AUGMENTED ESTIMATING EQUATIONS FOR SEMIPARAMETRIC PANEL COUNT REGRESSION WITH INFORMATIVE OBSERVATION TIMES AND CENSORING TIME

Xiaojing Wang, Shuangge Ma and Jun Yan

Google New York, Yale University and University of Connecticut

Abstract: We propose an augmented estimating equation (AEE) approach for a semiparametric mean regression model with panel count data under possibly informative observation schemes and censoring. On a grid of time points, counts in all the subintervals of each observation window are treated as missing values, and are imputed with a robust working model given the observed count in the window. The observation scheme and the event process are allowed to be dependent through covariates and an unobserved frailty, which enters the mean function multiplicatively. Conditional on covariates, the censoring time and the event process can be dependent through the frailty. Regression coefficients and the unspecified baseline mean function are estimated with an Expectation-Solving (ES) algorithm. Distributions of the observation times, censoring time, and frailty are all considered as nuisance and unspecified. With empirical process theory, estimators for both the parametric and nonparametric component are shown to be consistent. The regression coefficient estimator is shown to be asymptotically normal. The cumulative baseline estimator is self-consistent in that the estimator is automatically non-decreasing. In simulation studies, the estimator performs well for moderate sample sizes and appears to be competitive in comparison with existing estimators under a wide range of practical settings. The utility of the proposed methods is illustrated with a bladder tumor study.

Key words and phrases: Expectation-Solving algorithm, missing data, semiparametric regression.

1. Introduction

Panel count data arise when an event process is observed only at a finite number of, often random, observation time points. This frequently occurs in clinical or industrial studies when continuous monitoring of the subjects is infeasible or too costly. For instance, in many long-term studies, each subject can experience multiple recurrences of the same event, but observations are only recorded at several distinct time points and, hence, only the numbers of events between two consecutive observation times are available. The observation times

and the censoring (follow-up) time vary from subject to subject, and both may be associated with the event process, which further complicates the statistical inferences.

Statistical methods that properly address the challenges of panel count data have attracted considerable attention. When there is no covariate, the mean function of the event process is the target of the statistical inferences. Existing methods are isotonic regression (Sun and Kalbfleisch (1995)), nonparametric maximum likelihood and nonparametric maximum pseudolikelihood (Wellner and Zhang (2000); Lu, Zhang, and Huang (2007)), generalized least squares (Hu, Lagakos, and Lockhart (2009a)), and generalized estimating equations (Hu, Lagakos, and Lockhart (2009b)).

When covariate effects are of main interest, semiparametric models such as proportional means or proportional rates models are desired. Although having been studied in the recurrent event setting by many authors (see, e.g., Cook and Lawless (2007) and references therein), semiparametric regression models are much less developed for panel count. Sun and Wei (2000) proposed an estimating equation approach to situations with noninformative observation times and censoring time. The validity of their inference procedure, however, relies on correct modeling of the observation pattern and censoring time. Zhang (2002) and Wellner and Zhang (2007) constructed an easy-to-implement pseudolikelihood from a nonhomogeneous Poisson process, with the dependence of the cumulative counts within a subject ignored. At higher computing expense, Wellner and Zhang (2007) maximized the nonhomogeneous Poisson loglikelihood, resulting in an estimator more efficient than the maximum pseudologlikelihood estimator. The computing burden of the semiparametric maximum likelihood method is alleviated by Lu, Zhang, and Huang (2009), who approximated the logarithm of baseline mean function with monotone cubic B-splines. Noninformative observation times and censoring time are assumed, but no model specification is needed for them.

When the observation times, or the censoring time or both, and the event process are dependent after conditioning on covariates, the literature is very limited. Huang, Wang, and Zhang (2006) proposed an estimating equations approach that allows observation times to be associated with the event process through an unobserved multiplicative frailty. This method relaxes the conditional independence assumption between observation times and the event process given covariates, with no need to specify the dependence and to model the frailty. The asymptotic distribution of the estimator was not established.

We approach panel count data from a missing data perspective and propose an augmented estimating equations (AEE) approach. A fine time grid is constructed from all subjects' distinct observation times, and the counts in the

subintervals of any subject's observation window are treated as missing values. An Expectation–Solving (ES) algorithm (Heyde and Morton (1996); Elashoff and Ryan (2004)) is employed to solve the expected estimating equations conditional on the observed data. Precedents for this approach are Pan (2000) and Goetghebeur and Ryan (2000) in the context of interval censored data. The method handles informative observation times seamlessly. Similar to Huang, Wang, and Zhang (2006), the method can handle informative censoring time by further treating the panel counts beyond the censoring time as missing values.

In Section 2, we formally present the data structure, notation, and a semiparametric mean regression model for panel count data. In Section 3, we develop the AEE approach and a two-step iterative ES algorithm under conditional independent censoring and under informative censoring. In Section 4, we study the finite sample performance of the estimator via simulation, report comparison results with existing methods, and illustrate our method with data from a bladder tumor study (Byar (1980)). A discussion concludes in Section 5. Proofs of the asymptotic properties of the proposed estimator are relegated in Appendix A. Computational details of the variance estimator of the regression coefficient estimator under the Poisson assumption are sketched in Appendix B.

2. Data and Model

Consider a study involving n subjects who experience a single type of recurrent events under discrete monitoring. For subject i, let $N_i(t)$ be the number of events up to time t. We observe $N_i(t)$ only at $M_i \geq 1$ random time points, $0 < T_{i,1} < \cdots < T_{i,M_i}$, i.e., $N_i(T_{i,1}), \ldots, N_i(T_{i,M_i})$. The last observation time T_{i,M_i} is also the censoring time. Suppose that X_i is a p-dimensional covariate vector. Let $D_i = \{M_i, T_{i,j}, N_i(T_{i,j}), X_i; \ j = 1, \ldots, M_i\}$; the observed panel count data are taken to be independent and identically distributed copies $\{D_1, \ldots, D_n\}$.

Our semiparametric regression model for the panel count data specifies the mean of the event process $N_i(t)$, given covariate X_i, as

$$\Lambda(t; X_i) = E[N_i(t)|X_i] = \Lambda(t) \exp(X_i'\beta), \tag{2.1}$$

where β is a $p \times 1$ vector of covariate coefficient and $\Lambda(\cdot)$ is a completely unspecified baseline mean function. Model (2.1) characterizes the mean of the event process without fully specifying how the process evolves. It is the counterpart of the proportional means model for recurrent event data (Pepe and Cai (1993); Lawless and Nadeau (1995); Lin et al. (2000)) in a panel count setting. It covers many models as special cases, such as the nonhomogeneous Poisson processes and mixed Poisson processes. In particular, it covers the case where an unobserved nonnegative frailty variable Z_i with $E[Z_i|X_i] = 1$ enters the mean function

multiplicatively (Huang, Wang, and Zhang (2006)),

$$\Lambda(t; X_i, Z_i) = E[N_i(t)|X_i, Z_i] = Z_i\Lambda(t)\exp(X_i'\beta). \tag{2.2}$$

The distribution of Z_i is unspecified.

We consider two scenarios of censoring for Model (2.1):

1. Conditional independent censoring: the censoring time is independent of the event process given observed covariates (Sun and Wei (2000); Zhang (2002)).

2. Informative censoring through a frailty: after conditioning on observed covariates, the censoring time and the event process are still dependent through an unobserved multiplicative frailty (Huang, Wang, and Zhang (2006)).

Note that, in both scenarios, observation times may or may not be independent of the event process given covariates; no model specification for observation times or for the frailty is necessary. Our proposed methods only distinguish whether or not the censoring time is conditionally independent of the event process given observed covariates.

3. Augmented Estimating Equations

3.1. Conditional independent censoring

We first develop the estimating procedure under conditional independent censoring, i.e., T_{i,M_i} and $N_i(\cdot)$ are independent given observed covariate X_i. This assumption implies $E[N_i(t)|X_i, t \leq T_{i,M_i}] = E[N_i(t)|X_i]$, so that we can use the censored data to estimate quantities in Model (2.1) about $E[N_i(t)|X_i]$. Let $\{s_1, \ldots, s_m\}$ be the union of all observation times and censoring times in $[0, \tau]$. These points form a data-dependent grid $G = \{0 = s_0 < s_1 < \cdots < s_m = \tau\}$. Let $r_{ij} = I(s_j \leq T_{i,M_i})$ be the at-risk indicator. We write $\mathbb{N}_{ij} = N_i(s_j) - N_i(s_{j-1})$ as the number of events occurred in $(s_{j-1}, s_j]$. Only summations of $\mathbb{N}_{ij}$'s over those subintervals whose union coincides with an observation window are observed. Let $\lambda_j = \Lambda_0(s_j) - \Lambda_0(s_{j-1})$, the baseline mean number of events occurring in interval $(s_{j-1}, s_j]$. We treat $\mathbb{N}_{ij}$'s as missing values. If $\mathbb{N}_{ij}$'s were observed, under conditional independent censoring, Model (2.1) suggests a set of complete-data estimating equations:

$$\sum_{i=1}^{n}[\mathbb{N}_{ij} - \lambda_j\exp(X_i'\beta)]r_{ij} = 0, \quad j = 1, \ldots, m, \tag{3.1}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{m}[\mathbb{N}_{ij} - \lambda_j\exp(X_i'\beta)]X_i r_{ij} = 0. \tag{3.2}$$

Model (2.1) contains nonhomogeneous Poisson processes as a special case, which provides some insight about (3.1) and (3.2). The complete-data log likelihood for nonhomogeneous Poisson processes is

$$l(\Lambda, \beta) = \sum_{i=1}^{n} \sum_{j=1}^{m} \left[\mathbb{N}_{ij} \log(\lambda_j) + \mathbb{N}_{ij} X_i' \beta - \lambda_j \exp(X_i' \beta) \right] r_{ij}. \qquad (3.3)$$

Estimating equations (3.1) and (3.2) coincide with the score equations derived from (3.3). When the true processes are not nonhomogeneous Poisson, they remain valid as long as Model (2.1) is correctly specified.

Nonhomogeneous Poisson processes shed light on properties of the $\mathbb{N}_{ij}$'s given the observed data. For subject i, consider the observation window $(T_{i,k-1}, T_{i,k}]$. This interval is broken into H_{ik} subintervals by G, $(s_l, s_{l+1}], \ldots, (s_{l+H_{ik}-1}, s_{l+H_{ik}}]$ for some l such that $s_l = T_{i,k-1}$ and $s_{l+H_{ik}} = T_{i,k}$. We only observe the total number of counts $m_{ik} = N_i(T_{i,k}) - N_i(T_{i,k-1})$ in $(T_{i,k-1}, T_{i,k}]$. If $m_{ik} = 0$, there is no event times to be imputed. If $m_{ik} > 0$, under the nonhomogeneous Poisson assumption, the vector of counts in all H_{ik} subintervals given m_{ik} follows a multinomial distribution with m_{ik} trials and probability vector proportional to baseline mean vector $(\lambda_{l+1}, \ldots, \lambda_{l+H_{ik}})$. The conditional expectation of $\mathbb{N}_{ij}$, $l + 1 \leq j \leq l + H_{ik}$, is $\lambda_j m_{ik}/(\lambda_{l+1} + \cdots + \lambda_{l+H_{ik}})$. The multinomial distribution is unchanged when the event process is a mixed Poisson process with a subject level multiplicative frailty. In the general setting of Model (2.1) with only the marginal mean at each time t specified, the conditional expectation of $\mathbb{N}_{ij}$, $l + 1 \leq j \leq l + H_{ik}$, remains the same. It is the conditional expectation, instead of the distribution of the $\mathbb{N}_{ij}$'s, that is needed in the estimation.

To estimate β and λ_j, $j = 1, \ldots, m$, we solve the conditional expected version of (3.1) and (3.2) by adapting the ES algorithm of Elashoff and Ryan (2004). The ES algorithm iterates between an E-step which takes conditional expectation given the observed data, and an S-step which solves conditionally expected estimating equations:

E-step Calculate

$$e_{ij} = E[\mathbb{N}_{ij} | X_i, \{N_i(T_{i,1}), \cdots, N_i(T_{i,M_i})\}] = \frac{\lambda_j \sum_{k=1}^{M_i} I(T_{i,k-1} < s_j \leq T_{i,k}) m_{ik}}{\sum_{l=1}^{m} \sum_{k=1}^{M_i} I(T_{i,k-1} < s_l, s_j \leq T_{i,k}) \lambda_l}. \tag{3.4}$$

This essentially computes $\lambda_j m_{ik} / \sum_{l \in D_{ij}} \lambda_l$, where k is the index such that $s_j \in (T_{i,k-1}, T_{i,k}]$, and D_{ij} is the subset of the index of G containing all grid times in the same observation window $(T_{i,k-1}, T_{i,k}]$ as s_j.

S-step Replacing $\mathbb{N}_{ij}$ with e_{ij} in (3.1) and (3.2), the solution for the nonparametric component λ given β can be explicitly expressed as

$$\hat{\lambda}_j(\beta) = \frac{\sum\limits_{i=1}^{n} e_{ij} r_{ij}}{\sum\limits_{i=1}^{n} \exp(X_i'\beta) r_{ij}}, \quad j = 1, \ldots, m. \qquad (3.5)$$

Substituting (3.5) into (3.2), we get a nonlinear equation for β,

$$\sum_{i=1}^{n} \sum_{j=1}^{m} [e_{ij} - \hat{\lambda}_j(\beta) \exp(X_i'\beta)] X_i r_{ij} = 0, \qquad (3.6)$$

which can be solved for $\hat{\beta}$, using standard Newton-Raphson algorithm.

To initialize the ES algorithm, we set $\beta = 0$ and set e_{ij}'s proportional to the subinterval lengths. Then, an initial estimate of $\lambda_j(0)$, $j = 1, \ldots, m$, is obtained from (3.5) and the ES iteration continues until convergence. At convergence of $\hat{\beta}$, the baseline mean component λ_j is estimated as $\hat{\lambda}_j = \hat{\lambda}_j(\hat{\beta})$. The baseline mean function $\Lambda(t)$ is then estimated by a piecewise constant function $\hat{\Lambda}(t) = \sum_{j:s_j \leq t} \hat{\lambda}_j$. As the $\hat{\lambda}_j$'s are nonnegative, $\hat{\Lambda}(t)$ is automatically nondecreasing.

The ES algorithm encompasses the popular EM algorithm as a special case when the complete data estimating equations are the score equations. Local convergence of the ES algorithm is established with the general theory of iterative solutions to nonlinear equations (Ortega (1972)), by viewing ES as a block Newton–Gauss–Seidel algorithm or, more generally, a splitting algorithm (Elashoff and Ryan, 2004, Sec. 2.3). In our simulation studies, convergence usually occurred after a few iterations and was never an issue.

In Appendix A, we give some regularity conditions and prove the asymptotic properties of the proposed estimator. Under those conditions, $\hat{\beta}$ and $\hat{\Lambda}(\cdot)$ are consistent for the true coefficient vector β_0 and cumulative baseline mean function Λ_0, respectively. Further, the convergence rate of $\hat{\Lambda}(\cdot)$ is $n^{1/3}$, and $\sqrt{n}(\hat{\beta} - \beta_0)$ converges weakly to a normal distribution with mean zero and a variance that can be estimated.

It is practically challenging to obtain a closed-form variance estimator because of the missing $\mathbb{N}_{ij}$'s. When the event process is Poisson or mixed Poisson, an analytic sandwich variance estimator can be derived and the efficiency loss due to incompleteness can be calculated; see details in Appendix B. Because the "bread" part of the sandwich variance estimator, I_{obs}, is not enforced to be invertible for finite samples, other variance estimators are necessary in practice. A simpler alternative to the bootstrap variance estimator is to use multiple imputation (MI) as in Goetghebeur and Ryan (2000). If the exact event times

between two observation times were imputed appropriately, panel count data would reduce to recurrent event data, and many existing semiparametric regression methods for recurrent event data could then be applied to make inferences about regression coefficients and baseline mean function in a fixed time interval of interest $[0, \tau]$ (Pepe and Cai (1993); Lawless and Nadeau (1995); Lin et al. (2000)). Following Rubin and Schenker (1991), we impute event times using a working nonhomogeneous Poisson model after the convergence of the ES algorithm. Under conditional independent censoring, each imputed recurrent event data set yields a complete-data point estimate and variance estimate for β and $\Lambda(\cdot)$ from estimating equations (Lin et al. (2000)). In particular, the profile complete-data estimating equation for β in Lin et al. (2000) is the same as the score equation of Andersen and Gill (1982). This has been implemented in standard software, such as function `coxph` in `R` package `survival` (Therneau and Lumley (2009)). When a frailty variable enters the mean of event process multiplicatively, the robust variance estimate for $\hat{\beta}$ from `coxph` is still valid. A variance estimator for the ES estimator is a weighted sum of between-imputation variance and within-imputation variance with weights $1 + 1/R$ and 1, respectively, where R is the number of imputations (Tanner and Wong (1987); Schenker and Welsh (1988); Little and Rubin (2002)). In our numerical study we used $R = 50$, which seems to provide good approximation of the realized variation.

Once we have a robust variance estimator for the regression coefficient β, a standard Wald test can be applied to test hypotheses in the form $H_0 : C\beta = c$ for appropriately constructed contrast matrix C and vector c. This covers, for example, special cases of the significance test of $H_0 : \beta_j = 0$ for the jth component of β. The discussion of the significance of covariate effects in our data analysis is based on these tests.

3.2. Informative censoring

When the censoring time T_{i,M_i} and event process $N_i(\cdot)$ are dependent through an unobserved frailty Z_i after conditioning on observed covariates X_i, we no longer have $E[N_i(t)|X_i, t \leq T_{i,M_i}] = E[N_i(t)|X_i]$. Naively applying the AEE approach leads to biased estimators.

To adjust for the bias we introduce more missing values, conceptually. For each subject i with $T_{i,M_i} < \tau$, we treat the number of events in interval $(T_{i,M_i}, \tau]$ as a missing value. If this count were observed for each subject, the data would appear to be panel count data, with one extra count for those subjects with censoring time prior to τ. Further, the data would appear to have the same censoring time for all subjects and, hence, the dependence between the censoring time and the event process would have been artificially removed. This implies that, if we can impute the count between $(T_{i,M_i}, \tau]$, we reduce the problem to the

conditional independent censoring scenario, and the AEE approach in the last subsection can be applied.

Note that in the ES algorithm, it is not the full conditional distributions but the conditional expectations $E[\mathbb{N}_{ij}|X_i, \{N_i(T_{i,1}), \cdots, N_i(T_{i,M_i})\}]$, with $T_{i,M_i} \leq s_j \leq \tau$, that are involved. Model (2.1) implies that $E[N_i(T_{i,M_i})/\Lambda(T_{i,M_i}) = E[N_i(t)]/\Lambda(t)$ for any $t \in (T_{i,M_i}, \tau]$. Therefore, for subject i and $\mathbb{N}_{ij}$ with $T_{i,M_i} \leq s_j \leq \tau$, we have

$$e_{ij} = E[\mathbb{N}_{ij}|X_i, \{N_i(T_{i,1}), \cdots, N_i(T_{i,M_i})\}] = \frac{\lambda_j N_i(T_{i,M_i})}{\Lambda(T_{i,M_i})}. \tag{3.7}$$

This calculation is added to the E-step in the ES iteration.

In our study the conditional expectation at (3.7) was found to cause numerical problems when the censoring time T_{i,M_i} was too early and the count $N_i(T_{i,M_i})$ happened to be nonzero. In that case, more $\mathbb{N}_{ij}$'s were missing beyond T_{i,M_i} and all of them happen to have a larger conditional expectation, which can influence the numerical stability of the ES algorithm. In order to improve numerical stability, we propose a heuristic adjustment to the conditional expectation (3.7) with

$$e_{ij} = \frac{\lambda_j \{N_i(T_{i,M_i}) + a\}}{\Lambda(T_{i,M_i}) + a}. \tag{3.8}$$

for some small number $a > 0$. When $\Lambda(T_{i,M_i})$ is not small, e_{ij} in (3.8) and (3.7) are approximately equal; when $\Lambda(T_{i,M_i})$ is small relative to a, e_{ij} in (3.8) is bounded by $\{N_i(T_{i,M_i}) + a\}/a$.

An interpretation of a can be given in the special case of mixed nonhomogeneous Poisson event process with gamma frailty, where $N_i(T_{i,M_i})$ and $\mathbb{N}_{ij}$, $T_{i,M_i} < s_j \leq \tau$, are independent Poisson variables given frailty Z_i. If the frailty Z_i is gamma with both shape and rate parameter $a \exp(X_i'\beta)$ then, unconditioned on Z_i, $\mathbb{N}_{ij}$ is negative binomial with mean e_{ij} in (3.8) given $N_i(T_{i,M_i})$. From a Bayesian point of view, one can think of gamma as the prior distribution for frailty. If we choose a noninformative prior, a gamma distribution with shape and rate parameters approaching zero, then the conditional mean in (3.8) would approach that in (3.7). If the frailty is degenerate at one, i.e., both the shape and rate parameters approach infinity at the same speed, and e_{ij} approaches λ_j, which is the case of noninformative censoring.

With the adjusted e_{ij}, $T_{i,M_i} < s_j \leq \tau$, point estimation can be done by a slightly modified ES algorithm. In the E-step, in addition to (3.4), we compute conditional expectation e_{ij} for those subintervals $(s_{j-1}, s_j]$ that are between T_{i,M_i} and τ using (3.7). In the S-step, we can drop the at-risk indicator r_{ij} in (3.5) and (3.6) to obtain $\hat{\beta}$ and $\hat{\lambda}$, because each $r_{ij} = 1$ in $[0, \tau]$.

The tuning parameter a needs to be chosen close to zero to let the data speak, but it cannot be exactly zero to avoid numerical issues in estimation. Note that $\{N_i(T_{i,M_i}) + a\}/\{\Lambda(T_{i,M_i}) + a\}$ is a weighed average of $N_i(T_{i,M_i})/\Lambda(T_{i,M_i})$ and one. Consider the situation where subject i drops out of the study very early after one event; that is, the censoring time T_{i,M_i} is close to zero and $N_i(T_{i,M_i}) = 1$. If $\Lambda(T_{i,M_i})$ is of order $1/n$, then the conditional expectation e_{ij}, $T_{i,M_i} < s_j \leq \tau$, would be of order $n\lambda_j$ if (3.7) were used, which can cause divergence of the ES algorithm. If we choose $a = n^{-1/2}$ and use (3.8), then e_{ij} would be of the order $\sqrt{n}\lambda_j$, greatly reducing the variation in the iteration. For our simulation studies, setting $a = n^{-1/2}$ worked well and led to much better numerical stability than setting $a = 0$ for dataset containing subjects who dropped out very early.

The asymptotic properties of the estimator can be obtained similar to those under conditional independent censoring. Sandwich variance estimators can be derived accordingly. To obtain the variance estimate from MI, we propose to impute the count beyond censoring time based on the negative binomial working model that results from the nonhomogeneous Poisson event process with gamma frailty. This working model is a parametric device for MI to be used with parameters at the point estimate for variance estimation only after the ES algorithm has converged. In our simulations, it has worked well even when the frailty is not gamma.

4. Numerical Results

4.1. Simulation studies

We first consider two simulation studies, presented in Zhang (2002), under independent censoring and noninformative observation times. The focus here was the regression coefficient estimates.

- In Study 1, we took nonhomogeneous Poisson processes with independent censoring. There were three mutually independent covariates: $X_{i,1}$ was uniform over $(0,1)$, $X_{i,2}$ was $N(0,1)$, and $X_{i,3}$ was Bernoulli with success rate 0.5. The number of observation times M_i was generated uniformly over the set $\{1, 2, 3, 4, 5, 6\}$. The M_i observation times were the order statistics generated from uniform over $(1, 10)$ and rounded to the second decimal points. The event process $N_i(t)$ was Poisson with intensity $2\exp(X_i'\beta)$, where $\beta = (\beta_1, \beta_2, \beta_3)' = (-1.0, 0.5, 1.5)'$.

- In Study 2, we took mixed nonhomogeneous Poisson processes with independent censoring. The covariates and observation scheme were generated in the same fashion as in Study 1. Conditioning on α_i, the event process was a Poisson process with intensity $(2 + \alpha_i)\exp(X_i'\beta)$. The subject level frailty α_i was generated from a discrete set $\{-0.4, 0, 0.4\}$ with probabilities 0.25, 0.5,

and 0.25, respectively. This mixed Poisson process is equivalent to one with a multiplicative frailty in the mean, because the intensity can be expressed as $2Z_i \exp(X_i'\beta)$, with Z_i generated from $\{0.8, 1, 1.2\}$ with probabilities 0.25, 0.5, and 0.25. Therefore, Model (2.1) still held.

Table 1 summarizes the results from Study 1 and Study 2, with $n \in \{50, 100\}$, obtained from 1,000 replicates. We report our estimator under conditional independent censoring and the extended estimator under informative censoring, denoted by AEE and AEEX, respectively. Both estimators are virtually unbiased, but the AEE estimator appears to have smaller standard errors than AEEX estimator. This is not surprising because, although both estimators are consistent, AEEX does not use the information of conditional independent censoring and brings in more variation when making all subjects have the same censoring time. The standard errors of both AEE and AEEX decrease as sample size increases and the decreasing rate is approximately $\sqrt{n}$.

The proposed MI-based variance estimators for both AEE and AEEX seem to slightly under-estimate the true variation, especially in the Study 2 where a frailty is present. The variance under-estimation may stem from underestimation of the sandwich variance estimator for each imputed right censored data, which is not unusual for a sandwich variance estimator with small to moderate sample sizes (Mancl and DeRouen (2001)). As sample size increases, the agreement between the estimated standard errors and empirical standard deviation improves. The empirical coverage rates of the 95% confidence intervals are close to the nominal levels except for β_2. Nevertheless, as sample size increases, the coverage for β_2 gradually improves and in a study with $n = 400$, it was 92.1%.

We also report the estimator of Huang, Wang, and Zhang (2006) and the maximum pseudolikelihood estimator of Zhang (2002), denoted by HWZ and MPL, respectively. Results for HWZ were obtained from our own implementation, results for MPL were obtained from Table 1 in Zhang (2002). The two estimators are virtually unbiased as well. Their standard errors, however, appear to be higher than those of AEEX and, especially, AEE. This might be explained by the fact that HWZ only uses the cumulative count at the censoring time in estimating equations, and that MPL ignores the dependence within a subject.

Our next three simulation studies, adapted from Huang, Wang, and Zhang (2006), had informative observation times, and the latter two of them had informative censoring times. The focus was on both regression coefficients and baseline mean function.

- Study 3 was designed to have informative observation times but the censoring time was conditionally independent of the event process. The event process was Poisson with intensity $2Z_i \exp(X_i\beta)$, where covariate X_i was generated from a Bernoulli distribution with success rate 0.5, and frailty Z_i was

Table 1. Monte Carlo simulation results for Study 1 and Study 2 based on 1,000 samples. In Study 1, data were generated from Poisson processes with independent observation and censoring times. In Study 2, data were generated from mixed Poisson processes with independent observation and censoring times.

			Mean				SE				MESE		CP	
n	Param	True	AEE	AEEX	HWZ	MPL	AEE	AEEX	HWZ	MPL	AEE	AEEX	AEE	AEEX
							Study 1							
50	β_1	-1.0	-1.003	-1.000	-1.004	-0.994	0.101	0.109	0.122	0.120	0.090	0.100	0.911	0.928
	β_2	0.5	0.498	0.497	0.499	0.501	0.030	0.033	0.040	0.036	0.026	0.028	0.900	0.917
	β_3	1.5	1.500	1.494	1.500	1.502	0.069	0.077	0.086	0.083	0.067	0.071	0.912	0.921
100	β_1	-1.0	-1.002	-1.001	-1.003	-1.002	0.070	0.076	0.079	0.081	0.064	0.071	0.931	0.926
	β_2	0.5	0.500	0.499	0.500	0.500	0.019	0.020	0.021	0.024	0.019	0.020	0.934	0.944
	β_3	1.5	1.499	1.495	1.499	1.500	0.049	0.054	0.055	0.056	0.048	0.051	0.947	0.937
							Study 2							
50	β_1	-1.0	-0.995	-0.990	-0.991	-1.001	0.136	0.145	0.159	0.153	0.120	0.129	0.894	0.921
	β_2	0.5	0.501	0.499	0.501	0.498	0.042	0.043	0.044	0.049	0.034	0.037	0.864	0.878
	β_3	1.5	1.502	1.498	1.503	1.500	0.090	0.096	0.104	0.101	0.079	0.083	0.921	0.927
100	β_1	-1.0	-1.001	-0.999	-1.000	-1.002	0.099	0.103	0.111	0.106	0.089	0.094	0.909	0.935
	β_2	0.5	0.499	0.497	0.498	0.500	0.029	0.030	0.031	0.032	0.025	0.027	0.874	0.887
	β_3	1.5	1.500	1.496	1.500	1.501	0.061	0.065	0.068	0.067	0.057	0.059	0.922	0.923

AEE, augmented estimating equations estimator; AEEX, extended AEE estimator; HWZ, Huang–Wang–Zhang estimator; MPL, maximum pseudolikelihood estimator; Mean and SE, sample mean and sample standard deviation of the 1,000 estimates; MESE, square root of the mean of 1,000 variance estimates based on MI; CP, empirical coverage probability of 95% confidence intervals.

generated from a gamma distribution with both shape and rate equal to 2. The observation process was nonhomogeneous Poisson with cumulative mean $\log(1 + 2t)\exp(X_i/2)$, truncated in a fixed time interval $[0, 10]$. The number of observations in $[0, 10]$ depended on the observed covariate but not on the unobserved frailty.

- Study 4 was designed to have informative observation times and censoring time associated with the event process after conditioning on the observed co-variate. It used the same setting to generate covariate X_i and frailty Z_i as in Study 3. The event process was Poisson with intensity $2Z_i\exp(X_i\beta)$ in $[0, 10]$. If $X_i = 1$ and $Z_i > 1$, M_i was generated uniformly from $\{1, 2, \ldots, 8\}$, and $T_{i,1}, \ldots, T_{i,M_i}$ were the order statistics of M_i independent and identically distributed exponential random variables with mean 2; otherwise, M_i was generated uniformly from $\{1, 2, \ldots, 6\}$, and $T_{i,1}, \ldots, T_{i,M_i}$ were the order statistics of M_i independent and identically distributed uniform random variables on $[0, 10]$.

- Study 5 used the same setting as Study 4 except for the distribution of frailty Z_i, which was uniformly from $\{0.2, 1, 1.8\}$. This was designed to show that the proposed AEE and AEEX estimators do not rely on the gamma frailty assumption.

Table 2 summarizes the fitting results from Studies 3–5 for sample sizes $n \in \{50, 100, 200\}$ with 1,000 replicates. For all three estimators, AEE, AEEX, and HWZ, we present $\hat{\beta}$ and $\hat{\Lambda}(t)$ evaluated at $t \in 3, 7$. Although we have no asymptotic distributional results for $\hat{\Lambda}(t)$, the empirical results are still useful in assessing the estimators. In addition to the point estimate and empirical standard deviation, we also report the MI based variance estimator and the coverage probability of the 95% confidence intervals for the AEE and AEEX estimators.

In Study 3, as the observation times and censoring time were independent of the event process conditioning on the observed covariate, all three estimators were expected to be consistent. This is confirmed by negligible biases in all estimates. The AEE estimator has the smallest standard errors, which is as expected because it does not have the extra variation caused by imputing events after the censoring times. The difference between AEEX and HWZ is small, with AEEX having noticeably smaller standard errors. The MI-based variance estimator approximates the empirical variation reasonably well for the estimator of the regression parameter β. Consequently, the empirical coverage rates of the 95% confidence intervals are close to the nominal level, and the agreement improves as sample size increases.

In Study 4, since the censoring time is informative, the AEE estimator of β is biased downward by about 15%. The AEEX estimator of β is still nearly unbiased as it adjusts the complete-data estimating equations such that, artificially, all subjects have the same censoring time. The bias of the AEE estimator for $\Lambda(t)$ seems to be larger than that of the AEEX estimator, which is not surprising since the estimator of β is biased to start with. The difference between AEEX and HWZ is again small, with AEEX having slightly smaller standard errors most notably for smaller sample sizes. This suggests that even though AEEX has the potential to be more efficient than HWZ, the difference may be small.

In Study 5, the results are very similar to Study 4. The AEE estimator of β is biased, while the AEEX estimator is nearly unbiased. This study suggests that the proposed methods do not rely on the assumption of gamma frailty.

As a final illustration in Figure 1, we present graphical summaries of the AEEX estimates of β and $\Lambda(t)$ in Study 4, with sample size 100 from 1,000 replicates. The histogram of $\hat{\beta}$ suggests that the normal approximation of $\hat{\beta}$ is quite good. The mean of $\hat{\Lambda}(t)$ matches $\Lambda_0(t)$ closely. The pointwise 95% confidence intervals of $\hat{\Lambda}(t)$ were constructed from the realized 2.5th and 97.5th percentiles of the 1,000 estimates.

Table 2. Monte Carlo simulation results for Study 3–5 based on 1,000 samples. In Study 3, the event process was Poisson; observation times, censoring time, and event process were associated through observed covariates. In Study 4, the event process was mixed Poisson; observation times, censoring time, and event process were associated through both observed covariates and unobserved gamma frailties. In Study 5, the setup was the same as in Study 4, except that the frailty was a discrete variable.

n	Par	True	Mean			SE			MESE		CP	
			AEE	AEEX	HWZ	AEE	AEEX	HWZ	AEE	AEEX	AEE	AEEX
						Study 3						
50	β	-1	-1.005	-0.991	-1.014	0.120	0.156	0.173	0.117	0.158	0.928	0.943
	$\Lambda(3)$	6	5.935	5.996	6.027	0.765	0.905	1.005	0.650	0.844	0.861	0.914
	$\Lambda(7)$	14	13.819	14.175	14.187	1.271	1.843	2.070	1.154	1.882	0.889	0.942
100	β	-1	-1.004	-0.993	-1.009	0.079	0.112	0.122	0.082	0.111	0.943	0.934
	$\Lambda(3)$	6	6.000	6.044	6.067	0.522	0.635	0.695	0.471	0.605	0.892	0.934
	$\Lambda(7)$	14	13.938	14.130	14.142	0.943	1.326	1.461	0.851	1.277	0.900	0.948
200	β	-1	-1.003	-0.996	-1.006	0.058	0.083	0.084	0.057	0.080	0.928	0.935
	$\Lambda(3)$	6	6.010	6.049	6.050	0.368	0.463	0.454	0.341	0.439	0.924	0.937
	$\Lambda(7)$	14	14.022	14.187	14.172	0.660	0.967	0.979	0.606	0.910	0.904	0.939
						Study 4						
50	β	-1	-1.153	-0.979	-1.000	0.245	0.247	0.259	0.230	0.242	0.867	0.925
	$\Lambda(3)$	6	6.162	5.916	5.961	1.080	1.098	1.193	1.033	1.043	0.909	0.901
	$\Lambda(7)$	14	13.892	13.941	14.042	2.181	2.210	2.487	2.104	2.152	0.896	0.912
100	β	-1	-1.156	-0.988	-0.999	0.165	0.170	0.174	0.165	0.170	0.822	0.932
	$\Lambda(3)$	6	6.224	5.957	5.973	0.793	0.794	0.815	0.750	0.752	0.925	0.909
	$\Lambda(7)$	14	14.036	14.016	14.043	1.587	1.594	1.607	1.547	1.558	0.930	0.927
200	β	-1	-1.156	-0.992	-1.000	0.118	0.124	0.126	0.117	0.121	0.717	0.941
	$\Lambda(3)$	6	6.259	5.995	6.003	0.551	0.561	0.571	0.528	0.536	0.928	0.924
	$\Lambda(7)$	14	14.084	14.016	14.024	1.157	1.135	1.141	1.090	1.100	0.924	0.925
						Study 5						
50	β	-1	-1.151	-0.982	-0.998	0.236	0.234	0.241	0.231	0.235	0.881	0.944
	$\Lambda(3)$	6	6.175	5.914	5.945	1.084	1.066	1.088	1.006	1.011	0.913	0.918
	$\Lambda(7)$	14	14.008	14.028	14.075	2.111	2.071	2.103	2.050	2.076	0.922	0.935
100	β	-1	-1.146	-0.986	-0.997	0.166	0.166	0.170	0.161	0.161	0.832	0.942
	$\Lambda(3)$	6	6.207	5.948	5.968	0.765	0.778	0.799	0.714	0.719	0.914	0.913
	$\Lambda(7)$	14	14.031	13.995	14.038	1.550	1.522	1.567	1.457	1.455	0.911	0.934
200	β	-1	-1.147	-0.993	-1.000	0.112	0.112	0.115	0.113	0.115	0.744	0.943
	$\Lambda(3)$	6	6.217	5.963	5.976	0.520	0.529	0.534	0.509	0.513	0.930	0.927
	$\Lambda(7)$	14	14.078	14.027	14.041	1.090	1.077	1.079	1.043	1.038	0.933	0.935

AEE, augmented estimating equations estimator; AEEX, extended AEE estimator; HWZ, Huang–Wang–Zhang estimator; Mean and SE, sample mean and sample standard deviation of the 1,000 estimates; MESE, square root of the mean of 1,000 variance estimates based on MI; CP, empirical coverage probability of 95% confidence intervals.

4.2. Illustration with bladder tumor data

We applied our AEE method to the bladder tumor data (Byar (1980)) and compared with existing analyses in the literature. In the original study, patients
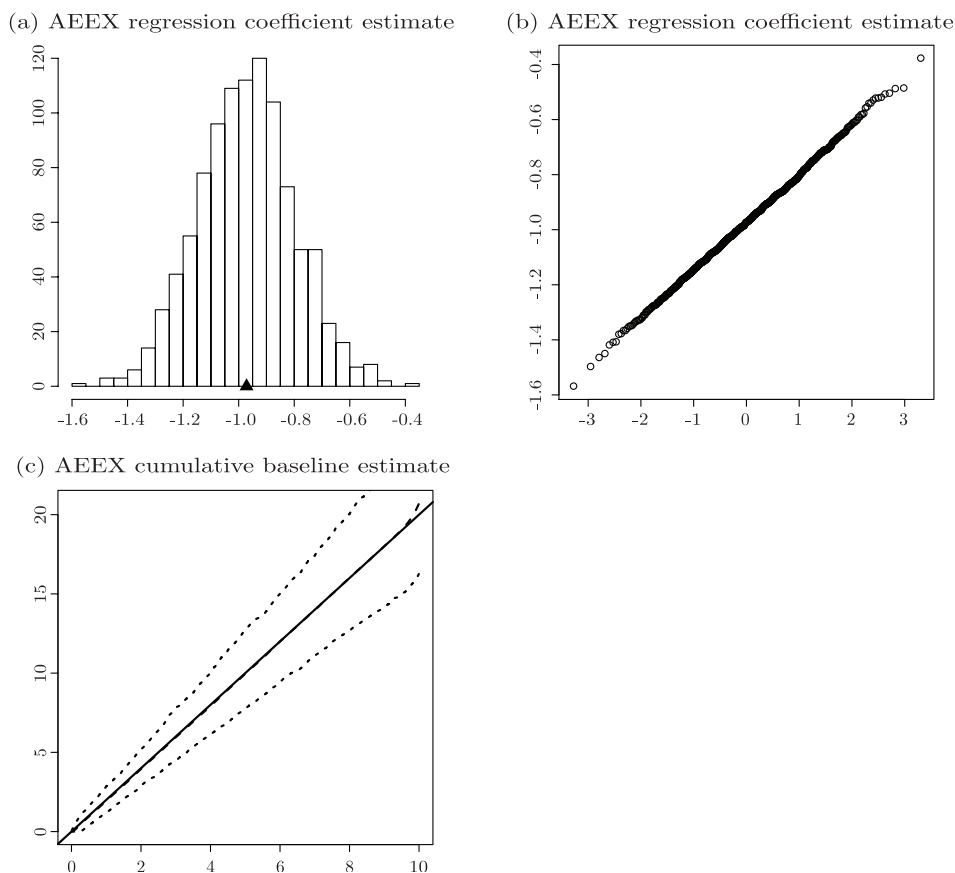
(a) AEEX regression coefficient estimate

(b) AEEX regression coefficient estimate

(c) AEEX cumulative baseline estimate

Figure 1. Graphical summaries of AEEX estimates $\hat{\beta}$ and $\hat{\Lambda}(t)$ in Study 4 for AEEX with sample size 100 from 1,000 replicates. (a) Histogram of $\hat{\beta}$; the true value β_0 is 1. (b) Normal Q-Q plot of $\hat{\beta}$; the true value β_0 is 1. (c) Mean of $\hat{\Lambda}(t)$ (dashed) and pointwise 95% confidence interval constructed from the realized 2.5th and 97.5th percentiles of the 1,000 estimates (dotted), overlaid with the true baseline $\Lambda(t) = 2t$ (solid).

who had superficial bladder tumors were randomized into three treatment arms: placebo, pyridoxine pills, and thiotepa. At each follow-up visit, tumors were counted, measured, and then removed transurethrally, and the treatment was continued. The objective of the study was to determine if the treatment reduces the recurrence of bladder tumor.

For comparison with existing analyses, we only analyzed data from two treatment arms, placebo and thiotepa, a total of 85 patients. Three covariates are available: initial number of tumors, size of the largest initial tumor, and treatment (1 = thiotepa). We fit a semiparametric mean regression Model (2.1) and calculated AEE under conditional independent censoring and AEEX under in-

Table 3. Regression coefficients estimates and standard errors from bladder tumor data analyses.

	AEE	AEEX	HWZ	MPL	SW[a]	SW[b]
$\hat{\beta}_1$	0.257	0.273	0.284	0.283	0.662	0.660
$\text{SE}(\hat{\beta}_1)$	0.071	0.083	0.101	0.083	0.213	0.225
$\hat{\beta}_2$	-0.028	0.030	0.033	-0.051	-0.130	-0.123
$\text{SE}(\hat{\beta}_2)$	0.097	0.119	0.144	0.101	0.203	0.204
$\hat{\beta}_3$	-0.789	-0.609	-0.602	-1.357	-2.025	-1.971
$\text{SE}(\hat{\beta}_3)$	0.303	0.376	0.379	0.369	0.450	0.442

AEE, augmented estimating equations estimator; AEEX, extended AEE estimator; HWZ, Huang–Wang–Zhang estimator; SW[a] and SW[b], Sun–Wei estimator with and without modeling the observation pattern; MPL, maximum pseudolikelihood estimator.

formative censoring, with β_1, β_2, and β_3 the regression coefficients for initial tumor number, largest initial tumor size, and treatment indicator, respectively. Estimates of regression coefficients and their standard errors are summarized in Table 3. Both AEE and AEEX results suggest that largest initial tumor size has no significant effect on tumor recurrence, initial tumor count implies higher risk of tumor recurrence, and the thiotepa treatment seems to reduce tumor recurrence. The close agreement between AEE estimates and AEEX estimates for all three regression coefficients may be an indication that censoring time and event process are conditionally independent. This is in line with the finding in Sun and Wei (2000).

Also in Table 3 are estimates from existing methods. The HWZ estimates were from our implementation of Huang, Wang, and Zhang (2006), the MPL and SW estimates were obtained from Sun (2006). Point estimates from AEE and AEEX are very close to those from HWZ. All these methods give a smaller thiotepa treatment effect in reducing tumor recurrence than MPL estimates and SW estimates. A similar finding was reported by Huang, Wang, and Zhang (2006), in an analysis with only treatment indicator in the regression model. The HWZ estimates seem to have slightly higher standard errors than the AEE estimates and AEEX estimates. The much larger magnitude of SW estimates in comparison with other estimates might be explained by its requirement of correct specification of the models for observation times and censoring time.

5. Discussion

Statistical inferences for semiparametric panel count regression is a challenging problem because of unobserved event times, informative observation times, and informative censoring time. Our approach uses robust working models to

impute event times up to a time grid and solves the conditionally expected version of complete-data estimating equations to estimate regression coefficients and the cumulative baseline mean function. The estimators, AEE under conditional independent censoring and AEEX under informative censoring, are applicable regardless of whether observation times are informative or not given covariates. Similar to Huang, Wang, and Zhang (2006), we do not need to model for observation times and censoring time, correct specifications of which are necessary to obtain consistent estimator in Sun and Wei (2000). We allow observation times and censoring time to be associated with the event process through an unobserved multiplicative frailty in the mean function, retaining the advantage of no need to model the frailty as in Huang, Wang, and Zhang (2006). When there is no covariate, our AEE resembles the first estimating equation of Hu, Lagakos, and Lockhart (2009b).

For informative censoring via frailty, our AEEX estimator does not require any different assumptions than the HWZ estimator. Nevertheless, our extended AEEs for regression coefficients use all imputed event times and, hence, have the potential for being more efficient than the estimating equations in Huang, Wang, and Zhang (2006). In fact, this was suggested by Wang, Qin, and Chiang (2001) for exploration in a recurrent event setting with informative censoring. When conditionally independent censoring is safe to assume, for instance, in situations where censoring times are administrative, AEE, AEEX, and HWZ are all valid, but AEE may be preferred for higher efficiency, as illustrated in simulation studies.

A comparison of the proposed methods and existing methods helps in understanding their differences and relative advantages. The MPL and ML estimators assume the the event processes are nonhomogeneous Poisson and are conditionally independent of the observation schemes and the censoring times given covariates. The MPL estimator trades efficiency for computing convenience. The SW, HWZ, AEE, and AEEX estimators are based on estimating equations without assuming that the event process is Poisson. The SW approach allows conditionally independent observation schemes and censoring times given covariates, but requires model specification for observation schemes and censoring times. The HWZ, AEE, and AEEX allow informative observation schemes. The HWZ and AEEX further allow informative censoring time through a multiplicative frailty with unspecified distribution. Most of these methods are implemented in the R package `spef` (Wang and Yan (2010)) available at the Comprehensive R Archive Network (`http://CRAN.R-project.org`). A companion paper of the package provides a comparative summary of their assumptions on the observation scheme and the censoring time (Wang and Yan, 2011, Table 1), in addition to illustrations of usages.

In our study, our main interest lies in the estimation and inference of the parametric parameter. The nonparametric parameter, although potentially of importance, is usually less interesting. For the nonparametric parameter, the convergence rate is slower than $\sqrt{n}$ and the asymptotic distribution is not normal. We suspect that a complex bootstrap procedure is needed for inference of the nonparametric parameter. We postpone this investigation to a future study.

Acknowledgements

Appendix A: Asymptotic Results

The cumulative baseline mean function Λ is estimable only on $\{s_1 \ldots s_m\}$. For uniqueness, we further specify that $\hat{\Lambda}$ is right-continuous, piecewise constant, and with possible discontinuities only at $\{s_1, \ldots, s_m\}$.

In what follows, we establish the asymptotic properties under conditional independent censoring. Under informative censoring, similar properties hold and the proofs are only slightly different. We first set out some assumptions.

A1. The observation times and censoring are conditionally independent of the event process given observed covariates.

A2. (a) The observation times are η-separable. That is, there exists $\eta > 0$ such that $\Pr(T_{i,j} - T_{i,j-1} > \eta) = 1$. (b) If $\mathcal{S}$ is the support of $\cup_{i,j} T_{i,j}$, there exist finite τ_0 and τ_1 such that $\tau_0 < \min \mathcal{S} < \max \mathcal{S} < \tau_\infty$. (c) There exists $\tau_0^* < \tau_1^*$ such that $[\tau_0^*, \tau_1^*] \subset \mathcal{S}$.

A3. The distribution of X is not concentrated on any proper subspace of $\mathbb{R}^p$; the support of the distribution is a compact subset of $\mathbb{R}^p$.

A4. (a) β_0 belongs to a compact subset of $\mathbb{R}^p$. (b) There exists κ_1 such that $\Lambda_0 < \kappa_1 < \infty$; Λ_0 is first-order differentiable, and there exists κ_2 such that $1/\kappa_2 < \Lambda_0^{(1)} < \kappa_2 < \infty$.

These assumptions are mild; comparable assumptions have been made in Wellner and Zhang (2007), Lu, Zhang, and Huang (2007) and other publications. Under (A2), any two observation times are separated by at least η; this usually holds in biomedical studies where continuous monitoring is unlikely. We note that this assumption rules out scenarios with accurately observable event times,

under which the estimate of Λ_0 is $n^{1/2}$ consistent. In our study, however, we can only observe the intervals of the event times. Because of this excessive loss of information, the estimate of Λ_0 is only $n^{1/3}$ consistent, as shown below in Theorem 2. A byproduct of (A2) is that M_i is bounded. Although it might be possible to relax the assumptions to allow $M_i \to \infty$ but $E(M_i) < \infty$, such a difference has a negligible impact in practice. The compactness assumptions (A3) and (A4) usually hold in practice; they are made for theoretical purposes and the real bounds may remain unknown.

For any $\beta_1 \in \mathbb{R}^p$ and $\beta_2 \in \mathbb{R}^p$, $\|\beta_1 - \beta_2\|_2$ is the L_2 norm of $\beta_1 - \beta_2$. For any nondecreasing function Λ_1 and Λ_2, let $m_\Delta(\Lambda_1, \Lambda_2) = \left[\int (\Delta\Lambda_1 - \Delta\Lambda_2)^2 d\mu\right]^{1/2}$, where $\Delta\Lambda = \Lambda(s_k) - \Lambda(s_j)$ for any $s_k \in G$, $s_j \in G$, and $s_k < s_j$. The consistency result can be summarized as follows.

Theorem A.1. Under (A1)–(A4), $\hat{\beta} \to_{a.s.} \beta_0$ and $m_\Delta(\hat{\Lambda}, \Lambda_0) = o_P(1)$.

Proof. First we note that under (A2c) and the compactness assumptions, the model we consider is identifiable. The proof follows from Wellner and Zhang (2007).

For a randomly selected subject, let $\mathbb{N}_{\cdot j}$, $j = 1, 2, \ldots$, be the missing counts, $r_{\cdot j}$ be the at-risk indicator, and X be the covariate vector. Consider an objective function

$$l(\beta, \Lambda) = \sum_j [\mathbb{N}_{\cdot j} \log(\lambda_j) + \mathbb{N}_{\cdot j} X'\beta - \lambda_j \exp(X'\beta)] r_{\cdot j}.$$

Take $\mathbb{P}_n$ as the empirical measure. If $\Delta\hat{\Lambda} \to 0$ or ∞ and/or $\hat{\beta} \to \infty$, $\mathbb{P}_n l(\hat{\beta}, \hat{\Lambda}) \to \infty$. Thus, we are able to focus on the set of bounded $(\hat{\beta}, \hat{\Lambda})$.

The functional set $\{\log(\lambda)\}$ is bounded and the functional set $\{\Lambda\}$ is monotone and bounded. They are compact with respect to the vague topology. Consistency thus follows from Theorem 5.14 of Van der Vaart (1998).

Theorem A.1 establishes the consistency of $\hat{\Lambda}$ in the L_2 sense. Consistency under other norms (for example the uniform consistency) that demands different, possibly stronger assumptions, is not pursued in this study.

To establish the convergence rate, we need an additional assumption.

A5. For (β, Λ) satisfying (A1)–(A4), $E[l(\beta, \Lambda) - l(\beta_0, \Lambda_0)] \leq -\kappa_3(\|\beta - \beta_0\|_2^2 + m_\Delta^2(\Lambda, \Lambda_0))$, where κ_3 is a fixed positive constant.

Here we assume that the maximizer of l is "well separated". This can be verified under the compactness conditions and differentiability of the objective function l.

We first insert the definition of bracketing number; see Van der Vaart and Wellner (1996) for more detailed descriptions. Let $(\mathbb{F}, \|\cdot\|)$ be a subset of a

normed space of real functions on some set. Given functions f_1 and f_2, the bracket $[f_1, f_2]$ is the set of all functions f with $f_1 \leq f \leq f_2$. An ϵ-bracket is a bracket $[f_1, f_2]$ with $\|f_1 - f_2\| \leq \epsilon$. The bracketing number $N_{[]}(\epsilon, \mathbb{F}, \|\cdot\|)$ is the minimum number of ϵ brackets needed to cover $\mathbb{F}$. The entropy with bracketing is the logarithm of the bracketing number.

Theorem A.2. Under (A1)–(A5), $\|\hat{\beta} - \beta_0\|_2^2 + m_\Delta^2(\hat{\Lambda}, \Lambda_0) = O_P(n^{-2/3})$.

Proof. Lemma 25.84 of Van der Vaart (1998) shows that, if (A4) is satisfied, there exists a constant κ_4 such that for every $\epsilon > 0$, $\log N_{[]}(\epsilon, \{\Lambda\}, L_2) \leq \kappa_4(\frac{1}{\epsilon})$. Since the objective function l is Hellinger differentiable, under (A2) and (A3), we have $\log N_{[]}(\epsilon, \{l(\beta, \Lambda)\}, L_2) \leq \kappa_5(\frac{1}{\epsilon})$ for a constant κ_5.

By Theorem 3.2.5 of Van der Vaart and Wellner (1996), for (β, Λ) satisfying $\|\beta - \beta_0\|_2^2 + +m_\Delta^2(\Lambda, \Lambda_0) < \xi$, we have

$$\mathbb{P}^* \sup |\sqrt{n}(\mathbb{P}_n - E)\left(l(\beta, \Lambda) - l(\beta_0, \Lambda_0)\right)| = O_p(1)\xi^{1/2}\left(1 + \frac{\xi^{1/2}}{\xi^2\sqrt{n}}\kappa_6\right) \text{ (A.1)}$$

with a constant κ_6, where $\mathbb{P}^*$ is the outer expectation. According to Theorem 3.2.1 of Van der Vaart and Wellner (1996), (A.1) and (A5) imply the desired result.

With semiparametric models, $\sqrt{n}$ consistency and asymptotic normality of estimates of parametric parameters requires non-singularity of the information matrix. We compute the information matrix for β as follows.

The score function for β is $\dot{l}_\beta = \sum_j [\mathbb{N}_{\cdot j} X - \lambda_j \exp(X'\beta)X]r_{\cdot j}$.

Consider a small perturbation of λ defined by $\lambda_a = \lambda + ah$ with $a \sim 0$ and $h \in L_2(P)$ such that $\Lambda_a = \int \lambda_a$ satisfies (A4). We can see that $h = \frac{\partial \lambda_a}{\partial a}|_{a=0}$. Thus, the score operator for λ is $\dot{l}_\lambda[h] = \sum_j [\mathbb{N}_{\cdot j}/\lambda_j - \exp(X'\beta)] r_{\cdot j}h$.

Computing the information matrix amounts to finding h^* such that, for any h defined above, $E((\dot{l}_\beta - \dot{l}_\lambda[h^*])\dot{l}_\lambda[h]) = 0$. It can be shown that one solution is

$$h^* = \frac{E\left(\dot{l}_\beta \sum_j \left[\frac{\mathbb{N}_{\cdot j}}{\lambda_j} - \exp(X'\beta)\right] r_{\cdot j}|t\right)}{E\left(\left\{\sum_j \left[\frac{\mathbb{N}_{\cdot j}}{\lambda_j} - \exp(X'\beta)\right] r_{\cdot j}\right\}^2 |t\right)}.$$

The information matrix is $E(\dot{l}_\beta - \dot{l}_\lambda[h^*])^{\otimes 2}$, which is assumed to be positive definite and component-wise bounded.

We now further establish that, despite the slow convergence rate of the estimate of Λ_0, the estimate of β_0 is still $\sqrt{n}$ consistent and asymptotically normal.

Theorem A.3. Under (A1)–(A5), $\sqrt{n}(\hat{\beta} - \beta_0) \rightarrow N(0, E^{-1}(\dot{l}_\beta - \dot{l}_\lambda[h^*])^{\otimes 2})$.

Proof. We list some relevant facts.

1. (Maximization of the objective function)
   $\mathbb{P}_n \dot{l}_\beta(\hat{\beta}, \hat{\Lambda}) = 0$ component-wise, and $\mathbb{P}_n \dot{l}_\lambda[h]|_{\beta=\hat{\beta}, \Lambda=\hat{\Lambda}} = 0$ for h defined above.

2. (Rate of convergence) $\|\hat{\beta} - \beta_0\|_2^2 + m_\Delta^2(\hat{\Lambda}, \Lambda_0) = O_P(n^{-2/3})$.

3. (Positive Information) The Fisher Information matrix is positive definite and component-wise bounded.

4. (Stochastic equicontinuity) For any $\delta_n \to 0$ and constant $\kappa_7 > 0$, within the neighborhood $\{\|\beta - \beta_0\| < \delta_n, m_\Delta(\Lambda, \Lambda_0) < \kappa_7 n^{-1/3}\}$,

$$\sup \sqrt{n} |(\mathbb{P}_n - E)(\dot{l}_\beta(\beta, \Lambda) - \dot{l}_\beta(\beta_0, \Lambda_0))| = o_P(1),$$
$$\sup \sqrt{n} |(\mathbb{P}_n - E)(\dot{l}_\lambda[h^*]|_{\beta, \Lambda} - \dot{l}_\lambda[h^*]|_{\beta_0, \Lambda_0})| = o_P(1).$$

   These equations can be established by applying Theorem 3.2.5 of Van der Vaart and Wellner (1996), and the entropy result.

5. (Smoothness of the model) Within the neighborhood $\{\|\beta - \beta_0\| < \delta_n, m_\Delta(\Lambda, \Lambda_0) < \kappa_7 n^{-1/3}\}$, the expectations of $\dot{l}_\beta$ and $\dot{l}_\lambda$ are Hellinger differentiable.

With these in hand, the desired result can be proved using Theorem 3.4 of Huang (1996).

Appendix B: Variance Estimation Under Poisson Assumption

Although the methods do not rely on the Poisson assumption of the event process, a simple closed-form variance estimator is easier to derive and may give some insights on efficiency loss for the parametric parameter estimation if we do assume it. For a more lucid view, we conduct the computation with finite sample data. When the sample size is finite, the number of distinct observation/censoring times is finite. We are hence able to treat the semiparametric model in a parametric manner. When the sample size converges to infinite, the computation below matches that in Appendix A. Specifically, the matrix inversion/projection "converges" to the functional projection of score function in Appendix A. We also note that this argument is only valid for the parametric parameter. For the nonparametric parameter, the variance matrix computed below has infinite dimension as $n \to \infty$. Thus, we are not able to make inference for the estimate of the nonparametric parameter based on this calculation.

From a computational point of view, the unknown parameter is an $(m+p) \times 1$ vector $\theta = (\lambda_1, \ldots, \lambda_m, \beta')'$. We stack equations (3.1) and (3.2) as $U(\theta) = \sum_{i=1}^n U_i(\theta) = 0$, where U is an $(m+p) \times 1$ vector with the first m components corresponding to (3.1) and the other p components corresponding to (3.2). This estimating equation is suggested by the fact that the conditional expectation of

U given observed data D_{obs} is zero. We first derive the observed information matrix of θ, and then give a sandwich variance estimator of β.

Differentiating $-U$ with respect to θ' and taking conditional expectation given D_{obs} yields the complete information matrix of θ. Taking the conditional expectation of the outer product of U given D_{obs} gives the missing information matrix of θ. The difference between the two information matrices leads to the observed information matrix:

$$I_{\text{obs}} = I_{\text{com}} - I_{\text{mis}} = -E\left[\frac{\partial U}{\partial \theta'}|D_{\text{obs}}\right] - E[UU'|D_{\text{obs}}]. \tag{A.2}$$

The complete information matrix can be written as:

$$I_{\text{com}} = \sum_{i=1}^{m} \begin{pmatrix} \exp(X_i'\beta)r_{i1} & \cdots & 0 & \lambda_1 \exp(X_i'\beta)X_i'r_{i1} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \exp(X_i'\beta)r_{im} & \lambda_m \exp(X_i'\beta)X_i'r_{im} \\ \exp(X_i'\beta)X_ir_{i1} & \cdots \exp(X_i'\beta)X_ir_{im} & \sum_{j=1}^{m} \lambda_j \exp(X_i'\beta)X_iX_i'r_{ij} \end{pmatrix},$$

where the upper left block is a $m \times m$ diagonal matrix.

We denote the missing information matrix as

$$I_{\text{mis}} = \sum_{i=1}^{n} \begin{pmatrix} B_{11,i} & B_{12,i} \\ B'_{12,i} & B_{22,i} \end{pmatrix},$$

where the upper-left $m \times m$ block $B_{11,i}$ has (j,k)th entry $B_{11,i}(j,k) = \text{Cov}(\mathbb{N}_{ij}, \mathbb{N}_{ik}|D_{\text{obs}})r_{ij}r_{ik}$, the upper-right $m \times p$ $B_{12,i}$ block has jth row $B_{12,i}(j,) = \text{Cov}(\mathbb{N}_{ij}r_{ij}, \sum_{j=1}^{m}\mathbb{N}_{ij}r_{ij}|D_{\text{obs}})X_i'$, and the lower-right $p \times p$ block $B_{22,i} = \text{Var}[\sum_{j=1}^{m}\mathbb{N}_{ij}r_{ij}|D_{\text{obs}}]X_iX_i'$. For either AEE or AEEX estimator, these covariance and variance terms are then computed based on the respective working models.

As $\hat{\theta}$ is consistent for θ, the covariance matrix of $\hat{\beta}$ is estimated by a sandwich estimator, which is the lower-right $p \times p$ block of

$$V = I_{\text{obs}}^{-1}(\hat{\theta})\left(\sum_{i=1}^{n} E[U_i(\hat{\theta})|D_{\text{obs}}]E[U_i(\hat{\theta})|D_{\text{obs}}]'\right) I_{\text{obs}}^{-1}(\hat{\theta}).$$

References

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study (Com: p1121-1124). *Ann. Statist.* **10**, 1100-1120.

Byar, D. P. (1980). The veterans administration study of chemoprophylaxis for recurrent stage I bladder tumors: Comparison of placebo, pyridoxine, and total thiotepa. In *Bladder Tumors and Other Topics in Urological Oncology* (Edited by M. Pavone-Macaluso, P. H. Smith, and F. Edsmyn), 363-370. Plenum, New York.

Cook, R. J. and Lawless, J. F. (2007). *The Statistical Analysis of Recurrent Events*. Springer.

Elashoff, M. and Ryan, L. (2004). An EM algorithm for estimating equations. *J. Comput. Graph. Statist.* **13**, 48-65.

Goetghebeur, E. and Ryan, L. (2000). Semiparametric regression analysis of interval-censored data. *Biometrics* **56**, 1139-1144.

Heyde, C. C. and Morton, R. (1996). Quasi-likelihood and generalizing the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **58**, 317-327.

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

Hu, X. J., Lagakos, S. W. and Lockhart, R. A. (2009b). Marginal analysis of panel counts through estimating functions. *Biometrika* **96**, 445-456.

Huang, C., Wang, M. and Zhang, Y. (2006). Analysing panel count data with informative observation times. *Biometrika* **93**, 763-776.

Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.* **24**, 540-568.

Lawless, J. F. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* **37**, 158-168.

Lin, D. Y., Wei, L. J., Yang, I. and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J. Roy. Statist. Soc. Ser. B: Statistical Methodology* **62**, 711-730.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. Wiley.

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 data using monotone B-Splines. *J. Amer. Statist. Assoc.* **104**, 1060-1070.

Mancl, L. A. and DeRouen, T. A. (2001). A covariance estimator for GEE with improved small-sample properties. *Biometrics* **57**, 126-134.

Ortega, J. M. (1972). *Numerical Analysis: A Second Course*. Academic Publisher, New York.

Pan, W. (2000). A multiple imputation approach to Cox regression with interval-censored data. *Biometrics* **56**, 199-203.

Pepe, M. S. and Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *J. Amer. Statist. Assoc.* **88**, 811-820.

Rubin, D. B. and Schenker, N. (1991). Multiple imputation in health-care databases: An overview and some applications. *Statist. Medicine* **10**, 585-598.

Schenker, N. and Welsh, A. H. (1988). Asymptotic results for multiple imputation. *Ann. Statist.* **16**, 1550-1566.

Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data*. Springer.

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

Sun, J. and Wei, L. J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *J. Roy. Statist. Soc. Ser. B: Statistical Methodology* **62**, 293-302.

Tanner, M. and Wong, W. (1987). An application of imputation to an estimation problem in grouped lifetime analysis. *Technometrics* **29**, 23-32.

Therneau, T. and Lumley, T. (2009). *survival: Survival analysis, including penalised likelihood.* R package version 2.35-5.

Van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New York.

Wang, M.-C., Qin, J. and Chiang, C.-T. (2001). Analyzing recurrent event data with informative censoring. *J. Amer. Statist. Assoc.* **96**, 1057-1065.

Wang, X. and Yan, J. (2010). *spef: Semiparametric Estimating Functions.* R package version 0.1-7.

Wang, X. and Yan, J. (2011). Fitting semiparametric regressions for panel count survival data with an R package spef. *Computer Methods and Programs in Biomedicine* **104**, 278-285.

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

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

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

Google New York, 76 Ninth Avenue, 4th Floor, New York, New York 10011, U.S.A.

E-mail: wangxj03@gmail.com

Division of Biostatistics, School of Public Health, Yale University, 60 College Street, P.O. Box 208034, New Haven, Connecticut 06520-8034, U.S.A.

E-mail: shuangge.ma@yale.edu

Department of Statistics, University of Connecticut, 215 Glenbrook Rd. U-4120, Storrs, Connecticut 06269-4120, U.S.A.

E-mail: jun.yan@uconn.edu