Appendix: Technical Arguments

Proof of Consistency and Asymptotic Normality of $\widehat{\Lambda}_0(\cdot)$.

Conditions. Before presenting the proof of the asymptotic properties of $\widehat{\Lambda}_0(\cdot)$, we first describe the conditions under which these large sample results have been proven.

C1. (Finite Interval) There exists a maximum follow-up time $\tau > 0$ and a constant c such that $\Pr(X_{ij} = \tau) \geq c > 0$.

C2. (Boundedness) The baseline hazard function $\lambda_0(\cdot)$ is twice differentiable with second derivatives over $[0, \tau]$ bounded by some fixed constant. Moreover,

$$\int_0^t \lambda_0(s)/\pi(s)ds < \infty \quad \forall t \in [0, \tau], \tag{A.1}$$

where $\pi(s) = \mathrm{E}[I(X_{i0} < s)\sum_{j=1}^{M} Y_{ij}(s)]$.

C3. (Identifiability) There is a positive probability that two or more members in a family fail in the interval $[0, \tau]$.

C4. (Differentiability) The copula function $h(t_0, t_1, \ldots, t_M; \theta)$ is continuously thrice differentiable with respect to each time component on $[c', 1]^{M+1}$, where $c' = \exp(-c)$ and c is defined in Condition C1. In addition, h and its second partial derivatives are twice differentiable with respect to θ, and the derivatives are also continuously differentiable with respect to each time component on $[c', 1]^{M+1}$.

C5. (Nondegeneracy) The function

$$\ell^*(\beta, \gamma) = (1 - \alpha)E\{\ell^\circ(\beta, \gamma, X_{i0}, Z_{i0})|\delta_{i0} = 0\} + \alpha E\{\ell^\circ(\beta, \gamma, X_{i0}, Z_{i0})|\delta_{i0} = 1\}$$

has a unique maximizer $(\bar{\beta}, \bar{\gamma})$. In addition, the limiting value of $\partial U(\theta, \Lambda)/\partial\theta$ is positive definite at θ_0.

C6. (Kernel regularity condition) The kernel $K(\cdot)$ is symmetric, has support on $[-1, 1]$ and is differentiable on $(-\infty, \infty)$ such that $K(\cdot)$ decreases smoothly to 0 in the neighborhood of ± 1. The bandwidth h satisfies $h \sim n^{-\nu}$ with $\nu \in (\frac{1}{4}, \frac{1}{2})$.

C7. (Density regularity condition) The conditional density $g(\cdot)$ of X_{i0} given $\delta_{i0} = 1$ is bounded below on $t \in [0, \tau]$ by some positive number g_{min}, and $g(t)$ is thrice differentiable w.r.t. t over $[0, \tau]$ with bounded third derivative.

Condition C1 is a standard condition survival analysis. The condition (A.1) in Condition C2 is needed for the first-stage estimator, to ensure that the denominator in (3) does not tend too quickly to zero as $t \downarrow 0$. Condition C3 is necessary for the dependence parameter to be identifiable. The same condition has also been assumed in Nielsen *et al.* (1992) and Murphy (1994, 1995). Condition C4 is usually satisfied for commonly used copula functions such as the Clayton-Oakes model (Oakes, 1989) or the normal transformation model (Li and Lin, 2006). Regarding Condition C5, $\ell^*(\beta, \gamma)$ has a maximizer because it is concave, and the condition that the maximizer is unique simply rules out a degenerate case. Likewise, the matrix $\partial U(\theta, \Lambda)/\partial\theta$ is automatically nonnegative definite, and the assumption that it be positive definite at θ_0 simply rules out a degenerate case. Regarding the assumptions on the kernel function in Condition C6, it is not difficult to find kernels meeting these assumptions; one example is the biweight kernel $K(u) = (15/16)(1-u^2)^2 I(|u| \leq 1)$.

For a function $\varphi : [0, \tau] \to \mathcal{R}$, let $\|\varphi\| = \sup_{t\in[0,\tau]} |\varphi(t)|$.

Asymptotic properties of $\widehat{\beta}$. Under Condition C5, it follows from standard theory (White, 1982; van der Vaart, 1998, Chapter 5), that $(\widehat{\beta}, \widehat{\gamma})$ converges almost surely to $(\bar{\beta}, \bar{\gamma})$ and that $n^{1/2}\{(\widehat{\beta}, \widehat{\gamma}) - (\bar{\beta}, \bar{\gamma})\}$ is asymptotically normal. For use in proving the asymptotics of $\widehat{\Lambda}_0(\cdot)$, we note that

$$\widehat{\beta} - \bar{\beta} = \frac{1}{n}\sum_{i=1}^{n} e_1(\bar{\beta}, \bar{\gamma}) \left[\Psi(\bar{\beta}, \bar{\gamma}, X_{i0}, Z_{i0}) - E\{\Psi(\bar{\beta}, \bar{\gamma}, X_{i0}, Z_{i0})|\delta_{i0}\}\right] + o_p(n^{-1/2}) \quad \text{(A.2)}$$

where $\Psi(\beta, \gamma, t, z)$ is the gradient vector of $\ell^\circ(\beta, \gamma, t, x)$ with respect to β and γ, and $e_1(\beta, \gamma)$ is -1 times the inverse of the Hessian matrix of $\ell^*(\beta, \gamma)$ with the last row deleted (the uniqueness of $(\bar{\beta}, \bar{\gamma})$ as a maximizer of ℓ^* implies that this Hessian matrix evaluated at $(\bar{\beta}, \bar{\gamma})$ is positive definite). This gives a representation of $\widehat{\beta} - \bar{\beta}$ as asymptotically equivalent to a quantity of the form (11).

Asymptotic properties of $\widehat{\Lambda}(\cdot, \widehat{\theta})$. We will show the consistency and asymptotic normality of $\widehat{\Lambda}(\cdot, \widehat{\theta})$ using some of the ideas presented in Gorfine *et al.* (2009) and Spiekerman and Lin (1998). Essentially, we can show that for a given θ,

$$n^{1/2}\{\widehat{\Lambda}(t, \theta) - \Lambda(t, \theta)\} = \mathcal{M}_1(t) + \mathcal{M}_2(t) + o_p(1) \quad \text{(A.3)}$$

with

$$\begin{aligned}
\mathcal{M}_1(t) &= n^{-\frac{1}{2}} p_1(t) \int_0^t \frac{1}{p_1(s)h(s,\Lambda)} \sum_{i=1}^{n} \sum_{j=1}^{M} M_{ij}(ds) \\
\mathcal{M}_2(t) &= n^{-\frac{1}{2}} p_1(t) \int_0^\tau \frac{\mathcal{A}_1(t,u)}{p(u)q(u,\Lambda)} \sum_{i=1}^{n} \sum_{j=1}^{M} I(X_{i0} < u) M_{ij}(du),
\end{aligned}$$

where the integrands in $\mathcal{M}_1(t)$ and $\mathcal{M}_2(t)$ are predictable processes. (The notation and detailed derivation are provided in Section A, the Supplementary Materials.) Hence by standard martingale theory, $\|\widehat{\Lambda}(\cdot,\theta) - \Lambda(\cdot,\theta)\| \to_p 0$ and $n^{1/2}\{\widehat{\Lambda}(t,\theta) - \Lambda(t,\theta)\}$ converges weakly to a zero-mean Gaussian process, with weak convergence defined with respect to the uniform metric on $D[0,\tau]$ (Pollard, 1984, Section VIII.2).

To establish the consistency of $\widehat{\theta}$, we note that under condition C4, $dh/d\theta$, $dh^{(01)}/d\theta$, $dh^{(10)}/d\theta$ and $dh^{(11)}/d\theta$ are also uniformly Lipschitz continuous with respect to $\Lambda(\cdot)$. Hence, from the uniform convergence of $\widehat{\Lambda}(\cdot,\theta)$ to $\Lambda(\cdot,\theta)$ and the Lipschitz continuity, we have that $U\{\theta, \widehat{\Lambda}(t,\theta)\}$ converges uniformly to $U\{\theta, \Lambda(t,\theta)\}$ in θ as $n \to \infty$. By the strong law of large numbers, $U\{\theta, \Lambda(\cdot,\theta)\}$ converges to a limit $u\{\theta, \Lambda(\cdot,\theta)\}$, which equals 0 at θ_0. Finally, under condition C5 and by the Foutz theorem (1977), we conclude that there exists a unique root $\widehat{\theta}$ to $U\{\widehat{\theta}, \widehat{\Lambda}(\cdot,\widehat{\theta})\} = 0$ and that $\widehat{\theta} \to \theta_0$ in probability, as $n \to \infty$. To show the asymptotic normality of $\widehat{\theta}$, we expand

$$\begin{aligned}
U\{\widehat{\theta}, \widehat{\Lambda}(\cdot,\widehat{\theta})\} &= U\{\theta_0, \Lambda(t)\} + [U\{\theta_0, \widehat{\Lambda}(\cdot,\theta_0)\} - U\{\theta_0, \Lambda(t)\}] \\
&\quad + [U\{\widehat{\theta}, \widehat{\Lambda}(\cdot,\widehat{\theta})\} - U\{\theta_0, \widehat{\Lambda}(\cdot,\theta_0)\}] \\
&= 0.
\end{aligned}$$

By a Taylor expansion, the second term is asymptotically equivalent to $\partial U\{\theta_0, \Lambda(\cdot)\}/\partial\Lambda(t)\{\widehat{\Lambda}(\cdot;\theta_0) - \Lambda(\cdot)\}$ and the third term is asymptotically equivalent to $\partial U\{\theta, \Lambda(\cdot,\theta)\}/\partial\theta|_{\theta=\theta_0}(\widehat{\theta} - \theta_0)$. By the martingale representation of $\widehat{\Lambda}(\cdot,\theta_0) - \Lambda(\cdot)$ and condition C5, we get that $n^{1/2}(\widehat{\theta} - \theta_0)$ is asymptotically equivalent to $n^{1/2}$ times a quantity of the form (11). By the central limit theorem, it is asymptotically normal with mean 0 and a covariance matrix that can be consistently estimated by a sandwich-type estimator. Similarly, we expand $\widehat{\Lambda}(\cdot,\widehat{\theta}) - \Lambda(\cdot) = \{\widehat{\Lambda}(\cdot,\widehat{\theta}) - \widehat{\Lambda}(\cdot,\theta_0)\} + \{\widehat{\Lambda}(\cdot,\theta_0) - \Lambda(\cdot)\}$, and approximate the first term by $\partial\widehat{\Lambda}(\cdot,\theta)\}/\partial\theta|_{\theta=\theta_0}(\widehat{\theta} - \theta_0)$. By the asymptotic results of $\widehat{\theta}$, the martingale representation of $\widehat{\Lambda}(\cdot,\theta)$ and condition C4, we also get that $\|\widehat{\Lambda}(\cdot,\widehat{\theta}) - \Lambda(\cdot)\| \to_p 0$ and $n^{1/2}\{\widehat{\Lambda}(\cdot,\widehat{\theta}) - \Lambda(\cdot)\}$ converges to a zero-mean Gaussian process.

Asymptotic results of $\widehat{\Lambda}_0(\cdot)$. We now prove that $n^{1/2}\{\widehat{\Lambda}_0(\cdot) - \Lambda_0(\cdot, \bar{\beta})\}$ converges weakly to a Gaussian random process (which implies uniform convergence in probability of $\widehat{\Lambda}_0(t)$ to $\Lambda_0(\cdot, \bar{\beta})$). Note that $\widehat{\Lambda}_0(\cdot)$ is a function of θ, since it is a function of the composite hazard function $\widehat{\Lambda}(\cdot, \theta)$, which depends on θ. For clarity of presentation, we now indicate θ explicitly in the notation, writing $\widehat{\Lambda}_0(\cdot)$ as $\Lambda_0(\cdot, \theta)$. We have

$$\begin{aligned}\widehat{\Lambda}_0(t) - \Lambda_0(t, \bar{\beta}) &= \int_0^t \phi(u, \bar{\beta})\{\widehat{\Lambda}(du, \widehat{\theta}) - \Lambda(du)\} + \int_0^t \{\widetilde{\phi}(u) - \phi(u, \bar{\beta})\}\Lambda(du) \\ &+ \int_0^t \{\widehat{\phi}(u) - \widetilde{\phi}(u)\}\Lambda(du) + \int_0^t \{\widehat{\phi}(u) - \phi(u, \bar{\beta})\}\{\widehat{\Lambda}(du, \widehat{\theta}) - \Lambda(du)\} \\ &\equiv \text{(I)} + \text{(II)} + \text{(III)} + \text{(IV)}. \end{aligned} \tag{A.4}$$

Term (IV) is $o_p(n_1^{-1/2})$, and thus asymptotically negligible (For details, see Section C, Supplementary Materials). Regarding (I), by integration by parts, we can write

$$\text{(I)} = \phi(t, \bar{\beta})\{\widehat{\Lambda}(t, \widehat{\theta}) - \Lambda(t)\} - \int_0^t \phi'(u, \bar{\beta})\{\widehat{\Lambda}(u, \widehat{\theta}) - \Lambda(u)\}du = \text{(Ia)} + \text{(Ib)} + o_p(n^{-1/2})$$

with

$$\text{(Ia)} = \left[\phi(t, \bar{\beta})\hat{\Lambda}'(t, \theta_0) - \int_0^t \phi'(u, \bar{\beta})\hat{\Lambda}'(u, \theta_0)du\right](\widehat{\theta} - \theta_0)$$

and

$$\text{(Ib)} = \phi(t, \bar{\beta})\{\widehat{\Lambda}(t, \theta_0) - \Lambda(t)\} - \int_0^t \phi'(u, \bar{\beta})\{\widehat{\Lambda}(u, \theta_0) - \Lambda(u)\}du,$$

where $\hat{\Lambda}'(t, \theta_0)$ is the derivative of $\hat{\Lambda}(t, \theta)$ with respect to θ and evaluated at θ_0, which can be shown to converge to $\Lambda'(t, \theta_0)$ uniformly in t using similar techniques as in the proof for $\widehat{\Lambda}(\cdot, \theta_0)$. The asymptotic results shown above for $\widehat{\theta}$ and $\widehat{\Lambda}(t, \theta)$, along with the continuous mapping theorem, imply that (I) is asymptotically equivalent to a quantity of the form (11) and that $n^{1/2}$ times (I) is weakly convergent.

We next show that (II) is asymptotically equivalent to the average of i.i.d. mean-zero terms involving the n_1 families only and that $n^{1/2}$ times (II) converges weakly to a Gaussian process. Here we sketch the argument; full details are provided in Section B, the Supplementary Materials.

Denoting $Y_i = \exp(-\bar{\beta}' Z_i)$ and $\Delta_i = Y_i - \phi(X_{i0}, \bar{\beta})$, it can be shown that

$$\int_0^t \{\widetilde{\phi}(u) - \phi(u, \bar{\beta})\}\Lambda(du) \quad = \quad \frac{1}{n_1}\sum_{i=1}^{n} \delta_{i0}\Omega^{\circ}(t, X_{i0})\Delta_i + o_p(n_1^{-1/2}), \tag{A.5}$$

where $\Omega^{\circ}(t,x) = \frac{\lambda(x)}{g(x)} I(x \leq t)$. It is easy to see that the main term, which is an average of mean 0 i.i.d. processes in t, converges weakly in $\ell^{\infty}[0,\tau]$ to $\mathbb{B}(V(t))$, where $\mathbb{B}$ is a Brownian motion process and the variance

$$V(t) = \int_0^t \left\{\frac{\lambda(u)}{g(u)}\right\}^2 \sigma^2(u) g(u) du$$

with $\sigma^2(u) = \mathrm{var}(Y_i | X_{i0} = u, \delta_{i0} = 1)$.

For term (III), a Taylor expansion gives (III) $= \mathcal{Q}(t,\bar{\beta})'(\widehat{\beta} - \bar{\beta}) + o_p(n^{-1/2})$, where

$$\mathcal{Q}(t,\bar{\beta}) = \int_0^t \left[\sum_{i=1}^{n} \{c_i(u) Z_i \exp(-\bar{\beta} Z_i)\}\right] \Lambda(du)$$

Using Claim 11 in Section B, the Supplementary Materials, we find that $\mathcal{Q}(t)$ converges to a limit $\bar{\mathcal{Q}}(t,\bar{\beta})$ uniformly in probability, and $\widehat{\beta} - \bar{\beta}$ has already been shown above to be asymptotically equivalent to a quantity of the form (11).

Taking the results for (I)-(IV) together, and noting that the sum of tight processes is tight, we find that

$$\widehat{\Lambda}_0(t) - \Lambda_0(t,\bar{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \Upsilon(X_i, t) + o_p(n^{-1/2})$$

uniformly in t, where $E[\Upsilon(X_i,t)|\delta_{i0}] = 0$, and that $n^{1/2}$ times the main term on the right side converges weakly to a Gaussian process, with weak convergence defined in terms of the uniform metric. Theorem 1 is thus proved.