*Biometrika* (2012), **99**, 1, *pp*. 1–8

*Printed in Great Britain*
Advance Access publication on 31 July 2012

# Supplementary material for 'Sieve maximum likelihood regression analysis of dependent current status data'

BY LING MA
*Department of Statistics, University of Missouri, Columbia, Missouri 65211, U.S.A.*
lingma@mizzou.edu

TAO HU
*School of Mathematical Sciences, Capital Normal University, Beijing 100048, China*
hutaomath@foxmail.com

AND JIANGUO SUN
*Department of Statistics, University of Missouri, Columbia, Missouri 65211, U.S.A.*
sunj@missouri.edu

SUMMARY

This supplementary material contains some additional simulation results and the proofs of the three theorems.

## S1. ADDITIONAL SIMULATION RESULTS

In this section, we present some additional results obtained from the simulation studies conducted to evaluate the performance of the proposed estimation procedure assuming that both the copula model and the degree of association $\tau$ are correctly specified. Here we use the same notation and also the data were generated in the same way as in Section 4 of the paper. Table S1 presents the results on estimation of regression parameters with the data generated under the Gumbel copula with $\tau = 0{\cdot}1$ or $0{\cdot}5$, $\beta_0 = (0,\ 0{\cdot}5)^{\mathrm{T}}$ or $(0{\cdot}5,\ 0{\cdot}5)^{\mathrm{T}}$, and $\gamma_0 = (0{\cdot}5,\ 1)^{\mathrm{T}}$ or $(0{\cdot}5,\ 0{\cdot}5)^{\mathrm{T}}$. The results here are similar to those given in the left panel of part I of Table 1 in the paper except using a different copula model. Table S2 and Table S3 give the same results as in Table S1 except under the Farlie–Gumbel–Morgenstern copula with $\tau = -\ 0{\cdot}1$ or $0{\cdot}1$, $\beta_0 = (0,\ -0{\cdot}5)^{\mathrm{T}}$ or $(0, 0{\cdot}5)^{\mathrm{T}}$, and $\gamma_0 = (0{\cdot}5,\ 1)^{\mathrm{T}}$, and the Frank copula with $\tau = -\ 0{\cdot}5$ or $0{\cdot}5$, $\beta_0 = (-0{\cdot}5,\ -0{\cdot}5)^{\mathrm{T}}$ or $(0{\cdot}5, 0{\cdot}5)^{\mathrm{T}}$, and $\gamma_0 = (0{\cdot}5, 0{\cdot}5)^{\mathrm{T}}$, respectively. To assess the normality of the estimates, we obtained and present in Fig. 1 the Q-Q plots when data were generated and analyzed using Frank copula with $\tau = 0{\cdot}5$, $\beta_0 = (0{\cdot}5,\ 0{\cdot}5)^{\mathrm{T}}$ and $\gamma_0 = (0{\cdot}5,\ 0{\cdot}5)^{\mathrm{T}}$. They suggest that the normal approximation seems to be reasonable.

## S2. PROOFS OF THE ASYMPTOTIC PROPERTIES

In this section, we sketch the proofs of Theorems 1–3 of the paper. To prove Theorems 1–3, we will employ the theory of empirical processes and some techniques commonly

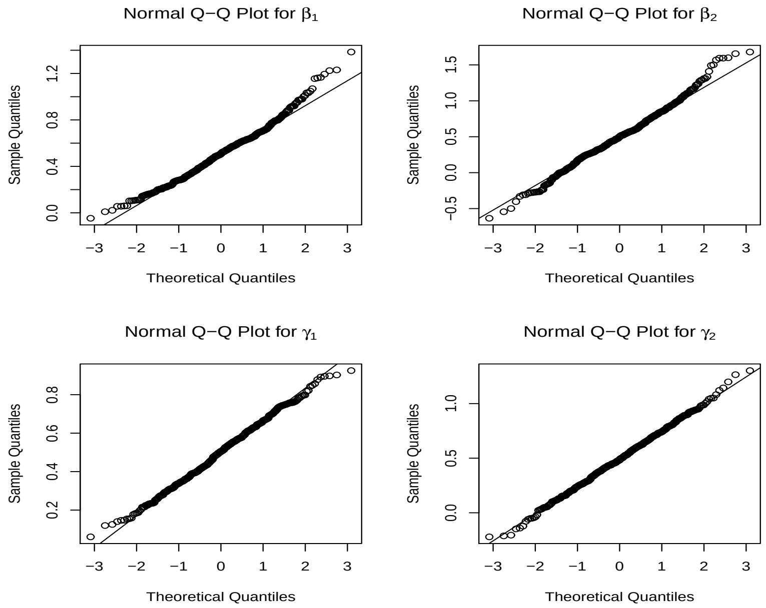


Fig. 1. The Q-Q plots for estimates of regression parameters when data were generated and analyzed using Frank copula with $\tau = 0{\cdot}5$.

Table S1. *Simulation results under Gumbel copula with correctly specified copula models and $\tau$.*

| | | | $\tau = 0{\cdot}1$ | | | | | $\tau = 0{\cdot}5$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | True | Bias $\times 10^2$ | SSE $\times 10^2$ | SEE $\times 10^2$ | CP $\times 10^2$ | True | Bias $\times 10^2$ | SSE $\times 10^2$ | SEE $\times 10^2$ | CP $\times 10^2$ |
| $\beta_1$ | 0 | $-3$ | 24 | 22 | 93 | 0·5 | 3 | 23 | 21 | 94 |
| $\beta_2$ | 0·5 | $-2$ | 39 | 40 | 96 | 0·5 | $-1$ | 38 | 35 | 94 |
| $\gamma_1$ | 0·5 | 1 | 16 | 15 | 95 | 0·5 | 0 | 16 | 15 | 95 |
| $\gamma_2$ | 1 | 0 | 28 | 27 | 95 | 0·5 | $-2$ | 25 | 26 | 96 |

Table S2. *Simulation results under Farlie–Gumbel–Morgenstern copula with correctly specified copula models and $\tau$.*

| | | | $\tau = -0{\cdot}1$ | | | | | $\tau = 0{\cdot}1$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | True | Bias $\times 10^2$ | SSE $\times 10^2$ | SEE $\times 10^2$ | CP $\times 10^2$ | True | Bias $\times 10^2$ | SSE $\times 10^2$ | SEE $\times 10^2$ | CP $\times 10^2$ |
| $\beta_1$ | 0 | 1 | 27 | 27 | 95 | 0 | $-1$ | 24 | 23 | 95 |
| $\beta_2$ | $-0{\cdot}5$ | $-3$ | 47 | 47 | 96 | 0·5 | 0 | 41 | 41 | 95 |
| $\gamma_1$ | 0·5 | 0 | 17 | 16 | 93 | 0·5 | 1 | 16 | 15 | 95 |
| $\gamma_2$ | 1 | 2 | 27 | 27 | 96 | 1 | 0 | 27 | 27 | 95 |

used in nonparametric literature. For ease of exposition, we denote $Pg = \int g(x)dP(x)$ and $P_n g = n^{-1}\sum_{i=1}^{n} g(X_i)$ with the empirical process indexed by function $g(X)$. Also denote $K$ as a constant that may differ from place to place in the proofs.

To prove Theorem 1, we will first define the covering number and establish two lemmas. Let $X = (X_1, \cdots, X_n)$ denote the observed data as before. For $\epsilon > 0$, define the covering number

Table S3. *Simulation results under Frank copula with correctly specified copula models and $\tau$.*

| | $\tau = -0{\cdot}5$ | | | | | $\tau = 0{\cdot}5$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | True | Bias $\times 10^2$ | SSE $\times 10^2$ | SEE $\times 10^2$ | CP $\times 10^2$ | True | Bias $\times 10^2$ | SSE $\times 10^2$ | SEE $\times 10^2$ | CP $\times 10^2$ |
| $\beta_1$ | $-0{\cdot}5$ | $-2$ | 31 | 29 | 95 | 0·5 | 1 | 22 | 21 | 93 |
| $\beta_2$ | $-0{\cdot}5$ | 0 | 51 | 0·51 | 96 | 0·5 | 0 | 37 | 35 | 93 |
| $\gamma_1$ | 0·5 | 1 | 15 | 16 | 94 | 0·5 | 0 | 16 | 15 | 95 |
| $\gamma_2$ | 0·5 | 0 | 27 | 27 | 95 | 0·5 | $-1$ | 26 | 26 | 95 |

$N(\epsilon, \mathcal{L}_n, L_1(P_n))$ as the smallest value of $\kappa$ for which there exists $\{\theta^{(1)}, \ldots, \theta^{(\kappa)}\}$ such that

$$\min_{j\in\{1,\cdots,\kappa\}} \frac{1}{n}\sum_{i=1}^{n} \left| l(\theta, X_i) - l(\theta^{(j)}, X_i)\right| < \epsilon$$

for all $\theta \in \Theta_n$, where $\theta^{(j)} = (\beta^{(j)\mathrm{T}}, \gamma^{(j)\mathrm{T}}, \Lambda_T^{(j)}, \Lambda_C^{(j)})^{\mathrm{T}} \in \Theta_n,\ j = 1, \cdots, \kappa$. We will define $N(\epsilon, \mathcal{L}_n, L_1(P_n)) = \infty$ if no such $\kappa$ exists.

LEMMA 1. *Calculation of the covering number: Let $l(\theta, X)$ denote the log likelihood function based on the single observation $X$ as defined in Section 2. Assume that Conditions 1 and 2 hold. Then the covering number of the class $\mathcal{L}_n = \{l(\theta, X) : \theta \in \Theta_n\}$ satisfies*

$$N\big\{\epsilon, \mathcal{L}_n, L_1(P_n)\big\} \leq K M^{2p} M_n^{2(m+k_n)} \epsilon^{-p_m},$$

*where $p_m = 2p + 2(m + k_n)$.*

*Proof.* For any $\theta^1 = (\beta^{1^{\mathrm{T}}}, \gamma^{1^{\mathrm{T}}}, \Lambda_T^1, \Lambda_C^1)^{\mathrm{T}},\ \theta^2 = (\beta^{2^{\mathrm{T}}}, \gamma^{2^{\mathrm{T}}}, \Lambda_T^2, \Lambda_C^2)^{\mathrm{T}} \in \Theta_n$, we can easily obtain

$$|l(\theta^1, X) - l(\theta^2, X)| \leq K\big(\|\beta^1 - \beta^2\| + \|\gamma^1 - \gamma^2\| + \|\Lambda_T^1 - \Lambda_T^2\|_\infty + \|\Lambda_C^1 - \Lambda_C^2\|_\infty\big)$$

for some constant $K$ using Taylor's series expansion under Conditions 1 and 2, where $\|f\|_\infty = \sup_t |f(t)|$ is defined as the supremum norm for a function $f$.

Let $\xi^j = (\xi_1^j, \cdots, \xi_{m+k_n}^j)^{\mathrm{T}}$ denote the I-spline coefficients corresponding to $\Lambda_T^j$ with $j = 1, 2$. Then one can show that

$$\begin{aligned}\|\Lambda_T^1 - \Lambda_T^2\|_\infty &= \sup_t \Big| \sum_{j=1}^{m+k_n} \xi_j^1 I_j(t) - \sum_{j=1}^{m+k_n} \xi_j^2 I_j(t)\Big| = \sup_t \Big| \sum_{j=1}^{m+k_n} (\xi_j^1 - \xi_j^2) I_j(t)\Big| \\ &\leq \max_{1\leq j\leq m+k_n} |\xi_j^1 - \xi_j^2| \sup_t \Big| \sum_{j=1}^{m+k_n} I_j(t)\Big| = K_1\|\xi^1 - \xi^2\|\end{aligned}$$

with $K_1 = \sup_t \big|\sum_{j=1}^{m+k_n} I_j(t)\big|$. Similarly by letting $\eta^j$ denote the I-spline coefficients corresponding to $\Lambda_C^j$, respectively, $j = 1, 2$, we will have $\|\Lambda_C^1 - \Lambda_C^2\|_\infty \leq K_1\|\eta^1 - \eta^2\|$. It follows that

$$|l(\theta^1, X) - l(\theta^2, X)| \leq K\|\beta^1 - \beta^2\| + K\|\gamma^1 - \gamma^2\| + K\|\xi^1 - \xi^2\| + K\|\eta^1 - \eta^2\|,$$

which gives

$$\frac{1}{n}\sum_{i=1}^{n} \left|l(\theta, X_i) - l(\theta^{(j)}, X_i)\right| \leq K\big[\|\beta - \beta^{(j)}\| + \|\gamma - \gamma^{(j)}\|\big] + K\|\xi - \xi^{(j)}\| + K\|\eta - \eta^{(j)}\|,$$

for any $\theta \in \Theta_n$.

By the calculation in page 94 of van der Vaart & Wellner (1996), one can show that $\{(\beta,\gamma)\in R^{2p}, \|\beta\|+\|\gamma\|\le M\}$ is covered by $[5M/\{\epsilon/(3K)\}]^p$ balls with radius $\epsilon/(3K)$. Similarly one can find the number of balls with radius $\epsilon/(3K)$ to cover $\{\xi\in R^{m+k_n}, \sum_1^{m+k_n}|\xi_j|\le M_n\}$ and $\{\eta\in R^{m+k_n}, \sum_1^{m+k_n}|\eta_j|\le M_n\}$, respectively. Thus,

$$\begin{aligned} N\big\{\epsilon,\mathcal{L}_n,L_1(P_n)\big\} &\le \Big(\frac{15MK}{\epsilon}\Big)^{2p}\cdot\Big(\frac{15M_nK}{\epsilon}\Big)^{m+k_n}\cdot\Big(\frac{15M_nK}{\epsilon}\Big)^{m+k_n}\\ &\le KM^{2p}M_n^{2(m+k_n)}\epsilon^{-p_m}. \end{aligned}$$

This completes the proof. □

LEMMA 2. *Uniform convergence: Assume that Conditions 1 and 2 hold. Then,*

$$\sup_{\theta\in\Theta_n}\big|P_nl(\theta,X)-Pl(\theta,X)\big|\to 0$$

*almost surely.*

*Proof.* Let $\delta_n=1$, $\nu/2<\phi_1<1/2$, and $\alpha_n=n^{-1/2+\phi_1}(\log n)^{1/2}$. The $\{\alpha_n\}$ is a nonincreasing sequence. Also for a fixed $\epsilon>0$, let $\epsilon_n=\epsilon\alpha_n$. Then for any $l(\theta,X)\in\mathcal{L}_n$ and sufficiently large $n$, we have

$$\mathrm{var}\{P_nl(\theta,X)\}/(4\epsilon_n)^2\le\frac{(1/n)Pl^2(\theta,X)}{16\epsilon^2\alpha_n^2}\le\frac{K}{16\epsilon^2n\alpha_n^2}<<\frac{1}{16\epsilon^2\log n}<\frac{1}{2},$$

where $a_n<<b_n$ means $a_n/b_n\to 0$.

Furthermore, by applying the inequality (31) and Lemma 33 of Pollard (1984) and the Lemma 1 above, we can show that

$$\begin{aligned} &P\big\{\sup_{\mathcal{L}_n}|P_nl(\theta,X)-Pl(\theta,X)|>8\epsilon_n\big\}\\ &\le 8N\{\epsilon_n,\mathcal{L}_n,L_1(P_n)\}\exp(-n\epsilon_n^2/128)P\{\sup_{\mathcal{L}_n}|P_nl^2(\theta,X)|\le 64\}+P\{\sup_{\mathcal{L}_n}|P_nl^2(\theta,X)|>64\}\\ &\le KM^{2p}M_n^{2(m+k_n)}\epsilon_n^{-p_m}\exp(-n\epsilon_n^2/128)\\ &=K\exp[(p_m-2p)a\log n-p_m\log\{\epsilon n^{-1/2+\phi_1}(\log n)^{1/2}\}-n\epsilon^2n^{-1+2\phi_1}\log n/128]\\ &\le K\exp[p_m\{(a+1/2-\phi_1)\log n-\log\log n/2-\log\epsilon\}-\epsilon^2n^{2\phi_1}\log n/128]\\ &\le K\exp\big(-Kn^{2\phi_1}\log n\big). \end{aligned}$$

It follows that $\sum_{n=1}^{\infty}P\{\sup_{\mathcal{L}_n}|P_nl(\theta,X)-Pl(\theta,X)|>8\epsilon_n\big\}<\infty$, and by the Borel-Cantelli lemma, we have $\sup_{\mathcal{L}_n}|P_nl(\theta,X)-Pl(\theta,X)|\to 0$ almost surely. This completes the proof. □

*Proof of Theorem 1.* Consider the class of functions $\mathcal{L}_n=\{l(\theta,X):\theta\in\Theta_n\}$. Note that by Lemma 1, we have

$$N\big\{\epsilon,\mathcal{L}_n,L_1(P_n)\big\}\le KM^{2p}M_n^{2(m+k_n)}\epsilon^{-\{2p+2(m+k_n)\}},$$

and according to Lemma 2, we have

$$\sup_{\theta\in\Theta_n}\big|P_nl(\theta,X)-Pl(\theta,X)\big|\to 0, \tag{S1}$$

almost surely.

Let $M(\theta, X) = -l(\theta, X)$ and

$$\zeta_{1n} = \sup_{\theta \in \Theta_n} |P_n M(\theta, X) - PM(\theta, X)|, \ \zeta_{2n} = P_n M(\theta_0, X) - PM(\theta_0, X).$$

Denote $K_\epsilon = \{\theta : d(\theta, \theta_0) \geq \epsilon, \theta \in \Theta_n\}$. One can show that

$$\begin{aligned} \inf_{K_\epsilon} PM(\theta, X) &= \inf_{K_\epsilon} \big\{PM(\theta, X) - P_n M(\theta, X) + P_n M(\theta, X)\big\} \\ &\leq \zeta_{1n} + \inf_{K_\epsilon} P_n M(\theta, X). \end{aligned}$$

If $\hat{\theta}_n \in K_\epsilon$, we have

$$\inf_{K_\epsilon} P_n M(\theta, X) = P_n M(\hat{\theta}_n, X) \leq P_n M(\theta_0, X) = \zeta_{2n} + PM(\theta_0, X).$$

By condition 3, we obtain that $\inf_{K_\epsilon} PM(\theta, X) - PM(\theta_0, X) = \delta_\epsilon > 0$. Thus,

$$\inf_{K_\epsilon} PM(\theta, X) \leq \zeta_{1n} + \zeta_{2n} + PM(\theta_0, X) = \zeta_n + PM(\theta_0, X)$$

with $\zeta_n = \zeta_{1n} + \zeta_{2n}$. Hence we obtain that $\zeta_n \geq \delta_\epsilon$ and furthermore, $\{\hat{\theta}_n \in K_\epsilon\} \subseteq \{\zeta_n \geq \delta_\epsilon\}$. By (S1) and Strong Law of Large Numbers, we have $\zeta_{1n} = o(1)$ and $\zeta_{2n} = o(1)$ almost surely. Therefore, by $\cup_{k=1}^{\infty} \cap_{n=k}^{\infty} \{\hat{\theta}_n \in K_\epsilon\} \subseteq \cup_{k=1}^{\infty} \cap_{n=k}^{\infty} \{\zeta_n \geq \delta_\epsilon\}$, we complete the proof. □

*Proof of Theorem 2.* To establish the convergence rate, note that by Lemma A1 of Lu et al. (2007), there exist I-splines $\Lambda_{Tn0}$ and $\Lambda_{Cn0}$ such that $\|\Lambda_{T0} - \Lambda_{Tn0}\|_\infty = O_p(n^{-r\nu})$ and $\|\Lambda_{C0} - \Lambda_{Cn0}\|_\infty = O_p(n^{-r\nu})$, respectively. For any $\eta > 0$, define the class $\mathcal{F}_\eta = \{l(\theta_{n0}, X) - l(\theta, X) : \theta \in \Theta_n, d(\theta, \theta_{n0}) \leq \eta\}$ with $\theta_{n0} = (\beta_0^{\mathrm{T}}, \gamma_0^{\mathrm{T}}, \Lambda_{Tn0}, \Lambda_{Cn0})^{\mathrm{T}}$. Following the calculation of Shen & Wong (1994) (p.597), we can establish that $\log N_{[]}(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2) \leq CN \log(\eta/\varepsilon)$ with $N = 2(m + k_n)$. Moreover, some algebraic calculations lead to $\|l(\theta_{n0}, X) - l(\theta, X)\|_2^2 \leq C\eta^2$ for any $l(\theta_{n0}, X) - l(\theta, X) \in \mathcal{F}_\eta$.

Therefore, by Lemma 3·4·2 of van der Vaart & Wellner (1996), we obtain

$$E_P \|n^{1/2}(P_n - P)\|_{\mathcal{F}_\eta} \leq CJ_\eta(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2)\bigg\{1 + \frac{J_\eta(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2)}{\eta^2 n^{1/2}}\bigg\}, \tag{S2}$$

where $J_\eta(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2) = \int_0^\eta \{1 + \log N_{[]}(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2)\}^{1/2} d\varepsilon \leq CN^{1/2}\eta$. The right-hand side of (S2) yields $\phi_n(\eta) = C(N^{1/2}\eta + N/n^{1/2})$. It is easy to see that $\phi_n(\eta)/\eta$ is decreasing in $\eta$, and $r_n^2 \phi_n(1/r_n) = r_n N^{1/2} + r_n^2 N/n^{1/2} < 2n^{1/2}$, where $r_n = N^{-1/2} n^{1/2} = n^{(1-\nu)/2}$ with 0< $\nu$ <0·5. Hence $n^{(1-\nu)/2} d(\hat{\theta}, \theta_{n0}) = O_P(1)$ by Theorem 3·4·1 of van der Vaart & Wellner (1996). This, together with $d(\theta_{n0}, \theta_0) = O_p(n^{-r\nu})$ (Lu et al., 2007), yields that $d(\hat{\theta}, \theta_0) = O_p(n^{-(1-\nu)/2} + n^{-r\nu})$. The choice of $\nu = 1/(1 + 2r)$ yields the rate of convergence $d(\hat{\theta}_n, \theta_0) = O_p(n^{-r/(1+2r)})$. □

*Proof of Theorem 3.* Denote $V$ as the linear span of $\Theta - \theta_0$, where $\theta_0$ denotes the true value of $\theta = (\beta^{\mathrm{T}}, \gamma^{\mathrm{T}}, \Lambda_T, \Lambda_C)^{\mathrm{T}}$. Recall that $l(\theta, X)$ is the log-likelihood for a sample of size one and $\delta_n = (n^{-(1-\nu)/2} + n^{-r\nu})$. For any $\theta \in \{\theta \in \Theta_0 : \|\theta - \theta_0\| = O(\delta_n)\}$, define the first order directional derivative of $l(\theta, X)$ at the direction $v \in V$ as

$$\dot{l}(\theta, X)[v] = \frac{dl(\theta + sv, X)}{ds} \, |_{s=0},$$

and the second order directional derivative as

$$\ddot{l}(\theta, X)[v, \tilde{v}] = \frac{d^2 l(\theta + sv + \tilde{s}\tilde{v}, X)}{d\tilde{s}ds} |_{s=0}|_{\tilde{s}=0} = \frac{d\dot{l}(\theta + \tilde{s}\tilde{v}, X)}{d\tilde{s}} |_{\tilde{s}=0} .$$

Also define the Fisher inner product on the space $V$ as

$$< v, \tilde{v} >= P\big\{\dot{l}(\theta, X)[v]\dot{l}(\theta, X)[\tilde{v}]\big\}$$

and the Fisher norm for $v \in V$ as $\|v\|^{1/2} =< v, v >$ . Let $\bar{V}$ be the closed linear span of $V$ under the Fisher norm. Then $(\bar{V}, \| \cdot \|)$ is a Hilbert space.

Furthermore, define the smooth functional of $\theta$ as

$$\gamma(\theta) = b_1^{\mathrm{T}}\beta + b_2^{\mathrm{T}}\gamma,$$

where $b = (b_1^{\mathrm{T}}, b_2^{\mathrm{T}})^{\mathrm{T}}$ is any vector of $2p$ dimension with $\|b\| \le 1$. For any $v \in V$, we denote

$$\dot{\gamma}(\theta_0)[v] = \frac{d\gamma(\theta_0 + sv)}{ds} |_{s=0} .$$

Note that $\gamma(\theta) - \gamma(\theta_0) = \dot{\gamma}(\theta_0)[\theta - \theta_0]$. It follows from the Riesz representation theorem that there exists $v^* \in \bar{V}$ such that $\dot{\gamma}(\theta_0)[v] =< v^*, v >$ for all $v \in \bar{V}$ and $\|v^*\|^2 = \|\dot{\gamma}(\theta_0)\|$. Thus it follows from the Cramér-Wold device that to prove Theorem 3, it suffices to show that

$$n^{1/2} < \hat{\theta} - \theta_0, v^* > \to N(0, b^{\mathrm{T}}\Sigma b)$$

in distribution, since $b^{\mathrm{T}}\{(\hat{\beta} - \beta_0)^{\mathrm{T}}, (\hat{\gamma} - \gamma_0)^{\mathrm{T}}\}^{\mathrm{T}} = \gamma(\hat{\theta}) - \gamma(\theta_0) = \dot{\gamma}(\theta_0)[\hat{\theta} - \theta_0] =< \hat{\theta} - \theta_0, v^* >$ . We will first prove that $n^{1/2} < \hat{\theta} - \theta_0, v^* > \to N(0, \|v^*\|^2)$ in distribution and then that $\|v^*\|^2 = b^{\mathrm{T}}\Sigma b$.

Let $\varepsilon_n$ be any positive sequence satisfying $\varepsilon_n = o(n^{-1/2})$. For any $v^* \in \Theta_0$, by Corollary 6·21 of Schumaker (1981), there exists $\Pi_n v^* \in \Theta_n$ such that $\|\Pi_n v^* - v^*\| = o(1)$ and $\varepsilon_n\|\Pi_n v^* - v^*\| = o(n^{-1/2})$. Define $g[\theta - \theta_0, X] = l(\theta, X) - l(\theta_0, X) - \dot{l}(\theta, X)[\theta - \theta_0]$. Then by definition of $\hat{\theta}$, we have

$$\begin{aligned}
0 &\le P_n\{l(\hat{\theta}, X) - l(\hat{\theta} \pm \varepsilon_n \Pi_n v^*, X)\} \\
&= (P_n - P)\{l(\hat{\theta}, X) - l(\hat{\theta} \pm \varepsilon_n \Pi_n v^*, X)\} + P\{l(\hat{\theta}, X) - l(\hat{\theta} \pm \varepsilon_n \Pi_n v^*, X)\} \\
&= \pm\varepsilon_n P_n \dot{l}(\theta, X)[\Pi_n v^*] + (P_n - P)\big(g[\hat{\theta} - \theta_0, X] - g[\hat{\theta} \pm \varepsilon_n \Pi_n v^* - \theta_0, X]\big) \\
&\quad + P\big(g[\hat{\theta} - \theta_0, X] - g[\hat{\theta} \pm \varepsilon_n \Pi_n v^* - \theta_0, X]\big) \\
&= \mp\varepsilon_n P_n \dot{l}(\theta, X)[v^*] \pm \varepsilon_n P_n \dot{l}(\theta, X)[\Pi_n v^* - v^*] + (P_n - P)\big(g[\hat{\theta} - \theta_0, X] \\
&\quad - g[\hat{\theta} \pm \varepsilon_n \Pi_n v^* - \theta_0, X]\big) + P\big(g[\hat{\theta} - \theta_0, X] - g[\hat{\theta} \pm \varepsilon_n \Pi_n v^* - \theta_0, X]\big) \\
&= \mp\varepsilon_n P_n \dot{l}(\theta, X)[v^*] + I_1 + I_2 + I_3.
\end{aligned}$$

Now we study the three terms on the right side of the equation above. For $I_1$, it follows from Conditions 1–2, Chebyshev inequality and $\|\Pi_n v^* - v^*\| = o(1)$ that $I_1 = \varepsilon_n \times o_p(n^{-1/2})$. For $I_2$, we have

$$\begin{aligned}
I_2 &= (P_n - P)\big\{l(\hat{\theta}, X) - l(\hat{\theta} \pm \varepsilon_n \Pi_n v^*, X) \pm \varepsilon_n \dot{l}(\theta_0, X)[\Pi_n v^*]\big\} \\
&= \mp\varepsilon_n (P_n - P)\big\{\dot{l}(\tilde{\theta}, X) - \dot{l}(\theta_0, X)[\Pi_n v^*]\big\},
\end{aligned}$$

where $\tilde{\theta}$ lies between $\hat{\theta}$ and $\hat{\theta} \pm \varepsilon_n \Pi_n v^*$. By Theorem 2·8·3 of van der Vaart & Wellner (1996), we know that $\{\dot{l}(\theta; X)[\Pi_n v^*] : \|\theta - \theta_0\| = O(\delta_n)\}$ is Donsker class. Therefore, by Corollary 2·3·12 of van der Vaart & Wellner (1996), we have $I_2 = \varepsilon_n \times o_p(n^{-1/2})$.

For $I_3$, note that

$$\begin{aligned}P(g[\theta-\theta_0,X]) &= P\{l(\theta,X)-l(\theta_0,X)-\dot{l}(\theta_0,X)[\theta-\theta_0]\}\\ &= 2^{-1}P\{\ddot{l}(\tilde{\theta},X)[\theta-\theta_0,\theta-\theta_0]-\ddot{l}(\theta_0,X)[\theta-\theta_0,\theta-\theta_0]\}\\ &\quad +2^{-1}P\{\ddot{l}(\theta_0,X)[\theta-\theta_0,\theta-\theta_0]\}\\ &= 2^{-1}P\{\ddot{l}(\theta_0,X)[\theta-\theta_0,\theta-\theta_0]\}+\varepsilon_n\times o_p(n^{-1/2}),\end{aligned}$$

where $\tilde{\theta}$ lies between $\theta_0$ and $\theta$ and the last equation is due to Taylor expansion, Conditions 1–2 and $r>2$. Therefore,

$$\begin{aligned}I_3 &= -2^{-1}(\|\hat{\theta}-\theta_0\|^2-\|\hat{\theta}\pm\varepsilon_n\Pi_n v^*-\theta_0\|^2)+\varepsilon_n\times o_p(n^{-1/2})\\ &= \pm\varepsilon_n<\hat{\theta}-\theta_0,\Pi_n v^*>+2^{-1}\|\varepsilon_n\Pi_n v^*\|^2+\varepsilon_n\times o_p(n^{-1/2})\\ &= \pm\varepsilon_n<\hat{\theta}-\theta_0,v^*>+2^{-1}\|\varepsilon_n\Pi_n v^*\|^2+\varepsilon_n\times o_p(n^{-1/2})\\ &= \pm\varepsilon_n<\hat{\theta}-\theta_0,v^*>+\varepsilon_n\times o_p(n^{-1/2}),\end{aligned}$$

where the last equality holds since $\varepsilon_n\|\Pi_n v^*-v^*\|=o(n^{-1/2})$, Cauchy–Schwartz inequality, and $\|\Pi_n v^*\|^2\to\|v^*\|^2$. Combing the above facts, together with $P\dot{l}(\theta_0,X)[v^*]=0$, we can establish that

$$\begin{aligned}0 &\le P_n\{l(\hat{\theta},X)-l(\hat{\theta}\pm\varepsilon_n\Pi_n v^*,X)\}\\ &= \mp\varepsilon_n P_n\dot{l}(\theta_0,X)[v^*]\pm\varepsilon_n<\hat{\theta}-\theta_0,v^*>+\varepsilon_n\times o_p(n^{-1/2})\\ &= \mp\varepsilon_n(P_n-P)\{\dot{l}(\theta_0,X)[v^*]\}\pm\varepsilon_n<\hat{\theta}-\theta_0,v^*>+\varepsilon_n\times o_p(n^{-1/2}).\end{aligned}$$

Therefore we have that $n^{1/2}<\hat{\theta}-\theta_0,v^*>=n^{1/2}(P_n-P)\{\dot{l}(\theta_0,X)[v^*]\}+o_p(1)\to N(0,\|v^*\|^2)$ in distribution, where the asymptotic normality is guaranteed by Central limits Theorem and the asymptotic variance being equal to $\|v^*\|^2=\|\dot{l}(\theta_0,X)[v^*]\|^2$. This implies $n^{1/2}\{\gamma(\hat{\theta})-\gamma(\theta_0)\}=n^{1/2}<\hat{\theta}-\theta_0,v^*>+o_p(1)\to N(0,\|v^*\|^2)$ in distribution. The semiparametric efficiency can be established by applying the result of Bickel & Kwon (2001) or Theorem 4 in Shen (1997).

Next we will prove that $\|v^*\|^2=b^{\mathrm{T}}\Sigma b$. Recall that $\vartheta=(\beta^{\mathrm{T}},\gamma^{\mathrm{T}})^{\mathrm{T}}$. For each component $\vartheta_q$, $q=1,2,\cdots,2p$, we denote by $\psi_q^*=(b_{1q}^*,b_{2q}^*)$ which minimizes

$$E\Big(l_\vartheta\cdot e_q-l_{b_1}[b_{1q}]-l_{b_2}[b_{2q}]\Big)^2,$$

where $l_\vartheta=(l_\beta^{\mathrm{T}},l_\gamma^{\mathrm{T}})^{\mathrm{T}}$, $e_q$ is a $2p$-dimensional vector of zeros except the $q$-th element equal to 1, $l_{b_1}[b_1]$ and $l_{b_2}[b_1]$ are the directional derivatives with respect to $\Lambda_T$ and $\Lambda_C$, respectively, and can be calculated as directional derivatives defined at the beginning of the proof of Theorem 3. Now let $\psi^*=(\psi_1^*,\cdots,\psi_{2p}^*)$. By the calculations of Chen et al. (2006), we have

$$\|v^*\|^2=\|\dot{\gamma}(\theta_0)\|=\sup_{v\in\bar{V}:\|v\|>0}\frac{|\dot{\gamma}(\theta_0)[v]|}{\|v\|}=b^{\mathrm{T}}\Sigma b,$$

where $\Sigma=\{E(S_\vartheta S_\vartheta^{\mathrm{T}})\}^{-1}$, $S_\vartheta=\big(l_\vartheta-l_{b_1^*}[b_1^*]-l_{b_2^*}[b_2^*]\big)$. Thus the conclusion of the theorem follows by $b^{\mathrm{T}}\{(\hat{\beta}-\beta_0)^{\mathrm{T}},(\hat{\gamma}-\gamma_0)^{\mathrm{T}}\}^{\mathrm{T}}=<\hat{\theta}-\theta_0,v^*>$ and the Cramér-Wold device. □

## REFERENCES

BICKEL, P. J., & KWON, J. (2001). Inference for semiparametric models: some questions and an answer. *Statist. Sinica.* **11**, 863–960.

CHEN, X., FAN, Y., & TSYRENNIKOV, V. (2006). Efficient estimation of semiparametric multivariate copula models. *J. Am. Statist. Assoc.* **101**, 1228–40.

LU, M., ZHANG, Y., & HUANG, J. (2007). Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika* **94**, 705–18.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. New York: Springer.

SCHUMAKER, L. (1981). *Spline Function: Basic Theory*. New York: Wiley.

SHEN, X. (1997). On methods of sieves and penalization. *Ann. Statist.* **25**, 2555–91.

SHEN, X., & WONG, W. H. (1994). Convergence rate of sieve estimates. *Ann. Statist.* **22**, 580–615.

VAN DER VAART, A. W., & WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Berlin: Springer.

ZHANG, Y., HUA, L., & HUANG, J. (2010). A spline-based semiparametric maximum likelihood estimation method for the Cox model with interval-censored data. *Scand. J. Statist.* **37**, 338–54.

ZHENG, M., & KLEIN, J. P. (1995). Estimates of marginal survival for dependent competing risk based on an assumed copula. *Biometrika* **82**, 127–38.

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