Statistica Sinica Preprint No: SS-2021-0353	
Title	Semiparametric Reversed Mean Model for Recurrent Event Process with Informative Terminal Event
Manuscript ID	SS-2021-0353
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202021.0353
Complete List of Authors	Wen Su, Li Liu, Guosheng Yin, Xingqiu Zhao and Ying Zhang
Corresponding Authors	Wen Su
E-mails	jenna.wen.su@connect.hku.hk

doi:10.5705/ss.202021.0353

Statistica Sinica

Semiparametric Reversed Mean Model for Recurrent Event Process with Informative Terminal Event

Wen Su[1*], Li Liu[2*], Guosheng Yin[1], Xingqiu Zhao[3] and Ying Zhang[4]

[1]*Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong*

[2]*School of Mathematics and Statistics, Wuhan University, Wuhan, China*

[3]*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong* [4]*Department of Biostatistics, University of Nebraska Medical Center, Omaha, NE, USA*

Abstract: We study semiparametric regression for a recurrent event process with an informative terminal event, where observations are taken only at discrete time points, rather than continuously over time. To account for the effect of a terminal event on the recurrent event process, we propose a semiparametric reversed mean model, for which we develop a two-stage sieve likelihood-based method to estimate the baseline mean function and the covariate effects. Our approach overcomes the computational difficulties arising from the nuisance functional parameter in the assumption that the likelihood is based on a Poisson process. We establish the consistency, convergence rate, and asymptotic normality of the

*Co-first author

proposed two-stage estimator, which is robust against the assumption of an underlying Poisson process. The proposed method is evaluated using extensive simulation studies, and demonstrated using panel count data from a longitudinal healthy longevity study and data from a bladder tumor study.

Key words and phrases: Counting process; Expected log-likelihood; Reversed mean model; Semiparametric M-estimator; Terminal event.

1. Introduction

Panel count data often arise in biomedical research, economics, social sciences, and reliability studies (Thall, 1988; Sun and Wei, 2000; Hu, Sun, and Wei, 2003; Wellner and Zhang, 2007; Lu, Zhang, and Huang, 2009; Zhao, Li, and Sun, 2013). During the follow-up, observations are taken at finite distinct time points, with researchers collecting the number of recurrent events that occurred between observation times, without information on the exact timing of the events. While both the observation and the follow-up times may vary between subjects, observations may be terminated by a terminal event. Examples of studies based on panel count data include the bladder cancer study conducted by the Veterans Administration Cooperative Urological Research Group (Andrews and Herzberg, 1985) and the Chinese Longitudinal Healthy Longevity Study (CLHLS) (Zeng et al., 2017). We aim to estimate the mean function of the underlying counting process and make inferences about the factors that affect the event occurrence rate in the setting of panel count data, subject to an informative terminal event. The unique characteristics of panel count data present additional challenges to statistical inference, requiring advanced modeling techniques.

Many methods have been developed for estimating the mean function of a counting process with panel count data but without considering the effect of a terminal event. These include the parametric methods by Kalbfleisch and Lawless (1985), Hinde (1982), Breslow (1984), and Thall (1988); the nonparametric methods of Sun and Kalbfleisch (1995), Wellner and Zhang (2000), Hu, Lagakos, and Lockhart (2009), Zhang and Jamshidian (2003), Huang, Wang, and Zhang (2006), and Lu, Zhang, and Huang (2007); and the semiparametric methods of Cheng and Wei (2000), Sun and Wei (2000), Zhang (2002), Hu, Sun, and Wei (2003), Wellner and Zhang (2007), and Lu, Zhang, and Huang (2009). More recently, Ma and Sundaram (2018) investigated gap times with panel count data, Zhu et al. (2018) developed a semiparametric likelihood-based method for a regression analysis of mixed panel count data, and Diao, Zeng, Hu, and Ibrahim (2019) studied semiparametric frailty models when dropouts are informative. Chiou, Huang, Xu, and Yan (2019) provide a comprehensive overview of existing methods for panel count data analysis and the corresponding software implementa-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0353

tions.

Despite the abundant research on modeling panel count data, few works include an informative terminal event. Two common methods used for counting process data to account for the effect of terminal events are joint modeling approaches that use frailty variables (Huang and Wang, 2004; Zeng and Cai, 2010; Sun, Song, Zhou and Liu, 2012; Zhou, Zhang, Sun and Sun, 2017; Diao, Zeng, Hu, and Ibrahim, 2017, 2019) and marginal modeling approaches that use the inverse probability weighting technique (Zhao, Zhou and Sun, 2011; Zhao, Li and Sun, 2013). However, both methods have limitations. In particular, the joint modeling approach cannot explicitly depict the relationship between a recurrent event process and the terminal event, although it is implied by the shared unknown latent variables, and the marginal modeling approach may not be feasible when a terminal event permanently stops the recurrent event process. On the other hand, the terminal event time usually induces direct effects on the recurrent event through a functional relationship. For example, Chan et al. (1995) found that the occurrence of AIDS-defending events increases prior to the death of HIV-infected individuals. Lunney et al. (2003) discovered that the functional decline changes significantly in the final year before the death of HIV-infected individuals. To the best of our knowledge, there are

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0353

no semiparametric methods that directly capture the correlation between the terminal event and the recurrent event process in the case of panel count data.

This study contributes to the literature in three ways. First, we develop a semiparametric reversed mean model anchoring at a terminal event to explicitly quantify its effect on the recurrent event process, while emphasizing the process near the anchoring event (Chan and Wang, 2010, 2017; Kong et al., 2018). Second, conditional on censored terminal event time data, we formulate the log-likelihood for panel count data as an objective function, and design a corresponding two-stage estimation procedure. Third, we establish a general framework for the asymptotic distributional theory of semiparametric M-estimators with a nuisance functional parameter that can be applied to derive the asymptotic normality of functionals of two-stage semiparametric M-estimators for panel count data. In particular, the asymptotic results of the proposed estimator do not rely on the working model following a Poisson process.

The remainder of this paper is organized as follows. In Section 2, we introduce the proposed semiparametric reversed mean model, and present the corresponding two-stage estimation procedure. In Section 3, we establish the asymptotic properties, including the consistency, convergence rate, and asymptotic normality of the proposed estimator. Section 4 reports the results of simulation studies to demonstrate the performance of the proposed estimator. In Section 5, we apply the proposed method to two real datasets. Section 6 concludes the paper. All technical proofs and a general theorem for semiparametric M-estimators are provided in the online Supplementary Material.

2. Methodology

2.1 Model Setting

We use $N(t)$ to denote the total number of occurrences of the event of interest up to time, t for $0 \le t \le \tau_0$, where τ_0 is the length of the study duration. Let U be the terminal event time with the counting process $N(\cdot)$, and let C be the censoring time for $(U, N(\cdot))$, after which the counting process or terminal event may still happen, but is unobserable. Let $Y = U \wedge C$ be the observed time for the terminal event, where $a \wedge b = \min(a, b)$, and let $\Delta = 1_{\{U \le C\}}$ be a censoring indicator. Suppose that $N(\cdot)$ is observed at discrete time points $0 < T_{K,1} < \cdots < T_{K,K}$, where K is the potential number of observation times. Let $\underline{T}_K = (T_{K,1}, \ldots, T_{K,K})$ denote panel observation times on the counting process, and let $\underline{N} = (N(T_{K,1}), \ldots, N(T_{K,K}))$ be the cumulative event counts observed at $\underline{T}_K$. Let Z be a d-dimensional vector of associated covariates at the baseline. The observed data consist of independent and identically distributed (i.i.d.) copies of $D = (Y, \Delta, \underline{N}, \underline{T}_K, K, Z)$.

Motivated by the works of Chan and Wang (2010, 2017) and Kong et al. (2018), to characterize the behavior of the counting process near the informative terminal event time U, we investigate the reversed counting process $\widetilde{N}(t; U)$, which is the event count from time t to the terminal event time U. Suppose that $\widetilde{N}(t; U)$ follows a semiparametric reversed mean model,

$$E(\widetilde{N}(t; U)|U = u, Z) = \Lambda(u - t)e^{\beta^\top Z}, \tag{2.1}$$

where $\Lambda(\cdot)$ is a nondecreasing function with $\Lambda(0) = 0$, and β is a d-dimensional vector of unknown regression coefficients. This reversed mean model indicates that the reversed counting process is associated with the random terminal time U only through the length to the terminal event $U - t$, and hence can be viewed as a homogeneous temporal model.

Because $\widetilde{N}(t; U)$ may not be observed at some t, owing to U being censored, model (2.1) is not immediately useful for estimating $\Lambda(\cdot)$ and β based on the observed data. Noting that $N(t) = \widetilde{N}(0; U) - \widetilde{N}(t; U)$, we have

$$E(N(t)|U = u, Z) = \{\Lambda(u) - \Lambda(u - t)\}e^{\beta^\top Z}, \tag{2.2}$$

which we can use to study the reversed mean function $\Lambda(\cdot)$ and the regression coefficient β using a likelihood method similar to those in Wellner and Zhang (2007) and Lu, Zhang, and Huang (2009).

2.2 Estimation Procedure

We assume that U and C are independent given Z, and both the distribution of the censoring time C and that of $(K, \underline{T}_K)$ are non-informative to Λ. Let $\triangle N_j = N(T_{K,j}) - N(T_{K,j-1})$, $\underline{\triangle N} = (\triangle N_1, \ldots, \triangle N_K)$, $\underline{\triangle n} = (\triangle n_1, \ldots, \triangle n_k)$, and $\underline{t}_k = (t_{k,1}, \ldots, t_{k,k})$, with $\triangle n_j = n_j - n_{j-1}$. Suppose that the counting process $N(\cdot)$ follows a conditional Poisson process, with

$$P(\underline{\triangle N} = \underline{\triangle n} | Z, U = u, K = k, \underline{T}_K = \underline{t}_k) = \prod_{j=1}^{k} \frac{\exp(-\triangle\Lambda_j(u) e^{\beta^\top Z})(\triangle\Lambda_j(u) e^{\beta^\top Z})^{\triangle n_j}}{(\triangle n_j)!}, \tag{2.3}$$

where $t_{k,0} \equiv 0$, $n_0 \equiv 0$, and $\triangle\Lambda_j(u) = \Lambda(u - t_{k,j-1}) - \Lambda(u - t_{k,j})$, for $j \geq 1$. We write $\theta = (\beta, \Lambda)$, and use the notation $\theta_0 = (\beta_0, \Lambda_0)$ and $\triangle\Lambda_{0j}(u)$ to denote the true values of θ and $\triangle\Lambda_j(u)$, respectively. The individual log-likelihood of θ based on the working model (2.3) is given by

$$l(\theta; U, \underline{\triangle N}, \underline{T}_K, K, Z) = \sum_{j=1}^{K} \left\{ \triangle N_j \log(\triangle\Lambda_j(U) e^{\beta^\top Z}) - \triangle\Lambda_j(U) e^{\beta^\top Z} \right\}.$$

However, we may not be able to evaluate this likelihood, because U is subject to censoring. To deal with this case, we consider the conditional expectation of the log-likelihood:

$E\{l(\theta; U, \underline{\triangle N}, \underline{T}_K, K, Z)|Y, \Delta, Z\}$. Let $F_0(u|z)$ be the conditional cumulative distribution function of U given the covariate $Z = z$. A direct calculation yields

$$
\begin{aligned}
& E\{l(\theta; U, \underline{\triangle N}, \underline{T}_K, K, Z)|Y, \Delta, Z\} \\
= \; & \Delta \sum_{j=1}^{K} \left\{ \triangle N_j \log(\triangle\Lambda_j(Y) e^{\beta^\top Z}) - \triangle\Lambda_j(Y) e^{\beta^\top Z} \right\} \\
& + \frac{1-\Delta}{\overline{F}_0(Y|Z)} \sum_{j=1}^{K} \int_Y^\tau \left\{ \triangle N_j \log(\triangle\Lambda_j(u) e^{\beta^\top Z}) - \triangle\Lambda_j(u) e^{\beta^\top Z} \right\} dF_0(u|Z),
\end{aligned}
$$

where $\overline{F}_0(u|Z) = 1 - F_0(u|Z)$, and τ is a finite time. Thus, the conditional expectation of log-likelihood (CELL) for i.i.d. samples $\underline{D} = \{D_i : i = 1, \ldots, n\}$ is

$$
\begin{aligned}
l_n(\theta, F; \underline{D}) = & \frac{1}{n} \sum_{i=1}^{n} \left[\Delta_i \sum_{j=1}^{K_i} \left\{ \triangle N_{i,j} \log(\triangle\Lambda_{i,j}(Y_i) e^{\beta^\top Z_i}) - \triangle\Lambda_{i,j}(Y_i) e^{\beta^\top Z_i} \right\} \right. \\
& \left. + \frac{1-\Delta_i}{\overline{F}(Y_i|Z_i)} \sum_{j=1}^{K_i} \int_{Y_i}^{\tau} \left\{ \triangle N_{i,j} \log(\triangle\Lambda_{i,j}(u) e^{\beta^\top Z_i}) - \triangle\Lambda_{i,j}(u) e^{\beta^\top Z_i} \right\} dF(u|Z_i) \right],
\end{aligned} \tag{2.4}
$$

after omitting the parts unrelated to (θ, F), where $T_{K_i,0} \equiv 0$, $\triangle N_{i,j} = N_i(T_{K_i,j}) - N_i(T_{K_i,j-1})$, and $\triangle\Lambda_{i,j}(u) = \Lambda(u - T_{K_i,j-1}) - \Lambda(u - T_{K_i,j})$. To obtain the estimators of θ_0 and F_0 based on (2.4), we consider a two-stage procedure, similar in spirit to a pseudo-likelihood estimation.

Stage 1: Obtain the estimator of F_0, $\hat{F}_n(u|Z)$, based on the right-censored data $\{(Y_i, \Delta_i, Z)$, for $i = 1, \ldots, n\}$.

Stage 2: Obtain the CELL estimator $\hat{\theta}_n = (\hat{\beta}_n, \hat{\Lambda}_n)$ by maximizing $l_n(\theta; \underline{D}) := l_n(\theta, \hat{F}_n; \underline{D})$.

In Stage 1, to estimate $F_0(t|Z)$, we can assume a survival model such as the Cox proportional hazards model (Cox, 1972). Suppose the hazard function of U given the covariate Z satisfies

$$\lambda(u|Z) = \upsilon(u)e^{\zeta^\top Z},$$

where $\upsilon(\cdot)$ is an unknown baseline hazard function, and ζ is a vector of unknown regression parameters. Denote the true values of $\upsilon(\cdot)$ and ζ in this model as υ_0 and ζ_0, respectively. We can estimate the regression coefficient ζ_0 using the partial likelihood, and estimate the cumulative baseline hazard $\Upsilon_0(u) = \int_0^u \upsilon_0(t)dt$ using the Breslow estimator (Breslow, 1972), denoted by $\hat{\zeta}_n$ and $\hat{\Upsilon}_n$, respectively. As a result, $F_0(u|z)$ is estimated by

$$\hat{F}_n(u|Z) = 1 - \exp\big(- \hat{\Upsilon}_n(u) e^{\hat{\zeta}_n^\top Z}\big). \tag{2.5}$$

In Stage 2, we estimate the smooth nondecreasing monotone function Λ_0 using the spline-based sieve method (Lu, Zhang, and Huang, 2009). Take $\mathcal{T} = \{s_i, i = 1, \dots, m_n + 2l\}$, with

$$0 = s_1 = \cdots = s_l < s_{l+1} < \cdots < s_{m_n+l} < s_{m_n+l+1} = \cdots = s_{m_n+2l} = \tau$$

being a sequence of knots that partition $[0, \tau]$ into $m_n + 1$ subintervals $I_i = [s_{l+i}, s_{l+i+1}]$, for $i = 0, 1, \dots, m_n$. Let Φ_n be a class of polynomial splines of order $l \geq 1$ with the knot sequence $\mathcal{T}$. Then Φ_n is linearly spanned by the normalized B-spline basis functions $\{B_l, l = 1, \dots, q_n\}$, with $q_n = m_n + l$ (Schumaker, 1981). Let

$$\Psi_n = \left\{ \sum_{l=1}^{q_n} \alpha_l B_l : 0 \leq \alpha_1 \leq \cdots \leq \alpha_{q_n}, \sum_{l=1}^{q_n} \alpha_l B_l(0) = 0 \right\}$$

be a subclass of Φ_n. For any $\Lambda(t) \in \Psi_n$, $\Lambda(t)$ is monotone nondecreasing and $\Lambda(0) = 0$. In addition, let $\mathcal{R}$ be a compact set of $\mathbb{R}^d$. The two-stage estimator of θ_0 is given by

$$\hat{\theta}_n = \arg \max_{\theta \in \mathcal{R} \times \Psi_n} l_n(\theta, \hat{F}_n; \underline{D}),$$

with $\hat{\Lambda}_n = \sum_{l=1}^{q_n} \hat{\alpha}_l B_l$. The proposed estimator can be computed in two steps using the usual profile likelihood method. First, for each fixed value of β, we take

$$\hat{\Lambda}_n(\beta) = \arg \max_{\Lambda \in \Psi_n} l_n(\theta, \hat{F}_n; \underline{D})$$

and define $l_n^{\text{Profile}}(\beta; \underline{D}) = l_n(\beta, \hat{\Lambda}(\beta), \hat{F}_n; \underline{D})$, yielding $\hat{\beta}_n = \arg \max_{\beta \in \mathcal{R}} l_n^{\text{Profile}}(\beta; \underline{D})$ and $\hat{\Lambda}_n = \hat{\Lambda}_n(\hat{\beta}_n)$. The proposed two-stage CELL estimator can be implemented using the constrained optimization package *constrOptim* in R.

3. Asymptotic Properties

3.1 Notation and Metrics

We expand the notation and metrics originally defined in Wellner and Zhang (2007) to study the asymptotic properties of the proposed estimator. Let $\mathcal{B}^d$ and $\mathcal{B}$ denote the collection of Borel sets in $\mathbb{R}^d$ and $\mathbb{R}$, respectively, and let $\mathcal{B}_{[0,\tau]} = \{B \cap [0, \tau] : B \in \mathcal{B}\}$. We then define the measures $\nu_1, \nu_2, \nu_{n1}, \mu_1$, and μ_2 as follows. For any $B_1, B_2, B \in \mathcal{B}_{[0,\tau]}$, and $B_3 \in \mathcal{B}^d$,

$$\begin{aligned}
\nu_1(B_1 \times B_2 \times B_3) = &\int_{B_3}\int_0^{\tau}\sum_{k=1}^{\infty} P(K=k|Z=z,U=u) \\
&\times \sum_{j=1}^{k} P(u-T_{k,j-1}\in B_1, u-T_{k,j}\in B_2|Z=z,K=k,U=u)dF_0(u|z)dH(z), \\
\nu_2(B \times B_3) = &\int_{B_3}\int_0^{\tau}\sum_{k=1}^{\infty} P(K=k|Z=z,U=u) \\
&\times \sum_{j=1}^{k} P(u-T_{k,j}\in B|Z=z,K=k,U=u)dF_0(u|z)dH(z), \\
\mu_1(B_1\times B_2) =&\nu_1(B_1\times B_2\times \mathbb{R}^d), \\
\mu_2(B) =&\nu_2(B\times \mathbb{R}^d), \\
\mu_3(B_1\times B_2) = &\int_{\mathbb{R}^d}\int\sum_{k=1}^{\infty} P(K=k|Z=z,U=u) \\
&\times P(u-T_{k,k}\in B_1, u\in B_2|Z=z,K=k,U=u)dF_0(u|z)dH(z),
\end{aligned}$$

where H is the distribution function of Z. When the distribution function F_0 in measure ν_1 is replaced by $\hat{F}_n$, we use the notation ν_{n1} instead. We also define the L_2-metrics $d_j(\theta_1,\theta_2), for j = 1, 20$, in the parameter space Θ as

$$d_j(\theta_1,\theta_2) =\Big\{\|\beta_1-\beta_2\|^2 + \|\Lambda_1-\Lambda_2\|^2_{L_2(\mu_j)}\Big\}^{1/2}.$$

Let $\mathcal{Z}$ be the support of H. Let M_j ($j = 1, \ldots, 7$) and c denote universal constants. Let $g^{(r)}$ be the rth derivative function of g. For $r \geq 1$, we define

$$\mathcal{H}_r = \{g : |g^{(r-1)}(s) - g^{(r-1)}(t)| \leq c|s-t| \text{ for all } 0 \leq s < t \leq \tau\},$$

$$\mathcal{F} = \{F : \ F(u|z) \text{ is a conditional distribution function on } [0,\tau] \times \mathcal{Z}\},$$

$$\mathcal{F}_\eta = \{F : \ \|F - F_0\|_\infty \leq \eta, F \in \mathcal{F}\},$$

$$\Psi = \{\Lambda : \ \Lambda(u) \text{ is a strictly increasing continuous function over } [0,\tau] \text{ with } \Lambda(0) = 0, \Lambda \in \mathcal{H}_r\},$$

$$\Theta = \mathcal{R} \times \Psi, \ \Theta_n = \mathcal{R} \times \Psi_n,$$

$$\Theta_{n\delta} = \{\theta : \ d_1(\theta, \theta_0) \leq \delta, \theta \in \Theta_n\}, \ \Theta_\delta = \{\theta : \ d_1(\theta,\theta_0) \leq \delta, \theta \in \Theta\},$$

where $\|F - F_0\|_\infty = \sup_{u,z} |F(u|z) - F_0(u|z)|$, for $F \in \mathcal{F}$.

3.2 Consistency and Convergence Rate

We impose the following regularity conditions in order to derive the consistency and convergence rate of the proposed estimator:

(C1) The true parameter $\theta_0 = (\beta_0, \Lambda_0) \in \mathcal{R}^0 \times \Psi$, with $0 < \Lambda_0(\tau) < \infty$, where $\mathcal{R}^0$ is the interior of $\mathcal{R}$.

(C2) $\mathcal{Z}$ is a bounded set in $\mathbb{R}^d$ and $P(K \leq M_1) = 1$, for some positive constant M_1.

(C3) The function $J_0(D) \equiv \sum_{j=1}^{K} \triangle N_j \log(\triangle N_j)$ satisfies $E\{J_0(D)\} < \infty$.

(C4) $\inf_{z \in \mathcal{Z}} P(U > Y|Z = z) = M_2$, for $0 < M_2 < 1$.

(C5) The conditional density function f_0 of the terminal event time U for given Z satisfies $f_0(u|z) \geq M_3$ on its support set, a measurable subset of $[u_0, \tau] \times \mathcal{Z}$ with $u_0 > 0$ and $M_3 > 0$.

(C6) The measure $\mu_j \times H$ on $([0, \tau]^j \times \mathbb{R}^d, \mathcal{B}[0, \tau]^j \times \mathcal{B}^d)$ is absolutely continuous with respect to ν_j, for $j = 1, 2$.

(C7) For all $\beta \in \mathbb{R}^d$ with $\beta \neq 0$, $P(\beta^\top Z \neq c) > 0$.

(C8) The maximum spacing of the knots, $\max_{1 \leq i \leq q_n+1} |t_i - t_{i-1}| = O(n^{-\gamma})$, with $q_n = O(n^\gamma)$, for $0 < \gamma < 1/2$, and $\frac{\max_{1 \leq i \leq q_n+1} |t_i - t_{i-1}|}{\min_{1 \leq i \leq q_n+1} |t_i - t_{i-1}|} \leq M_4$ uniformly for n.

Condition (C1) is a regular condition on the true parameter; (C2) and (C3) impose bounded conditions on the number of observation times K and the recurrent process N, respectively; (C4) implies that the censoring rate falls between zero and one; (C5) is a condition on the distribution of the terminal event time, and is satisfied by most continuous random variables; (C6) and (C7) are needed to establish the identifiability of the semiparametric model; and (C8) is a technical condition to ensure the approximation for the monotone function (Lu, Zhang, and Huang, 2007, 2009).

Theorem 3.1. *(Consistency) Suppose that* $\|\hat{F}_n - F_0\|_\infty \to 0$ *almost surely and* $\mu_3(\{0\} \times \{\tau\}) > 0$. *Under Conditions (C1)–(C8),* $d_1(\hat{\theta}_n, \theta_0) \to 0$ *and* $d_2(\hat{\theta}_n, \theta_0) \to 0$ *almost surely.*

To derive the rate of convergence, we need additional conditions:

(C9) The observation time points are s_0-separated; that is, there exists a constant $s_0 > 0$ such that $P(T_{K,j} - T_{K,j-1} \geq s_0$ for all $j = 1, \ldots, K) = 1$. Furthermore, μ_2 is absolutely continuous with respect to the Lebesgue measure, with a derivative $\dot{\mu}_2(t) \geq M_5 > 0$, for some positive constant M_5.

(C10) The true baseline function Λ_0 is differentiable and the derivative has positive and finite lower and upper bounds in the observation interval; that is, there exists a constant $0 < M_6 < \infty$ such that $1/M_6 \leq \dot{\Lambda}_0(t) \leq M_6$, for $t \in [0, \tau]$.

(C11) $P\big(e^{cN(\tau)}\big) \leq M_7$, for some positive constants c and M_7.

(C12) For some $\varpi \in (0, 1)$, $a^\top Var(Z|U_1, U_2)a \geq \varpi a^\top E(ZZ^\top|U_1, U_2)a$ almost surely for all $a \in \mathbb{R}^d$, where (U_1, U_2, Z) has distribution $\nu_1/\nu_1([0, \tau]^2 \times \mathcal{Z})$.

These conditions are similar to those required by Wellner and Zhang (2007) and Lu, Zhang, and Huang (2009) when studying semiparametric models for panel count data. Condition (C9) requires that adjacent observation times be at least s_0 apart, and that the intensity of the measure μ_2 be strictly positive. Condition (C10) requires that the true baseline function Λ_0 be absolutely continuous, with a bounded intensity function. This assumption is used mainly for technical convenience in the proofs. Condition (C11) is satisfied easily by a uniformly bounded process $N(t)$ or a Poisson-type process, and (C12) is justified by the arguments for conditions (C13) and (C14) in Wellner and Zhang (2007). These conditions, while mainly for technical purposes, are quite mild in practice.

Theorem 3.2. *(Convergence rate) Suppose that* $\|\hat{F}_n - F_0\|_\infty = O_p(n^{-r/(1+2r)})$. *Under Conditions (C1)–(C12), we have*

$$d_1(\hat{\theta}_n, \theta_0) = O_p(n^{-r/(1+2r)}).$$

Theorem 3.2 shows that the convergence rate of the semiparametric two-stage estimator $\hat{\theta}_n$ is of order $n^{-r/(1+2r)}$, even if the convergence rate of the nonparametric estimator $\hat{F}_n$ for the nuisance functional parameter F_0 is below $n^{\frac{1}{2}}$.

3.3 Asymptotic Normality

Define $\boldsymbol{H}_1 = \Big\{\boldsymbol{h} = (h_1, h_2) : h_1 \in \mathcal{R}, h_2 \in \mathcal{H}_r, \|h_1\| \leq 1, \|h_2\|_\infty \leq 1\Big\}$. Writing

$$A_j(u) = \Big(\frac{\triangle N_j}{\triangle\Lambda_j(u)e^{\beta^\top Z}} - 1\Big)e^{\beta^\top Z},$$

$$S(u) = \sum_{j=1}^{K}\Big[\triangle N_j \log(\triangle\Lambda_j(u)e^{\beta^\top Z}) - \triangle\Lambda_j(u)e^{\beta^\top Z}\Big],$$

we define, for any $\boldsymbol{h} = (h_1, h_2) \in \boldsymbol{H}_1$,

$$m(\theta, F; D) = \Delta S(Y) + \frac{1-\Delta}{\overline{F}(Y|Z)}\int_Y^\infty S(u)dF(u|Z),$$

$$m_1(\theta, F; D)[h_1] = \Delta h_1^\top Z\sum_{j=1}^{K} A_j(Y)\triangle\Lambda_j(Y) + \frac{1-\Delta}{\overline{F}(Y|Z)}h_1^\top Z\int_Y^\infty \sum_{j=1}^{K} A_j(u)\triangle\Lambda_j(u)dF(u|Z),$$

$$m_2(\theta, F; D)[h_2] = \Delta\sum_{j=1}^{K} A_j(Y)\triangle h_{2j}(Y) + \frac{1-\Delta}{\overline{F}(Y|Z)}\int_Y^\infty \sum_{j=1}^{K} A_j(u)\triangle h_{2j}(u)dF(u|Z),$$

and

$$m^{(1)}(\theta, F; D)[\boldsymbol{h}] = m_1(\theta, F; D)[h_1] + m_2(\theta, F; D)[h_2],$$

with $\triangle h_j(u) = h(u - T_{K,j-1}) - h(u - T_{K,j})$. For ease of exposition, we use $\{\tilde{D}_i, i = 1, \ldots n\}$ to represent the i.i.d. sample for estimating F_0 in Stage 1.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0353

Theorem 3.3. *Suppose that Conditions (C1)–(C12) hold, and* $\|\hat{F}_n - F_0\|_\infty = O_p(n^{-r/(1+2r)})$. *Assume that there exists a uniformly bounded process* O *and a Lipschitz function* g *such that*

$$\sqrt{n}\int_0^\tau \psi(u;D)d[\hat{F}_n(u|Z) - F_0(u|Z)] = \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^\tau \tilde{\psi}(u;D)dO(u;D;\tilde{D}_i)$$

is distributed asymptotically as a normal distribution with mean zero for the integrable function ψ, *where* $\tilde{\psi} = g \circ \psi$, *with* $g \circ \psi$ *denoting the composite of the functions* g *and* ψ. *Then, we have*

(i)

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{\mathcal{D}} N_d(0, \Sigma_1^{-1}\Sigma_2(\Sigma_1^{-1})^\top),$$

where

$$\Sigma_1 = E\left(\sum_{j=1}^K \triangle\Lambda_{0j}(U)e^{\beta_0^\top Z}\big[Z - R(U,K,T_{K,j-1},T_{K,j})\big]^{\otimes 2}\right),$$
$$\Sigma_2 = E\Big(m^*(\theta_0, F_0; D)^{\otimes 2}\Big),$$

with $R(U,K,T_{K,j},T_{K,j'}) = E(Ze^{\beta_0^\top Z}|U,K,T_{K,j},T_{K,j'})/E(e^{\beta_0^\top Z}|U,K,T_{K,j},T_{K,j'})$;

(ii) For $h \in \mathcal{H}_r$,

$$\sqrt{n}\int \frac{\triangle\hat{\Lambda}_n(t,s)-\triangle\Lambda_0(t,s)}{\triangle\Lambda_0(t,s)} e^{\beta_0^\top z}\triangle h(t,s)d\nu_1(s,t,z) \xrightarrow{\mathcal{D}} N(0,\sigma_1^2[h]),$$

where $\triangle\Lambda(t,s)=\Lambda(t)-\Lambda(s)$, $\triangle h(t,s)=h(t)-h(s)$ *and* $\sigma_1^2[h]=E(m^\dagger(\theta_0,F_0;D)[h])^2$, *with* $m^\dagger(\theta_0,F_0;D)[h]=m_2(\theta_0,F_0;D)[h]-\tilde{m}^*(\theta_0,F_0;D)[h]+m^{**}(\theta_0,F_0;D)[h]$.

Here, for some complicated expression $m^*(\theta,F;D)$, $\tilde{m}^*(\theta,F;D)$ and $m^{**}(\theta,F;D)$ are given in the Supplementary Material for the proof of Theorem 3.3. The second part of this theorem is developed to construct test statistics for testing the null hypothesis: $\Lambda=\Lambda_0$ and $\beta=\beta_0$.

To derive the asymptotic normality of the two-stage estimator under the Cox proportional hazards model for the terminal event time, we need an additional assumption.

(C13) The information matrix of the partial likelihood for the Cox regression model at the true parameter values is positive definite.

Corollary 3.1. *Suppose Conditions (C1)–(C13) hold. If there exists some positive constant* M *such that* $\inf_{z\in\mathcal{Z}} P(C\geq\tau|Z=z)=M$, *then Theorem 3.3 holds for* $\Lambda_0\in\mathcal{H}_r$, *for* $r\geq 2$, *when* F_0 *is estimated using* $\hat{F}_n(u|Z)$ *in (2.5).*

Remark 3.1. *The assumption that* $\inf_{z\in\mathcal{Z}} P(C\geq\tau|Z=z)=M>0$ *is*

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0353

a common technical condition for the weak convergence result of the baseline hazard function estimator on an interval $[0, \tau]$ *under Cox regression models (e.g., Condition C1(b) of Kalbfleisch and Prentice 2002, page 175; Condition D of Andersen and Gill 1982; Condition (2.5) of Fleming and Harrington 1991, page 290; Condition 5 in Kong et al. 2018). In addition, as pointed out by Kalbfleisch and Prentice (2002, page 178), with right-censoring, the asymptotic result can be extended to hold for data on the entire interval* $[0, \infty)$ *by placing somewhat stronger conditions on the covariates (see, e.g., Argas and Haara, 1988).*

4. Simulation Studies

We conducted extensive numerical studies to evaluate the finite-sample performance of the proposed likelihood-based two-stage estimator. To simulate panel count data truncated by a terminal event, we first generated the number of observation times K_i from a uniform distribution with an equal probability of 1/6 on $\{1, 2, 3, 4, 5, 6\}$. We generated the censoring time C_i from the exponential distribution with rate parameter $0.1e^{-1.5z_{1i}+0.5z_{2i}}$, truncated at τ_0. Here, we chose τ_0 to achieve censoring rates of 20% and 40%, in the two simulation settings, respectively, and assume that the terminal event time U_i follows an exponential distribution with rate parameter

$e^{z_{1i}-z_{2i}}$, where z_{1i} and z_{2i} are two covariates, with $z_{1i} \sim \text{Uniform}(0,1)$ and $z_{2i} \sim \text{Bernoulli}(0.5)$. The observed data for the terminal event time U_i consist of $Y_i = \min\{U_i, C_i\}$ and $\Delta_i = I(U_i \leq C_i)$, where $I(\cdot)$ is the indicator function. For subject i with observed data Y_i and K_i, we set the observation times T_{ij}, for $j = 1, \cdots, K_i$, as ordered variates from $\text{Uniform}(0, Y_i)$.

We consider three distributions for generating panel counts: (i) N_i follows a standard Poisson process, with conditional mean function $E(N_i(t)|U_i, Z_i) = \{\Lambda_0(U_i) - \Lambda_0(U_i - t)\}e^{\beta_0^\top Z_i}$; (ii) N_i follows a mixed Poisson process, with conditional mean function $E(N_i(t)|U_i, Z_i, \gamma_i) = \gamma_i\{\Lambda_0(U_i) - \Lambda_0(U_i - t)\}e^{\beta_0^\top Z_i}$, in which $\gamma_i \sim Gamma(2, 1/2)$; and (iii) N_i follows a negative binomial process, with conditional mean function $E(N_i(t)|U_i, Z_i) = \{\Lambda_0(U_i) - \Lambda_0(U_i - t)\}e^{\beta_0^\top Z_i}$. For all cases, we set $\Lambda_0(u) = 8(1 - e^{-u})$ and $\beta_0 = (1, 1)^\top$.

For Cases (i) and (ii), we generated $\triangle N_{i,j}$ from $\text{Poisson}(\lambda_{i,j})$, with $\lambda_{i,j} = \{\Lambda_0(U_i - T_{i,j-1}) - \Lambda_0(U_i - T_{i,j})\}e^{\beta_0^\top Z_i}$ and $\lambda_{i,j} = \gamma_i\{\Lambda_0(U_i - T_{i,j-1}) - \Lambda_0(U_i - T_{i,j})\}e^{\beta_0^\top Z_i}$, respectively; for Case (iii), we generated $\triangle N_{i,j}$ from $\text{Neg-Binomial}(\lambda_{i,j}, 0.5)$, with $\lambda_{i,j} = \{\Lambda_0(U_i - T_{i,j-1}) - \Lambda_0(U_i - T_{i,j})\}e^{\beta_0^\top Z_i}$.

For each of the three counting processes, we conducted a Monte Carlo simulation with 500 repetitions for each combination of sample sizes $n = 100, 200$ and censoring rates. Although we established the asymptotic normality in Theorem 3.3, the standard errors could not be obtained easily using an empirical estimation, owing to the complicated form of the asymptotic variance-covariance matrix of the estimators. Hence, we propose estimating the standard errors using the bootstrapping technique. The estimated standard error and the coverage probability were computed based on 100 bootstrap samples. Figure 1 displays the average estimated $\Lambda(u)$ based on 500 repetitions for the three cases with a censoring rate of 20%, where the pointwise estimates are close to the corresponding values of the true function, on average. Table 1 shows the simulation results, including the estimation bias, sample standard deviation (SSD), average of the bootstrap-based standard errors (ASE), and 95% coverage probability (CP) for the three scenarios. The results show that the proposed method yields asymptotically unbiased estimators, the estimated standard errors are close to the corresponding sample standard deviations, especially when the sample size increased to 200, and the 95% confidence intervals exhibit reasonably accurate coverage probability. As expected, the data with a low censoring rate of terminal events led to less biased estimations in finite samples and smaller estimation variability. Although the counting processes in Cases (ii) and (iii) violate the assumption of the conditional Poisson process used in the working model to derive the likelihood, the inferences remain valid. However, the estimation variability of the regression parameters appears

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0353

to be larger relative to Case (i), for which the working model was indeed a correct model. The simulation results attest that the proposed reversed mean model under the working Poisson process is robust against the underlying stochastic process used to generate the observed panel count data, as shown for the ordinary mean model for panel count data without considering an informative terminal event (Wellner and Zhang, 2007; Lu, Zhang, and Huang, 2009).

5. Applications

5.1 Chinese Longitudinal Healthy Longevity Survey Data

We applied the proposed likelihood-based two-stage estimation procedure under the reversed mean model to the Chinese Longitudinal Healthy Longevity Survey (CLHLS) data. The CLHLS conducted a survey study on elderly people aged 65 or older with health and quality of life related questionnaires from 1998 to 2014 (Zeng et al., 2017). Follow-up interviews were carried out every two to three years with new individuals recruited as replacement for those lost to follow-up or deceased. We only included individuals of the 1998 enrollment because these individuals provided the longest observation window. We noted there were some seemingly erroneous or missing data records, likely due to administrative mistakes. After removing those erroneous or missing records, data from 2904 individuals remained for analysis.

We considered the reported number of serious illnesses as the recurrent event process of interest, which could be truncated by the terminal event of death. The censored subjects include those lost to follow-up or alive at the end of study period, and the sample yielded a censoring rate 27% with the longest follow-up time of $\tau = 197$ months. We investigated two binary covariates including gender and residence with the goal to explore the covariate effects on the mean function of serious illness counts under consideration. For this purpose, we applied the proposed two-stage estimation procedure to panel count data with right-censored terminal event data, where the Cox proportional hazards model was used in Stage 1 to estimate the conditional distribution function of survival time. The diagnostic test for the proportional hazards assumption (Grambsch and Therneau, 1994) for the U variable yielded p-value 0.19, not suggesting violation of the Cox model. Table 2 shows the results from applying the two-stage estimation procedure to the CLHLS dataset, where the standard errors of the estimates were obtained by using 100 bootstrap samples. The effect size of RESIDENCE was -0.2820, which implies that participants living in rural regions experienced $1 - e^{-0.2820} = 24.6\%$ less serious illnesses compared to participants from urban regions, controlling for gender. In general, individuals living in urban areas may experience higher stress from social and physical environments, which may lead to more recurrences of serious illnesses. Additionally, easy access to medical facilities helps to identify more occurrences of serious illnesses as well as providing more timely and better quality of medical care, contributing to the fact that urban residents tend to report serious illnesses more frequently than those residing in rural regions but live longer. Gender did not show any significant effect on the number of serious illnesses occurred during the observation period.

5.2 Bladder Tumor Data

We also applied the proposed method to another data set from the bladder tumor study conducted by the Veterans Administration Co-operative Urological Research Group (Andrews and Herzberg, 1985). The study included 118 patients initially diagnosed with bladder cancer and subsequently randomized to one of the three treatments: thiotepa, pyridoxine, or placebo. During each follow-up clinical visit, the number of new tumors since the last visit was recorded and resected. For this analysis, we considered 85 patients in the placebo group and the thiotepa treatment group to study the efficacy of thiorepa treatment on suppressing the tumor recurrence. The recurrent event process was defined as the number of tumors accumulated since the initial diagnosis excluding the prior tumors and the terminal event time was defined as months elapsed from the date of diagnosis to death. The longest observation time was 64 months. For patients who lost to follow-up or were alive at the end of the study period, the terminal event times were regarded as censored and the censoring rate was 65%. To analyze the effect of thiotepa treatment, we also adjusted for the number of initial tumors at diagnosis. The observed data were typical panel count data with death as the right-censored terminal event. Table 3 shows the results from our two-stage estimation method. In particular, the results from Stage 2 show that the initial number of tumors at diagnosis was positively associated with the recurrence of subsequent tumors with an effect size of 0.2326 (p-value < 0.01), which indicates that one additional initial tumor is expected to increase subsequent recurrence of tumors by 26.2% , controlling for treatment. The thiotepa treatment was found to significantly suppress the tumor recurrence with the effect size of -0.7045 (p-value < 0.05). In other words, thiotepa is expected to reduce the tumor recurrences by 50.1%, controlling for the number of initial tumors. The results are comparable with those in previous studies (Sun and Wei, 2000; Wellner and Zhang, 2007; Lu, Zhang and Huang, 2009).

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0353

6. Conclusion

In joint analysis of longitudinal count and survival data, the effect of longitudinal markers on the survival is often the main study of interest (Huang and Wang, 2004; Sun et al., 2012). In the recent literature of joint analysis of longitudinal and survival data, there has been an increasing interest in studying the behavior of longitudinal data near the terminal event for addressing more relevant scientific questions (Chan and Wang, 2010, 2017; Kong et al., 2018). To the best of our knowledge, there is no such effort for panel count data. We propose a semiparametric reversed conditional mean model to characterize the behavior of a recurrent event process near an informative terminal event and develope a novel M-estimation procedure based on the conditional expectation of log-likelihood derived from a non-homogeneous Poisson process.

The estimation procedure is implemented through a two-stage mechanism for numerical convenience, for which the first stage estimates the conditional distribution function of the terminal event. We fit the Cox model to terminal event time data due to its popularity and well-established asymptotic properties that fulfill the required conditions for the nuisance parameter. If the proportional hazards assumption fails, we may consider alternative survival models such as additive hazards or accelerated failure time models, or even the local Kaplan–Meier estimator (Dabrowska, 1989) to estimate the conditional distribution function of the terminal event time.

For the adequacy of model (2.1), we can similarly develop some graphical and numerical procedures using cumulative sums of the residuals following the ideas in Lin et al. (2000) and Zhao and Tong (2011). However, the procedure is more complicated. An easy-to-implement model-checking procedure needs to be explored.

Acknowledgments

The authors thank the editor, associate editor, and two reviewers for their constructive and insightful comments and suggestions. This research was supported, in part, by the Research Grant Council of Hong Kong (15301218, 17308420), the Natural Science Foundation of China (12271459, 12171374), and National Institutes of Health of USA (U54GM115458).

Supplementary Material

The online Supplementary Material contains all technical proofs for Theorems 3.1–3.3 and Corollary 3.1.

References

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Annals of Statistics* **10**, 1100-1120.

Andrews, D. F. and Herzberg, A. M. (1985). *Data: A Collection of Problems from Many Fields for the Student and Research Worker*. Springer, New York.

Arjas, E. (1988). A graphical method for assessing goodness of fit in Cox's proportional hazards model. *Journal of the American Statistical Association* **83**, 204–212.

Breslow, N. E. (1972). Discussion of "Regression models and life-tables" by D. R. Cox. *Journal of the Royal Statistical Society, Series B* **34**, 216–217.

Breslow, N. E. (1984). Extra-Poisson variation in log-linear models. *Applied Statistics* **33**, 38–44.

Chan, K-C. and Wang, M-C. (2010). Backward estimation of stochastic processes with failure events as time origins. *Annals of Applied Statistics* **4**, 1602–1620.

Chan, K-C. and Wang, M-C. (2017). Semiparametric modeling and estimation of the terminal behavior of recurrent marker processes before failure events. *Journal of the American Statistical Association* **112**, 351–362.

Cheng, S. C. and Wei, L. J. (2000). Inferences for a semiparametric model with panel data. *Biometrika* **87**, 89–97.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34**, 187–220.

Chiou, S. H., Huang, C. Y., Xu, G. and Yan, J. (2019). Semiparametric regression analysis of panel count data: a practical review. *International Statistical Review* **87**, 24–43.

Dabrowska, D. M. (1989). Uniform consistency of the kernel conditional Kaplan- Meier estimate. *Annals of Statistics* **19**, 1157–1167.

Diao, G., Zeng, D., Hu, K. and Ibrahimb, J.G. (2017). Modeling event count data in the presence of informative dropout with application to bleeding and transfusion events in myelodysplastic syndrome. *Statistics in Medicine* **36**, 3475–3494.

Diao, G., Zeng, D., Hu, K. and Ibrahim J. G. (2019). Semiparametric frailty models for zero-inflated event count data in the presence of informative dropout. *Biometrics* **75**, 1168–1178.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. John Wiley & Sons.

Grambsch, P. M. and Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* **81**, 515–526.

Hinde. J. (1982). Compound Poisson regression models. In *GLIM 82: Proceedings of the International Conference in Generalized Linear Models*, R. Gilchrist (ed), pp. 109–121, Berlin: Springer-Verlag.

Hu X. J., Lagakos, S. W. and Lockhart, R. A. (2009). Generalized least squares estimation of the mean function of a counting process based on panel counts. *Statistica Sinica* **19**, 561–580.

Hu, X. J., Sun, J. and Wei, L. J. (2003). Regression parameter estimation from panel counts. *Scandinavian Journal of Statistics* **30**, 25–43.

Huang, C.-Y. and Wang, M.-C. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association* **99**, 1153–1165

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

Kalbfleisch, J. D. and Lawless, J. F. (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association* **80**, 863–871.

Kalbfleisch, J. D., and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data.* Second edition. New York: Wiley.

Kong, S., Nan, B., Kalbfleisch, J. D., Saran, R. and Hirth, R. (2018). Conditional modeling of longitudinal data with terminal event. *Journal of the American Statistical Association* **113**, 357–368.

Lu, M., Zhang, Y. and Huang, J. (2007). Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika* **94**, 705–718.

Lu, M., Zhang, Y. and Huang, J. (2009). Semiparametric estimation methods for panel count cata using monotone B-splines. *Journal of the American Statistical Association* **104**, 1060–1070.

Ma, L. and Sundaram, R. (2018). Analysis of gap times based on panel count data with informative observation times and unknown start time. *Journal of the American Statistical Association* **113**, 294–305.

Schumaker, L. (2007). *Spline Functions: Basic Theory*. Cambridge University Press.

Sun, J. and Kalbfleisch, J. D. (1995). Estimation of the mean function of point processes based on panel data. *Statistica Sinica* **5**, 279–289.

Sun, J., Tong, X. and He, X. (2007). Regression analysis of panel count data with dependent observation times. *Biometrics* **63**, 1053–1059.

Sun, J. and Wei, L. J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *Journal of the Royal Statistical Society Series B* **62**, 293–302.

Sun, L., Song, X., Zhou, J. and Liu, L. (2012). Joint analysis of longitudinal data with informative observation times and a dependent terminal event. *Journal of the American Statistical Association* **107**, 688–700.

Thall, P. F. (1988). Mixed Poisson likelihood regression models for longitudinal interval count data. *Biometrics* **44**, 197–209.

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

Wellner, J. A. and Zhang, Y. (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *Annals of Statistics* **35**, 2106–2142.

Zeng, D. and Cai, J. (2010). Asemiparametric additive rate model for recurrent events with an informative terminal event. *Biometrika* **97**, 699–712.

Zeng, Y., Vaupel, J., Xiao, Z., Liu, Y. and Zhang, C. (2017). Chinese Longitudinal Healthy Longevity Survey (CLHLS), 1998-2014. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2017-04-11.

Zhang, Y. (2002). A semiparametric pseudolikelihood estimation method for panel count data. *Biometrika* **89**, 39–48.

Zhang, Y. and Jamshidian, M. (2003). The gamma-frailty Poisson model for the nonparametric estimation of panel count data. *Biometrics* **59**, 1099–1106.

Zhao, H., Li, Y. and Sun, J. (2013). Analyzing panel count data with a dependent observation process and a terminal event. *Canadian Journal of Statistics* **41**, 174–191.

Zhao, X., Zhou, J. and Sun, L. (2011). Semiparametric transformation models with time-varying coefficients for recurrent and terminal events. *Biometrics* **67**, 404–414.

Zhou, J., Zhang, H., Sun, L. and Sun, J. (2017). Joint analysis of panel count data with an informative observation process and a dependent terminal event. *Lifetime Data Analysis* **23**, 560–584.

Zhu, L., Zhang, Y., Li, Y., Sun, J. and Robison, L. L. (2018). Semiparametric likelihood-based method for regression analysis of mixed panel-count data. *Biometrics* **74** 488–497.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0353

Table 1: Simulation results for the recurrent event process following Poisson process, mixed Poisson process, and negative binomial process

Poisson Process

CR		$n = 100$		$n = 200$	
		β_1	β_2	β_1	β_2
20%	Bias	0.0516	-0.0197	0.0454	-0.0180
	SSE	0.1756	0.1044	0.1131	0.0708
	ESE	0.1711	0.1018	0.1174	0.0709
	CP	0.9280	0.9360	0.9460	0.9520
40%	Bias	0.0510	-0.0076	0.0283	-0.0107
	SSE	0.2272	0.1206	0.1583	0.0845
	ESE	0.2302	0.1284	0.1564	0.0870
	CP	0.9380	0.9640	0.9320	0.9540

Mixed Poisson Process

CR		$n = 100$		$n = 200$	
		β_1	β_2	β_1	β_2
20%	Bias	0.0440	-0.0307	0.0383	-0.0197
	SSE	0.4188	0.2137	0.2876	0.1499
	ESE	0.3976	0.2119	0.2809	0.1492
	CP	0.9380	0.9500	0.9520	0.9580
40%	Bias	0.0408	-0.0152	0.0454	-0.0067
	SSE	0.5103	0.2592	0.3405	0.1759
	ESE	0.4686	0.2538	0.3263	0.1759
	CP	0.9340	0.9480	0.9440	0.9520

Negative Binomial Process

CR		$n = 100$		$n = 200$	
		β_1	β_2	β_1	β_2
20%	Bias	0.0367	-0.0091	0.0428	-0.0196
	SSE	0.2187	0.1404	0.1612	0.0962
	ESE	0.2216	0.1355	0.1532	0.0937
	CP	0.9440	0.9340	0.9300	0.9420
40%	Bias	0.0364	-0.0117	0.0261	-0.0180
	SSE	0.2961	0.1716	0.2005	0.1162
	ESE	0.2977	0.1694	0.2054	0.1153
	CP	0.9440	0.9520	0.9480	0.9440

Note: CR represents the censoring rate.

Table 2: Estimation results for CLHLS data

Stage 1: Cox proportional hazards model

Variable	Estimate (95% CI)	p-value
Gender (female = 1, male = 0)	0.0257 (-0.0590, 0.1104)	0.5520
Residence (rural = 1, urban = 0)	0.2000 (0.1156, 0.2844)	3.38e-06***

Stage 2: Reversed mean model

Variable	Estimate (95% CI)	p-value
Gender (female = 1, male = 0)	0.0217 (-0.1013, 0.1448)	0.7294
Residence (rural = 1, urban = 0)	-0.2820 (-0.4164, -0.1476)	3.9155e-05***

Level of significance: * 0.05 ** 0.01 *** 0.005

Table 3: Estimation results for bladder tumor data

Stage 1: Cox proportional hazards model

Variable	Estimate (95% CI)	p-value
Number of initial tumors	0.1717 (-0.0154, 0.3588)	0.0721
Treatment (thiotepa = 1, placebo = 0)	-0.0735 (-0.8099, 0.6630)	0.8450

Stage 2: Reversed mean model

Variable	Estimate (95% CI)	p-value
Number of initial tumors	0.2326 (0.0788, 0.3865)	0.0030***
Treatment (thiotepa = 1, placebo = 0)	-0.7045 (-1.3188, -0.0901)	0.0246*

Level of significance: * 0.05 ** 0.01 *** 0.005

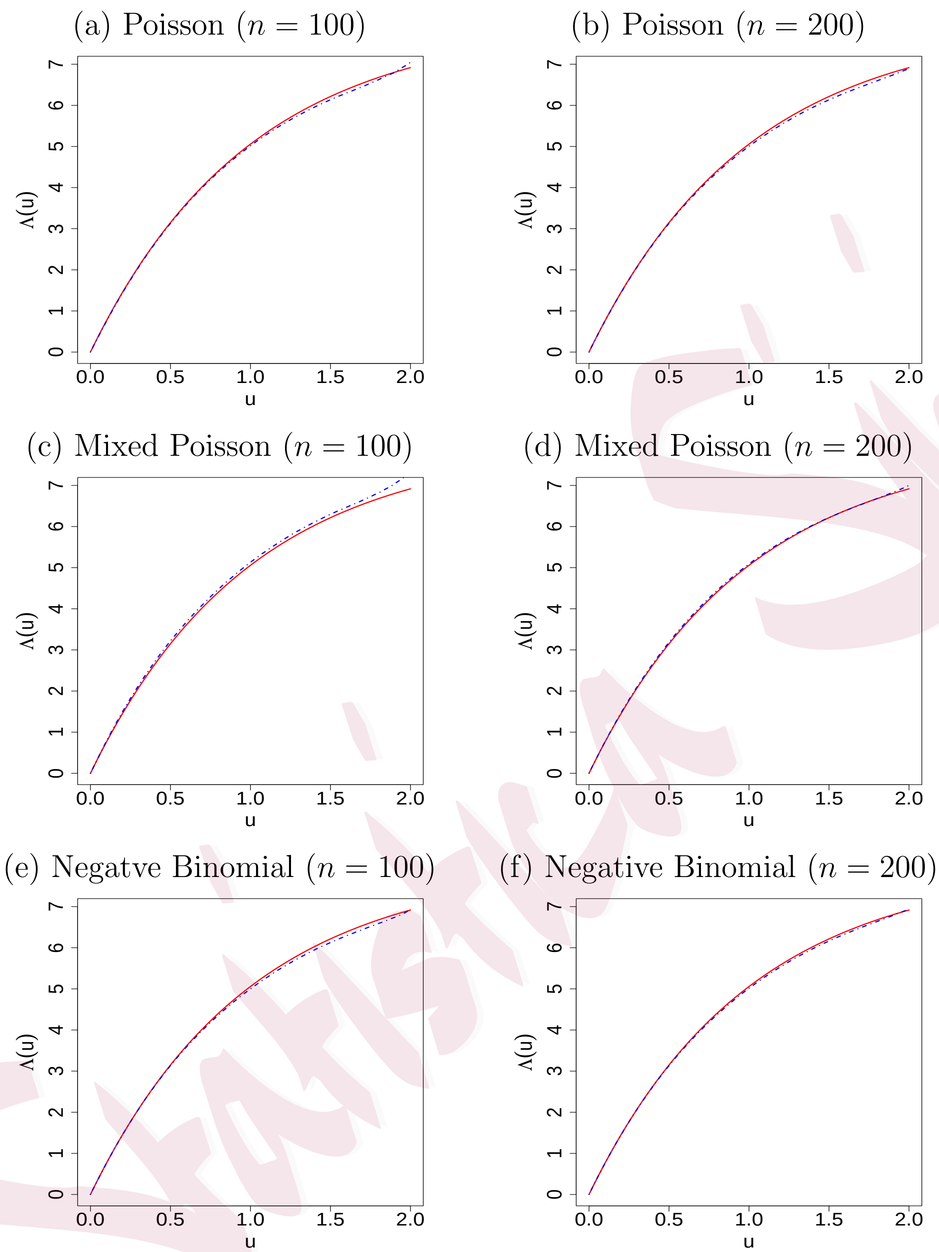

Figure 1: Plots of estimates for $\Lambda(u)$ with censoring rate 20% under three different counting processes. Red solid lines represent the true function and the blue dotted lines represent the point-wise average of estimated baseline mean functions based on 500 repetitions.