) :\| \boldsymbol{\varphi} - \boldsymbol{\varphi}_0 \|_{\Theta\times\bar{\mathbb{H}}} < \epsilon\}$ is P_0-Donsker. Now, because $\boldsymbol{U}(\boldsymbol{\varphi})(t)$ converges to $\boldsymbol{U}(\boldsymbol{\varphi}_0)(t)$ as $\| \boldsymbol{\varphi} - \boldsymbol{\varphi}_0 \|_{\Theta\times\bar{\mathbb{H}}} \to 0$ for any $t \in [0, \tau]$ and the convergence also holds in the square moment by the dominated convergence theorem, we have $\sup_{t\in[0,\tau]} E_0 \| \boldsymbol{U}(\boldsymbol{\varphi})(t) - \boldsymbol{U}(\boldsymbol{\varphi}_0)(t) \|^2_{\Theta\times\bar{\mathbb{H}}} \to 0$. Thus it follows from Lemma 3.3.5 of van der Vaart and Wellner (1996) that $n^{1/2}(\boldsymbol{U}_n - \boldsymbol{\mathcal{U}})(\hat{\boldsymbol{\varphi}}_n) - \sqrt{n}(\boldsymbol{U}_n - \boldsymbol{\mathcal{U}})(\boldsymbol{\varphi}_0) = o_p(1)$. The weak convergence of $n^{1/2}(\widehat{\boldsymbol{\varphi}}_n - \boldsymbol{\varphi}_0)$ to the mean zero Gaussian process $-\dot{\boldsymbol{\mathcal{U}}}_0^{-1}(\mathbb{W})$ now follows Theorem 3.3.1 of van der Vaart and Wellner (1996).

Let ϕ be the transformation from $(\boldsymbol{\theta}, \Lambda)$ to $(\boldsymbol{\theta}, F)$ with $\phi(\boldsymbol{\theta}_0, \Lambda_0) = (\boldsymbol{\theta}_0, F_0)$. It is known that the mapping is Hadamard differentiable. Applying the functional delta method, we can show that $n^{1/2}\{(\widehat{\boldsymbol{\theta}}_n, \widehat{F}_n) - (\boldsymbol{\theta}_0, F_0)\}$ converges weakly to a tight mean zero Gaussian process $-\boldsymbol{\phi}_0'\{\dot{\boldsymbol{\mathcal{U}}}^{-1}(\mathbb{W})\}$, where $\boldsymbol{\phi}_0'$ is the Hadamard derivative of $\boldsymbol{\phi}$ evaluated at $\boldsymbol{\psi}_0$.

2. Additional Simulation Results

We evaluated the power of the proposed semiparametric likelihood ratio test under various scenarios. We simulated survival time T^0 from a truncated exponential distribution with density function $\exp(-t)/\{1 - \exp(-10)\}$ for $t \in (0, 10]$ and $T^0/10$ from a beta distribution with parameters 0.5 and 5. The underlying truncation times were generated so that $A^0/10$ followed the uniform distribution on $[0, 1]$ and a beta distribution with shape parameters 0.75 and 1. The censoring times were generated from uniform distributions

so that the proportions of uncensored subjects were $100\%, 75\%, 50\%$. For each set of simulations, we considered different sample sizes $n = 100, 200$. The significance level of the semiparametric likelihood ratio test was set at 0.05. Table S.1 summaries the estimated size and power of the proposed semiparametric likelihood ratio test for testing $H_0 : \theta_1 = \theta_2 = \theta_3 = 0$ in the smooth alternative density (3.5).

For comparison, we also applied the paired logrank test proposed by Mandel and Betensky (2007) that compares the truncation time distribution and the residual survival time distribution, and reported the size and power of the test in Table S.1. When the underlying truncation time is uniformly distributed, the estimated sizes of both tests are close to the predetermined significance level (0.05). As expected, when the truncation time distribution is not uniform, the power to reject the null hypothesis increases with the sample size but decreases with the proportion of censored subjects. The proposed test is more powerful than the paired logrank test when the proportion of censored subjects is low, and is as efficient as its competitor when the censoring proportion is high.

Table S.1. Summary of power for the proposed test. LRT, the proposed semiparametric likelihood ratio test; PLR, paired log-rank test; Uniform, $A^0/10$ is generated from the uniform$(0, 1)$ distribution; Beta, $A^0/10$ is generated from the Beta$(0.75, 1)$ distribution.

		$T^0 \sim$ exp(1)				$T^0 \sim$ Beta$(0.5, 5)$			
		Uniform		Beta		Uniform		Beta	
n	pr$(\Delta = 1)$	LRT	PLR	LRT	PLR	LRT	PLR	LRT	PLR
100	1	0.06	0.05	0.60	0.50	0.06	0.05	0.57	0.47
	0.75	0.06	0.06	0.58	0.51	0.06	0.05	0.54	0.49
	0.50	0.06	0.05	0.53	0.53	0.05	0.05	0.50	0.50
200	1	0.05	0.05	0.88	0.78	0.06	0.05	0.87	0.78
	0.75	0.05	0.05	0.85	0.80	0.05	0.05	0.85	0.79
	0.50	0.05	0.04	0.82	0.81	0.05	0.05	0.80	0.80

REFERENCES

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

MANDEL, M AND BETENSKY, REBECCA A. (2007). Testing goodness of fit of a uniform truncation

model. *Biometrics* **63**, 405–412.

MURPHY, S A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23**, 182–198.

PARNER, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *Ann. Statist.* **26**, 183–214.

TRICOMI, F G. (1985). *Integral Equations*. New York: Dover Publications.

VAN DER VAART, A W AND WELLNER, JON A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer–Verlag.

VARDI, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. *Biometrika* **76**, 751–761.

ZEIDLER, E. (1995). *Applied Functional Analysis: Main Principles and Their Applications*. New York: Springer–Verlag.