Biometrics 000, 000–000 DOI: 000
000 0000

Web-based Supplementary Materials for "Semiparametric Estimation of the Accelerated Mean Model with Panel Count Data under Informative Examination Times" by

Sy Han Chiou[1,*], Gongjun Xu[2], Jun Yan[3], and Chiung-Yu Huang[4]

[1] Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75080, U.S.A.

[2] Department of Statistics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

[3] Department of Statistics, University of Connecticut, Storrs, Connecticut 06269, U.S.A.

[4] Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California 94158, U.S.A.

**email:* schiou@utdallas.edu

Summary: The Supplementary Material contains additional simulation results from the main article as well as the proof of consistency result.

This paper has been submitted for consideration for publication in *Biometrics*

1. Smoothing method used in simulation and data analysis

Throughout the paper, we use $\widetilde{\Lambda}_n(\boldsymbol{a}, \cdot)$ to denote a smoothed version of $\widehat{\Lambda}_n(\boldsymbol{a}, \cdot)$. In general, $\widetilde{\Lambda}_n(\boldsymbol{a}, \cdot)$ could be obtained from conventional smoothing techniques such as the smoothing splines or kernel regression. Within each smoothing approach, different tuning parameters yield a large number of possibilities. In the main article, $\widehat{\Lambda}_n(\boldsymbol{a}, \cdot)$ is estimated at the ordered, distinct values of the observed examination times, $t = t_{(0)}, t_{(1)}, \ldots, t_{(L)}$. We computed the smoothed value of $\widehat{\Lambda}_n(\boldsymbol{a}, \cdot)$, at the $t_{(i)}$'s using the Nadaraya–Watson kernel regression (Nadaraya, 1964; Watson, 1964), which has the form:

$$\widetilde{\Lambda}_n(\boldsymbol{a}, t) = \frac{\sum_{i=1}^{L} K_h(t, t_{(i)})\widehat{\Lambda}_n(\boldsymbol{a}, t)}{\sum_{i=1}^{L} K_h(t, t_{(i)})},$$

where $K_h(s, t) = \exp\{-(s-t)^2/2h^2\}$ is the Gaussian kernel and h is the bandwidth parameter. The Nadaraya–Watson kernel regression is readily available via `ksmooth` function in `R` (R Core Team, 2017). We specified the bandwidth parameter via an unbiased cross-validation (Bowman, 1984), which is a cross-validation method minimizing the integrated squared error defined by

$$\int_0^\infty \{\widetilde{\Lambda}_n(\boldsymbol{a}, t) - \widehat{\Lambda}_n(\boldsymbol{a}, t)\}\, dt.$$

The unbiased cross-validation is available via `ucv` function of `R` package `MASS` (Venables and Ripley, 2002).

2. Additional Simulation Specifications

2.1 *Timing results*

The proposed estimation procedure requires iteration between estimating the cumulative baseline rate function and estimating the regression parameters. We compared the computing time for the proposed estimation procedure with and without the SQUAREM acceleration in estimating the cumulative baseline rate function. In the former case, the standard expectation-maximization (EM) algorithm was carried out in the estimation of the cumulative baseline rate function. In all scenarios, we used ℓ-2 norm convergence criteria with a prefixed tolerance of 0.001 in estimation.

Table 1 displays the computing time (in seconds) required to obtain the estimate of regression parameters using a Linux machine with 8 cores Intel i7-6700 CPU at 3.40 GHz and 16GB memory. The point estimates from both procedures are very close (point estimates using EM are not reported), but the procedure with the SQUAREM is much faster in all scenarios considered. In particular, the SQUAREM procedure yields a computing time 5.3 times faster than the EM procedure under Poisson scenario with $n = 100$, $Z \sim \text{Gamma}(2, 2)$ and $\lambda_0(t) = 2$. As the sample size doubles, the computing times do not double linearly for both procedures. However, of the two procedures, the EM procedure suffers more from sample size increases. Thus, we expect the SQUAREM procedure to be even more beneficial with larger sample sizes. For these reasons, we used the SQUAREM procedure for the rest of the simulation study.

[Table 1 about here.]

2.2 *Association between recurrent event and examination time processes*

To have a better understanding of the effect of strength and direction of the association between the underlying recurrent event and examination time processes, we carried out additional simulation studies with different specifications. Since the primary objective is to investigate the robustness of the proposed method against different frailty distributions, we only report results with $n = 100$.

To investigate the impact of the association strength, we generated Z_i from whether a $\text{Gamma}(0.5, 0.5)$ or a $\text{Normal}(1, 0.2^2)$ while holding other variable specifications the same. These settings yield similar observed recurrent events per subject but the latter scenario yield a higher examination frequency. The association between the underlying recurrent event and examination time processes remains positive under these settings. Each of these frailty distributions has mean 1 as required by the identifiability assumption but the variances are different allowing comparisons across scenarios. The results are presented at Table 2. In all scenarios, the proposed estimate continues to be virtually unbiased. Both bootstrap estimates are reasonably close to the empirical standard error. The magnitude of standard error increases with the variance of Z_i; the standard error is the smallest when $Z_i \sim$Normal$(0, 0.2^2)$ and the largest when $Z_i \sim$ Gamma$(0.5, 0.5)$. Most importantly, the coverage probability remains satisfactory, with the proposed smoothed bootstrap estimate closer to the 95% nominal level. These results suggest that the strength of the association between the recurrent event and the examination times influence the variability of the proposed estimate but does not influence the consistency.

We next investigate the impact of the direction of association. In particular, we reverse the generation of K_i to generate the simulated data, so the recurrent event process and the examination time process are negatively associated. More specifically, holding all specifications the same, we generated K_i from a discrete uniform distribution on $\{1, \ldots, 6\}$ when $Z_i > 1$ and a discrete uniform distribution on $\{1, \ldots, 8\}$ when $Z_i \leqslant 1$. With this modification, subjects with $Z_i \leqslant 1$ have higher event rate and tend to be examined more frequently than subjects with $Z_i > 1$. We considered all four frailty distributions aforementioned; Gamma$(2, 2)$, Uniform$(0, 2)$, Gamma$(0.5, 0.5)$, and Normal$(1, 0.2^2)$. The results are summarized in Table 3. As in the case of the positive association, the proposed methods perform reasonably well with small bias, close agreement between the bootstrap estimates and justifiable coverage probability. These observations suggest that the proposed estimator is fairly robust against the direction of association between the underlying recurrent event and examination time processes.

[Table 2 about here.]

[Table 3 about here.]

3. Proof of Consistency Result for $\widehat{\Lambda}_n(\boldsymbol{a}, \cdot)$ of $\Lambda(\boldsymbol{a}, \cdot)$

To establish the consistency results, we first introduce a proper metric on the class of functions defined by $\mathcal{F}_\tau = \{\Lambda : [0, \tau_\alpha] \to [0, \infty); \Lambda \text{ is nondecreasing and } \Lambda(0) = 0\}$. Consider a subject with observed data $\{t_j, K, N_i(t_j), X; j = 1, \ldots, K\}$, $m_j = N_i(t_j) - N_i(t_{j-1})$ and $Y = N_i(t_K)$, for $\Lambda_1, \Lambda_2 \in \mathcal{F}_\tau$, we define $d(\Lambda_1, \Lambda_2) = \int \mid \Lambda_1(t) - \Lambda_2(t) \mid^2 \mathrm{d}v(t)$, where v is a measure defined by $v(B) = E[E\{\sum_{j=1}^K I(t_j \in B) \mid K\}]$ for $B \in \mathcal{B}_\tau$ with $\mathcal{B}_\tau$ being the Borel sets in $[0, \tau]$. We write $d(\Lambda_1, \Lambda_2) = E[E\{\sum_{j=1}^K \mid \Lambda_1(t_j) - \Lambda_2(t_j) \mid^2 \mid K\}]$ and assume the following regularity conditions.

C1 There exists an integer $k_0 < \infty$ such that $\mathrm{pr}(K \leqslant k_0) = 1$ and $\mathrm{pr}(K > 1) > 0$.

C2 The distribution of X has bounded support and the baseline cumulative rate function $\Lambda_0(\cdot)$ is bounded and positive on $[0, C]$ for any $C > 0$.

C3 The random variable $M_0 = \sum_{j=1}^K m_j \log m_j$ has bounded expectation.

C4 Variable Y has positive continuous density (positive probability mass) at τ.

The consistency of the estimator $\widehat{\Lambda}_n(\boldsymbol{a}, \cdot)$ of $\Lambda(\boldsymbol{a}, \cdot)$ follows a similar argument as the proofs in Wellner and Zhang (2000, Theorem 4.2) and in Huang et al. (2006, Theorem 1). We first consider the nonparametric distribution estimator $\widehat{\Phi}_n(\boldsymbol{a}, \cdot)$ for any $\boldsymbol{a}$ in a neighborhood of the true parameter $\boldsymbol{\alpha}$. Let $D = \{t_1, \ldots, t_K, K, Y; m_1, \ldots, m_K, m\}$ be a subject's observation vector and the working log-likelihood function $q(F, \boldsymbol{a}, D) = \sum_{j=1}^K m_j \log[F\{t_j^*(\boldsymbol{a})\} - F\{t_{j-1}^*(\boldsymbol{a})\}] - m \log F\{Y^*(\boldsymbol{a})\}$. Further define

$$\mathbb{P}_n(F, \boldsymbol{a}) = n^{-1} \sum_{i=1}^n q(F, \boldsymbol{a}, D_i) \text{ and } \mathrm{pr}(F, \boldsymbol{a}) = E\{q(F, \boldsymbol{a}, D)\}.$$

For a fixed $\boldsymbol{a}$, let $\Phi(\boldsymbol{a}, \cdot)$ be the maximizer (with respect to F) of $\mathrm{pr}(F, \boldsymbol{a})$ with the form

$$E\left(\sum_{j=1}^K \left[\Phi\{t_j^*(\boldsymbol{\alpha})\} - \Phi\{t_{j-1}^*(\boldsymbol{\alpha})\}\right] \log[F\{t_j^*(\boldsymbol{a})\} - F\{t_{j-1}^*(\boldsymbol{a})\}] - \Phi\{Y^*(\boldsymbol{\alpha})\} \log F\{Y^*(\boldsymbol{a})\}\right).$$

Note that estimated distribution function $\widehat{\Phi}_n(\boldsymbol{a}, \cdot)$ is a step function. Since for any k and positive vectors of $(x_1, \ldots, x_k)$ and $(a_1, \ldots, a_k)$, the function $g(x) = \sum_{j=1}^k a_j \log(x_j) -$

$(\sum_{j=1}^{k} a_j)\log(\sum_{j=1}^{k} x_j)$ has the maximum $\sum_{j=1}^{k} a_j \log(a_j) - (\sum_{j=1}^{k} a_j)\log(\sum_{j=1}^{k} a_j)$ when $x_j = ca_j$ for any j and some positive constant c. Therefore, we have an upper envelope for the set of functions $\mathcal{Q} = [q\{\Phi(\boldsymbol{a},\cdot), D\}; \Phi(\boldsymbol{a},\cdot) \in \mathcal{F}_{\tau_{\boldsymbol{a}}}]$ as $M_0 = \sum_j m_j \log m_j$. It then follows from the one-sided Glivenko–Cantelli Theorem that $\limsup_{n\to\infty} \sup_{F\in\mathcal{F}_{\tau_{\boldsymbol{a}}}} (\mathbb{P}-\mathrm{pr})F \leqslant 0$ almost surely. The Helly's selection theorem gives that for any sequence of $\widehat{\Phi}_n(\boldsymbol{a},\cdot)$, there exists a subsequence (indexed by n') converging to a limit function $\Phi^*(\boldsymbol{a},\cdot)$. Thus, $\limsup_{n'\to\infty} \mathbb{P}\{\widehat{\Phi}_{n'}(\boldsymbol{a},\cdot), \boldsymbol{a}\} \leqslant \mathrm{pr}\{\Phi^*(\boldsymbol{a},\cdot), \boldsymbol{a}\}$. Note that $\widehat{\Phi}_n(\boldsymbol{a},\cdot)$ is the maximizer of $\mathbb{P}(F, \boldsymbol{a})$, which implies that $\mathbb{P}\{\widehat{\Phi}_n(\boldsymbol{a},\cdot), \boldsymbol{a}\} \geqslant \mathbb{P}\{\Phi(\boldsymbol{a},\cdot), \boldsymbol{a}\}$. The law of large number further implies that $\liminf_{n\to\infty} \mathbb{P}\{\widehat{\Phi}_n(\boldsymbol{a},\cdot), \boldsymbol{a}\} \geqslant \mathrm{pr}\{\Phi(\boldsymbol{a},\cdot), \boldsymbol{a}\}$. The above argument then gives

$$
\begin{aligned}
0 &\geqslant \mathrm{pr}\{\Phi(\boldsymbol{a},\cdot), \boldsymbol{a}\} - \mathrm{pr}\{\Phi^*(\boldsymbol{a},\cdot), \boldsymbol{a}\} \\
&= E\left[\sum_j m_j \log \frac{\Phi\{\boldsymbol{a}, t_j^*(\boldsymbol{a})\} - \Phi\{\boldsymbol{a}, t_{j-1}^*(\boldsymbol{a})\}}{\Phi^*\{\boldsymbol{a}, t_j^*(\boldsymbol{a})\} - \Phi^*\{\boldsymbol{a}, t_{j-1}^*(\boldsymbol{a})\}} - m \log \frac{\Phi\{\boldsymbol{a}, Y^*(\boldsymbol{a})\}}{\Phi^*\{\boldsymbol{a}, Y^*(\boldsymbol{a})\}}\right] \geqslant 0.
\end{aligned}
$$

Therefore, we know that for some constant b, $d\{\widehat{\Phi}_n(\boldsymbol{a},\cdot), b\Phi(\boldsymbol{a},\cdot)\} \to 0$ almost surely and uniformly in $\boldsymbol{a}$; and furthermore $d\{\widehat{\Lambda}_n(\boldsymbol{a},\cdot), \Lambda(\boldsymbol{a},\cdot)\} \to 0$. The consistency of $\widehat{\boldsymbol{\alpha}}_n$ is obtained by solving the estimating function (5) of the main manuscript due to the fact that the estimating function S_n goes to 0 almost surely at $\boldsymbol{\alpha}$ while not 0 when $\boldsymbol{a} \neq \boldsymbol{\alpha}$. This further implies that $\widehat{\Lambda}_n(\widehat{\boldsymbol{\alpha}},\cdot)$ is consistent.

References

Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika* **71,** 353–360.

Huang, C.-Y., Wang, M.-C., and Zhang, Y. (2006). Analysing panel count data with informative observation times. *Biometrika* **93,** 763–775.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications* **9,** 141–142.

R Core Team (2017). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S.* Fourth edition. Springer, New York.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A* pages 359–372.

Wellner, J. A. and Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *The Annals of Statistics* **28,** 779–814.

Table 1

Summary of timing required for point estimation. Timing is recorded in seconds and averaged from 100 replicates.

		$n = 50$				$n = 100$			
		$Z \sim \text{Gamma}(2,2)$		$Z \sim \text{Uniform}(0,2)$		$Z \sim \text{Gamma}(2,2)$		$Z \sim \text{Uniform}(0,2)$	
	$\lambda_0(t)$	EM	SQUAREM	EM	SQUAREM	EM	SQUAREM	EM	SQUAREM
Poisson	2	98	20	117	23	523	96	670	130
	$2t$	53	17	64	22	457	135	644	184
non-	2	131	30	73	14	489	95	816	155
Poisson	$2t$	52	23	53	28	393	134	550	117

Table 2

Summary of the additional simulation data with positive association between the recurrent event process and the examination time process; ESE is the empirical standard error; ASE and ASE are the average standard error based on the standard bootstrap and the smoothed bootstrap procedure, respectively; CP and CP* are the empirical coverage probability (%) based on the standard bootstrap and the smoothed bootstrap procedure, respectively. Cases I–IV reflects the four combinations between the two choices of $\lambda_0(t)$ and whether the recurrent event process is a Poisson counting process; Case I: $\lambda_0(t) = 2$, Poisson process; Case II: $\lambda_0(t) = 2t$, Poisson process; Case III: $\lambda_0(t) = 2$, non-Poisson process; Case IV: $\lambda_0(t) = 2t$, non-Poisson process.*

		$Z \sim$ Gamma$(0.5, 0.5)$						$Z \sim$ Normal$(1, 0.2^2)$					
case	α	bias	ESE	ASE	ASE*	CP	CP*	bias	ESE	ASE	ASE*	CP	CP*
I	α_1	0.014	0.333	0.341	0.348	96.3	96.8	0.007	0.168	0.166	0.170	94.6	95.5
	α_2	0.057	0.595	0.567	0.614	94.7	95.9	0.065	0.299	0.251	0.306	91.3	95.3
II	α_1	−0.059	0.213	0.218	0.214	95.3	95.3	−0.048	0.131	0.128	0.134	93.8	95.6
	α_2	−0.064	0.372	0.348	0.372	92.9	95.0	−0.064	0.211	0.194	0.228	92.1	95.6
III	α_1	0.007	0.329	0.322	0.345	94.9	95.5	−0.002	0.146	0.139	0.148	95.3	95.1
	α_2	0.072	0.581	0.530	0.591	93.6	95.4	0.030	0.279	0.224	0.286	90.6	95.6
IV	α_1	−0.041	0.221	0.219	0.232	94.7	95.8	−0.060	0.119	0.111	0.127	92.1	96.1
	α_2	−0.056	0.382	0.355	0.410	92.1	95.6	−0.086	0.192	0.184	0.202	92.1	95.6

Table 3

Summary of simulation data with negative association between the recurrent event and examination time process; ESE is the empirical standard error; ASE and ASE are the average standard error based on the standard bootstrap and the smoothed bootstrap procedure, respectively; CP and CP* are the empirical coverage probability (%) based on the standard bootstrap and the smoothed bootstrap procedure, respectively. Cases I–IV reflects the four combinations between the two choices of $\lambda_0(t)$ and whether the recurrent event process is a Poisson counting process; Case I: $\lambda_0(t) = 2$, Poisson process; Case II: $\lambda_0(t) = 2t$, Poisson process; Case III: $\lambda_0(t) = 2$, non-Poisson process; Case IV: $\lambda_0(t) = 2t$, non-Poisson process.*

case	α	bias	ESE	ASE	ASE*	CP	CP*	bias	ESE	ASE	ASE*	CP	CP*
		$Z \sim$ Gamma$(2, 2)$						$Z \sim$ Uniform$(0, 2)$					
I	α_1	−0.012	0.210	0.205	0.216	95.6	95.7	−0.004	0.185	0.184	0.192	95.7	96.2
	α_2	0.002	0.355	0.348	0.361	94.4	95.4	−0.002	0.316	0.314	0.320	95.3	95.4
II	α_1	−0.031	0.136	0.136	0.145	95.8	96.3	−0.032	0.124	0.121	0.127	94.8	95.8
	α_2	−0.048	0.232	0.226	0.245	94.6	96.0	−0.054	0.216	0.204	0.221	95.2	96.0
III	α_1	−0.013	0.198	0.189	0.206	94.6	95.9	−0.024	0.159	0.164	0.166	95.6	95.6
	α_2	−0.029	0.359	0.329	0.373	93.2	96.0	−0.038	0.307	0.288	0.312	94.5	96.1
IV	α_1	−0.041	0.132	0.129	0.141	94.5	96.1	−0.044	0.117	0.114	0.125	94.3	95.8
	α_2	−0.061	0.229	0.221	0.236	94.9	96.3	−0.054	0.194	0.195	0.209	94.7	96.4
		$Z \sim$ Gamma$(0.5, 0.5)$						$Z \sim$ Normal$(1, 0.2^2)$					
I	α_1	−0.004	0.323	0.335	0.333	95.5	95.7	−0.006	0.157	0.162	0.164	95.6	95.6
	α_2	0.046	0.592	0.543	0.615	94.1	96.0	0.043	0.307	0.286	0.323	93.2	96.0
II	α_1	−0.014	0.186	0.202	0.209	95.5	95.7	−0.046	0.114	0.123	0.128	95.4	95.3
	α_2	−0.035	0.333	0.337	0.335	95.5	95.2	−0.058	0.216	0.181	0.220	92.9	95.6
III	α_1	0.004	0.318	0.324	0.328	95.7	95.1	−0.017	0.131	0.143	0.141	95.8	95.4
	α_2	0.045	0.564	0.535	0.571	95.1	95.8	0.042	0.270	0.253	0.280	92.6	96.4
IV	α_1	−0.032	0.186	0.199	0.205	96.0	95.9	−0.037	0.116	0.111	0.121	92.8	95.6
	α_2	−0.059	0.326	0.334	0.332	95.4	95.1	−0.056	0.184	0.173	0.191	93.8	96.0