Web-based Supplementary Materials for "Assessing Association for Bivariate Survival Data with Interval Sampling: A Copula Model Approach with Application to AIDS Study" by Hong Zhu and Mei-Cheng Wang

The following conditions are assumed throughout the paper.
(1) Assume that standard regularity conditions for maximum likelihood estimate hold.
(2) Define functions

$$W_\alpha\{\alpha, S_Y(y), S_Z(z)\} = \frac{\partial \log l(\alpha, u, v)}{\partial \alpha}, \quad V_\alpha\{\alpha, S_Y(y), S_Z(z)\} = \frac{\partial^2 \log l(\alpha, u, v)}{\partial \alpha^2}$$

$$V_{\alpha,1}\{\alpha, S_Y(y), S_Z(z)\} = \frac{\partial^2 \log l(\alpha, u, v)}{\partial \alpha \partial u}, \quad V_{\alpha,2}\{\alpha, S_Y(y), S_Z(z)\} = \frac{\partial^2 \log l(\alpha, u, v)}{\partial \alpha \partial v}$$

Assume that they are continuous and bounded for $(y, z) \in \mathcal{A} = [y_-, y_+] \times [z_-, z_+]$.

Web Appendix: Proof of Theorem 1

Under the assumed conditions listed above, we derive the asymptotic properties of $\hat{\alpha}(\theta)$. For simplicity of discussion, we denote $\hat{\alpha}(\theta)$ by $\hat{\alpha}$ since θ is known. First, we show the consistency of $\hat{\alpha}$. The score function of α is $U_\alpha^{(c)}(\alpha, S_Y, S_Z) = \frac{\partial}{\partial \alpha} \sum_{i=1}^n \log\, l(\alpha, S_Y, S_Z)$ and pseudo score function is $U_\alpha^{(p)}(\alpha, \hat{S}_Y, \hat{S}_Z) = \frac{\partial}{\partial \alpha} \sum_{i=1}^n \log\, l(\alpha, \hat{S}_Y, \hat{S}_Z)$ by substituting $S_Y(y, \theta)$ and $S_Z(y, \theta)$ in score function by $\hat{S}_Y$ and $\hat{S}_Z$. We have $\hat{S}_Y(\cdot)$ converges in probability to $S_Y(\cdot)$ uniformly in $[y_-, y_+]$, $\hat{S}_Z(\cdot)$ converges to $S_Z(\cdot)$ uniformly in $[z_-, z_+]$, and $U_\alpha^{(c)}(\alpha, u, v)$ is a continuous function of u and v. Therefore, for every α, $U_\alpha^{(p)}(\alpha, \hat{S}_Y, \hat{S}_Z)$ converge to $U_\alpha^{(c)}(\alpha, S_Y, S_Z)$ in probability. This pointwise convergence implies that the solution to $U_\alpha^{(p)}(\alpha, \hat{S}_Y, \hat{S}_Z)$, $\hat{\alpha}$, is consistent by the similar arguments to those in Samuelsen (1997).

Next, we show the asymptotic normality of $\hat{\alpha}$. By Taylor expansion on the pseudo score function $U_\alpha^{(p)}(\alpha, \hat{S}_Y, \hat{S}_Z)$ around α_0, rearranging and evaluating it at $\alpha = \hat{\alpha}$, we get

$$n^{1/2}(\hat{\alpha} - \alpha_0) \cong \frac{-U_\alpha^{(p)}\{\alpha_0, \hat{S}_Y(Y_i), \hat{S}_Z(X_i)\}/\sqrt{n}}{\sum_{i=1}^n V_\alpha\{\alpha_0, \hat{S}_Y(Y_i), \hat{S}_Z(X_i)\}/n}$$

Since $\hat{S}_Y(\cdot)$ converges in probability to $S_Y(\cdot)$ uniformly in $[y_-, y_+]$, $\hat{S}_Z(\cdot)$ converges to $S_Z(\cdot)$ uniformly in $[z_-, z_+]$, and $V_\alpha(\alpha, u, v)$ is a continous function of u and v, $|V_\alpha\{\alpha_0, \hat{S}_Y(y), \hat{S}_Z(z)\} - V_\alpha\{\alpha_0, S_Y(y), S_Z(z)\}|$ converges in probability to zero for $(y, z) \in \mathcal{A} = [y_-, y_+] \times [z_-, z_+]$. Thus $\sum_{i=1}^n -V_\alpha\{\alpha_0, \hat{S}_Y(Y_i), \hat{S}_Z(X_i)\}/n$ and $\sum_{i=1}^n -V_\alpha\{\alpha_0, S_Y(Y_i), S_Z(X_i)\}/n$ are asymptotically equivalent, which by the law of large numbers converges to ρ_1^2, specified as

$$\rho_1^2 = E[-V_\alpha\{\alpha_0, S_Y(Y_i), S_Z(X_i)\}] = \int_{\mathcal{A}} -V_\alpha\{\alpha_0, S_Y(y), S_Z(z)\} dJ_{\alpha_0}(y, z, \delta)$$

where J_{α_0} is the joint distribution of (Y, X, δ). Next, we have

$$\begin{aligned} n^{-1/2} U_\alpha^p(\alpha_0, \hat{S}_Y, \hat{S}_Z) &= n^{1/2} \int_{\mathcal{A}} W_\alpha\{\alpha_0, \hat{S}_Y(y), \hat{S}_Z(z)\} dJ_n(y, z, \delta) \\ &= n^{1/2} \int_{\mathcal{A}} W_\alpha\{\alpha_0, \hat{S}_Y(y), \hat{S}_Z(z)\} dJ_{\alpha_0}(y, z, \delta) \\ &+ n^{1/2} \int_{\mathcal{A}} W_\alpha\{\alpha_0, \hat{S}_Y(y), \hat{S}_Z(z)\} (dJ_n - dJ_{\alpha_0})(y, z, \delta) \\ &= \pi_n(\alpha_0, \hat{S}_Y, \hat{S}_Z) + \eta_n(\alpha_0, \hat{S}_Y, \hat{S}_Z) \end{aligned}$$

where J_n is the empirical distribution of J_{α_0}. We further decompose η_n into two terms,

$$\begin{aligned} \eta_n(\alpha_0, \hat{S}_Y, \hat{S}_Z) &= n^{1/2} \int_{\mathcal{A}} [W_\alpha\{\alpha_0, \hat{S}_Y(y), \hat{S}_Z(z)\} - W_\alpha\{\alpha_0, S_Y(y), S_Z(z)\}] \\ & (dJ_n - dJ_{\alpha_0})(y, z, \delta) \\ &+ n^{1/2} \int_{\mathcal{A}} W_\alpha\{\alpha_0, S_Y(y), S_Z(z)\} (dJ_n - dJ_{\alpha_0})(y, z, \delta) \end{aligned}$$

Because $\hat{S}_Y \to S_Y$, $\hat{S}_Z \to S_Z$, $n^{1/2}(J_n - J) \to O_p(1)$, and W_α is continuous and bounded, by the dominated convergence theorem, the first term in η_n convergence to 0. The second term of η_n is a sum of n i.i.d. random variables of mean zero and variance ρ_1^2, so it converges to normal with mean zero and variance ρ_1^2 by the central limit theorem. Using Von Mises expansion on $\pi_n(\alpha_0, \hat{S}_Y, \hat{S}_Z)$ around S_Y and S_Z, we get

$$\begin{aligned}
\pi_n(\alpha_0, \hat{S}_Y, \hat{S}_Z) &\cong \pi_n(\alpha_0, S_Y, S_Z) + n^{1/2}\int IC_Y(y)d(\hat{S}_Y - S_Y)(y) \\
&+ n^{1/2}\int IC_Z(z)d(\hat{S}_Z - S_Z)(z) \\
&= 0 + n^{1/2}\int IC_Y(y)d(\hat{S}_Y - S_Y)(y) \\
&+ n^{1/2}\int IC_Z(z)d(\hat{S}_Z - S_Z)(z)
\end{aligned}$$

where IC_Y and IC_Z are obtained by differentiating $\pi\{\alpha_0, (1-\varepsilon_1)S_Y + \varepsilon_1\hat{S}_Y, (1-\varepsilon_2)S_Z + \varepsilon_2\hat{S}_Z\}$ with respect to ε_1 and ε_2 and evaluating at $\varepsilon_1 = \varepsilon_2 = 0$, and $IC_Y(y) = -\int_0^y\int_0^{z_0} V_{\alpha,1}\{\alpha_0, S_Y(u), S_Z(z)\}j_{\alpha_0}(u, z, \delta)dzdu$ and $IC_Z(z) = -\int_0^z\int_0^{y_0} V_{\alpha,2}\{\alpha_0, S_Y(y), S_Z(u)\}j_{\alpha_0}(y, u, \delta)dydu$. By the counting process asymptotic techniques, $n^{1/2}\{\hat{S}_Y(y) - S_Y(y)\}$ is asymptotically equivalent to a sum of n i.i.d. random variables as $\sum_i n^{-1/2}I_1^0(Y_i)(y)$, and $n^{1/2}\{\hat{S}_Z(z) - S_Z(z)\}$ is asymptotically equivalent to a sum of n i.i.d. random variables as $\sum_i n^{-1/2}I_2^0(X_i, \delta_i)(z)$. I_1^0 and I_2^0 are martingales, defined as $I_1^0(Y_i)(y) = -S_Y(y)\{\int_0^y \frac{dN_{1i}(u)}{p(Y\geq u)} - \int_0^y \frac{I(Y_i\geq u)d\Lambda_1(u)}{p(Y\geq u)}\}$ and $I_2^0(X_i, \delta_i)(z) = -S_Z(z)\{\int_0^z \frac{dN_{2i}(u)}{p(Z\geq u, C_2\geq u)} - \int_0^z \frac{I(X_i\geq u)d\Lambda_2(u)}{p(Z\geq u, C_2\geq u)}\}$ where $C_2 = C - T - Y$, $N_{1i}(u) = I(Y_i \leq u)$, $N_{2i}(u) = I(Z_i \leq u, \delta_i = 1)$, and Λ_1 and Λ_2 are the cumulative hazard functions for Y and Z. Then we have

$$\begin{aligned}
\pi_n(\alpha_0, \hat{S}_Y, \hat{S}_Z) &\cong n^{-1/2}\Big[\sum_i \int_{\mathcal{A}} V_{\alpha,1}\{\alpha_0, S_Y(y), S_Z(z)\}I_1^0(Y_i)(y)dJ_{\alpha_0}(y, z, \delta) \\
&+ \int_{\mathcal{A}} V_{\alpha,2}\{\alpha_0, S_Y(y), S_Z(z)\}I_2^0(X_i, \delta_i)(z)dJ_{\alpha_0}(y, z, \delta)\Big] \\
&= n^{-1/2}\{\sum_i I_1(Y_i, \alpha_0) + I_2(X_i, \delta_i, \alpha_0)\}
\end{aligned}$$

which is a sum of n i.i.d. random variables. Since IC_Y and IC_Z are deterministic functions, the expectations of I_1 and I_2 are 0. By the central limit theorem, $\pi_n(\alpha_0, \hat{S}_Y, \hat{S}_Z)$ converges to normal with mean 0 and variance ρ_2^2, specified as

$$\rho_2^2 = E[\{I_1(Y, \alpha_0) + I_2(X, \delta, \alpha_0)\}^2] = \int_{\mathcal{A}} \{I_1(y, \alpha_0) + I_2(z, \delta, \alpha_0)\}^2 dJ_{\alpha_0}(y, z, \delta)$$

Note that we have proved that $\pi_n(\alpha_0, \hat{S}_Y, \hat{S}_Z)$ is asymptotically equivalent to $n^{1/2}\{\sum_i I_1(Y_i, \alpha_0) + I_2(X_i, \delta_i, \alpha_0)\}$, and $\eta_n(\alpha_0, \hat{S}_Y, \hat{S}_Z)$ is asymptotically equivalent to $n^{1/2}\sum_i W_\alpha\{\alpha_0, S_Y(Y_i), S_Z(X_i)\}$. Since π_n and η_n are asymptotically independent by the similar arguments in the proof of Theorem 1 in Shih and Louis (1995), $n^{1/2}(\hat{\alpha}-\alpha_0)$ converges to normal with mean zero and variance $\sigma^2 = (\rho_1^2+\rho_2^2)/\rho_1^4$.

The variance estimator $\hat{\sigma}^2$ can be obtained by replacing J by its empirical distribution function J_n, and S_Y, S_Z, α by $\hat{S}_Y$, $\hat{S}_Z$, $\hat{\alpha}$. Specifically,

$$\hat{\rho}_1^2 = \int_{\mathcal{A}} -V_\alpha\{\alpha_0, S_Y(y), S_Z(z)\} dJ_n(y,z,\delta) = \frac{1}{n}\sum_{i=1}^{n} -V_\alpha\{\hat{\alpha}, \hat{S}_Y(Y_i), \hat{S}_Z(X_i)\}.$$

$$\begin{aligned}
\hat{\rho}_2^2 &= \int_{\mathcal{A}} \{\hat{I}_1(y,\hat{\alpha}) + \hat{I}_2(z,\delta,\hat{\alpha})\}^2 dJ_n(y,z,\delta) = \frac{1}{n}\sum_{i=1}^{n}\{\hat{I}_1(Y_i,\hat{\alpha}) + \hat{I}_2(X_i,\delta_i,\hat{\alpha})\}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\Big[\frac{1}{n}\sum_{j=1}^{n} V_{\alpha,1}\{\hat{\alpha}, \hat{S}_Y(Y_j), \hat{S}_Z(X_j)\}\hat{I}_1^0(Y_i)(Y_j) \\
&+ \frac{1}{n}\sum_{j=1}^{n} V_{\alpha,2}\{\hat{\alpha}, \hat{S}_Y(Y_j), \hat{S}_Z(X_j)\}\hat{I}_2^0(X_i,\delta_i)(X_j)\Big]^2
\end{aligned}$$

where $\hat{I}_1^0(Y_i)(Y_j)$ and $\hat{I}_2^0(X_i,\delta_i)(X_j)$ are the corresponding empirical estimators. Since $\hat{\alpha}\to\alpha_0$, $\hat{S}_Y \to S_Y$, $\hat{S}_Z\to S_Z$, and V_α, $V_{\alpha,1}$, $V_{\alpha,2}$ are continuous functions, $\hat{\rho}_1^2$ converges to ρ_1^2, $\hat{\rho}_2^2$ converges to ρ_2^2, and $\hat{\sigma}^2$ converges to σ^2 in probability respectively.

References

[1] Samuelsen, S. O. (1997). A pseudolikelihood approach to analysis of nested case-control studies. *Biometrika* **84,** 379–394.

[2] Shih, J. H. and Louis, T. A. (1995). Inferences on the association parameters in copula models for bivariate survival data. *Biometrics* **51,** 1384–1399.