Statistica Sinica: Supplement

REGRESSION ANALYSIS OF MULTIVARIATE CURRENT STATUS DATA WITH SEMIPARAMETRIC TRANSFORMATION FRAILTY MODELS

Shuwei Li[1] Tao Hu[2] Shishun Zhao[3] and Jianguo Sun[4]

[1] *School of Economics and Statistics, Guangzhou University, Guangzhou, China*

[2] *School of Mathematical Sciences, Capital Normal University, Beijing, China*

[3] *Center for Applied Statistical Research, School of Mathematics, Jilin University, Changchun, China*

[4] *Department of Statistics, University of Missouri, Columbia, Missouri, USA*

Supplementary Material

PROOFS OF THE ASYMPTOTIC PROPERTIES.

In this part, we sketch the proofs for Theorems 1, 2 and 3. For this, we will mainly use some results about empirical processes given in van der Vaart and Wellner (1996).

Proof of Theorem 1.

To prove the consistency, we will verify the conditions of Theorem 5.7 of van der Vaart (1998). Let $BV[\tau_1, \tau_2]$ denote the functions whose total variation in $[\tau_1, \tau_2]$ are bounded by a given constant. Define $\Theta =$

$\mathcal{A} \otimes \mathcal{B} \otimes BV^{\otimes K}[\tau_1, \tau_2]$. Define the metric $\rho(\theta, \tilde{\theta})$ on the parameter space Θ as $\rho(\theta, \tilde{\theta}) = \|\zeta - \tilde{\zeta}\|^2 + \sum_{k=1}^{K} \sup_{t \in [\tau_1, \tau_2]} |\Lambda_k(t) - \tilde{\Lambda}_k(t)|$. Then the class of functions

$$\mathcal{F} = \left\{ \Lambda_k(C_k) e^{X_k^T \beta} : \beta \in \mathcal{B}, \Lambda_k \in BV[\tau_1, \tau_2] \right\}$$

is a Donsker class. By the condition (A4), we know that

$$\int_b \prod_{k=1}^{K} \left\{ 1 - \exp\left[-G_k\left(\Lambda_k(C_k)\, e^{X_k^T \beta}\, b \right) \right] \right\}^{\Delta_k}$$

$$\times \exp\left[-G_k\left(\Lambda_k(C_k)\, e^{X_k^T \beta}\, b \right) \right]^{1-\Delta_k} \, p(b|\gamma)\, \mathrm{d}b.$$

is bounded away from zero. Therefore, $l(\theta, O) = \log L(\theta, O)$ ($L(\theta, O)$ defined in (2))belongs to some Donsker class due to the preservation property of the Donsker class under the Lipschitz-continuous transformations. Then we can conclude that $\sup_{\theta \in \Theta} |\mathbb{P}_n l(\theta, O) - \mathbb{P} l(\theta_0, O)|$ converges in probability to 0 as $n \to \infty$.

Now, we verify that another condition of Theorem 5.7 of van der Vaart (1998) hold. That is, for any $\epsilon > 0$,

$$\sup_{\rho(\theta, \theta_0) > \epsilon} \mathbb{P} l(\theta, O) < \mathbb{P} l(\theta_0, O).$$

Note that this condition is satisfied if we can prove the model is identifiable. By condition (A5) and similar arguments to the proof of Theorem 2.1 of Chang et al. (2007), we can show the identifiability of the model parameter-

s. Now, by Theorem 5.7 of van der Vaart (1998), we have $\rho(\hat{\theta}_n, \theta_0) = o_p(1)$, which completes the proof of Theorem 1.

To prove the convergence rate of the proposed estimator in Theorem 2, we need the following lemma.

Lemma 1. Let $\mathcal{L} = \{\log\, L(\beta, \gamma, \Lambda) \mid \theta = (\beta, \gamma, \Lambda) \in \Theta\}$ and P be any probability measure on the sample space, then $\log N_{[]}(\varepsilon, \mathcal{L}, L_2(P)) = O(1/\varepsilon)$ as ε decreases to 0.

Proof of Lemma 1.

Let $\mathcal{A} = \{\Lambda = (\Lambda_1, \ldots, \Lambda_K) : M^{-1} < \Lambda_k < M, \text{for } k = 1, \ldots, K\}$, its $\varepsilon-$bracketing number is of the order of $O(e^{1/\varepsilon})$. That is, for each k, there exists a set of functions, $(\Lambda_{kj}^L, \Lambda_{kj}^U)$, such that $\|\Lambda_{kj}^U - \Lambda_{kj}^L\|_2 \leq \varepsilon$ and $\Lambda_{kj}^L \leq \Lambda_k \leq \Lambda_{kj}^U$ for some j.

We define

$$l_{(j,\zeta)}^L(O) = \log \int_b \prod_{k=1}^K \left\{\Delta_k + (-1)^{\Delta_k} \exp\left[-G_k\left(e^{X_k^T\beta}\left[\Delta_k \Lambda_{kj}^L(C_k) + (1-\Delta_k)\Lambda_{kj}^U(c_k)\right] b\right)\right]\right\} p(b|\gamma)\, \mathrm{d}b.$$

$$l_{(j,\zeta)}^U(O) = \log \int_b \prod_{k=1}^K \left\{\Delta_k + (-1)^{\Delta_k} \exp\left[-G_k\left(e^{X_k^T\beta}\left[\Delta_k \Lambda_{kj}^U(c_k) + (1-\Delta_k)\Lambda_{kj}^L(C_k)\right] b\right)\right]\right\} p(b|\gamma)\, \mathrm{d}b.$$

Here, O is a vector whose components are covariates, censoring time and the censoring indicators.

$$f(\vartheta, O) = \log \int_b \prod_{k=1}^{K} \left\{ \Delta_k + (-1)^{\Delta_k} \exp\left[-G_k \left(\Lambda_k(C_k)\, e^{X_k^T \beta}\, b \right) \right] \right\} p(b|\gamma)\, \mathrm{d}b.$$

where $\vartheta = (\zeta, \Lambda_1, \ldots, \Lambda_K)$. Let $\vartheta' = (\zeta', \Delta_1 \Lambda_{1j}^U + (1-\Delta_1)\Lambda_{1j}^L, \ldots, \Delta_K \Lambda_{Kj}^U + (1-\Delta_K)\Lambda_{Kj}^L)$.

Further, by mean value theorem, we have

$$l_{(j,\zeta')}^U(O) - l(\theta, O) = f(\vartheta', O) - f(\vartheta, O) = \sum_{i=1}^{d+1} f_i(\tilde{\vartheta}, O)(\zeta_i' - \zeta_i)$$
$$+ \sum_{k=1}^{K} f_{i+d+1}(\tilde{\vartheta}, O) \left\{ \Delta_k(\Lambda_{kj}^U - \Lambda_k)(C_k) + (1-\Delta_k)(\Lambda_{kj}^L - \Lambda_k)(C_k) \right\}$$

Here, f_i denotes the partial derivative of f with respect to ϑ_i and ϑ_i is the ith component of ϑ. For $(d+1) \le i \le (K+d+1)$, note that $f_i(\vartheta, x)/(2\Delta_i - 1)$ is positive, we have

$$l_{(j,\zeta')}^U(O) - l(\theta, O) \ge -a_0\|\zeta' - \zeta\| + a_1 \sum_{i=1}^{K} (2\Delta_i - 1)[\Delta_k(\Lambda_{kj}^U - \Lambda_k)(C_k) + (1-\Delta_k)(\Lambda_{kj}^L - \Lambda_k)(C_k)],$$

where a_0, a_1 are some positive constants. By the definition of Λ_{kj}^L and Λ_{kj}^U, we know that $l_{(j,\zeta')}^U(O) - l(\theta, O) \ge -a_0\|\zeta' - \zeta\| + a_1\varepsilon$. The same arguments lead to $l_{(j,\zeta')}^L(O) - l(\theta, O) \le a_0\|\zeta' - \zeta\| - b_1\varepsilon$ for a positive constant b_1.

Let $\zeta_1, \ldots, \zeta_N$ be the points in $\mathcal{N}_{\zeta_0}$ such that for each $\zeta \in \mathcal{N}_{\zeta_0}$, $\|\zeta - \zeta_m\| \le \min\{a_1\varepsilon/a_0, b_1\varepsilon/a_0\}$ for some m, where $1 \le m \le N$ and N is the bracketing number of $\mathcal{N}_{\zeta_0}$ which is of the order of $O(1/\varepsilon^{d+1})$. Therefore, for each (ζ, Λ) in $\mathcal{N}_{\zeta_0} \times [1/M, M]^{\otimes K}$, there exists ζ_m, Λ_{kj}^L and Λ_{kj}^U such that $l_{(j,\zeta)}^L(O) \le l(\theta, O) \le l_{(j,\zeta)}^U(O)$.

Furthermore, by mean value theorem, we have

$$\|l^U_{(j,\zeta)}(O)-l^L_{(j,\zeta)}(O)\|^2_{2,P} \leq \mathrm{E}\left(b_2 \sum_{i=1}^{K} |\Lambda^U_{kj}(C_k) - \Lambda^L_{kj}(C_k)|^2\right) = b_2 m \|\Lambda^U_{kj}(C_k)-\Lambda^L_{kj}(C_k)\|^2_{2,Q} < b_2 m \varepsilon^2,$$

where b_2 is a positive constant and Q denotes the distribution function for the observation time C.

Therefore, by Example 19.11 of van der Vaart (1998), we get that $N_{[\,]}(\varepsilon, \mathcal{L}, L_2(P))$ is of the order of $O(e^{1/\varepsilon}\varepsilon^{-d-1})$ and $\log N_{[\,]}(\varepsilon, \mathcal{L}, L_2(P)) = O(1/\varepsilon)$ as ε decreases to 0. This completes the proof of Lemma 1.

Proof of Theorem 2.

By the Kullback-Leibler inequality, we know that $\mathbb{P}l(\theta, O)$ is maximized at $\theta = \theta_0$. So its first derivative at θ_0 is equal to 0, then by Taylor expansion, we know that for every θ in a neighborhood of θ_0, $\mathbb{P}(l(\theta, O) - l(\theta_0, O)) \leq -C\, d^2(\theta, \theta_0)$, where C is a positive constant. From the Lemma 1, we know that the bracketing integral $J_{[\,]}(\eta, \mathcal{L}, L_2(P))$, defined as $\int_0^\eta \sqrt{\log N_{[\,]}(\varepsilon, \mathcal{L}, L_2(P))}\, \mathrm{d}\varepsilon$, is of the order of $\eta^{1/2}$. Then the lemma 19.36 of van der Vaart (1998) gives

$$E^* \sup_{d(\theta,\theta_0)<\eta} \|\sqrt{n}(\mathbb{P}_n - \mathbb{P})(l(\theta, O) - l(\theta_0, O))\| = O(1)\eta^{1/2}\left(1 + \frac{\eta^{1/2}}{\eta^2\sqrt{n}} M_1\right),$$

where E^* is the outer expectation and M_1 is a positive constant. Let $\phi_n(\eta) = \eta^{1/2}\left(1 + \frac{\eta^{1/2}}{\eta^2\sqrt{n}} M_1\right)$. Then $\phi_n(\eta)/\eta$ is a decreasing function, and

$n^{2/3}\phi_n(n^{-1/3}) = O(\sqrt{n})$ for large n. Furthermore, by Theorem 1, we know that $\hat{\theta}_n$ is consistent. According the theorem 3.4.1 of van der Vaart and Wellner (1996), we can conclude that $d(\hat{\theta}_n, \theta_0) = \left\{\|\hat{\zeta}_n - \zeta_0\|^2 + \sum_{k=1}^{K} \int [\hat{\Lambda}_{kn}(c) - \Lambda_{k0}(c)]^2 f_k(c) dc\right\}^{1/2} = O_p(n^{-1/3})$, which completes the proof of Theorem 2.

Proof of Theorem 3.

The score functions for β and γ are denoted by $S_\beta(\theta)$ and $S_\gamma(\theta)$, respectively, where $S_\beta(\theta) = \frac{\partial l(\beta,\gamma,\Lambda)}{\partial\beta}$ and $S_\gamma(\theta) = \frac{\partial l(\beta,\gamma,\Lambda)}{\partial\gamma}$. For $k = 1, \ldots, K$, we let $h_k(t)$ be a nonnegative and nondecreasing function on $[\tau_1, \tau_2]$. Define $\mathcal{H} = \{h = (h_1(t), \ldots, h_K(t))\}$. Consider parametric submodels $\Lambda_\epsilon(t) = (\Lambda_{1,\epsilon}(t), \ldots, \Lambda_{K,\epsilon}(t))$, where $\Lambda_{k,\epsilon}(t) = \Lambda_k(t) + \epsilon h_k(t)$. For each k, the score function along the kth submodels is given by,

$$S_{\Lambda_k}(\theta)[h_k] = \left.\frac{\partial l(\beta, \gamma, \Lambda_{k,\epsilon})}{\partial \epsilon}\right|_{\epsilon=0}$$

$$= L(\theta, O)^{-1} \int_b \prod_{k=1}^{K} A_k(\beta, \Lambda, O) \sum_{k=1}^{K} B_k(\beta, \Lambda, O) \left\{ \frac{\Delta_k}{1 - \exp\left[-G_k\left(\Lambda_k(C_k)\, e^{X_k^T\beta}\, b\right)\right]} - 1 \right\} p(b|\gamma)\, \mathrm{d}b,$$

where

$$L(\theta, O) = \int_b \prod_{k=1}^{K} \left\{ \Delta_k + (-1)^{\Delta_k} \exp\left[-G_k\left(\Lambda_k(C_k)\, e^{X_k^T\beta}\, b\right)\right] \right\} p(b|\gamma)\, \mathrm{d}b,$$

$$A_k(\beta, \Lambda, O) = \Delta_k + (-1)^{\Delta_k} \exp\left[-G_k\left(\Lambda_k(C_k)\, e^{X_k^T\beta}\, b\right)\right],$$

$$B_k(\beta, \Lambda, O) = G'_k\left(\Lambda_k(C_k)\, e^{X_k^T\beta}\, b\right)\, e^{X_k^T\beta} h_k(C_k) b.$$

The efficient score for ζ at (ζ_0, Λ_0) is $\tilde{l}(\zeta_0, \Lambda_0) = S_\zeta(\zeta_0, \Lambda_0) - \sum_{k=1}^{K} S_{\Lambda_k}(\zeta_0, \Lambda_0)[h_k^*]$, where $S_\zeta(\zeta_0, \Lambda_0) = (S_\beta(\theta_0)^T, S_\gamma(\theta_0))^T$, h_k^* is a $(d+1)$-vector function satisfying

$$\mathbb{P}\left[\left(S_\zeta(\zeta_0, \Lambda_0) - \sum_{k=1}^{K} S_{\Lambda_k}(\zeta_0, \Lambda_0)[h_k^*]\right)^T \left(\sum_{k=1}^{K} S_{\Lambda_k}(\zeta_0, \Lambda_0)[h_k]\right)\right] = 0,$$

for each h_k in $\mathcal{H}$. By following similar calculations in Section 3 of Chang et al. (2007), we can establish the existence of h_k^* in the above equation.

The efficient Fisher information matrix I_0 for ζ at (ζ_0, Λ_0) is defined as $\mathbb{P}(\tilde{l}(\zeta_0, \Lambda_0)\tilde{l}(\zeta_0, \Lambda_0)^T)$. In the following, we will show that I_0 is positive definite. If the I_0 is singular, then there exists a nonzero vector $\nu \in R^{(d+1)}$ such that $\nu^T I_0 \nu = 0$. It follows that, with probability one, the score function along the submodel $\{\zeta_0 + \epsilon\nu, \Lambda_{10} + \epsilon\nu^T h_1^*, \ldots, \Lambda_{K0} + \epsilon\nu^T h_K^*\}$ is zero. Therefore,

$$\nu^T \left(\frac{\partial}{\partial\zeta} + \sum_{k=1}^{K} h_k^* \frac{\partial}{\partial y_k}\right)\Bigg|_{(\zeta, y_1, \ldots, y_K) = (\zeta_0, \Lambda_{10}(c_1), \ldots, \Lambda_{K0}(c_K))}$$
$$\log \int_b \prod_{k=1}^{K} \left\{\Delta_k + (-1)^{\Delta_k} \exp\left[-G_k\left(y_k\, e^{X_k^T\beta}\, b\right)\right]\right\} p(b|\gamma)\, \mathrm{d}b = 0.$$

Using the condition (A5), we know that $\nu = 0$, this is a contradiction. Therefore, we can conclude that $\nu^T I_0 \nu = 0$ implies $\nu = 0$. That is, the efficient Fisher information matrix is positive.

Define

$$S_{\zeta,k}(\theta)[h_k] = \frac{\partial}{\partial \epsilon}\bigg|_{\epsilon=0} S_\zeta(\theta; \Lambda_k = \Lambda_{k\epsilon}),$$

$$S_{k,j}(\theta)[\tilde{h}_k, h_j] = \frac{\partial}{\partial \epsilon}\bigg|_{\epsilon=0} S_{\Lambda_k}(\theta; \Lambda_j = \Lambda_{j\epsilon})[\tilde{h}_k]$$

for $k = 1, \ldots, K$ and $j = 1, \ldots, K$, where $\partial/\partial\epsilon|_{\epsilon=0}\Lambda_{j\epsilon} = h_j$. By Taylor expansion, we can obtain

$$\mathbb{P}\tilde{l}(\zeta_0, \Lambda) = \mathbb{P}\tilde{l}(\zeta_0, \Lambda_0) + \mathbb{P}\left\{\sum_{k=1}^{K} S_{\zeta,k}(\theta)[\Lambda_k - \Lambda_{k0}] - \sum_{k=1}^{K}\sum_{j=1}^{K} S_{k,j}(\theta)[h_k^*, \Lambda_k - \Lambda_{k0}]\right\}$$

$$+ O_p\left(\sum_{k=1}^{K} \|\Lambda_k - \Lambda_{k0}\|^2\right).$$

Note that $\mathbb{P}\tilde{l}(\zeta_0, \Lambda_0) = 0$, $\mathbb{P}(S_\zeta(\theta) S_{\Lambda_k}(\theta)[h_k]) = -\mathbb{P}(S_{\zeta,k}(\theta)[h_k])$, $\mathbb{P}(S_{\Lambda_k}(\theta)[\tilde{h}_k] S_{\Lambda_j}(\theta)[h_j]) = -\mathbb{P}(S_{k,j}(\zeta)[\tilde{h}_k, h_j])$, by the consistency and the convergence rate of $\hat{\Lambda}_n$, we can conclude that $\mathbb{P}\tilde{l}(\zeta_0, \hat{\Lambda}_n) = O_p(n^{-2/3})$, which implies $\sqrt{n}\,\mathbb{P}\tilde{l}(\zeta_0, \hat{\Lambda}_n) = o_p(1)$. We know from Example 19.11 of van der Vaart (1998) that the class of uniformly bounded functions with bounded variations is a Donsker class. By using Theorem 2.10.6 of van der Vaart and Wellner (1996), we can verify that $\tilde{l}(\zeta, \Lambda)$ is a uniformly bounded Donsker class. In addition, we have proved that $\hat{\theta}_n$ is consistency. Therefore, $\sqrt{n}(\mathbb{P}_n - \mathbb{P})(\tilde{l}(\hat{\zeta}_n, \hat{\Lambda}_n) - \tilde{l}(\zeta_0, \Lambda_0)) = o_p(1)$. Due to the fact that

$\mathbb{P}_n\tilde{l}(\hat{\theta}_n) = \mathbb{P}\tilde{l}(\theta_0) = 0$ and $\mathbb{P}\tilde{l}(\zeta_0, \hat{\Lambda}) = o_p(1)$, we can have

$$-\sqrt{n}\mathbb{P}(\tilde{l}(\hat{\theta}_n) - \tilde{l}(\zeta_0, \hat{\Lambda}_n)) = \sqrt{n}\mathbb{P}_n\tilde{l}(\theta_0) + o_p(1).$$

By the mean value theorem, we have

$$-\sqrt{n}\mathbb{P}\frac{\partial}{\partial\zeta}\tilde{l}(\zeta^{'}, \hat{\Lambda}_n)(\hat{\zeta}_n - \zeta_0) = \sqrt{n}\mathbb{P}_n\tilde{l}(\theta_0) + o_p(1),$$

where $\zeta^{'}$ is a point between $\hat{\zeta}_n$ and ζ_0. Since $\hat{\theta}_n$ is consistency and $\mathbb{P}(-\frac{\partial}{\partial\zeta}\tilde{l}(\theta_0)) = \mathbb{P}(\tilde{l}(\theta_0)\tilde{l}(\theta_0)^T) = I_0$, we can conclude that

$$\sqrt{n}(\hat{\zeta}_n - \zeta_0) = I_0^{-1}\sqrt{n}\mathbb{P}_n\tilde{l}(\theta_0) + o_p(1) \xrightarrow{d} N(0, I_0^{-1}).$$

This completes the proof of Theorem 3.