Ann. Inst. Statist. Math.
Vol. 54, No. 3, 476–499 (2002)

ASYMPTOTIC THEORY FOR THE GAMMA FRAILTY MODEL WITH DEPENDENT CENSORING*

MICHAEL R. KOSOROK[1], JASON P. FINE[1], HONGYU JIANG[2] AND RICK CHAPPELL[1]

[1]*Department of Statistics, University of Wisconsin, 600 Highland Avenue, Box 4675, Madison, WI 53792, U.S.A.*
[2]*Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, MA 02115, U.S.A.*

(Received October 23, 2000; revised June 4, 2001)

Abstract. In many clinical studies, there are two dependent event times with one of the events being terminal, such as death, and the other being nonfatal, such as myocardial infarction or cancer relapse. Morbidity can be dependently censored by mortality, but not vice versa. Asymptotic theory is developed for simultaneous estimation of the marginal distribution functions in this *semi-competing risks* setting. We specify the joint distribution of the event times in the upper wedge, where the nonfatal event happens before the terminal event, with the popular gamma frailty model. The estimators are based on an adaptation of the self-consistency principle. To study their properties, we employ a modification of the functional delta-method applied to Z-estimators. This approach to weak convergence leads naturally to asymptotic validity of both the nonparametric and multiplier bootstraps, facilitating inference in spite of the complexity of the limiting distribution.

Key words and phrases: Bootstrap, dependent censoring, empirical processes, functional delta-method, gamma frailty model, U-statistics, weak convergence, Z-estimators.

1. Introduction

In many clinical studies, there are two dependent event times with one of the events (say Y) being terminal, such as death, and the other being a nonfatal event (say X) such as myocardial infarction or cancer relapse. Often, these studies also have an independent right-censoring time (say U) caused by random loss to follow-up. Because X can be dependently censored by Y but not vice versa, these data pose a *semi-competing risks* problem. In a recent multi-center clinical trial of allogenic marrow transplants in patients with acute leukemia, the primary endpoint was time to death while an important secondary endpoint was time to relapse (Copelan *et al.* (1991); Klein and Moeschberger (1997)). An important scientific question is how to estimate the distribution of the relapse times in the presence of the dependent censoring caused by death.

Jiang, Fine, Kosorok and Chappell (2001) (hereafter JFKC) propose a *pseudo self-consistency* method of estimation under the popular gamma frailty model (Clayton (1978)). However, the details of the asymptotic theory were not provided. The assumed

*This research was partially supported (M. R. Kosorok) by grant CA75142 from the U.S.A. National Cancer Institute.

model has the form $\mathbb{P}(X > x, Y > y) = C_{\theta_0}(S_0(x), R_0(y))$, where S_0 and R_0 satisfy the definition of survival functions, and where, for $\theta \geq 0$ and $u, v \in [0,1]$,

$$C_\theta(u,v) \equiv [(u^{1-\theta} + v^{1-\theta} - 1) \vee 0]^{1/(1-\theta)}. \tag{1.1}$$

Here we define $C_1(u,v) \equiv \lim_{\theta\to 1} C_\theta(u,v) = uv$ and $C_\infty(u,v) \equiv \lim_{\theta\to\infty} C_\theta(u,v) = u\wedge v$, where $a \vee b$ is the maximum and $a \wedge b$ the minimum of a and b. The model is only on the upper wedge where $x \leq y$. This is weaker than hypothesizing a parametric model for $x > y$, as in traditional competing risks analyses. The pseudo self-consistency methodology proposed in JFKC is insensitive to $\mathbb{P}(X > x, Y > y)$ on the lower wedge.

This robustness is important because the model on the lower wedge is nonidentifiable just as in the case of competing risks data (Tsiatis (1975)). In fact, the parameter S_0 is the marginal distribution of X only if $S_0(x) = \mathbb{P}(X > x, Y > 0)$. A class of distributions with this property follows. Let $\mathbb{P}(X > x, Y > y) = D\{S_0(x), R_0(y)\}$ for $x > y$, where $D(s,r) = \mathbb{P}(A > s, B > r)$, A, B are uniform(0,1) variates with unspecified joint distribution, and $D\{S_0(u), R_0(u)\} = C_{\theta_0}\{S_0(u), R_0(u)\}$, for all $u > 0$. Then $\mathbb{P}(X > x, Y > y)$ has the same S_0 and R_0 on both wedges. The copula D is nonparametric and X, Y may be dependent on the lower wedge. Observe that $R_0(y) = \mathbb{P}(X > 0, Y > y)$ is the marginal distribution of Y, regardless of the model for $x > y$.

Following Day *et al.* (1997), θ in the model on the upper wedge is interpretable as the predictive (Oakes (1989)) hazard ratio. For $x \leq y$, $\lambda(Y \mid \{X\})/\lambda(y \mid (x,\infty]) = \theta$, where $\lambda(y \mid A) = \lim_{\epsilon\to 0} d\{\mathbb{P}(Y < y + \epsilon \mid Y \geq y, X \in A)\}/d\epsilon$ and $A \subset (0,\infty)$. When $\theta = 1$, X and Y are independent on the upper wedge. Consider the following interpretation of the predictive hazard ratio on the upper wedge in the context of the leukemia example. Take two patients at time t: one that has just relapsed and one that has not yet relapsed. There is a θ-fold increase in the probability of death for the relapsed patient relative to the non-relapsed patient at all times $s > t$. This knowledge has clinical implications for disease management. Further details on the interpretation of this model are given in JFKC.

The basic idea of pseudo self-consistency is to first construct self-consistency equations (Efron (1967)) for S_0 and R_0 assuming that the association parameter, $\theta_0 \in [1,\infty)$, is known. Next, a consistent estimate of θ_0, $\hat{\theta}_n$, is substituted for θ_0 in these equations and the equations are then solved for the marginal distribution functions. A U-statistic estimating function can be used to estimate $\hat{\theta}_n$ separately from the marginals (Jiang *et al.* (1999)). An obvious alternative approach would be to maximize a nonparametric likelihood for the data to estimate θ_0, S_0 and R_0. The key challenge for establishing consistency would be to demonstrate that the Kullback-Leibler distance uniquely identifies the true parameters (see, for example, Murphy (1994)). Because pseudo self-consistency circumvents estimation of θ_0 via joint maximum likelihood, establishing identifiability of the marginal distributions is simplified. Of course, showing that the equations have a unique solution in the limit is still a formidable task. An added benefit of pseudo self-consistency is that computation of the estimates is relatively straightforward and converges reliably. Furthermore, simulation studies in JFKC demonstrate that the pseudo self-consistency estimates may be more efficient than maximum likelihood estimates for small and moderate sample sizes.

The fact that the pseudo self-consistency equations contain an estimated parameter precludes the use of existing theory for self-consistency equations (see Tsai and Crowley (1985); and Vardi and Zhang (1992)) and previous asymptotic results for the gamma frailty model (see Murphy (1994); Murphy (1995); and Parner (1998)). New techniques are required. We employ a modification of the functional delta-method applied to Z-estimators (van der Vaart and Wellner (1996), hereafter abbreviated VW). This modification involves using the Hoeffding decomposition to reduce U-statistics to sums of independent terms and using the identifiability of the score operators to facilitate verification of the Hadamard-differentiability of the map which assigns zeroes to the estimating equations. Modern empirical process theory then gives weak convergence. The approach leads naturally to the asymptotic validity of both the nonparametric and the multiplier bootstraps, facilitating inference in spite of the complexity of the limiting distribution.

In Section 2, we formalize the semi-competing risks data and precisely specify the gamma frailty model. We also present the pseudo self-consistency equations and prove existence of solutions. Uniform consistency of the estimators of the marginal distribution functions is demonstrated in Section 3. Weak convergence of the estimators to tight Gaussian elements is given in Section 4, while validity of both the nonparametric and multiplier bootstraps is established in Section 5.

2. The statistical model and pseudo self-consistency estimators

In this section, we present the data model and assumptions, describe estimation of the association parameter θ_0, and present the pseudo self-consistency estimators.

2.1 *The model*

Model 1. The data $\{Z_i, i = 1,\ldots,n\}$ consist of n i.i.d. realizations of $Z \equiv (X', Y', \eta, \delta)$, where $X' \equiv X \wedge Y \wedge U$, $Y' \equiv Y \wedge U$, $\eta \equiv \{X' = X\}$, $\delta \equiv \{Y' = Y\}$, and $\{B\}$ is the indicator of B. X and Y have bivariate decrement function $\mathbb{P}(X > x, Y > y) = C_{\theta_0}(S_0(x), R_0(y))$ in the upper wedge ($x \leq y$), where $1 \leq \theta_0 < \infty$, C_θ is defined in (1.1), and S_0 and R_0 are continuous survivor functions. U is independent of X and Y with continuous survival function L. Let P_0 denote the probability measure for Z.

ASSUMPTION 1. $t_0 < \infty$ is a positive time such that $L(t_0) > 0$, $S_0(t_0) > 0$, and $R_0(t_0) > 0$. Accordingly, let ϵ_0, $\epsilon_1 > 0$ be chosen so that $\epsilon_1 < [\log 2/\log(1/\epsilon_0)] \wedge (1/2)$ and $L(t_0)S_0(t_0)R_0(t_0) > 2\epsilon_0$.

Remark 1. An observation generated by Model 1 exists on the probability space $(\Omega, \mathcal{B}, P_0)$, where $\Omega = \mathbb{R}^2 \times \{0,1\}^2$ and where $\mathcal{B}$ is the Borel σ-algebra on Ω. Data generated according to this model exist on the product probability space $(\Omega^\infty, \mathcal{B}^\infty, \mathcal{P}_0)$, where Ω^∞ is the space of all infinite sequences $Z_1, Z_2, \ldots$ with Borel σ-algebra $\mathcal{B}^\infty$ generated by the product topology on $\prod_{i=1}^\infty \Omega$, and where $\mathcal{P}_0$ is the product measure based on P_0. Let $\mathbb{P}_n$ denote the corresponding empirical probability measure for a sample of size n.

2.2 *Estimating θ_0*

We will use the family of estimators described in Jiang *et al.* (1999). These estimators generalize the concordance estimator of Oakes (1986) to the semi-competing risks setting. For any two independent pairs of failure times (X_i, Y_i) and (X_j, Y_j), $1 \leq i, j \leq n$, define

$$\Delta_{ij} = \begin{cases} 1 & \text{if} \quad (X_i - X_j)(Y_i - Y_j) > 0, \\ 0 & \text{if} \quad (X_i - X_j)(Y_i - Y_j) \leq 0. \end{cases}$$

Let $\tilde{X}_{ij} = X_i \wedge X_j$, $\tilde{Y}_{ij} = Y_i \wedge Y_j$, $\tilde{U}_{ij} = U_i \wedge U_j$, $\tilde{X}'_{ij} = X'_i \wedge X'_j$, $\tilde{Y}'_{ij} = Y'_i \wedge Y'_j$, and $D_{ij} = \{\tilde{X}_{ij} \leq \tilde{Y}_{ij} \leq \tilde{U}_{ij}\}$. Note that Δ_{ij} is observable if and only if $D_{ij} = 1$.

Remark 2. By arguments given in Oakes (1986), we have, for $1 \leq i < j \leq n$, that Δ_{ij} is unconditionally independent of $(\tilde{X}'_{ij}, \tilde{Y}'_{ij})$ and $\mathbb{E}\Delta_{ij} = \theta_0/(1+\theta_0)$.

Remark 3. By arguments given in Jiang *et al.* (1999), we have, for $1 \leq i < j \leq n$, that D_{ij} can always be determined from the censored data of Model 1.

The estimator of θ_0 proposed by Jiang *et al.* (1999) is a zero of the U-statistic estimating function

$$G_n(\theta) \equiv n^{-2} \sum_{1 \leq i < j \leq n} W_n(\tilde{X}'_{ij}, \tilde{Y}'_{ij}) D_{ij} [\Delta_{ij} - \theta/(\theta+1)], \tag{2.1}$$

where W_n is a weight function depending on the data which satisfies Assumption 2 below. For a set $\mathcal{F}$, let $\ell^\infty(\mathcal{F})$ be the set of all uniformly bounded real functions on $\mathcal{F}$. We assume this space has the uniform topology, unless otherwise stated.

ASSUMPTION 2. W_n in (2.1) is non-negative and has the form $W_n(u,v) = \nu\{\mathbb{P}_n(\cdot)\}(u,v)$, where $(\cdot)$ ranges over a P_0-Glivenko-Cantelli class $\mathcal{X}$ and $\nu : \ell^\infty(\mathcal{X}) \mapsto \ell^\infty([0,\infty) \times [0,\infty))$ is continuous, with $W_0(u,v) \equiv \nu\{P_0(\cdot)\}(u,v)$ uniformly bounded.

Consider $W'_n(u,v) = [n^{-1} \sum_{i=1}^n \{X'_i \geq \hat{a} \wedge u, Y'_i \geq \hat{b} \wedge v\}]^{-1}$, where $\hat{a}$ is the $1-\alpha$ quantile of $X'_1, \ldots, X'_n$ and $\hat{b}$ is the $1-\alpha$ quantile of $Y'_1, \ldots, Y'_n$. One can show that this weight is a continuous mapping of $\mathbb{P}_n(\cdot)$, where $(\cdot)$ ranges over the P_0-Glivenko-Cantelli class $\mathcal{X} \equiv (\{X' \geq x, Y' \geq y\} : x, y \geq 0)$.

Remark 4. Although Assumption 2 requires specifying a fixed class of functions, its generality permits the use of adaptive weights like W'_n.

The proof of consistency for the following is given in Jiang *et al.* (1999):

LEMMA 1. *Under Model* 1 *and Assumptions* 1 *and* 2, *let $\hat{\theta}_n$ be the unique solution of $G_n(\theta) = 0$ (allowing $\hat{\theta}_n = \infty$ if necessary), if such a solution exists; if no such solution exists (i.e., $\sum_{1 \leq i < j \leq n} D_{ij} = 0$), take $\hat{\theta}_n = 1 - \epsilon_1/2$. Then $\hat{\theta}_n$ is almost surely consistent for θ_0.*

2.3 *Pseudo self-consistency*

JFKC demonstrate that the self-consistency equations for S_0 and R_0 based on maximum likelihood with θ_0 assumed known are

$$\mathbb{P}_n \psi_j^{\theta_0}(Z; S, R, t) = 0 \qquad (j = 1, 2), \tag{2.2}$$

where t ranges over a subset of the nonnegative reals, S and R are survival functions,

$$\psi_1^\theta(Z;S,R,t) \equiv S(t) - \left[\{X'>t\} + (1-\eta)\{Y'\le t\}\left((1-\delta)\frac{C_\theta(S(t),R(Y'))}{C_\theta(S(Y'),R(Y'))} + \delta\frac{C_\theta^\theta(S(t),R(Y'))}{C_\theta^\theta(S(Y'),R(Y'))}\right)\right],$$

and

$$\psi_2^\theta(Z;S,R,t) \equiv R(t) - \left[\{Y'>t\} + (1-\delta)\{Y'\le t\}\left((1-\eta)\frac{C_\theta(S(Y'),R(t))}{C_\theta(S(Y'),R(Y'))} + \eta\frac{C_\theta^\theta(S(X'),R(t))}{C_\theta^\theta(S(X'),R(Y'))}\right)\right].$$

When (2.2) is solved, the solutions $\hat{S}$ and $\hat{R}$ are maximum likelihood estimators with respect to model 1 by Theorem 3 of Robertson and Uppuluri (1984). Denote

$$\psi^\theta(Z;S,R,(s,t)) \equiv \begin{pmatrix} \psi_1^\theta(Z;S,R,s) \\ \psi_2^\theta(Z;S,R,t) \end{pmatrix}.$$

For pseudo self-consistency, we substitute a consistent estimator $\hat{\theta}_n$ for θ_0 in ψ^{θ_0} and solve $\mathbb{P}_n\psi^{\hat{\theta}_n}(Z;S,R,(s,t)) = 0$, for s,t ranging over a subset (specified below) of the positive reals.

For $0 \le \epsilon \le 1$ and $t \ge 0$, let $\boldsymbol{H}_\epsilon^t$ be the space of non-increasing real functions $f:[0,t]\mapsto[\epsilon,1]$ with $f(0)=1$. Define $T_n \equiv 0 \vee \sup(t : \mathbb{P}_n\{X'>t\} > r_n)$ and

$$r_n \equiv \begin{cases} 0 & \text{if } \hat{\theta}_n \ge 1, \\ (1/2)^{1/(1-\hat{\theta}_n)} & \text{if } \hat{\theta}_n < 1. \end{cases}$$

Define $\boldsymbol{A}_n, \boldsymbol{B}_n \subset \boldsymbol{H}_0^\infty$ to be all piecewise constant functions with jumps only at $\{X_i' : \eta_i = 1, X_i' \le T_n \wedge t_0, i=1,\ldots,n\}$ for $\boldsymbol{A}_n$ and at $\{Y_i' : \delta_i = 1, Y_i' \le T_n \wedge t_0, i=1,\ldots,n\}$ for $\boldsymbol{B}_n$.

The following lemma defines and establishes the existence of the pseudo self-consistency estimators $\hat{S}_n$ and $\hat{R}_n$:

LEMMA 2. *Under Model* 1, *Assumption* 1, *and provided* $\hat{\theta}_n \to \theta_0$ *almost surely, we have with probability* 1 *that* $T_n \ge t_0$ *for all* n *large enough. Furthermore, there exist solutions* $\hat{S}_n \in \boldsymbol{H}_0^{t_0} \cap \boldsymbol{A}_n$ *and* $\hat{R}_n \in \boldsymbol{H}_0^{t_0} \cap \boldsymbol{B}_n$ *of* $\mathbb{P}_n\psi^{\hat{\theta}_n}(Z;S_n,R_n,(s\wedge T_n, t\wedge T_n)) = 0$ *for all* $s,t \in [0,t_0]$.

The next lemma is needed for the proof of Lemma 2:

LEMMA 3. *Suppose for some* $t \ge 0$, $\hat{S} \in \boldsymbol{H}_0^t \cap \boldsymbol{A}_n$, $\hat{R} \in \boldsymbol{H}_0^t \cap \boldsymbol{B}_n$, *and* $S,R \in \boldsymbol{H}_0^t$. *Let* $\hat{S}'(s) = \hat{S}(s) - \mathbb{P}_n\psi_1^{\hat{\theta}_n}(Z;\hat{S},\hat{R},s)$, $\hat{R}'(s) = \hat{R}(s) - \mathbb{P}_n\psi_2^{\hat{\theta}_n}(Z;\hat{S},\hat{R},s)$, $S'(s) = S(s) - P_0\psi_1^{\theta_0}(Z;S,R,s)$ *and* $R'(s) = R(s) - P_0\psi_2^{\theta_0}(Z;S,R,s)$, *for all* $s \in [0,t]$. *Then* $\hat{S}' \in \boldsymbol{H}_{\tau_1}^t \cap \boldsymbol{A}_n$, $\hat{R}' \in \boldsymbol{H}_{\tau_1}^t \cap \boldsymbol{B}_n$, *and* $S',R' \in \boldsymbol{H}_{\tau_2}^t$, *where* $\tau_1 = \mathbb{P}_n\{X'>t\}$ *and* $\tau_2 = P_0\{X'>t\}$.

PROOF. For a cadlag (right-continuous with left-hand limits) function F, let $d_sF(s) \equiv F(s+ds) - F(s-)$, where $F(s-) = \lim_{t\uparrow s} F(t)$. For any $s \in (0,t]$,

$$
\begin{aligned}
nd_s\hat{S}'(s) = &\sum_{i=1}^{n} d_s\{X_i' > s\} + \sum_{i=1}^{n}(1-\eta_i)(1-\delta_i)\left[d_s\{X_i' \le s\}\frac{C_\theta(\hat{S}(s),\hat{R}(Y_i'))}{C_\theta(\hat{S}(X_i'),\hat{R}(Y_i'))}\right.\\
&\qquad\qquad\qquad\qquad\qquad\qquad\left. + \{X_i' < s\}\frac{d_sC_\theta(\hat{S}(s),\hat{R}(Y_i'))}{C_\theta(\hat{S}(X_i'),\hat{R}(Y_i'))}\right]\\
&+\sum_{i=1}^{n}(1-\eta_i)\delta_i\left[d_s\{X_i' \le s\}\frac{C_\theta^\theta(\hat{S}(s),\hat{R}(Y_i'))}{C_\theta^\theta(\hat{S}(X_i'),\hat{R}(Y_i'))}\right.\\
&\qquad\qquad\qquad\left. + \{X_i' < s\}\frac{d_sC_\theta^\theta(\hat{S}(s),\hat{R}(Y_i'))}{C_\theta^\theta(\hat{S}(X_i'),\hat{R}(Y_i'))}\right].
\end{aligned}
$$

Thus $\hat{S}'$ is in $\boldsymbol{H}_0^t \cap \boldsymbol{A}_n$ when $\hat{S} \in \boldsymbol{H}_0^t \cap \boldsymbol{A}_n$ and $\hat{R} \in \boldsymbol{H}_0^t$. Moreover, $\hat{S}'(0) = 1$ and $\hat{S}'(t) \ge \mathbb{P}_n\{X' > t\} = \tau_1$ by definition of ψ_1^θ. Hence $\hat{S}' \in \boldsymbol{H}_{\tau_1}^t \cap \boldsymbol{A}_n$. Similar arguments establish the result for $\hat{R}'$ and, after taking expectations, also for S' and R'. □

PROOF OF LEMMA 2. The almost sure consistency of $\hat{\theta}_n$ combined with Assumption 1 and the almost sure consistency of $\mathbb{P}_n\{X' > t_0\}$ gives, with probability 1, that $T_n \ge t_0$ for all n large enough. Assume $S, R \in \boldsymbol{H}_0^{T_n \wedge t_0}$ with $S(t)$, $R(t) \ge \mathbb{P}_n\{X' > t\}$ for all $t \in [0, T_n \wedge t_0]$. Further restrict S to have mass only at the values $\{X_i' : \eta_i = 1, X_i' \le T_n \wedge t_0, i = 1, \ldots, n\}$, and denote these masses Z_{1j}, $j = 1, \ldots, m_1$ (m_1 can be zero). Similarly restrict R to have mass only at the values $\{Y_i' : \delta_i = 1, Y_i' \le T_n \wedge t_0, i = 1, \ldots, n\}$, and denote these masses Z_{2j}, $j = 1, \ldots, m_2$. Since $\mathbb{P}_n\{X' > t\} > (1/2)^{1/(1-\hat{\theta}_n \wedge 1)}$ for all $t \in [0, T_n \wedge t_0)$, $C_{\hat{\theta}_n}(S(x), R(y))$ is positive for all $x, y \in [0, T_n \wedge t_0)$. If either $x = T_n$ or $y = T_n$, $C_{\hat{\theta}_n}(S(x), R(y))$ might be zero. But, because S only jumps at uncensored values of X_i' and R only jumps at uncensored values of Y_i', no divisions by zero occur when computing $\mathbb{P}_n\psi^{\hat{\theta}_n}(Z; S, R, (s,t))$ for all $s, t \in [0, T_n \wedge t_0]$. Thus $C_{\hat{\theta}_n}(S(x), R(y))$ is well defined for all $x, y \in [0, T_n \wedge t_0]$.

Let $\boldsymbol{H}_{(k)} \equiv \{u_{kj}, \ldots, u_{km_k} : u_{kj} \ge 0$, for $j = 1, \ldots, m_k, \sum_{j=1}^{m_k} u_{kj} \le 1\}$, for $k = 1, 2$, and define the map

$$
M\begin{pmatrix} S \\ R \end{pmatrix} \equiv \begin{pmatrix} S \\ R \end{pmatrix} - \mathbb{P}_n\begin{pmatrix} \psi_1^{\hat{\theta}_n}(Z; S, R, (\cdot) \wedge T_n) \\ \psi_2^{\hat{\theta}_n}(Z; S, R, (\cdot) \wedge T_n) \end{pmatrix},
$$

where $(\cdot)$ is used to denote the argument t as it ranges over $[0, t_0]$. The existence of solutions $\hat{S}_n(t)$ and $\hat{R}_n(t)$, for all $t \in [0, T_n \wedge t_0]$, will follow from Brouwer's finite-dimensional fixed-point theorem if the map $M : \boldsymbol{H}_{(1)} \times \boldsymbol{H}_{(2)} \mapsto \boldsymbol{H}_{(1)} \times \boldsymbol{H}_{(2)}$ is both continuous and into. Since M has domain $\boldsymbol{H}_{(1)} \times \boldsymbol{H}_{(2)}$, continuity follows from the form of ψ^θ and the fact that, for any $\theta \ge 0$ and $v \in [0,1]$, $C_\theta(u, v)$ is a continuous function of u over $[0,1]$. The into part is a direct consequence of Lemma 3. The result obtains if we assign, for $t \in [T_n \wedge t_0, t_0]$, $\hat{S}_n(t) = \hat{S}_n(T_n \wedge t_0)$ and $\hat{R}_n(t) = \hat{R}_n(T_n \wedge t_0)$. □

Lemma 2 permits:

ASSUMPTION 3. Let $\hat{S}_n$ and $\hat{R}_n$ be solutions of the pseudo self-consistency equations stopped at $T_n \wedge t_0$, as defined in Lemma 2.

Remark 5. Lemma 2 yields existence of solutions of the pseudo self-consistency equations but says nothing about uniqueness of those solutions. We later show that the solutions are asymptotically unique.

Remark 6. The fact that $\hat{S}_n$ jumps only at uncensored values of X' and $\hat{R}_n$ jumps only at uncensored values of Y' greatly simplifies computation. This feature is shared with the Kaplan-Meier estimator which jumps only at observed failure times.

3. Uniform consistency

In the sequel, we utilize the concepts of outer and inner probability and related ideas of weak convergence, convergence in outer probability, and outer almost sure convergence, as described in VW. The following theorem is the main result of this section:

THEOREM 1. *Assume Model* 1 *and Assumptions* 1 *and* 3 *hold and that $\hat{\theta}_n$ is almost surely consistent for θ_0. Then $\hat{S}_n$ and $\hat{R}_n$ are, uniformly on $[0, t_0]$, outer almost surely consistent for S_0 and R_0, respectively.*

The following lemma and theorem, the proofs of which will be given at the end of this section, are needed for the proof of Theorem 1:

LEMMA 4. *Assume Model* 1 *and Assumption* 1 *hold. Let ϵ_0, ϵ_1 be as defined in Assumption* 1. *Then*

$$\sup_{\theta\in[1-\epsilon_1,\theta_0+1]}\ \sup_{S,R\in \boldsymbol{H}_{\epsilon_0}^{t_0}}\ \sup_{Z\in\Omega}\ \sup_{s,t\in[0,t_0]}\left|\frac{\partial}{\partial\theta}\psi^\theta(Z;S,R,(s,t))\right| < \infty; \tag{3.1}$$

and the class of functions $\mathcal{F} \equiv \{\psi^\theta(Z;S,R,(s,t)) : \theta \in [1-\epsilon_1, \theta_0+1], s,t \in [0,t_0]$, and $S, R \in \boldsymbol{H}_{\epsilon_0}^{t_0}\}$ is P_0-Glivenko-Cantelli and P_0-Donsker.

THEOREM 2. *Assume Model* 1 *and Assumption* 1 *hold. Then for $S, R \in \boldsymbol{H}_{\epsilon_0}^{t_0}$, the following are equivalent:*

(i) *$P_0\psi_j^{\theta_0}(Z;S,R,t) = 0$ for all $t \in [0,t_0]$ and for $j = 1, 2$.*

(ii) *$S(t) = S_0(t)$ and $R(t) = R_0(t)$ for all $t \in [0,t_0]$.*

PROOF OF THEOREM 1. By Lemma 2, for large enough n, $\hat{S}_n$ and $\hat{R}_n$ are elements of $\boldsymbol{H}_{\epsilon_0}^{t_0}$ with inner probability one. By the almost sure consistency of $\hat{\theta}_n$, we have with probability 1 that $\hat{\theta}_n \in [1-\epsilon_1, \theta_0+1]$ for large enough n. Lemma 4 yields that as $n \to \infty$

$$\sup_{S,R\in \boldsymbol{H}_{\epsilon_0}^{t_0}}\ \sup_{t\in[0,t_0]} |\mathbb{P}_n\psi_j^{\hat{\theta}_n}(Z;S,R,t) - P_0\psi_j^{\theta_0}(Z;S,R,t)| \to 0$$

outer almost surely, $j = 1, 2$. Thus, by the Helly selection theorem, for every $\epsilon > 0$ there exists a $B \in \mathcal{B}^\infty$ such that $\mathcal{P}_0(B) > 1-\epsilon$ and, for each $\omega \in B$ with corresponding sequence of solutions $\{\hat{S}_{n\omega}, \hat{R}_{n\omega}, n \geq 1\}$, there exists a subsequence $\{n_k\}$ with $\lim_{k\to\infty}\int_0^{t_0}|\hat{S}_{n_k\omega}(t) - S(t)|dt = 0$ and $\lim_{k\to\infty}\int_0^{t_0}|\hat{R}_{n_k\omega}(t) - R(t)|dt = 0$ for some $S, R \in \boldsymbol{H}_{\epsilon_0}^{t_0}$ satisfying $P_0\psi^{\theta_0}(Z;S,R,(s,t)) = 0$, all $s, t \in [0,t_0]$. (The fact that $S, R \in \boldsymbol{H}_{\epsilon_0}^{t_0}$ is a consequence of Lemma 2.) Theorem 2 implies $S = S_0$ and $R = R_0$. Since S_0 and R_0 are continuous on $[0,t_0]$, L_1-convergence implies uniform convergence. Since these conclusions hold for every L_1-convergent subsequence, the uniform outer almost sure convergence follows. □

PROOF OF LEMMA 4. Fix $S, R \in \boldsymbol{H}_{\epsilon_0}^{t_0}$ and $x, y \in [0, t_0]$. For $u \in [-\epsilon_1, \theta_0]$, let $\zeta(u) \equiv \log[S^{-u}(x) + R^{-u}(y) - 1]$. Clearly,

$$\frac{\partial}{\partial u} C_{u+1}(S(x), R(y)) = C_{u+1}(S(x), R(y)) \left[\frac{\zeta(u)}{u^2} - \frac{\dot{\zeta}(u)}{u}\right],$$

where

$$\dot{\zeta}(u) \equiv \frac{\partial}{\partial u}\zeta(u) = -\frac{\log(S(x))S^{-u}(x) + \log(R(y))R^{-u}(y)}{S^{-u}(x) + R^{-u}(y) - 1}.$$

Let

$$\ddot{\zeta}(u) \equiv \frac{\partial}{\partial u}\dot{\zeta}(u) = \frac{\log^2(S(x))S^{-u}(x) + \log^2(R(y))R^{-u}(y)}{S^{-u}(x) + R^{-u}(y) - 1} - [\dot{\zeta}(u)]^2.$$

We have $\zeta(0) = 0$ and for all $u \in [-\epsilon_1, \theta_0]$,

$$\begin{aligned}
|\zeta(u)| &\leq \log 2 + \theta_0 \log(1/\epsilon_0) + 4\epsilon_1 \log(1/\epsilon_0), \\
|\dot{\zeta}(u)| &\leq \frac{2\log(1/\epsilon_0)\epsilon_0^{-\theta_0}}{2\epsilon_0^{\epsilon_1} - 1}, \\
|\ddot{\zeta}(u)| &\leq \frac{2\log^2(1/\epsilon_0)\epsilon_0^{-\theta_0}}{2\epsilon_0^{\epsilon_1} - 1} + |\dot{\zeta}(u)|^2.
\end{aligned}$$

These bounds also apply at $u = 0$ after taking limits. Thus

$$\begin{aligned}
\frac{\zeta(u)}{u^2} - \frac{\dot{\zeta}(u)}{u} &= \frac{\zeta(0) + \dot{\zeta}(0)u + \ddot{\zeta}(u')u^2/2 - u\dot{\zeta}(0) - \ddot{\zeta}(u'')u^2}{u^2} \\
&= \ddot{\zeta}(u')/2 - \ddot{\zeta}(u''),
\end{aligned}$$

where u' and u'' are between 0 and u. Now, for some $k_0 < \infty$ not depending on $S, R \in \boldsymbol{H}_{\epsilon_0}^{t_0}$ or $x, y \in [0, t_0]$,

$$\sup_{u \in [-\epsilon_1, \theta_0]} \left|\frac{\zeta(u)}{u^2} - \frac{\dot{\zeta}(u)}{u}\right| \leq k_0.$$

Since $C_{u+1}(S(x), R(y)) \leq 1$ and $C_{u+1}^{-1}(S(x), R(y)) \leq (2\epsilon_0^{\epsilon_1} - 1)^{-1/\epsilon_1}$ for all $u \in [-\epsilon_1, \theta_0]$, all $S, R \in \boldsymbol{H}_{\epsilon_0}^{t_0}$, and all $x, y \in [0, t_0]$, we have established (3.1).

Now, for $S_1, S_2, R_1, R_2 \in \boldsymbol{H}_{\epsilon_0}^{t_0}$, let $S_\lambda(t) = S_1(t) + \lambda(S_2(t) - S_1(t))$ and $R_\lambda(t) = R_1(t) + \lambda(R_2(t) - R_1(t))$. For $a, b, c, d \in [0, t_0]$ and $\theta \in (0, \infty)$,

$$\begin{aligned}
\frac{\partial}{\partial\lambda}\left[\frac{C_\theta(S_\lambda(a), R_\lambda(b))}{C_\theta(S_\lambda(c), R_\lambda(d))}\right] &= \frac{C_\theta(S_\lambda(a), R_\lambda(b))}{C_\theta(S_\lambda(c), R_\lambda(d))} \\
&\times (C_\theta^{\theta-1}(S_\lambda(a), R_\lambda(b))[S_\lambda^{-\theta}(a)(S_2(a) - S_1(a)) + R_\lambda^{-\theta}(b)(R_2(b) - R_1(b))] \\
&\quad - C_\theta^{\theta-1}(S_\lambda(c), R_\lambda(d))[S_\lambda^{-\theta}(c)(S_2(c) - S_1(c)) + R_\lambda^{-\theta}(d)(R_2(d) - R_1(d))]).
\end{aligned} \tag{3.2}$$

This implies that there exists a $k_1 < \infty$ such that

$$(3.3) \qquad |\psi_j^\theta(Z; S_2, R_2, t) - \psi_j^\theta(Z; S_1, R_1, t)| \le k_1 \left(\sup_{s\in[0,t]} |S_2(s) - S_1(s)| + \sup_{s\in[0,t]} |R_2(s) - R_1(s)| \right),$$

for $j = 1, 2$, all $t \in [0, t_0]$, and for all $\theta \in [1 - \epsilon_1, \theta_0 + 1]$.

By Theorem 2.7.5 of VW, we know for $r \ge 1$ that the log $L_r(Q)$-bracketing number for $\boldsymbol{H}_{\epsilon_0}^{t_0}$ at distance τ is of order K_r/τ, for every probability measure Q and for some $0 < K_r < \infty$ depending only on r. Thus, by (3.1) and (3.3), $\mathcal{F}$ is a Lipschitz function of several bounded Donsker classes. Hence $\mathcal{F}$ is Donsker (Theorem 2.10.6, VW). Since Donsker classes are also Glivenko-Cantelli, we are done. $\square$

PROOF OF THEOREM 2. The fact that (ii) implies (i) follows directly from the definitions. The challenge is to show that (i) implies (ii). Let $\bar{L} \equiv 1 - L$, $\bar{S}_0 \equiv 1 - S_0$, and $\bar{R}_0 \equiv 1 - R_0$. Fix $S, R \in \boldsymbol{H}_{\epsilon_0}^{t_0}$ and define S' and R' as in Lemma 3. Also let $\Delta_1 \equiv S - S_0$ and $\Delta_2 \equiv R - R_0$. For $\lambda \in [0, 1]$, define $S_\lambda = S_0 + \lambda(S - S_0)$ and $R_\lambda \equiv R_0 + \lambda(R - R_0)$. Using formula (3.2) and rearranging terms, we obtain for all $t \in [0, t_0]$,

$$
\begin{aligned}
(3.4) \quad & S'(t) - S_0(t) \\
&= \Bigg[\int_0^t \left(\int_0^1 \frac{C_{\theta_0}^{\theta_0}(S_\lambda(t), R_\lambda(U))}{C_{\theta_0}(S_\lambda(U), R_\lambda(U))} S_\lambda^{-\theta_0}(t) d\lambda \right) C_{\theta_0}(S_0(U), R_0(U)) d\bar{L}(U) \\
&\quad + \theta_0 \int_0^t \left(\int_0^1 \frac{C_{\theta_0}^{2\theta_0 - 1}(S_\lambda(t), R_\lambda(Y))}{C_{\theta_0}^{\theta_0}(S_\lambda(Y), R_\lambda(Y))} S_\lambda^{-\theta_0}(t) d\lambda \right) \\
&\quad \times C_{\theta_0}^{\theta_0}(S_0(Y), R_0(Y)) R_0^{-\theta_0}(Y) L(Y) d\bar{R}_0(Y) \Bigg] \Delta_1(t) \\
&\quad + \int_0^t \Bigg(\int_0^1 \frac{C_{\theta_0}(S_\lambda(t), R_\lambda(U))}{C_{\theta_0}(S_\lambda(U), R_\lambda(U))} [C_{\theta_0}^{\theta_0 - 1}(S_\lambda(t), R_\lambda(U)) R_\lambda^{-\theta_0}(U) \Delta_2(U) \\
&\qquad - C_{\theta_0}^{\theta_0 - 1}(S_\lambda(U), R_\lambda(U))(S_\lambda^{-\theta_0}(U)\Delta_1(U) + R_\lambda^{-\theta_0}(U)\Delta_2(U))] d\lambda \Bigg) \\
&\quad \times C_{\theta_0}(S_0(U), R_0(U)) d\bar{L}(U) \\
&\quad + \theta_0 \int_0^t \Bigg(\int_0^1 \frac{C_{\theta_0}^{\theta_0}(S_\lambda(t), R_\lambda(Y))}{C_{\theta_0}^{\theta_0}(S_\lambda(Y), R_\lambda(Y))} [C_{\theta_0}^{\theta_0 - 1}(S_\lambda(t), R_\lambda(Y)) R_\lambda^{-\theta_0}(Y) \Delta_2(Y) \\
&\qquad - C_{\theta_0}^{\theta_0 - 1}(S_\lambda(Y), R_\lambda(Y))(S_\lambda^{-\theta_0}(Y)\Delta_1(Y) + R_\lambda^{-\theta_0}(Y)\Delta_2(Y))] d\lambda \Bigg) \\
&\quad \times C_{\theta_0}^{\theta_0}(S_0(Y), R_0(Y)) R_0^{-\theta_0}(Y) L(Y) d\bar{R}_0(Y).
\end{aligned}
$$

Let the terms in the first pair of square brackets on the right-hand-side of (3.4) be denoted $K_1(t)$. Note, by the convexity of $\boldsymbol{H}_{\epsilon_0}^{t_0}$, that $S_\lambda, R_\lambda \in \boldsymbol{H}_{\epsilon_0}^{t_0}$ for all $\lambda \in [0, 1]$. Thus there exists $c_1 < \infty$ such that

$$(3.5) \quad |S'(t) - S(t)| \le K_1(t)|\Delta_1(t)| + c_1 \int_0^t (|\Delta_1(s)| + |\Delta_2(s)|)(d\bar{L}(s) + d\bar{R}_0(s))$$

for all $t \in [0, t_0]$. There also exists $c_2 < \infty$ such that $K_1(t) \leq c_2 \int_0^t (d\bar{L}(s) + d\bar{R}_0(s))$ for all $t \in [0, t_0]$. Assume $S(t) = S_0(t)$ for all $t \in [0, t_1]$, where $t_1 \leq t_0$. Then

$$
\begin{aligned}
(3.6) \qquad K_1(t_1) = & \int_0^{t_1} C_{\theta_0}^{\theta_0}(S_0(t), R_0(U)) S_0^{-\theta_0}(t) d\bar{R}_0(U) \\
& + \theta_0 \int_0^{t_1} C_{\theta_0}^{2\theta_0 - 1}(S_0(t), R_0(Y)) S_0^{-\theta_0}(t) R_0^{-\theta_0}(Y) L(Y) d\bar{R}_0(Y).
\end{aligned}
$$

The first term on the right-hand-side of (3.6) is $\int_0^{t_1} P_0(Y > U \mid X = t) d\bar{R}_0(U)$. Since

$$
\begin{aligned}
\theta_0 & \int_0^v C_{\theta_0}^{2\theta_0-1}(S_0(t), R_0(Y)) S_0^{-\theta_0}(t) R_0^{-\theta_0}(Y) d\bar{R}_0(Y) \\
& = \theta_0 \int_0^v [S_0(t)^{1-\theta_0} + R_0(Y)^{1-\theta_0} - 1]^{(2\theta_0-1)/(1-\theta_0)} S_0^{\theta_0}(t) R_0^{-\theta_0}(Y) d\bar{R}_0(Y) \\
& = \theta_0(\theta_0 - 1)^{-1} \int_1^{R_0(v)^{1-\theta_0}} [S_0^{1-\theta_0}(t) + w - 1]^{(2\theta_0-1)/(1-\theta_0)} dw S_0^{-\theta_0}(t) \\
& = 1 - C_{\theta_0}^{\theta_0}(S_0(t), R_0(v)) S_0^{-\theta_0}(t) \\
& = P_0(Y \leq v \mid X = t),
\end{aligned}
$$

the second term on the right-hand-side of (3.6) equals $\int_0^{t_1} L(v) d_v P_0(Y \leq v \mid X = t)$. Now (3.6) simplifies to $K_1(t_1) = \bar{L}(t_1) + L(t_1) P_0 (Y \leq t_1 \mid X = t_1)$. One can show that under Model 1, $P_0 (Y \leq t \mid X = t) \leq \bar{R}_0(t)$ for all $t \in [0, t_0]$ and thus $K_1(t_1) \leq \bar{L}(t_0) + L(t_0)\bar{R}_0(t_0) \equiv 1 - \tau_1$ for some $\tau_1 > 0$.

Using similar methods, for all $t \in [0, t_0]$,

$$
\begin{aligned}
(3.7) \; & R'(t) - R_0(t) \\
= & \Bigg[\int_0^t \left(\int_0^1 \frac{C_{\theta_0}^{\theta_0}(S_\lambda(U), R_\lambda(t))}{C_{\theta_0}(S_\lambda(U), R_\lambda(U))} R_\lambda^{-\theta_0}(t) d\lambda \right) C_{\theta_0}(S_0(U), R_0(U)) d\bar{L}(U) \\
& + \theta_0 \int_0^t \int_0^U \left(\int_0^1 \frac{C_{\theta_0}^{2\theta_0-1}(S_\lambda(X), R_\lambda(t))}{C_{\theta_0}^{\theta_0}(S_\lambda(X), R_\lambda(U))} R_\lambda^{-\theta_0}(t) d\lambda \right) \\
& \times C_{\theta_0}^{\theta_0}(S_0(X), R_0(U)) S_0^{-\theta_0}(X) d\bar{S}_0(X) d\bar{L}(U) \Bigg] \Delta_2(t) \\
& + \int_0^t \left(\int_0^1 \frac{C_{\theta_0}(S_\lambda(U), R_\lambda(t))}{C_{\theta_0}(S_\lambda(U), R_\lambda(U))} [C_{\theta_0}^{\theta_0-1}(S_\lambda(U), R_\lambda(t)) S_\lambda^{-\theta_0}(U) \Delta_1(U) \right. \\
& \qquad \left. - C_{\theta_0}^{\theta_0-1}(S_\lambda(U), R_\lambda(U))(S_\lambda^{-\theta_0}(U)\Delta_1(U) + R_\lambda^{-\theta_0}(U)\Delta_2(U))] d\lambda \right) \\
& \times C_{\theta_0}(S_0(U), R_0(U)) d\bar{L}(U) \\
& + \theta_0 \int_0^t \int_0^U \left(\int_0^1 \frac{C_{\theta_0}^{\theta_0}(S_\lambda(X), R_\lambda(t))}{C_{\theta_0}^{\theta_0}(S_\lambda(X), R_\lambda(U))} [C_{\theta_0}^{\theta_0-1}(S_\lambda(X), R_\lambda(t)) S_\lambda^{-\theta_0}(X) \Delta_1(X) \right. \\
& \qquad \left. - C_{\theta_0}^{\theta_0-1}(S_\lambda(X), R_\lambda(U))(S_\lambda^{-\theta_0}(X)\Delta_1(X) + R_\lambda^{-\theta_0}(U)\Delta_2(U))] d\lambda \right) \\
& \times C_{\theta_0}^{\theta_0}(S_0(X), R_0(U)) S_0^{-\theta_0}(X) d\bar{S}_0(X) d\bar{L}(U).
\end{aligned}
$$

Let the terms in the first pair of square brackets on the right-hand-side of (3.7) be denoted $K_2(t)$. We can find $c_3 < \infty$ such that

$$(3.8)\quad |R'(t) - R(t)| \le K_2(t)|\Delta_2(t)| + c_3 \int_0^t (|\Delta_1(s)| + |\Delta_2(s)|)\,(d\bar{L}(s) + d\bar{S}_0(s))$$

for all $t \in [0, t_0]$. There also exists $c_4 < \infty$ such that $K_2(t) \le c_4 \int_0^t (d\bar{L}(s) + d\bar{S}_0(s))$ for all $t \in [0, t_0]$. Assume $R(t) = R_0(t)$ for all $t \in [0, t_2]$, where $t_2 \le t_0$. Then

$$(3.9)\quad K_2(t_2) = \int_0^{t_2} C_{\theta_0}^{\theta_0}(S_0(U), R_0(t)) R_0^{-\theta_0}(t) d\bar{L}_0(U)$$
$$+\theta_0 \int_0^{t_2} \int_0^U C_{\theta_0}^{2\theta_0 - 1}(S_0(X), R_0(t)) R_0^{-\theta_0}(t) S_0^{-\theta_0}(X) d\bar{S}_0(Y) d\bar{L}(U).$$

After several change-of-variable steps similar to those used to evaluate $K_1(t_1)$, (3.9) becomes $K_2(t_2) = \bar{L}(t_2) \le \bar{L}(t_0) \equiv 1 - \tau_2$ for some $\tau_2 > 0$.

By the continuity of L, S_0, and R_0, we can construct a finite partition of $[0, t_0]$, $s_0, \ldots, s_k$, where $0 = s_0 < s_1 < \cdots < s_k = t_0$, such that

$$(c_1 + c_2 + c_3 + c_4) \int_{s_{j-1}}^{s_j} (d\bar{L}(s) + d\bar{S}_0(s) + d\bar{R}_0(s)) \le (\tau_1 \wedge \tau_2)/2,$$

for $j = 1, \ldots, k$. This gives that

$$\sup_{t\in[0,s_1]} |S'(t) - S_0(t)| + \sup_{t\in[0,s_1]} |R'(t) - R_0(t)|$$
$$\le \frac{\tau_1 \wedge \tau_2}{2} \left[\sup_{t\in[0,s_1]} |S(t) - S_0(t)| + \sup_{t\in[0,s_1]} |R(t) - R_0(t)| \right],$$

which means $S(t) = S_0(t)$ and $R(t) = R_0(t)$ for all $t \in [0, s_1]$ if S and R satisfy (i). If $S(t) = S_0(t)$ and $R(t) = R_0(t)$ for all $t \in [0, s_j]$ and $j < k$, and if S and R satisfy (i), then the restrictions on K_1 and K_2 imply

$$\sup_{t\in[s_j,s_{j+1}]} |S'(t) - S_0(t)| + \sup_{t\in[s_j,s_{j+1}]} |R'(t) - R_0(t)|$$
$$\le (1 - \tau_1/2) \sup_{t\in[s_j,s_{j+1}]} |S(t) - S_0(t)| + (1 - \tau_2/2) \sup_{t\in[s_j,s_{j+1}]} |R(t) - R_0(t)|.$$

Hence $S(t) = S_0(t)$ and $R(t) = R_0(t)$ for all $t \in [0, s_{j+1}]$. So, by induction, (i) implies (ii). □

4. Weak convergence

Let $\hat{\gamma}_n \equiv (\hat{\theta}_n, \hat{S}_n, \hat{R}_n)$ and $\gamma_0 \equiv (\theta_0, S_0, R_0)$. In this section, we establish weak convergence of $\sqrt{n}(\hat{\gamma}_n - \gamma_0)$ in the uniform topology on $\mathcal{L} \equiv \mathbb{R}^\Delta \times \mathbf{D}[0, t_0] \times \mathbf{D}[0, t_0]$, where $\mathbb{R}^\Delta = \mathbb{R} \cup \{-\infty, \infty\}$ and $\mathbf{D}[0, t_0]$ is the space of cadlag functions on $[0, t_0]$.

THEOREM 3. *Assume Model* 1 *obtains and Assumptions* 1, 2, *and* 3 *hold with* $\hat{\theta}_n$ *as defined in Lemma* 1. *Then* $\sqrt{n}(\hat{\gamma}_n - \gamma_0)$ *converges weakly in the uniform topology on* $\mathcal{L}$ *to a tight, mean zero Gaussian element* Y_0.

Asymptotic theory for M-estimators or Z-estimators (Chapters 3.2 and 3.3, VW) could possibly be used to prove this result. However, the functional delta method applied to Z-estimators (Section 3.9.4.7, VW) leads most naturally to weak convergence and consistency of the bootstrap in Section 5. A key ingredient of our proof is a modification of Lemma 3.9.34 of VW, Lemma 5 below, which takes advantage of the identifiability of the estimation procedure as reflected in Theorem 2. To state the lemma, we need several definitions. Let $\gamma \equiv (\theta, S, R)$,

$$\Psi_n(\gamma) \equiv \begin{pmatrix} \hat{\theta}_n - \theta \\ \mathbb{P}_n \psi^\theta(Z; S, R, (\cdot,\cdot) \wedge T_n) \end{pmatrix},$$

and

$$\Psi(\gamma) \equiv \begin{pmatrix} \theta_0 - \theta \\ P_0 \psi^\theta(Z; S, R, (\cdot,\cdot)) \end{pmatrix},$$

where $(\cdot,\cdot) \wedge T_n$ denotes s and t ranging over $[0, t_0 \wedge T_n]$. Denote $\Gamma \equiv [0,\infty] \times \boldsymbol{H}_0^{t_0} \times \boldsymbol{H}_0^{t_0}$ and $\Gamma_0 \equiv [1-\epsilon_1, \theta_0+1] \times \boldsymbol{H}_{\epsilon_0}^{t_0} \times \boldsymbol{H}_{\epsilon_0}^{t_0}$. Let $\ell^\infty(\Gamma, \mathcal{L})$ be the Banach space of all uniformly bounded functions $z : \Gamma \mapsto \mathcal{L}$ with $Z(\Gamma, \mathcal{L})$ being the subset consisting of all maps with at least one zero. Define $\phi : Z(\Gamma, \mathcal{L}) \mapsto \Gamma$ to be a map that assigns a zero $\phi(z)$ to each $z \in Z(\Gamma, \mathcal{L})$. Let $\ell^\infty(\Gamma_0, \mathcal{L})$ and $Z(\Gamma_0, \mathcal{L})$ be similarly defined, and let $\check{\phi}$ be the restriction of ϕ to $Z(\Gamma_0, \mathcal{L})$. Without loss of generality, assume that $\phi\Psi_n = \hat{\gamma}_n$ and $\phi(0) = \gamma_1 \equiv (\theta_1, S_1, R_1)$, where $\theta_1 = 1 - \epsilon_1/2$, and $S_1(t) = R_1(t) = 1$ for all $t \in [0, t_0]$.

We have the following key lemma:

LEMMA 5. *Under Model 1 and Assumption 1, the map $\check{\phi} : Z(\Gamma_0, \mathcal{L}) \subset \ell^\infty(\Gamma_0, \mathcal{L}) \mapsto \Gamma_0$ is Hadamard-differentiable at Ψ tangentially to the set of $z \in \ell^\infty(\Gamma_0, \mathcal{L})$ that are continuous at γ_0. The derivative is given by $\check{\phi}'_\Psi(z) = -\dot{\Psi}_{\gamma_0}^{-1}(z(\gamma_0))$, where, for $h = (h_1, h_2, h_3) \in \mathcal{L}$ with $h_1 \in \mathbb{R}$, $h_2 \in \mathbf{D}[0, t_0]$ and $h_3 \in \mathbf{D}[0, t_0]$,*

$$(4.1) \quad \dot{\Psi}_{\gamma_0}^{-1} h \equiv \begin{pmatrix} -1 & 0 & 0 \\ \dot{\Psi}_{(10)}^{-1} & \dot{\Psi}_{(11)}^{-1} & \dot{\Psi}_{(12)}^{-1} \\ \dot{\Psi}_{(20)}^{-1} & \dot{\Psi}_{(21)}^{-1} & \dot{\Psi}_{(22)}^{-1} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} \equiv \begin{pmatrix} -h_1 \\ \dot{\Psi}_{(10)}^{-1} \times h_1 + \dot{\Psi}_{(11)}^{-1} h_2 + \dot{\Psi}_{(12)}^{-1} h_3 \\ \dot{\Psi}_{(20)}^{-1} \times h_1 + \dot{\Psi}_{(21)}^{-1} h_2 + \dot{\Psi}_{(22)}^{-1} h_3 \end{pmatrix};$$

where for $1 \le j, k \le 2$, $\dot{\Psi}_{(jk)}^{-1} : \mathbf{D}[0, t_0] \mapsto \mathbf{D}[0, t_0]$ are linear operators and $\dot{\Psi}_{(j0)}^{-1} \in \mathbf{D}[0, t_0]$;

$$(4.2) \quad \begin{pmatrix} \dot{\Psi}_{(11)}^{-1} & \dot{\Psi}_{(12)}^{-1} \\ \dot{\Psi}_{(21)}^{-1} & \dot{\Psi}_{(22)}^{-1} \end{pmatrix} \equiv \sum_{k=0}^{\infty} \begin{pmatrix} I - \dot{\Psi}_{11} & -\dot{\Psi}_{12} \\ -\dot{\Psi}_{21} & I - \dot{\Psi}_{22} \end{pmatrix}^k;$$

where I is the identity operator on $\mathbf{D}[0, t_0]$; and for $j = 1, 2$,

$$(4.3) \quad \dot{\Psi}_{(j0)}^{-1} = \dot{\Psi}_{(j1)}^{-1} \dot{\Psi}_{10} + \dot{\Psi}_{(j2)}^{-1} \dot{\Psi}_{20}.$$

For $1 \le j, k \le 2$, $\dot{\Psi}_{j0} \in \mathbf{D}[0, t_0]$ are functions and $\dot{\Psi}_{jk} : \mathbf{D}[0, t_0] \mapsto \mathbf{D}[0, t_0]$ are linear operators defined as follows:

$$(4.4) \quad \dot{\Psi}_{10}(t) = -P_0 \Bigg[\{U \le t, X > U, Y > U\} \frac{C_{\theta_0}(S_0(t), R_0(U))}{C_{\theta_0}(S_0(U), R_0(U))} \times (\dot{C}_{\theta_0}(S_0(t), R_0(U)) - \dot{C}_{\theta_0}(S_0(U), R_0(U)))$$

$$+\{Y \le t, X > Y, Y \le U\} \frac{C_{\theta_0}^{\theta_0}(S_0(t), R_0(Y))}{C_{\theta_0}^{\theta_0}(S_0(Y), R_0(Y))}$$
$$\times \left(\log \left[\frac{C_{\theta_0}(S_0(t), R_0(Y))}{C_{\theta_0}(S_0(Y), R_0(Y))} \right] + \theta_0 [\dot{C}_{\theta_0}(S_0(t), R_0(Y))\right.$$
$$\left. \left. -\dot{C}_{\theta_0}(S_0(Y), R_0(Y))] \right) \right],$$

$$\dot{\Psi}_{20}(t) = -P_0 \left[\{U \le t, X > U, Y > U\} \frac{C_{\theta_0}(S_0(U), R_0(t))}{C_{\theta_0}(S_0(U), R_0(U))} \right.$$
$$\times (\dot{C}_{\theta_0}(S_0(U), R_0(t)) - \dot{C}_{\theta_0}(S_0(U), R_0(U)))$$
$$+ \{U \le t, X \le U, Y > U\} \frac{C_{\theta_0}^{\theta_0}(S_0(X), R_0(t))}{C_{\theta_0}^{\theta_0}(S_0(X), R_0(U))}$$
$$\times \left(\log \left[\frac{C_{\theta_0}(S_0(X), R_0(t))}{C_{\theta_0}(S_0(X), R_0(U))} \right] \right.$$
$$\left. \left. + \theta_0 [\dot{C}_{\theta_0}(S_0(X), R_0(t)) - \dot{C}_{\theta_0}(S_0(X), R_0(U))] \right) \right], \tag{4.5}$$

where

(4.6)
$$\dot{C}_{\theta_0}(u,v) \equiv \begin{cases} \dfrac{\log(u^{1-\theta_0} + v^{1-\theta_0} - 1)}{(\theta_0 - 1)^2} + \dfrac{u^{1-\theta_0}\log u + v^{1-\theta_0}\log v}{(\theta_0 - 1)(u^{1-\theta_0} + v^{1-\theta_0} - 1)}, & \textit{if} \quad \theta_0 > 1, \\ \log u \log v, & \textit{if} \quad \theta_0 = 1. \end{cases}$$

For $h \in \mathbf{D}[0, t_0]$ *and all* $t \in [0, t_0]$,

$$(\dot{\Psi}_{11}h)(t) = h(t) - \left[\int_0^t C_{\theta_0}^{\theta_0}(S_0(t), R_0(U)) S_0^{-\theta_0}(t) d\bar{L}(U) \right.$$
$$\left. + \theta_0 \int_0^t C_{\theta_0}^{2\theta_0 - 1}(S_0(t), R_0(U)) S_0^{-\theta_0}(t) R_0^{-\theta_0}(Y) L(Y) d\bar{R}_0(Y) \right] h(t)$$
$$+ \int_0^t C_{\theta_0}(S_0(t), R_0(U)) C_{\theta_0}^{\theta_0 - 1}(S_0(U), R_0(U)) S_0^{-\theta_0}(U) h(U) d\bar{L}(U)$$
$$+ \theta_0 \int_0^t C_{\theta_0}^{\theta_0}(S_0(t), R_0(Y)) C_{\theta_0}^{\theta_0 - 1}(S_0(Y), R_0(Y)) S_0^{-\theta_0}(Y) R_0^{-\theta_0}(Y) h(Y) d\bar{R}_0(Y), \tag{4.7}$$

$$(\dot{\Psi}_{12}h)(t) = -\int_0^t C_{\theta_0}(S_0(t), R_0(U))$$
$$\times [C_{\theta_0}^{\theta_0 - 1}(S_0(t), R_0(U)) - C_{\theta_0}^{\theta_0 - 1}(S_0(U), R_0(U))] R_0^{-\theta_0}(U) h(U) d\bar{L}(U)$$
$$- \theta_0 \int_0^t C_{\theta_0}^{\theta_0}(S_0(t), R_0(Y))$$
$$\times [C_{\theta_0}^{\theta_0 - 1}(S_0(t), R_0(Y)) - C_{\theta_0}^{\theta_0 - 1}(S_0(Y), R_0(Y))] R_0^{-2\theta_0}(Y) h(Y) L(Y) d\bar{R}_0(Y), \tag{4.8}$$

$$(\dot{\Psi}_{21}h)(t) = -\int_0^t C_{\theta_0}(S_0(U), R_0(t))$$
$$\times [C_{\theta_0}^{\theta_0 - 1}(S_0(U), R_0(t)) - C_{\theta_0}^{\theta_0 - 1}(S_0(U), R_0(U))] S_0^{-\theta_0}(U) h(U) d\bar{L}(U) \tag{4.9}$$

$$-\theta_0 \int_0^t \int_0^U C_{\theta_0}^{\theta_0}(S_0(X), R_0(t))$$
$$\times [C_{\theta_0}^{\theta_0 - 1}(S_0(X), R_0(t)) - C_{\theta_0}^{\theta_0 - 1}(S_0(X), R_0(U))] S_0^{-2\theta_0}(X) h(X) d\bar{S}_0(X) d\bar{L}(U),$$

$$(\dot{\Psi}_{22} h)(t) = h(t) - \left[\int_0^t C_{\theta_0}^{\theta_0}(S_0(U), R_0(t)) R_0^{-\theta_0}(t) d\bar{L}(U)\right. \tag{4.10}$$
$$\left. + \theta_0 \int_0^t \int_0^U C_{\theta_0}^{2\theta_0 - 1}(S_0(X), R_0(t)) R_0^{-\theta_0}(t) S_0^{-\theta_0}(X) d\bar{S}_0(X) d\bar{L}(U)\right] h(t)$$
$$+ \int_0^t C_{\theta_0}(S_0(U), R_0(t)) C_{\theta_0}^{\theta_0 - 1}(S_0(U), R_0(U)) R_0^{-\theta_0}(U) h(U) d\bar{L}(U)$$
$$+ \theta_0 \int_0^t \int_0^U C_{\theta_0}^{\theta_0}(S_0(X), R_0(t)) C_{\theta_0}^{\theta_0 - 1}(S_0(X), R_0(U))$$
$$\times R_0^{-\theta_0}(U) S_0^{-\theta_0}(X) d\bar{S}_0(X) h(U) d\bar{L}(U).$$

PROOF. Using the derivatives from the proof of Lemma 4, one can show that there exists an $M < \infty$ such that

$$\left\| \frac{\partial^2}{\partial \lambda \partial \mu} \Psi(\gamma_0 + \mu(\gamma - \gamma_0) + \lambda(\gamma - \gamma_0)) \right\| \le M \|\gamma - \gamma_0\|^2,$$

for all $\gamma \in \Gamma_0$, where $\|\cdot\|$ is the uniform norm. Thus, Ψ is Fréchet-differentiable at γ_0. Combining the proof of Lemma 4 with (3.4) and (3.7) gives that this derivative is the linear operator $\dot{\Psi}_{\gamma_0} : \mathcal{L} \mapsto \mathcal{L}$, where

$$\dot{\Psi}_{\gamma_0} \equiv \begin{pmatrix} -1 & 0 & 0 \\ \dot{\Psi}_{10} & \dot{\Psi}_{11} & \dot{\Psi}_{12} \\ \dot{\Psi}_{20} & \dot{\Psi}_{21} & \dot{\Psi}_{22} \end{pmatrix},$$

and where the components are defined in (4.4), (4.5), and (4.7)–(4.10).

Let

$$B = \begin{pmatrix} I - \dot{\Psi}_{11} & -\dot{\Psi}_{12} \\ -\dot{\Psi}_{21} & I - \dot{\Psi}_{22} \end{pmatrix}$$

and $D_1 = \{h \in \mathbf{D}[0, t_0] \times \mathbf{D}[0, t_0] : \|h\| \le 1\}$. The arguments used to prove Lemma 4 imply that B is a bounded linear operator on $\mathbf{D}[0, t_0] \times \mathbf{D}[0, t_0]$ and that there exists an $\epsilon > 0$ and a positive integer $r < \infty$ such that for all $h \in D_1$, $\|B^r h\| \le (1 - \epsilon)\|h\|$. Thus, for all $h \in D_1$,

$$\left\| \sum_{k=0}^{\infty} B^k h \right\| = \left\| \sum_{j=0}^{r-1} B^j \sum_{k=0}^{\infty} B^{rk} h \right\| \le \left\| \sum_{j=0}^{r-1} B^j \right\| \times \epsilon^{-1} \|h\|.$$

This implies that

$$\begin{pmatrix} \dot{\Psi}_{11} & \dot{\Psi}_{12} \\ \dot{\Psi}_{21} & \dot{\Psi}_{22} \end{pmatrix}^{-1} = (I - B)^{-1}$$

exists, is a bounded continuous linear operator on $\mathbf{D}[0,t_0] \times \mathbf{D}[0,t_0]$, and has the form given in (4.2). After a few additional calculations, we have that $\dot{\Psi}_{\gamma_0}$ is continuously invertible on the linear span of Γ_0, with inverse $\dot{\Psi}_{\gamma_0}^{-1}$ given in (4.1).

Let $z_t \to z$ uniformly on Γ_0, as $t \downarrow 0$, where $z : \Gamma_0 \mapsto \mathcal{L}$ is continuous at γ_0 and where $\{\Psi + tz_t, t > 0\} \in Z(\Gamma_0, \mathcal{L})$. By definition, the element $\gamma_t = \check{\phi}(\Psi + tz_t)$ satisfies $\Psi(\gamma_t) + tz_t(\gamma_t) = 0$. Hence $\Psi(\gamma_t) = O(t)$; and, by Theorem 2, $\gamma_t \to \gamma_0$. The remaining steps are identical to the last steps of the proof of Lemma 3.9.34 of VW. □

Before proving Theorem 3, we show that the estimators $\hat{\theta}_n$ satisfying Lemma 1 are asymptotically equivalent to i.i.d. sums. This is achieved by establishing equivalence to a second order U-statistic so that a Hoeffding representation applies. The difficulty is that the kernel is estimated from the data rather than fixed. This is resolved in Lemma 6 below, the proof of which will be given at the end of this section.

Remark 7. The weight functions permitted by Assumption 2 are more general than those in Oakes (1986). New techniques for obtaining the asymptotic structure of $\hat{\theta}_n$ are needed.

LEMMA 6. *Assume Model* 1 *obtains and Assumptions* 1 *and* 2 *hold. Then*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = I_0^{-1} n^{-1/2} \sum_{i=1}^{n} A_i + Q_n,$$

where Q_n goes to zero in probability as $n \to \infty$,

$$\begin{aligned} &I_0 \equiv (1+\theta_0)^{-2}\mathbb{E}[W_0(\tilde{X}'_{12}, \tilde{Y}'_{12})D_{12}]/2, \\ &A_i \equiv \mathbb{E}[W_0(\tilde{X}'_{ij}, \tilde{Y}'_{ij})D_{ij}[\Delta_{ij} - \theta_0/(1+\theta_0)] \mid Z_i], \qquad \text{any } j \neq i, \end{aligned} \tag{4.11}$$

and $\{A_i, i \geq 1\}$ is a sequence of i.i.d. real random variables. Furthermore, A_1 has mean zero, depends only on Z_1, and satisfies $|A_1| \leq \sup_{u,v \geq 0} W_0(u,v) < \infty$ almost surely.

PROOF OF THEOREM 3. Note that $\gamma_0 = \phi\Psi$ and $\hat{\gamma}_n = \phi\Psi_n$ even in the case that $G_n(\theta) = 0$ has no unique solution and $\hat{\theta}_n = 1 - \epsilon_1/2$ as in Lemma 1. By Lemma 2, we have with probability 1 that $T_n \geq t_0$ for all n large enough. By Lemma 1 and Theorem 1, we have with probability 1 that $\Psi_n \in Z(\Gamma_0, \mathcal{L})$ for all n large enough. Therefore,

$$\sqrt{n}\,(\phi\Psi_n - \phi\Psi) - \sqrt{n}\,(\phi\Psi_n - \phi\Psi)\,\{\Psi_n \in Z(\Gamma_0, \mathcal{L})\} \to 0$$

in outer probability. In addition,

$$\begin{aligned} \sqrt{n}(\phi\Psi_n - \phi\Psi)\{\Psi_n \in Z(\Gamma_0,\mathcal{L})\} &= \sqrt{n}(\phi[\Psi_n\{\Psi_n \in Z(\Gamma_0,\mathcal{L})\}] - \phi\Psi)\{\Psi_n \in Z(\Gamma_0,\mathcal{L})\} \\ &= \sqrt{n}(\check{\phi}[\Psi_n\{\Psi_n \in Z(\Gamma_0,\mathcal{L})\}] - \check{\phi}\Psi)\{\Psi_n \in Z(\Gamma_0,\mathcal{L})\}. \end{aligned}$$

By Lemmas 4 and 6, $\sqrt{n}\,(\Psi_n\,\{\Psi_n \in Z(\Gamma_0,\mathcal{L})\} - \Psi)$ converges HJD-weakly in the uniform topology on $\ell^\infty(\Gamma_0,\mathcal{L})$ to a tight—and hence separable—Gaussian process Z_0. Combining this with Lemma 5 and the functional delta method (Theorem 3.9.4, VW) then yields that $\sqrt{n}(\check{\phi}[\Psi_n\{\Psi_n \in Z(\Gamma_0,\mathcal{L})\}] - \check{\phi}\Psi)\{\Psi_n \in Z(\Gamma_0,\mathcal{L})\}$ converges HJD-weakly to $Y_0 \equiv \check{\phi}'_\Psi(Z_0(\gamma_0))$, where Y_0 is tight by the continuity of $\check{\phi}'_\Psi$. □

PROOF OF LEMMA 6. Let $H_n(x,y) = W_n(x,y) - W_0(x,y)$. Define $\mathcal{G}_n = \sigma\{\mathcal{F}_n, \tilde{X}_{ij}, \tilde{Y}_{ij}, D_{ij}, 1 \leq i < j \leq n\}$, where $\mathcal{F}_n$ is the exchangeable σ-field based on the data $\{Z_i, i = 1, \ldots, n\}$ (exchangeable with respect to permuting the subscripts), and where $\sigma\{B\}$ is the smallest σ-field making all of B measurable. Denote $\alpha_{ij} = \Delta_{ij} - \theta_0/(1+\theta_0)$. Now,

$$
\begin{aligned}
(4.12)\quad & \left(n^{-3/2} \sum_{1\leq i<j\leq n} H_n(\tilde{X}_{ij}, \tilde{Y}_{ij}) D_{ij}\alpha_{ij}\right)^2 \\
&= n^{-3} \sum_{1\leq i<j\leq n} H_n^2(\tilde{X}_{ij}, \tilde{Y}_{ij}) D_{ij}\alpha_{ij}^2 \\
&\quad +2n^{-3} \sum_{1\leq i<j<k\leq n} [H_n(\tilde{X}_{ij}, \tilde{Y}_{ij}) H_n(\tilde{X}_{ik}, \tilde{Y}_{ik}) D_{ij} D_{ik}\alpha_{ij}\alpha_{ik} \\
&\qquad\qquad + H_n(\tilde{X}_{ij}, \tilde{Y}_{ij}) H_n(\tilde{X}_{jk}, \tilde{Y}_{jk}) D_{ij} D_{jk}\alpha_{ij}\alpha_{jk} \\
&\qquad\qquad + H_n(\tilde{X}_{ik}, \tilde{Y}_{ik}) H_n(\tilde{X}_{jk}, \tilde{Y}_{jk}) D_{ik} D_{jk}\alpha_{ik}\alpha_{jk}] \\
&\quad + \sum_{(i,j,k,l)\in N_n(4)} H_n(\tilde{X}_{ij}, \tilde{Y}_{ij}) H_n(\tilde{X}_{kl}, \tilde{Y}_{kl}) D_{ij} D_{kl}\alpha_{ij}\alpha_{kl},
\end{aligned}
$$

where $N_n(4)$ represents all quadruples (i,j,k,l) such that $1 \leq i < j \leq n$, $1 \leq k < l \leq n$, $i \neq k$, $j \neq k$, $i \neq l$, and $j \neq l$. The first and second sums on the right-hand-side of (4.12) go to zero almost surely since, as $n \to \infty$,

$$
(4.13)\qquad \sup_{x,y\geq 0} |H_n(x,y)| \to 0
$$

almost surely by Assumption 2. For the third term on the right-hand-side of (4.12),

$$
\begin{aligned}
(4.14)\quad & \mathbb{E}\left(n^{-3} \sum_{(i,j,k,l)\in N_n(4)} H_n(\tilde{X}_{ij}, \tilde{Y}_{ij}) H_n(\tilde{X}_{kl}, \tilde{Y}_{kl}) D_{ij} D_{kl}\alpha_{ij}\alpha_{kl} \mid \mathcal{G}_n\right) \\
&= [n\mathbb{E}(\alpha_{12}\alpha_{34} \mid \mathcal{F}_n)] \times \left[n^{-4} \sum_{(i,j,k,l)\in N_n(4)} H_n(\tilde{X}_{ij}, \tilde{Y}_{ij}) H_n(\tilde{X}_{kl}, \tilde{Y}_{kl}) D_{ij} D_{kl}\right],
\end{aligned}
$$

by the permutation invariance of H_n, the fact that $\{\tilde{X}_{ij}, \tilde{Y}_{ij}, D_{ij}, 1 \leq i < j \leq n\}$ are in $\mathcal{G}_n$, the fact that $\{\alpha_{ij}, 1 \leq i < j \leq n\}$ and $\{\tilde{X}_{ij}, \tilde{Y}_{ij}, D_{ij}, 1 \leq i < j \leq n\}$ are independent by Remark 2, and the fact that $\mathbb{E}(\alpha_{12}\alpha_{34} \mid \mathcal{F}_n)$ is constant over all $(i,j,k,l) \in N_n(4)$. Now, again by exchangeability,

$$
\begin{aligned}
& n\mathbb{E}(\alpha_{12}\alpha_{34} \mid \mathcal{F}_n) \\
&= \frac{n[(\sum_{1\leq i<j\leq n}\alpha_{ij})^2 - \sum_{1\leq i<j\leq n}\alpha_{ij}^2 - 2\sum_{1\leq i<j<k\leq n}(\alpha_{ij}\alpha_{ik} + \alpha_{ij}\alpha_{jk} + \alpha_{ik}\alpha_{jk})]}{n(n-1)(n-2)(n-3)/4},
\end{aligned}
$$

which is $O_p(1)$ by the boundedness of α_{ij}. Thus (4.14) goes to zero in probability, as $n \to \infty$, by (4.13). Hence (4.12) also converges to zero in probability.

For some θ' between $\hat{\theta}_n$ and θ_0,

$$
0 = \sqrt{n}G_n(\hat{\theta}_n)
$$

$$
\begin{aligned}
&= n^{-3/2} \sum_{1\le i<j\le n} W_n(\tilde{X}_{ij}, \tilde{Y}_{ij}) D_{ij} \left(\Delta_{ij} - \theta_0/(1+\theta_0)\right) \\
&\quad -(1+\theta')^{-2} n^{-2} \sum_{1\le i<j\le n} W_n(\tilde{X}_{ij}, \tilde{Y}_{ij}) D_{ij} \sqrt{n}(\hat{\theta}_n - \theta_0) \\
&= n^{-3/2} \sum_{1\le i<j\le n} W_0(\tilde{X}_{ij}, \tilde{Y}_{ij}) D_{ij} \left[\Delta_{ij} - \theta_0/(1+\theta_0)\right] - I_0 \sqrt{n}(\hat{\theta}_n - \theta_0) + o_p(1),
\end{aligned}
$$

by the convergence of (4.12) to zero, Remark 2, and Assumption 2. The result now follows by the Hoeffding representation for U-statistics (Theorem 5.3.2, Serfling (1980)). □

Remark 8. Using the U-statistic equivalent for $n^{1/2}(\hat{\theta}_n - \theta_0)$ leads naturally to the variance estimator $\hat{\sigma}_n^2 \equiv \hat{J}_n / \hat{I}_n^2$, where

$$
\begin{aligned}
\hat{I}_n &\equiv -n^{-2} \sum_{1\le i<j\le n} W_n(\tilde{X}'_{ij}, \tilde{Y}'_{ij}) D_{ij} (1+\hat{\theta}_n)^{-2}, \\
\hat{J}_n &\equiv 2n^{-3} \sum_{1\le i<j<k\le n} \hat{Q}_{ij}\hat{Q}_{ik} + \hat{Q}_{ij}\hat{Q}_{jk} + \hat{Q}_{ik}\hat{Q}_{jk}, \qquad \text{and} \\
\hat{Q}_{ij} &\equiv W_n(\tilde{X}'_{ij}, \tilde{Y}'_{ij}) D_{ij} \left(\Delta_{ij} - \frac{\hat{\theta}_n}{1+\hat{\theta}_n}\right), \qquad \text{for } \ 1 \le i < j \le n.
\end{aligned}
$$

This can be used to construct confidence intervals for θ_0 without the bootstrap.

5. Consistency of the bootstrap

The complexity of the limiting distribution of $\hat{S}_n$ and $\hat{R}_n$ precludes inference by analytical means, and bootstrapping the joint distribution of $(\hat{\theta}_n, \hat{S}_n, \hat{R}_n)$ is needed. The nonparametric bootstrap (Efron (1979)) is commonly used and often possesses second or higher order accuracy (Chapter 2, Hall (1992)). In our setting, a computational issue is the presence of ties in the X and Y values in the bootstrap samples. An alternative bootstrap which avoids this difficulty is the multiplier bootstrap, also known as the wild bootstrap (Praestegaard and Wellner (1993)). We study both bootstraps but present the multiplier bootstrap first.

Let $\mathcal{Z}_n \equiv \{Z_i, 1 \le i \le n\}$ denote the data of model 1, $n \ge 1$. For the multiplier bootstrap, we use random multiplier weights $\{\xi_i, i \ge 1\}$ which are independent of $\mathcal{Z}_n$, for all $n \ge 1$, and which satisfy:

ASSUMPTION 4. $\{\xi_i, i \ge 1\}$ is a sequence of i.i.d. nonnegative random variables with $\mathbb{E}\xi_1 = \operatorname{var}\xi_1 = 1$ and $\int_0^\infty \sqrt{\mathbb{P}(|\xi_1| > x)}dx < \infty$.

Let $\hat{\theta}_n^\circ$ be the solution of

$$
G_n^\circ(\theta) \equiv n^{-2} \sum_{1\le i<j\le n} \xi_i \xi_j W_n(\tilde{X}'_{ij}, \tilde{Y}'_{ij}) D_{ij} \left[\Delta_{ij} - \theta/(\theta+1)\right] = 0
$$

(allowing $\hat{\theta}_n^\circ = \infty$ if needed), if such a solution exists; if no such solution exists, take $\hat{\theta}_n^\circ = 1 - \epsilon_1/2$. Now define, for any $f : \Omega \mapsto \mathbb{R}$,

$$
\mathbb{P}_n^\circ f(Z) \equiv \left\{\sum_{i=1}^n \xi_i > 0\right\} \left[\sum_{i=1}^n \xi_i\right]^{-1} \sum_{i=1}^n \xi_i f(Z_i).
$$

That is, $\mathbb{P}_n^\circ$ is the multiplier-weighted empirical measure. Also define $T_n^\circ \equiv 0 \vee \sup(t : \mathbb{P}_n^\circ \{X' > t\} > r_n^\circ)$, where

$$r_n^\circ = \begin{cases} 0 & \text{if } \hat{\theta}_n^\circ \geq 1, \\ (1/2)^{1/(1-\hat{\theta}_n^\circ)} & \text{if } \hat{\theta}_n^\circ < 1. \end{cases}$$

The multiplier bootstrap of $\hat{S}_n$ and $\hat{R}_n$, $\hat{S}_n^\circ$ and $\hat{R}_n^\circ$, is a solution to $\mathbb{P}_n^\circ \psi^{\hat{\theta}_n^\circ}(Z; S, R, (s \wedge T_n^\circ, t \wedge T_n^\circ)) = 0$, for all $s, t \in [0, t_0]$. Define $\hat{\gamma}_n^\circ \equiv (\hat{\theta}_n^\circ, \hat{S}_n^\circ, \hat{R}_n^\circ)$.

For the nonparametric bootstrap, let $\mathcal{Z}_n^\bullet = \{Z_i^\bullet, 1 \leq i \leq n\}$ be a random resample of $\mathcal{Z}_n$ with replacement. Let the number of times observation Z_i appears in $\mathcal{Z}_n^\bullet$ be denoted $\xi_{ni}^\bullet$, $i = 1, \ldots, n$. Note that $\{\xi_{ni}^\bullet, 1 \leq i \leq n\}$ are the multinomial bootstrap weights and are independent of $\mathcal{Z}_n$, for all $n \geq 1$. In general, let superscript $\bullet$ denote variables based on the bootstrap sample. Define $\hat{\theta}_n^\bullet$ to be the solution of

$$\hat{G}_n^\bullet(\theta) \equiv n^{-2} \sum_{1 \leq i < j \leq n} W_n^\bullet(\tilde{X}_{ij}'^\bullet, \tilde{Y}_{ij}'^\bullet) D_{ij}^\bullet [\Delta_{ij}^\bullet - \theta/(\theta + 1)] = 0$$

(allowing $\hat{\theta}_n^\bullet = \infty$ if necessary), if such a solution exists; if no such solution exists, take $\hat{\theta}_n^\bullet = 1 - \epsilon_1/2$. Define, for any $f : \Omega \mapsto \mathbb{R}$, $\mathbb{P}_n^\bullet f(Z) \equiv n^{-1} \sum_{i=1}^n \xi_{ni}^\bullet f(Z_i)$. That is, $\mathbb{P}_n^\bullet$ is the bootstrap-weighted empirical measure. Let $T_n^\bullet \equiv 0 \vee \sup(t : \mathbb{P}_n^\bullet\{X' > t\} > r_n^\bullet)$, where

$$r_n^\bullet = \begin{cases} 0 & \text{if } \hat{\theta}_n^\bullet \geq 1, \\ (1/2)^{1/(1-\hat{\theta}_n^\bullet)} & \text{if } \hat{\theta}_n^\bullet < 1. \end{cases}$$

The nonparametric bootstrap of $\hat{S}_n$ and $\hat{R}_n$, $\hat{S}_n^\bullet$ and $\hat{R}_n^\bullet$, is a solution to $\mathbb{P}_n^\bullet \psi^{\hat{\theta}_n^\bullet}(Z; S, R, (s \wedge T_n^\bullet, t \wedge T_n^\bullet)) = 0$, for all $s, t \in [0, t_0]$. Let $\hat{\gamma}_n^\bullet \equiv (\hat{\theta}_n^\bullet, \hat{S}_n^\bullet, \hat{R}_n^\bullet)$.

Consistency results for these bootstraps are presented next. For a uniform metric space D, let $BL_1(D)$ be the space of real valued functions on D with uniform Lipshitz norm bounded by 1. As in Theorem 1.12.2 of VW, a stochastic process H_n converges HJD-weakly in the uniform topology on D to a Borel measurable and separable H if and only if

$$\sup_{f \in BL_1(D)} |\mathbb{E}^* f(H_n) - \mathbb{E} f(H)| \to 0,$$

as $n \to \infty$, where superscript $*$ denotes outer expectation. We adopt this approach for establishing consistency because it avoids measurability issues.

5.1 *Consistency of the multiplier bootstrap*

The multiplier bootstrap $\hat{\gamma}_n^\circ$ is a solution of

$$\hat{\Psi}_n^\circ(\gamma) \equiv \begin{pmatrix} \hat{\theta}_n^\circ - \theta \\ \mathbb{P}_n^\circ \psi^\theta (Z; S, R, (\cdot, \cdot) \wedge T_n^\circ \wedge t_0) \end{pmatrix} \{T_n^\circ > 0\} = 0.$$

THEOREM 4. *Assume the conditions of Theorem 3 with $\hat{\gamma}_n$, Z_0, and Y_0 defined accordingly. Let the random sequence $\{\xi_i, i \geq 1\}$ satisfy Assumption 4 and be independent of $\mathcal{Z}_n$ for all $n \geq 1$. Then there exists a solution $\hat{\gamma}_n^\circ$ of $\Psi_n^\circ(\gamma) = 0$, $\sqrt{n}(\hat{\gamma}_n^\circ - \hat{\gamma}_n)$ is asymptotically measurable, and*

$$\sup_{f \in BL_1(\mathcal{L})} |\mathbb{E}_\xi f(\sqrt{n}(\hat{\gamma}_n^\circ - \hat{\gamma}_n)) - \mathbb{E} f(Y_0)| \to 0$$

in outer probability as $n \to \infty$, where $\mathbb{E}_\xi$ is the expectation with respect to the multipliers conditional on $\mathcal{Z}_n$. Hence, the multiplier bootstrap is consistent.

Before giving the proof of this theorem, we need two lemmas, the proofs of which will be given later. The first lemma establishes key properties of $\hat{\theta}_n^\circ$, while the second lemma gives existence of $\hat{\gamma}_n^\circ$.

LEMMA 7. *Assume the conditions of Theorem 4 obtain. Then*

$$\mathcal{P}_0\left(\omega \in \Omega^\infty : \mathbb{P}\left\{\lim_{n\to\infty} \hat{\theta}_n^\circ = \theta_0 \mid \omega\right\} = 1\right) = 1$$

and

$$\sqrt{n}(\hat{\theta}_n^\circ - \hat{\theta}_n) = I_0^{-1} n^{-1/2} \sum_{i=1}^n (\xi_i - 1) A_i + Q_n^\circ,$$

where $\mathbb{E}[(Q_n^\circ)^2 \mid \mathcal{Z}_n]$ converges to zero in probability as $n \to \infty$.

LEMMA 8. *Assume the conditions of Theorem 4 obtain. Then*

$$\mathcal{P}_0(\omega \in \Omega^\infty : \mathbb{P}\{T_n^\circ \geq t_0 \text{ for all } n \text{ large enough} \mid \omega\} = 1) = 1. \tag{5.1}$$

Furthermore, there exists a solution $\hat{\gamma}_n^\circ$ of $\Psi_n^\circ(\gamma) = 0$, where $\hat{S}_n^\circ, \hat{R}_n^\circ \in \boldsymbol{H}_0^{t_0}$ are piecewise constant with possible jumps occurring only at $\{X_i' : \eta_i \xi_i > 0, X_i' \leq T_n^\circ \wedge t_0, i = 1, \ldots, n\}$ for $\hat{S}_n^\circ$ and at $\{Y_i' : \delta_i \xi_i > 0, Y_i' \leq T_n^\circ \wedge t_0, i = 1, \ldots, n\}$ for $\hat{R}_n^\circ$.

PROOF OF THEOREM 4. Existence of a solution $\hat{\gamma}_n^\circ$ follows from Lemma 8. Take ϕ and $\check{\phi}$ as in Section 5. Without loss of generality assume $\phi\Psi_n^\circ = \hat{\gamma}_n^\circ$. Using arguments from the proof of Lemma 8, we can show that if B_3 is the set of all $\omega \in \Omega^\infty$ such that

$$\sqrt{n}(\phi\Psi_n^\circ - \phi\Psi_n) - \sqrt{n}(\phi\Psi_n^\circ - \phi\Psi_n)\{\Psi_n^\circ, \Psi_n \in Z(\Gamma_0, \mathcal{L})\} = 0$$

for all n large enough, then $\mathcal{P}_0(B_3)_* = 1$, where subscript $*$ denotes the maximal measurable minorant. In addition,

$$\begin{aligned}&\sqrt{n}(\phi\Psi_n^\circ - \phi\Psi_n)\{\Psi_n^\circ, \Psi_n \in Z(\Gamma_0, \mathcal{L})\}\\ &\quad = \sqrt{n}(\check{\phi}[\Psi_n^\circ\{\Psi_n^\circ \in Z(\Gamma_0, \mathcal{L})\}] - \check{\phi}[\Psi_n\{\Psi_n \in Z(\Gamma_0, \mathcal{L})\}])\{\Psi_n^\circ, \Psi_n \in Z(\Gamma_0, \mathcal{L})\}.\end{aligned}$$

By Lemmas 4, 6, and 7 above and by Theorem 2.9.6 of VW, we have for

$$H_n^\circ = \sqrt{n}(\Psi_n^\circ\{\Psi_n^\circ \in Z(\Gamma_0, \mathcal{L})\} - \Psi_n\{\Psi_n \in Z(\Gamma_0, \mathcal{L})\})$$

that H_n° is asymptotically measurable and as $n \to \infty$

$$\sup_{h \in BL_1(\ell^\infty(\Gamma_0, \mathcal{L}))} |\mathbb{E}_\xi h(H_n^\circ) - \mathbb{E}h(Z_0)| \to 0$$

in outer probability. Lemma 5 above and Theorem 3.9.11 of VW yield the desired result. □

PROOF OF LEMMA 7. The conditional almost sure consistency of $\hat{\theta}_n^\circ$ follows from Assumption 2 and standard probability arguments. Now, for some θ' between $\hat{\theta}_n^\circ$ and

$\hat{\theta}_n$,

$$
\begin{aligned}
0 &= G_n^\circ(\hat{\theta}_n^\circ) \\
&= n^{-3/2} \sum_{1\le i<j\le n} W_n(\tilde{X}'_{ij}, \tilde{Y}'_{ij}) D_{ij}[\Delta_{ij} - \hat{\theta}_n/(1+\hat{\theta}_n)]\xi_i\xi_j \\
&\quad -\sqrt{n}(\hat{\theta}_n^\circ - \hat{\theta}_n)(1-\theta')^{-2} n^{-2} \sum_{1\le i<j\le n} W_n(\tilde{X}'_{ij}, \tilde{Y}'_{ij}) D_{ij},
\end{aligned}
$$

which can be expressed

$$
\begin{aligned}
(5.2)\quad & \sqrt{n}(\hat{\theta}_n^\circ - \hat{\theta}_n) \\
&= I_0^{-1} n^{-3/2} \sum_{1\le i<j\le n} W_0(\tilde{X}'_{ij}, \tilde{Y}'_{ij}) D_{ij}[\Delta_{ij} - \theta_0/(1+\theta_0)](\xi_i\xi_j - 1) + Q_{n(1)}^\circ.
\end{aligned}
$$

If we let

$$
F_{ij} \equiv H_n(\tilde{X}'_{ij}, \tilde{Y}'_{ij}) D_{ij}[\Delta_{ij} - \hat{\theta}_n/(1+\hat{\theta}_n)] - W_0(\tilde{X}'_{ij}, \tilde{Y}'_{ij}) D_{ij}[\hat{\theta}_n/(1+\hat{\theta}_n) - \theta_0/(1+\theta_0)],
$$

where $H_n = W_n - W_0$, then as $n \to \infty$

$$
\begin{aligned}
& \mathbb{E}\left[\left(n^{-3/2} \sum_{1\le i<j\le n} F_{ij}\xi_i\xi_j\right)^2 \middle| \mathcal{Z}_n\right] \\
&= n^{-3}\left(4 \sum_{1\le i<j\le n} F_{ij}^2 + 2 \sum_{1\le i<j<k\le n} F_{ij}F_{ik} + F_{ij}F_{jk} + F_{ik}F_{jk}\right)
\end{aligned}
$$

goes to 0 almost surely by previous arguments. Thus $\mathbb{E}([Q_{n(1)}^\circ]^2 \mid \mathcal{Z}_n) \to 0$ almost surely.

We next show that (5.2) can be approximated by the desired sum. To do this, let $B_{ij} \equiv W_0(\tilde{X}'_{ij}, \tilde{Y}'_{ij}) D_{ij}\,[\Delta_{ij} - \theta_0/(1+\theta_0)]$. Note that

$$
\begin{aligned}
& \mathbb{E}\left[\left(n^{-3/2} \sum_{1\le i<j\le n} B_{ij}(\xi_i\xi_j - 1) - n^{-1/2}\sum_{i=1}^n A_i(\xi_i - 1)\right)^2 \middle| \mathcal{Z}_n\right] \\
&= \mathbb{E}\left[\left(n^{-3/2} \sum_{1\le i<j\le n} [B_{ij}(\xi_i\xi_j - 1) - A_i(\xi_i - 1) - A_j(\xi_j - 1)]\right)^2 \middle| \mathcal{Z}_n\right] + o_p(1) \\
&= o_p(1) + n^{-3} \sum_{1\le i<j\le n} [3B_{ij}^2 - 2(A_i + A_j)B_{ij} + A_i^2 + A_j^2] \\
&\quad + 2n^{-3} \sum_{1\le i<j<k\le n} [B_{ij}B_{ik} - A_i(B_{ij} + B_{ik}) + A_i^2 + B_{ij}B_{jk} - A_j(B_{ij} + B_{jk}) \\
&\qquad\qquad + A_j^2 + B_{ik}B_{jk} - A_k(B_{ik} + B_{jk}) + A_k^2]
\end{aligned}
$$

goes to 0 in probability, as $n \to \infty$, since $\mathbb{E}(B_{ij} \mid Z_i) = A_i$ for $i \ne j$, $1 \le i \ne j \le n$. □

PROOF OF LEMMA 8. Let $B_1 = \{\omega \in \Omega^\infty : \hat{\theta}_n^\circ(\omega)$ is almost surely consistent for $\theta_0\}$, and note that $\mathcal{P}_0(B_1) = 1$ by Lemma 7. Let B_2 be all $\omega \in \Omega^\infty$ such $\mathbb{P}_n^\circ\{X' > t_0\}$

converges almost surely to $P_0\{X' > t_0\}$, and note that $\mathcal{P}_0(B_2) = 1$ by standard probability arguments. Thus (5.1) holds. When $T_n^\circ > 0$, $\hat{\gamma}_n^\circ$ exists by Brouwer's fixed point theorem and arguments in Lemma 2 if, for $t \in [t_0 \wedge T_n^\circ, t_0]$, we set $\hat{S}_n^\circ(t) = \hat{S}_n^\circ(t_0 \wedge T_n^\circ)$ and $\hat{R}_n^\circ(t) = \hat{R}_n^\circ(t_0 \wedge T_n^\circ)$. When $T_n^\circ = 0$, take $\hat{\gamma}_n^\circ = \phi(0)$ as in Theorem 3 so that $\hat{\gamma}_n^\circ$ is always defined. □

5.2 *Consistency of the nonparametric bootstrap*

The nonparametric bootstrap $\hat{\gamma}_n^\bullet$ is a solution to

$$\Psi_n^\bullet(\gamma) \equiv \begin{pmatrix} \hat{\theta}_n^\bullet - \theta \\ \mathbb{P}_n^\bullet \psi^\theta (Z; S, R, (\cdot,\cdot) \wedge T_n^\bullet \wedge t_0) \end{pmatrix} \{T_n^\bullet > 0\} = 0.$$

THEOREM 5. *Assume the conditions of Theorem 3 with $\hat{\gamma}_n$, Z_0, and Y_0 defined accordingly. Let $\mathcal{Z}_n^\bullet \equiv \{Z_i^\bullet, 1 \le i \le n\}$ be a nonparametric bootstrap of $\mathcal{Z}_n$, with bootstrap weights $\{\xi_{ni}^\bullet, 1 \le i \le n\}$, $n \ge 1$. Then there exists a solution $\hat{\gamma}_n^\bullet$ of $\Psi_n^\bullet(\gamma) = 0$, $\sqrt{n}(\hat{\gamma}_n^\bullet - \hat{\gamma}_n)$ is asymptotically measurable, and*

$$\sup_{f \in BL_1(\mathcal{L})} |\mathbb{E}_{\xi^\bullet} f(\sqrt{n}(\hat{\gamma}_n^\bullet - \hat{\gamma}_n)) - \mathbb{E} f(Y_0)| \to 0$$

in outer probability as $n \to \infty$, where $\mathbb{E}_{\xi^\bullet}$ is the expectation with respect to the nonparametric bootstrap conditional on $\mathcal{Z}_n$. Hence, the nonparametric bootstrap is consistent.

Before giving the proof of this theorem, we need two lemmas, the proofs of which will be given later. The first lemma establishes key properties of $\hat{\theta}_n^\bullet$, while the second lemma gives existence of $\hat{\gamma}_n^\bullet$.

LEMMA 9. *Assume the conditions of Theorem 5 obtain. Then*

$$\mathcal{P}_0\left(\omega \in \Omega^\infty : \mathbb{P}\left\{\lim_{n\to\infty} \hat{\theta}_n^\bullet = \theta_0 \mid \omega\right\} = 1\right) = 1$$

and

$$\sqrt{n}(\hat{\theta}_n^\bullet - \hat{\theta}_n) = I_0^{-1} n^{-1/2} \sum_{i=1}^n (\xi_{ni}^\bullet - 1) A_i + Q_n^\bullet,$$

where $\mathbb{E}[(Q_n^\bullet)^2 \mid \mathcal{Z}_n]$ converges to zero in probability as $n \to \infty$.

LEMMA 10. *Assume the conditions of Theorem 5 obtain. Then*

$$\mathcal{P}_0(\omega \in \Omega^\infty : P\{T_n^\bullet \ge t_0 \text{ for all } n \text{ large enough} \mid \omega\} = 1) = 1. \tag{5.3}$$

Furthermore, there exists a solution $\hat{\gamma}_n^\bullet$ of $\Psi_n^\bullet(\gamma) = 0$, where $\hat{S}_n^\bullet, \hat{R}_n^\bullet \in \boldsymbol{H}_0^{t_0}$ are piecewise constant with possible jumps occurring only at $\{X_i' : \eta_i \xi_{ni}^\bullet > 0, X_i' \le T_n^\bullet \wedge t_0, i = 1, \ldots, n\}$ for $\hat{S}_n^\bullet$ and at $\{Y_i' : \delta_i \xi_{ni}^\bullet > 0, Y_i' \le T_n^\bullet \wedge t_0, i = 1, \ldots, n\}$ for $\hat{R}_n^\bullet$.

PROOF OF THEOREM 5. Existence of $\hat{\gamma}_n^\bullet$ follows from Lemma 10. Take ϕ and $\check{\phi}$ as in Section 5. Without loss of generality $\phi\Psi_n^\bullet = \hat{\gamma}_n^\bullet$. Using arguments from the proof of Theorem 4, $\sqrt{n}(\hat{\gamma}_n^\bullet - \hat{\gamma}_n)$ is asymptotically equivalent, outer almost surely conditional on $\omega \in \Omega^\infty$, to

$$\sqrt{n}(\check{\phi}[\Psi_n^\bullet\{\Psi_n^\bullet \in Z(\Gamma_0, \mathcal{L})\}] - \check{\phi}[\Psi_n\{\Psi_n \in Z(\Gamma_0, \mathcal{L})\}])\{\Psi_n^\bullet, \Psi_n \in Z(\Gamma_0, \mathcal{L})\}.$$

By Lemmas 4, 6, and 9 above and by Theorem 3.6.1 of VW, we have for

$$H_n^\bullet = \sqrt{n}(\Psi_n^\bullet\{\Psi_n^\bullet \in Z(\Gamma_0, \mathcal{L})\} - \Psi_n\{\Psi_n \in Z(\Gamma_0, \mathcal{L})\})$$

that $H_n^\bullet$ is asymptotically measurable and as $n \to \infty$

$$\sup_{h \in BL_1(\ell^\infty(\Gamma_0, \mathcal{L}))} |\mathbb{E}_{\xi^\bullet} h(H_n^\bullet) - \mathbb{E}h(Z_0)| \to 0$$

in outer probability. Lemma 5 above and Theorem 3.9.11 of VW yield the result. □

PROOF OF LEMMA 9. Note that

$$G_n^\bullet(\theta) = n^{-2} \sum_{1 \le i < j \le n} \xi_{ni}^\bullet \xi_{nj}^\bullet W_n^\bullet(\tilde{X}'_{ij}, \tilde{Y}'_{ij}) D_{ij}[\Delta_{ij} - \theta/(1+\theta)] \tag{5.4}$$

which implies

$$\hat{\theta}_n^\bullet = \frac{\sum_{1\le i<j\le n} \xi_{ni}^\bullet \xi_{nj}^\bullet W_0(\tilde{X}'_{ij}, \tilde{Y}'_{ij}) D_{ij} \Delta_{ij}}{\sum_{1\le i<j\le n} \xi_{ni}^\bullet \xi_{nj}^\bullet W_0(\tilde{X}'_{ij}, \tilde{Y}'_{ij}) D_{ij}(1-\Delta_{ij})} + Q_{n(1)}^\bullet,$$

where $\mathcal{P}_0(\omega \in \Omega^\infty : \mathbb{P}\{\lim_{n\to\infty} Q_{n(1)}^\bullet = 0 \mid \omega\} = 1) = 1$ by arguments used in the proof of Lemma 10. As in the proofs of Theorems 3.6.1 and 3.6.2 of VW, let N_n be Poisson with mean n. Let $\{m_n^{(k)}, k \ge 1\}$ be an infinite sequence of i.i.d. multinomial $(1, 1/n, \ldots, 1/n)$ vectors in $\mathbb{R}^n$, independent of N_n. Take $M_n = \sum_{k=1}^n m_n^{(k)}$ and $M_{N_n} = \sum_{k=1}^{N_n} m_n^{(k)}$. Let $(\xi_1, \ldots, \xi_n) \equiv M_{N_n}$ and, without loss of generality, $(\xi_{n1}^\bullet, \ldots, \xi_{nn}^\bullet) \equiv M_n$. Note $\xi_1, \ldots, \xi_n$ are i.i.d. Poisson with mean 1 and satisfy Assumption 4. Also $\sum_{i=1}^n |\xi_i - \xi_{ni}^\bullet| = |N_n - n|$. If the triangular array $\{h_{ij}, j > i \ge 1\}$ is uniformly bounded, with $\sup_{j>i\ge 1} |h_{ij}| \le K$, then

$$\begin{aligned}\left| n^{-2} \sum_{1\le i<j\le n} h_{ij}(\xi_{ni}^\bullet \xi_{nj}^\bullet - \xi_i \xi_j) \right| &\le K \sum_{1\le i<j\le n} (\xi_i |\xi_{nj}^\bullet - \xi_j| + \xi_{nj}^\bullet |\xi_{ni}^\bullet - \xi_i|) \\ &\le K n^{-2}(N_n |N_n - n| + n|N_n - n|)\end{aligned}$$

converges to 0 almost surely. Thus

$$\hat{\theta}_n^\bullet = \frac{\sum_{1\le i<j\le n} \xi_i \xi_j W_0(\tilde{X}'_{ij}, \tilde{Y}'_{ij}) D_{ij} \Delta_{ij}}{\sum_{1\le i<j\le n} \xi_i \xi_j W_0(\tilde{X}'_{ij}, \tilde{Y}'_{ij}) D_{ij}(1-\Delta_{ij})} + Q_{n(2)}^\bullet,$$

where $\mathcal{P}_0(\omega \in \Omega^\infty : \mathbb{P}\{\lim_{n\to\infty} Q_{n(2)}^\bullet = 0 \mid \omega\} = 1) = 1$. Consistency now follows from standard probability arguments.

The fact that $\sqrt{n} G_n^\bullet(\hat{\theta}_n^\bullet) = 0$, combined with (5.4), implies that

$$\sqrt{n}(\hat{\theta}_n^\bullet - \hat{\theta}_n) = I_0^{-1} n^{-3/2} \sum_{1\le i<j\le n} W_0(\tilde{X}'_{ij}, \tilde{Y}'_{ij}) D_{ij} [\Delta_{ij} - \theta_0/(1+\theta_0)] + Q_{n(3)}^\bullet,$$

where $\mathcal{P}_0(\omega \in \Omega^\infty : \mathbb{P}\{\lim_{n\to\infty} Q_{n(3)}^\bullet = 0 \mid \omega\} = 1) = 1$ by previous arguments. Let $\{B_{ij}, j > i \ge 1\}$ be as defined in the proof of Lemma 7. Based on the multinomial structure of the bootstrap weights and on previous arguments, it is easy to show

$$\mathbb{E}\left[\left(n^{-3/2}\sum_{1\le i<j\le n} B_{ij}(\xi^\bullet_{ni}\xi^\bullet_{nj}-1)-n^{-1/2}\sum_{i=1}^n A_i(\xi^\bullet_{ni}-1)\right)^2 \,\middle|\, \mathcal{Z}_n\right]$$
$$= o_p(1)+O_p(n^{-1})+2n^{-3}(1+O_p(n^{-1}))\sum_{1\le i<j<k\le n}[B_{ij}B_{ik}-A_i(B_{ij}+B_{ik})+A_i^2$$
$$+B_{ij}B_{jk}-A_j(B_{ij}+B_{jk})+A_j^2+B_{ik}B_{jk}-A_k(B_{ik}+B_{jk})+A_k^2]$$

goes to 0 in probability, as $n \to \infty$. □

Proof of Lemma 10. As a consequence of the outer almost sure equivalence of the nonparametric bootstrap and a multiplier bootstrap having mean 1 Poisson weights, the arguments are essentially identical to the arguments used in the proof of Lemma 8. The difference is that Lemma 9 is used in place of Lemma 7. □

References

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**, 141–151.

Copelan, E. A., Biggs, J. C., Thompson, J. M., Crilley, P., Szer, J., Klein, J. P., Kapoor, N., Avalos, B. R., Cunningham, I., Atkinson, K., Kowns, K., Harmon, G. S., Daly, M. B., Brodsky, I., Bulova, S. K. and Tutschka, P. J. (1991). Treatment for acute myelocytic leukemia with allogenic bone marrow transplantation following preparation with Bu/Cy, *Blood*, **78**, 838–843.

Day, R., Bryant, J. and Lefkopoulou, M. (1997). Adaptation of bivariate frailty models for prediction, with application to biological markers as prognostic factors, *Biometrika*, **84**, 45–56.

Efron, B. (1967). The two sample problem with censored data, *Proc. Fifth Berkeley Symp. on Math. Statist. Prob.*, **4**, 831–855, Univ. of California Press, Berkeley.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife, *Ann. Statist.*, **7**, 1–26.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer, New York.

Jiang, H., Fine, J. P. and Chappell, R. J. (1999). Estimation of the association between bivariate failure times in the semi-competing risks problem, 1999 *Proceedings of the Biometrics Section*, 192–194, American Statistical Association Alexandria, Virginia.

Jiang, H., Fine, J. P., Kosorok, M. R., and Chappell, R. J. (2001). On self-consistency in the gamma frailty model with dependent censoring. Tech. Report, No. 137, Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison.

Klein, J. P. and Moeschberger, M. L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*, Springer, New York.

Murphy, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect, *Ann. Statist.*, **22**, 712–731.

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

Oakes, D. (1986). Semiparametric inference in a model for association in bivariate survival data, *Biometrika*, **73**, 353–361.

Oakes, D. (1989). Bivariate survival models induced by frailties, *J. Amer. Statist. Assoc.*, **84**, 487–493.

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

Praestegaard, J. and Wellner, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process, *Ann. Probab.*, **21**, 2053–2086.

Robertson, J. B. and Uppuluri, V. R. (1984). A generalized Kaplan-Meier estimator, *Ann. Statist.*, **12**, 366–371.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Tsai, W.-Y. and Crowley, J. (1985). A large sample study of generalized maximum likelihood estimators from incomplete data via self-consistency, *Ann. Statist.*, **13**, 1317–1334.

Tsiatis, A. A. (1975). A nonidentifiability aspect of the problem of competing risks, *Proc. Nat. Acad. Sci. U.S.A.*, **72**, 20–22.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*, Springer, New York.

Vardi, Y. and Zhang, C.-H. (1992). Large sample study of empirical distributions in random multiplicative censoring model, *Ann. Statist.*, **20**, 1022–1039.