A Semiparametric Regression Model for Panel Count Data: When Do Pseudo-likelihood Estimators Become Badly Inefficient?

Jon A. Wellner[1], Ying Zhang[2], and Hao Liu[3]

[1] Statistics, University of Washington jaw@stat.washington.edu
[2] Statistics, University of Central Florida zhang@pegasus.cc.ucf.edu
[3] Biostatistics, University of Washington hliu@u.washington.edu

ABSTRACT. We consider estimation in a particular semiparametric regression model for the mean of a counting process under the assumption of "panel count" data. The basic model assumption is that the conditional mean function of the counting process is of the form $E\{\mathbb{N}(t)|Z\} = \exp(\theta' Z)\Lambda(t)$ where Z is a vector of covariates and Λ is the baseline mean function. The "panel count" observation scheme involves observation of the counting process $\mathbb{N}$ for an individual at a random number K of random time points; both the number and the locations of these time points may differ across individuals.

We study maximum pseudo-likelihood and maximum likelihood estimators $\widehat{\theta}_n^{ps}$ and $\widehat{\theta}_n$ of the regression parameter θ. The pseudo-likelihood estimators are fairly easy to compute, while the full maximum likelihood estimators pose more challenges from the computational perspective. We derive expressions for the asymptotic variances of both estimators under the proportional mean model. Our primary aim is to understand when the pseudo-likelihood estimators have very low efficiency relative to the full maximum likelihood estimators. The upshot is that the pseudo-likelihood estimators can have arbitrarily small efficiency relative to the full maximum likelihood estimators when the distribution of K, the number of observation time points per individual, is very heavy-tailed.

1 Introduction

Our goal in this paper is to study efficiency aspects of two types of estimators for a particular semiparametric model for *panel count data*. Panel count data arise in many fields including demographic studies, industrial reliability, and clinical trials; see for example [19], [7], [31], [32], [29], and [33] where the estimation of either the intensity of event recurrence or the mean function of a counting process with panel count data was studied. Many applications involve covariates whose effects on the underlying counting process are of interest. While there is considerable work on regression modeling for recurrent events based on continuous observations (see, for example [22], [5], and [23]), regression analysis with panel count data for counting processes has just started recently. [30] proposed estimating equation methods, while [35] and [36] proposed a pseudo-likelihood method for studying the multiplicative mean model (1) with panel count data.

Here is a description of the model and the observation scheme. Suppose that $\mathbb{N} = \{\mathbb{N}(t) : t \geq 0\}$ is a univariate counting process. In many applications, it is important to estimate the expected number of events $E\{\mathbb{N}(t)|Z\}$ which will occur by the time t, conditionally on a covariate vector Z.

In this paper we consider the proportional mean regression model given by

$$\Lambda(t|Z) \equiv E\{\mathbb{N}(t)|Z\} = e^{\theta' Z}\Lambda(t)\,, \tag{1}$$

where the monotone increasing function Λ is the *baseline mean function*. The parameters of primary interest are θ and Λ.

The observation scheme we want to study is as follows: suppose that we observe the counting process $\mathbb{N}$ at a random number K of random times

$$0 \equiv T_{K,0} < T_{K,1} < \cdots < T_{K,K}.$$

We write $\underline{T}_K \equiv (T_{K,1}, \ldots, T_{K,K})$, and we assume that $(K, \underline{T}_K|Z) \sim G(\cdot|Z)$ is conditionally independent of the counting process $\mathbb{N}$ given the covariate vector Z. We further assume that $Z \sim H$ on R^d, but we will make no further assumptions about G or H (modulo mild integrability and boundedness requirements).

The data for each individual will consist of

$$X = (Z, K, \underline{T}_K, \mathbb{N}(T_{K,1}), \ldots, \mathbb{N}(T_{K,K})) \equiv (Z, K, \underline{T}_K, \underline{\mathbb{N}}_K)\,. \tag{2}$$

We will assume that the data consist of $X_1, \ldots, X_n$ i.i.d. as X.

To derive useful estimators for this model we will often assume, in addition to (1), that the counting process $\mathbb{N}$, conditionally on Z, is a non-homogeneous Poisson process. But our general perspective will be to study the estimators and other procedures when the Poisson assumption *fails to hold* and we assume *only* that the proportional mean assumption (1) holds.

Such a program was carried out for estimation of Λ without any covariates for this panel count observation model by [33]. Briefly, [33] studied both the maximum likelihood estimator $\widehat{\Lambda}_n$ and the pseudo-maximum likelihood estimator $\widehat{\Lambda}_n^{ps}$ of Λ. They showed that both estimators are consistent in $L_2(\mu)$ where μ is the measure defined (in terms of the observation process) by

$$\mu(B) = \sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k} P(T_{k,j} \in B | K = k) \,.$$

[33] also succeeded in showing (under additional smoothness and boundedness assumptions) that

$$n^{1/3}(\widehat{\Lambda}_n^{ps}(t_0) - \Lambda_0(t_0)) \to_d \left\{ \frac{\sigma^2(t_0)\Lambda_0'(t_0)}{2G'(t_0)} \right\}^{1/3} 2\mathbb{Z}$$

where $\sigma^2(t) = Var[\mathbb{N}(t)]$, $G'(t) = \sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k} G_{k,j}'(t)$, and $\mathbb{Z} = \text{argmax}\{W(t) - t^2\}$ for a two-sided Brownian motion process W starting at 0. They also proved a corresponding result for a "toy estimator" version of the maximum likelihood estimator $\widehat{\Lambda}_n$ under the Poisson process assumption, and made efficiency comparisons between the two estimators based on Monte-Carlo studies.

The outline of the rest of the present paper is as follows: In section 2, we describe two methods of estimation, namely *maximum pseudo-likelihood estimators* and *maximum likelihood estimators* of θ and Λ. The basic picture is that the pseudo-likelihood estimators are computationally relatively straightforward and easy to implement, while the (full, semiparametric) maximum likelihood estimators are considerably more difficult, requiring an iterative algorithm in the computation of the profile likelihood. For other examples of the use of pseudo-likelihood to obtain computationally simple methods, see e.g. [3] and [27].

In section 3 we present information calculations for the semiparametric model described by the proportional mean function assumption (1) together with the non-homogeneous Poisson process assumption on $\mathbb{N}$. This provides a baseline for comparisons of variances with the best possible asymptotic variance under the Poisson and proportional mean model assumptions. In section 4 we describe asymptotic normality results for the pseudo-likelihood and full maximum likelihood estimators $\widehat{\theta}_n^{ps}$ and $\widehat{\theta}_n$ of θ assuming only the proportional mean structure (1), but *not assuming* that $\mathbb{N}$ is a Poisson process. Proofs of these results will be presented in detail in [34]. Finally, in section 4 we compare the pseudo-likelihood and full likelihood estimators of θ under three different scenarios with the goal of determining situations under which the pseudo-likelihood estimators will lose considerable efficiency relative to the full maximum likelihood estimators.

As will be seen, the rough upshot of the calculations here is that the efficiency of the pseudo-likelihood estimators relative to the full maximum likelihood estimators can be low when the distribution of K, the number of observation times per subject, is heavy-tailed.

2 Two Methods of Estimation

Maximum Pseudo-likelihood Estimation: The natural pseudo-likelihood estimators for this model use the marginal distributions of $\mathbb{N}$, conditional on Z,

$$P(\mathbb{N}(t) = k|Z) = \frac{\Lambda(t|Z)^k}{k!} \exp(-\Lambda(t|Z))$$

and *ignore dependence* between $\mathbb{N}(t_1)$, $\mathbb{N}(t_2)$ to obtain the *pseudo-likelihood*:

$$l_n^{ps}(\theta, \Lambda) = \sum_{i=1}^{n} \sum_{j=1}^{K_i} \Big\{ \mathbb{N}^{(i)}(T_{K_i,j}^{(i)}) \log \Lambda(T_{K_i,j}^{(i)}) + \mathbb{N}^{(i)}(T_{K_i,j}^{(i)})\theta' Z_i - e^{\theta' Z_i} \Lambda(T_{K_i,j}^{(i)}) \Big\} .$$

Then the maximum pseudo-likelihood estimator $(\widehat{\theta}_n^{ps}, \widehat{\Lambda}_n^{ps})$ of (θ, Λ) is given by

$$(\widehat{\theta}_n^{ps}, \widehat{\Lambda}_n^{ps}) \equiv \mathrm{argmax}_{\theta,\Lambda}\, l_n^{ps}(\theta, \Lambda) .$$

This can be implemented in two steps via the usual (pseudo-) profile likelihood. For each fixed value of θ we set

$$\widehat{\Lambda}_n^{ps}(\cdot, \theta) \equiv \mathrm{argmax}_{\Lambda}\, l_n^{ps}(\theta, \Lambda) , \tag{3}$$

and define

$$l_n^{ps,profile}(\theta) \equiv l_n^{ps}(\theta, \widehat{\Lambda}_n^{ps}(\cdot, \theta)) .$$

Then

$$\widehat{\theta}_n^{ps} = \mathrm{argmax}_{\theta}\, l_n^{ps,profile}(\theta) , \qquad \text{and} \qquad \widehat{\Lambda}_n^{ps} = \widehat{\Lambda}_n^{ps}(\cdot, \widehat{\theta}_n^{ps}) .$$

In fact, the optimization problem in (3) is easily solved as follows: Let $t_1 < \ldots < t_m$ denote the ordered distinct observation time points in the collection of all observations times, $\{T_{K_i,j}^{(i)},\ j = 1, \ldots, K_i,\ i = 1, \ldots, n\}$, let $\mathbb{N}_{K_i,j}^{(i)} \equiv \mathbb{N}^{(i)}(T_{K_i,j}^{(i)})$, and set

$$w_l = \sum_{i=1}^{n} \sum_{j=1}^{K_i} 1_{[T_{K_i,j}^{(i)} = t_l]} , \quad \overline{N}_l = \frac{1}{w_l} \sum_{i=1}^{n} \sum_{j=1}^{K_i} \mathbb{N}_{K_i,j}^{(i)} 1_{[T_{K_i,j}^{(i)} = t_l]} ,$$

$$\overline{A}_l(\theta) = \frac{1}{w_l} \sum_{i=1}^{n} \sum_{j=1}^{K_i} \exp(\theta' Z^{(i)})\, 1_{[T_{K_i,j}^{(i)} = t_l]} .$$

Then it is easily shown that

$$\widehat{\Lambda}_n^{ps}(\cdot,\theta) = \text{left-derivative of Greatest Convex Minorant of}$$
$$\{(\sum_{l\le i} w_l \overline{A}_l(\theta), \sum_{l\le i} w_l \overline{N}_l)\}_{i=1}^m$$
$$= \max_{i\le l}\min_{j\ge l} \frac{\sum_{i\le p} \le w_p \overline{N}_p}{\sum_{i\le p} \le w_p \overline{A}_p(\theta)} \quad \text{at} \quad t_l \,,$$

which is straightforward to compute.

Maximum Likelihood Estimation: Under the assumption that $\mathbb{N}$ is (conditionally, given Z) a non-homogeneous Poisson process, the likelihood can be calculated using the (conditional) independence of the increments of $\mathbb{N}$, $\Delta\mathbb{N}(s,t] \equiv \mathbb{N}(t) - \mathbb{N}(s)$, and the Poisson distribution of these increments:

$$P(\Delta\mathbb{N}(s,t] = k|Z) = \frac{[\Delta\Lambda((s,t]|Z)]^k}{k!} \exp(-\Delta\Lambda((s,t]|Z))$$

to obtain the *log-likelihood*:

$$l_n(\theta,\Lambda) = \sum_{i=1}^n \sum_{j=1}^{K_i} \left\{ \Delta\mathbb{N}_{K_i,j}^{(i)} \cdot \log \Delta\Lambda_{K_i,j} + \ \Delta\mathbb{N}_{K_i,j}^{(i)} \theta' Z_i - \ e^{\theta' Z_i} \Delta\Lambda_{K_i,j} \right\}$$

where

$$\Delta\mathbb{N}_{Kj} \equiv \mathbb{N}(T_{K,j}) - \mathbb{N}(T_{K,j-1}), \qquad j = 1,\ldots,K\,,$$
$$\Delta\Lambda_{Kj} \equiv \Lambda(T_{K,j}) - \Lambda(T_{K,j-1}), \qquad j = 1,\ldots,K\,,$$

with $\mathbb{N}(T_{K,0}) = 0$ and $\Lambda(T_{K,0}) = 0$. Then

$$(\widehat{\theta}_n, \widehat{\Lambda}_n) \equiv \text{argmax}_{\theta,\Lambda}\, l_n(\theta,\Lambda)\,.$$

This maximization can also be carried out in two steps via profile likelihood. For each fixed value of θ we set

$$\widehat{\Lambda}_n(\cdot,\theta) \equiv \text{argmax}_{\Lambda}\, l_n(\theta,\Lambda)\,,$$

and define

$$l_n^{profile}(\theta) \equiv l_n(\theta, \widehat{\Lambda}_n(\cdot,\theta))\,.$$

Then

$$\widehat{\theta}_n = \text{argmax}_{\theta}\, l_n^{profile}(\theta)\,, \qquad \text{and} \qquad \widehat{\Lambda}_n = \widehat{\Lambda}_n(\cdot,\widehat{\theta}_n)\,.$$

Computation of the (profile) "estimator" $\widehat{\Lambda}_n(\cdot,\theta)$ is computationally involved, but possible via the iterative convex minorant algorithm; see e.g. [17]. For more on computation without covariates see [33].

3 Information bounds for θ under the Poisson model.

We first compute information bounds for estimation of θ under the proportional mean (non-homogeneous) Poisson process model.

Suppose that $(\mathbb{N}|Z) \sim \text{Poisson}(\Lambda(\cdot|Z))$, and $((K, T_K)|Z) \sim G(\cdot|Z)$ are conditionally independent given Z. We will assume here that $\mathbb{N}$ is conditionally a nonhomogeneous Poisson process with conditional mean function

$$E[\mathbb{N}(t)|Z] = \Lambda(t|Z) \equiv e^{\theta_0' Z} \Lambda_0(t) . \tag{4}$$

The second equality expresses the proportional mean regression model assumption.

The likelihood for one observation is, using the same notation introduced in Section 2,

$$p(X; \theta, \Lambda) = \prod_{j=1}^{K} \exp(-\Delta\Lambda_{Kj}) \frac{(\Delta\Lambda_{Kj})^{\Delta\mathbb{N}_{Kj}}}{(\Delta\mathbb{N}_{Kj})!} .$$

Thus the log-likelihood for (θ, Λ) for one observation is given by

$$\log p(X; \theta, \Lambda) = \sum_{j=1}^{K} \{\Delta\mathbb{N}_{Kj} \log \Delta\Lambda_{Kj} - \Delta\Lambda_{Kj} - \log(\Delta\mathbb{N}_{Kj}!)\} .$$

Differentiating this with respect to θ and Λ respectively, the scores for θ and Λ are easily seen to be

$$\dot{\mathrm{l}}_\theta(x) = \sum_{j=1}^{K} Z(\Delta\mathbb{N}_{Kj} - e^{\theta_0' Z} \Delta\Lambda_{0Kj}) , \tag{5}$$

while

$$\begin{aligned} \dot{\mathrm{l}}_\Lambda a(x) &= \sum_{j=1}^{K} \left\{ \frac{\Delta\mathbb{N}_{Kj}}{\Delta\Lambda_{0Kj}} - e^{\theta_0' Z} \right\} \Delta a_{Kj} \\ &= \sum_{j=1}^{K} \left\{ \Delta\mathbb{N}_{Kj} - e^{\theta_0' Z} \Delta\Lambda_{0Kj} \right\} \frac{\Delta a_{Kj}}{\Delta\Lambda_{0Kj}} , \end{aligned}$$

where

$$\Delta a_{Kj} = \int_{T_{K,j-1}}^{T_{K,j}} a d\Lambda_0 , \qquad a \in L_2(\Lambda_0) .$$

To compute the information bound for estimation of θ it follows from the results of [2] and [4] that we want to find a^* so that

$$\dot{\mathrm{l}}_\theta - \dot{\mathrm{l}}_\Lambda a^* \perp \dot{\mathrm{l}}_\Lambda a$$

for all $a \in L_2(\Lambda_0)$; i.e.

$$
\begin{aligned}
0 &= E\left\{(\dot{\mathrm{l}}_\theta - \dot{\mathrm{l}}_\Lambda a^*)\dot{\mathrm{l}}_\Lambda a\right\} \\
&= E\left\{\sum_{j=1}^{K}(\Delta\mathbb{N}_{Kj} - e^{\theta_0' Z}\Delta\Lambda_{0Kj})\left(Z - \frac{\Delta a^*_{Kj}}{\Delta\Lambda_{0Kj}}\right)\dot{\mathrm{l}}_\Lambda a\right\} \\
&= E\left\{\sum_{j=1}^{K}(\Delta\mathbb{N}_{Kj} - e^{\theta_0' Z}\Delta\Lambda_{0Kj})^2\left(Z - \frac{\Delta a^*_{Kj}}{\Delta\Lambda_{0Kj}}\right)\frac{\Delta a_{Kj}}{\Delta\Lambda_{0Kj}}\right\} \\
&= E\left\{\sum_{j=1}^{K} e^{\theta_0' Z}\Delta\Lambda_{0Kj}\left(Z - \frac{\Delta a^*_{Kj}}{\Delta\Lambda_{0Kj}}\right)\frac{\Delta a_{Kj}}{\Delta\Lambda_{0Kj}}\right\} \\
&\qquad\qquad \text{by conditioning on } \ K, T_K, Z \\
&= E_K\left\{\sum_{j=1}^{K} E\left\{e^{\theta_0' Z}\Delta\Lambda_{0Kj}\left(Z - \frac{\Delta a^*_{Kj}}{\Delta\Lambda_{0Kj}}\right)\frac{\Delta a_{Kj}}{\Delta\Lambda_{0Kj}}\Big| K\right\}\right\} \\
&= E_K\left\{\sum_{j=1}^{K} E\left\{E\left\{e^{\theta_0' Z}\Delta\Lambda_{0Kj}\right.\right.\right. \\
&\qquad\qquad \left.\left.\left.\left(Z - \frac{\Delta a^*_{Kj}}{\Delta\Lambda_{0Kj}}\right)\frac{\Delta a_{Kj}}{\Delta\Lambda_{0Kj}}\Big| K, T_{K,j-1}, T_{K,j}\right\}\Big| K\right\}\right\} \\
&= E_K\left\{\sum_{j=1}^{K} E\left\{\Delta\Lambda_{0Kj}\frac{\Delta a_{Kj}}{\Delta\Lambda_{0Kj}}\left(E\left\{Ze^{\theta_0' Z}|K, T_{K,j-1}, T_{K,j}\right\}\right.\right.\right. \\
&\qquad\qquad \left.\left.\left.- \frac{\Delta a^*_{Kj}}{\Delta\Lambda_{0Kj}}E\left\{e^{\theta_0' Z}|K, T_{K,j-1}, T_{K,j}\right\}\right)\Big| K\right\}\right\}.
\end{aligned}
$$

Thus we see that the desired orthogonality holds with

$$
\frac{\Delta a^*_{Kj}}{\Delta\Lambda_{0Kj}} = \frac{E\left\{Ze^{\theta_0' Z}|K, T_{K,j-1}, T_{K,j}\right\}}{E\left\{e^{\theta_0' Z}|K, T_{K,j-1}, T_{K,j}\right\}}. \tag{6}
$$

Hence the efficient score function for θ is given by

$$
\begin{aligned}
\mathrm{l}^*_\theta(x) &= \dot{\mathrm{l}}_\theta(x) - \dot{\mathrm{l}}_\Lambda a^*(x) \\
&= \sum_{j=1}^{K}(\Delta\mathbb{N}_{Kj} - e^{\theta' Z}\Delta\Lambda_{0Kj})\left(Z - \frac{E\left\{Ze^{\theta' Z}|K, T_{K,j-1}, T_{K,j}\right\}}{E\left\{e^{\theta' Z}|K, T_{K,j-1}, T_{K,j}\right\}}\right),
\end{aligned}
$$

and the information for θ is, by computing conditionally on Z, K, T_K,

$$
I(\theta) = E_0\left\{\dot{\mathrm{l}}^*_\theta(X)^{\otimes 2}\right\}
$$

$$= E_0 \left\{ \sum_{j=1}^{K} e^{\theta_0' Z} \Delta \Lambda_{0Kj} \left(Z - \frac{E\left\{ Z e^{\theta_0' Z} | K, T_{K,j-1}, T_{K,j} \right\}}{E\left\{ e^{\theta_0' Z} | K, T_{K,j-1}, T_{K,j} \right\}} \right)^{\otimes 2} \right\} .$$

In particular, we have the following corollary:

Corollary 1. (Current status data). If $P(K = 1) = 1$ (so that the only T_{Kj} of relevance is $T_{1,1} \equiv T$ while $T_{1,0} = 0$), then the efficient score function is

$$\mathbf{l}_\theta^*(x) = \dot{\mathbf{l}}_\theta(x) - \dot{\mathbf{l}}_\Lambda a^*(x) = (\mathbb{N}(T) - e^{\theta_0' Z} \Lambda_0(T)) \left(Z - \frac{E\left\{ Z e^{\theta_0' Z} | T \right\}}{E\left\{ e^{\theta_0' Z} | T \right\}} \right) .$$

and the information for θ is given by

$$I(\theta) = E_0 \left\{ \mathbf{l}_\theta^*(X)^{\otimes 2} \right\} = E_0 \left\{ e^{\theta_0' Z} \Lambda_0(T) \left(Z - \frac{E\left\{ Z e^{\theta_0' Z} | T \right\}}{E\left\{ e^{\theta_0' Z} | T \right\}} \right)^{\otimes 2} \right\} .$$

This can be compared with the information for θ for the Cox proportional hazards model with current status data given by [15], page 547 (with Huang's Y replaced by our present T for comparison):

$$I(\theta) = E \left\{ R(T, Z) \left\{ Z - \frac{E(Z R(T, Z) | T)}{E(R(T, Z) | T)} \right\}^{\otimes 2} \right\}$$

where $R(T, Z) = \Lambda^2(T, Z) O(T|Z)$ and

$$O(t|z) = \frac{\overline{F}(t|z)}{1 - \overline{F}(t|z)} = \frac{(1 - F_0(t))^{\exp(\theta_0' z)}}{1 - (1 - F_0(t))^{\exp(\theta_0' z)}} .$$

Corollary 2. (Case 2 Interval-censored data). If $P(K = 2) = 1$ (so that the only T_{Kj}'s of relevance are $T_{2,1} \equiv T_1$ and $T_{2,2} = T_2$, while $T_{2,0} = 0$), then the efficient score function is

$$\begin{aligned} \mathbf{l}_\theta^*(x) &= \dot{\mathbf{l}}_\theta(x) - \dot{\mathbf{l}}_\Lambda a^*(x) \\ &= (\mathbb{N}(T_1) - e^{\theta_0' Z} \Lambda_0(T_1)) \left(Z - \frac{E\left\{ Z e^{\theta_0' Z} | T_1 \right\}}{E\left\{ e^{\theta_0' Z} | T_1 \right\}} \right) \\ &\quad + \mathbb{N}(T_2) - \mathbb{N}(T_1) \\ &\quad - e^{\theta_0' Z} (\Lambda_0(T_2) - \Lambda_0(T_1))) \left(Z - \frac{E\left\{ Z e^{\theta_0' Z} | T_1, T_2 \right\}}{E\left\{ e^{\theta_0' Z} | T_1, T_2 \right\}} \right) , \end{aligned}$$

and the information for θ is given by

$$
\begin{aligned}
I(\theta) &= E_0\left\{\dot{\mathbf{l}}_\theta^*(X)^{\otimes 2}\right\} \\
&= E_0\left\{e^{\theta_0' Z}\Lambda_0(T_1)\left(Z - \frac{E\left\{Ze^{\theta_0' Z}|T_1\right\}}{E\left\{e^{\theta_0' Z}|T_1\right\}}\right)^{\otimes 2}\right\} \\
&\quad + E_0\left\{e^{\theta_0' Z}(\Lambda_0(T_2) - \Lambda_0(T_2))\left(Z - \frac{E\left\{Ze^{\theta_0' Z}|T_1, T_2\right\}}{E\left\{e^{\theta_0' Z}|T_1, T_2\right\}}\right)^{\otimes 2}\right\}.
\end{aligned}
$$

This is much simpler than the information for θ for the Cox proportional hazards model with interval censored case II data given by [16]. The calculations of Huang and Wellner resulted in an integral equation to be solved, analogously to the results for the mean functional considered by [8], [9], and [10].

4 Asymptotic normality of the two estimators of $\boldsymbol{\theta}$.

Here is the crucial theorem concerning the asymptotic behavior of the maximum pseudo-likelihood and maximum likelihood estimators of θ when the proportional mean model holds, but the Poisson assumption concerning $\mathbb{N}$ may fail.

Theorem 1. Under suitable regularity and integrability conditions, the estimators $\widehat{\theta}_n^{ps}$ and $\widehat{\theta}_n$ are asymptotically normal:

$$
\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d Z \sim N_d\left(0, A^{-1}B\left(A^{-1}\right)'\right), \tag{7}
$$

and

$$
\sqrt{n}(\widehat{\theta}_n^{ps} - \theta_0) \to_d Z^{ps} \sim N_d\left(0, (A^{ps})^{-1}B^{ps}\left((A^{ps})^{-1}\right)'\right) \tag{8}
$$

where

$$
\begin{aligned}
B &\equiv Em^*(\theta_0, \Lambda_0; X)^{\otimes 2} \\
&= E\left\{\sum_{j,j'=1}^{K} C_{j,j'}(Z)\left[Z - \frac{E\left(Ze^{\theta_0' Z}|K, T_{K,j}, T_{K,j'}\right)}{E\left(e^{\theta_0' Z}|K, T_{K,j}, T_{K,j'}\right)}\right]^{\otimes 2}\right\}, \\
A &= E\left\{\sum_{j=1}^{K} \Delta\Lambda_{0Kj}e^{\theta_0' Z}\left[Z - \frac{E\left(Ze^{\theta_0' Z}|K, T_{K,j-1}, T_{K,j}\right)}{E\left(e^{\theta_0' Z}|K, T_{K,j-1}, T_{K,j}\right)}\right]^{\otimes 2}\right\}, \\
C_{j,j'}(Z) &= \mathrm{Cov}\left[\Delta N_{Kj}, \Delta N_{Kj'}|Z, K, \underline{T}_K\right], \\
B^{ps} &= Em^{*ps}(\theta_0, \Lambda_0; X)^{\otimes 2}
\end{aligned}
$$

$$= E\left\{\sum_{j,j'=1}^{K} C^{ps}_{j,j'}(Z)\left[Z - \frac{E\left(Ze^{\theta_0' Z}|K, T_{K,j}\right)}{E\left(e^{\theta_0' Z}|K, T_{K,j}\right)}\right]\right.$$

$$\left.\left[Z - \frac{E\left(Ze^{\theta_0' Z}|K, T_{K,j'}\right)}{E\left(e^{\theta_0' Z}|K, T_{K,j'}\right)}\right]'\right\},$$

$$A^{ps} = E\left\{\sum_{j=1}^{K} \Lambda_{0Kj} e^{\theta_0' Z}\left[Z - \frac{E\left(Ze^{\theta_0' Z}|K, T_{K,j}\right)}{E\left(e^{\theta_0' Z}|K, T_{K,j}\right)}\right]^{\otimes 2}\right\},$$

$$C^{ps}_{j,j'}(Z) = \mathrm{Cov}\left[N_{Kj}, N_{Kj'}|Z, K, T_{K,j}, T_{K,j'}\right] .$$

Our proof of this theorem is based on the results of [35]. While we will not give the proof in detail, we will present here a sketch of the computation of the asymptotic variances given in (7) and (8).

Based on the Poisson model, the log-likelihood for (θ, Λ) with one observation is given by

$$m(\theta, \Lambda; X) = \log p(X; \theta, \Lambda) = \sum_{j=1}^{K}\{\Delta\mathbb{N}_{Kj}\log\Delta\Lambda_{Kj} - \Delta\Lambda_{Kj} - \log(\Delta\mathbb{N}_{Kj}!)\}$$

$$= \sum_{j=1}^{K}\Big\{\Delta\mathbb{N}_{Kj}\log\Delta\Lambda_{Kj} + \Delta\mathbb{N}_{Kj}\theta' Z - e^{\theta' Z}\Delta\Lambda_{0Kj} - \log(\Delta\mathbb{N}_{Kj}!)\Big\} . \qquad (9)$$

Thus the log-likelihood $l_n(\theta, \Lambda)$ for n i.i.d. observations is given by

$$l_n(\theta, \Lambda) = n\mathbb{P}_n m(\theta, \Lambda; \cdot) . \qquad (10)$$

The maximum likelihood estimators $(\widehat{\theta}, \widehat{\Lambda})$ are obtained by maximizing (10).

A natural pseudo-likelihood is obtained by simply taking the product of the marginal distributions of the observed counts at the successive observation times. Thus a log-pseudo-likelihood for one observation is given by

$$m^{ps}(\theta, \Lambda; X) = \sum_{j=1}^{K}\Big\{\mathbb{N}_{Kj}\log\Lambda_{Kj} + \mathbb{N}_{Kj}\theta' Z - e^{\theta' Z}\Lambda_{Kj} - \log(\mathbb{N}_{Kj}!)\Big\} \qquad (11)$$

with $\Lambda_{K,j} = \Lambda(T_{K,j})$, and the log-pseudo-likelihood $l^{ps}_n(\theta, \Lambda)$ for n i.i.d. observations is given by

$$l^{ps}_n(\theta, \Lambda) = n\mathbb{P}_n m^{ps}(\theta, \Lambda; \cdot) , \qquad (12)$$

and the corresponding pseudo-MLE's $(\widehat{\theta}^{ps}, \widehat{\Lambda}^{ps})$ are obtained by maximizing (12).

4.1 Asymptotic variance of the MLE

Based on the Poisson model, the log-likelihood for (θ, Λ) with one observation is given by (9). Using the notation of [35], page 29, we have

$$
\begin{aligned}
m_1(\theta, \Lambda; X) &= \sum_{j=1}^{K} Z \left[\Delta\mathbb{N}_{Kj} - \Delta\Lambda_{Kj} e^{\theta' Z} \right], \\
m_2(\theta, \Lambda; X)[h] &= \sum_{j=1}^{K} \left[\frac{\Delta\mathbb{N}_{Kj}}{\Delta\Lambda_{Kj}} - e^{\theta' Z} \right] \Delta h_{Kj}, \\
m_{11}(\theta, \Lambda; X) &= -\sum_{j=1}^{K} \Delta\Lambda_{Kj} Z Z' e^{\theta' Z}, \\
m_{12}(\theta, \Lambda; X)[h] &= m_{21}^T(\theta, \Lambda; X)[h] = -\sum_{j=1}^{K} Z e^{\theta' Z} \Delta h_{Kj}, \\
m_{22}(\theta, \Lambda; X)[\mathbf{h}, h] &= -\sum_{j=1}^{K} \frac{\Delta\mathbb{N}_{Kj}}{(\Delta\Lambda_{Kj})^2} \Delta\mathbf{h}_{Kj} \Delta h_{Kj},
\end{aligned}
$$

where $\Delta h_{Kj} = \int_{T_{K,j-1}}^{T_{K,j}} h d\Lambda_0$ for $h \in L_2(\Lambda)$. By A2 of [35], page 30, we need to find a $\mathbf{h}^*$ such that

$$
\begin{aligned}
&\dot{S}_{12}(\theta_0, \Lambda_0)[h] - \dot{S}_{22}(\theta_0, \Lambda_0)[\mathbf{h}^*, h] \\
&= P\left\{ m_{12}(\theta_0, \Lambda_0; X)[h] - m_{22}(\theta_0, \Lambda_0; X)[\mathbf{h}^*, h] \right\} = 0,
\end{aligned}
$$

for all $h \in L_2(\Lambda_0)$. Note that

$$
\begin{aligned}
&P\left\{ m_{12}(\theta_0, \Lambda_0; X)[h] - m_{22}(\theta_0, \Lambda_0; X)[\mathbf{h}^*, h] \right\} \\
&= -\, E\left\{ \sum_{j=1}^{K} \left[Z e^{\theta_0' Z} - \frac{\Delta\mathbb{N}_{Kj}}{(\Delta\Lambda_{0Kj})^2} \Delta\mathbf{h}_{Kj}^* \right] \Delta h_{Kj} \right\} \\
&= -\, E_{(K, T_K, Z)} \left\{ \sum_{j=1}^{K} \left[Z e^{\theta_0' Z} - \frac{e^{\theta_0' Z} \Delta\mathbf{h}_{Kj}^*}{\Delta\Lambda_{0Kj}} \right] \Delta h_{Kj} \right\}.
\end{aligned}
$$

Therefore, an obvious choice of $\mathbf{h}^*$ is

$$
\Delta\mathbf{h}_{Kj}^* = \Delta\Lambda_{0Kj} \frac{E\left(Z e^{\theta_0' Z} | K, T_{K,j-1}, T_{K,j} \right)}{E\left(e^{\theta_0' Z} | K, T_{K,j-1}, T_{K,j} \right)}.
$$

Hence

$$
m^*(\theta_0, \Lambda_0; X)
$$

$$
\begin{aligned}
&= m_1(\theta_0, \Lambda_0; X) - m_2(\theta_0, \Lambda_0; X)[\mathbf{h}^*] \\
&= \sum_{j=1}^{K} \Bigg\{ Z \left(\Delta\mathbb{N}_{Kj} - e^{\theta_0' Z} \Delta\Lambda_{0Kj} \right) \\
&\qquad - \left(\frac{\Delta\mathbb{N}_{Kj}}{\Delta\Lambda_{0Kj}} - e^{\theta_0' Z} \right) \Delta\Lambda_{0Kj} \frac{E\left(Ze^{\theta_0' Z} | K, T_{K,j-1}, T_{K,j} \right)}{E\left(e^{\theta_0' Z} | K, T_{K,j-1}, T_{K,j} \right)} \Bigg\} \\
&= \sum_{j=1}^{K} \left(\Delta\mathbb{N}_{Kj} - e^{\theta_0' Z} \Delta\Lambda_{0Kj} \right) \left[Z - \frac{E\left(Ze^{\theta_0' Z} | K, T_{K,j-1}, T_{K,j} \right)}{E\left(e^{\theta_0' Z} | K, T_{K,j-1}, T_{K,j} \right)} \right].
\end{aligned}
$$

By Theorem 2.3.5 of [35], page 32, the asymptotic variance will be $A^{-1}B\left(A^{-1}\right)'$, where

$$
\begin{aligned}
B &= Em^*(\theta_0, \Lambda_0; X)^{\otimes 2} \\
&= E_{(K, T_K, Z)} \Bigg\{ \sum_{j,j'=1}^{K} C(T_{K,j}, T_{K,j'}, T_{K,j-1}, T_{K,j'-1}; Z) \\
&\qquad \left[Z - \frac{E\left(Ze^{\theta_0' Z} | K, T_{K,j-1}, T_{K,j} \right)}{E\left(e^{\theta_0' Z} | K, T_{K,j-1}, T_{K,j} \right)} \right] \\
&\qquad \left[Z - \frac{E\left(Ze^{\theta_0' Z} | K, T_{K,j'-1}, T_{K,j'} \right)}{E\left(e^{\theta_0' Z} | K, T_{K,j'-1}, T_{K,j'} \right)} \right]' \Bigg\},
\end{aligned}
$$

$$
\begin{aligned}
A &= -\dot{S}_{11}(\theta_0, \Lambda_0) + \dot{S}_{21}(\theta_0, \Lambda_0)[\mathbf{h}^*] \\
&= E \Bigg\{ \sum_{j=1}^{K} \Bigg[\Delta\Lambda_{0Kj} e^{\theta_0' Z} Z Z' \\
&\qquad - e^{\theta_0' Z} \Delta\Lambda_{0Kj} \frac{E\left(Ze^{\theta_0' Z} | K, T_{K,j'-1}, T_{K,j'} \right)}{E\left(e^{\theta_0' Z} | K, T_{K,j'-1}, T_{K,j'} \right)} Z' \Bigg] \Bigg\} \\
&= E_{(K, T_K, Z)} \left\{ \sum_{j=1}^{K} \Delta\Lambda_{0Kj} e^{\theta_0' Z} \left[Z - \frac{E\left(Ze^{\theta_0' Z} | K, T_{K,j-1}, T_{K,j} \right)}{E\left(e^{\theta_0' Z} | K, T_{K,j-1}, T_{K,j} \right)} \right] Z' \right\} \\
&= E_{(K, T_K, Z)} \left\{ \sum_{j=1}^{K} \Delta\Lambda_{0Kj} e^{\theta_0' Z} \left[Z - \frac{E\left(Ze^{\theta_0' Z} | K, T_{K,j-1}, T_{K,j} \right)}{E\left(e^{\theta_0' Z} | K, T_{K,j-1}, T_{K,j} \right)} \right]^{\otimes 2} \right\},
\end{aligned}
$$

and

$$\begin{aligned} &C(T_{K,j}, T_{K,j'}, T_{K,j-1}, T_{K,j'-1}; Z) \\ &= E\left[\left(\Delta\mathbb{N}_{Kj} - e^{\theta_0' Z}\Delta\Lambda_{0Kj}\right)\right. \\ &\qquad \left.\left(\Delta\mathbb{N}_{Kj'} - e^{\theta_0' Z}\Delta\Lambda_{0Kj'}\right)|Z, K, T_{K,j-1}, T_{K,j}, T_{K,j'-1}, T_{K,j'}\right]. \end{aligned}$$

Note that if the counting process is indeed a conditional Poisson process with the mean function given as specified,

$$C(T_{K,j}, T_{K,j'}, T_{K,j-1}, T_{K,j'-1}; Z) = \begin{cases} \Delta\Lambda_{0Kj} e^{\theta_0' Z} & , \text{ if } j = j' \\ 0 & , \text{ if } j \neq j'. \end{cases}$$

This yields $B = A = I(\theta_0)$ and thus $A^{-1}B\left(A^{-1}\right)' = I^{-1}(\theta_0)$.

4.2 Asymptotic Variance of the Pseudo-MLE

Based on the Poisson model, the pseudo log-likelihood for (θ, Λ) with one observation is given by (11). Using the notation of Zhang (1998), page 29, we have

$$\begin{aligned} m_1^{ps}(\theta, \Lambda; X) &= \sum_{j=1}^{K} Z\left[\mathbb{N}_{Kj} - \Lambda_{Kj} e^{\theta' Z}\right], \\ m_2^{ps}(\theta, \Lambda; X)[h] &= \sum_{j=1}^{K}\left[\frac{\mathbb{N}_{Kj}}{\Lambda_{Kj}} - e^{\theta' Z}\right] h_{Kj}, \\ m_{11}^{ps}(\theta, \Lambda; X) &= -\sum_{j=1}^{K} \Lambda_{Kj} Z Z' e^{\theta' Z}, \\ m_{12}^{ps}(\theta, \Lambda; X)[h] &= m_{21}^{T}(\theta, \Lambda; X)[h] = -\sum_{j=1}^{K} Z e^{\theta' Z} h_{Kj}, \\ m_{22}^{ps}(\theta, \Lambda; X)[\mathbf{h}, h] &= -\sum_{j=1}^{K} \frac{\mathbb{N}_{Kj}}{\left(\Lambda_{Kj}\right)^2} \mathbf{h}_{Kj} h_{Kj}, \end{aligned}$$

where $h_{Kj} = \int_0^{T_{K,j}} h d\Lambda$ for $h \in L_2(\Lambda)$. By A2 of [35], page 30, we need to find a $\mathbf{h}^*$ such that

$$\begin{aligned} &\dot{S}_{12}^{ps}(\theta_0, \Lambda_0)[h] - \dot{S}_{22}^{ps}(\theta_0, \Lambda_0)[\mathbf{h}^*, h] \\ &= P\left\{m_{12}^{ps}(\theta_0, \Lambda_0; X)[h] - m_{22}^{ps}(\theta_0, \Lambda_0; X)[\mathbf{h}^*, h]\right\} = 0, \end{aligned}$$

for all $h \in L_2(\Lambda_0)$. Note that

$$
\begin{aligned}
&P\left\{m_{12}^{ps}(\theta_0,\Lambda_0;X)[h]-m_{22}^{ps}(\theta_0,\Lambda_0;X)[\mathbf{h}^*,h]\right\}\\
&=-\,E\left\{\sum_{j=1}^{K}\left[Ze^{\theta_0'Z}-\frac{\mathbb{N}_{Kj}}{(\Lambda_{0Kj})^2}\mathbf{h}_{Kj}^*\right]h_{Kj}\right\}\\
&=-\,E_{(K,T_K,Z)}\left\{\sum_{j=1}^{K}\left[Ze^{\theta_0'Z}-\frac{e^{\theta_0'Z}\mathbf{h}_{Kj}^*}{\Lambda_{0Kj}}\right]h_{Kj}\right\}.
\end{aligned}
$$

Therefore, an obvious choice of $\mathbf{h}^*$ is

$$
\mathbf{h}_{Kj}^*=\Lambda_{0Kj}\frac{E\left(Ze^{\theta_0'Z}|K,T_{K,j}\right)}{E\left(e^{\theta_0'Z}|K,T_{K,j}\right)}.
$$

Hence

$$
\begin{aligned}
&m^{*ps}(\theta_0,\Lambda_0;X)\\
&=m_1^{ps}(\theta_0,\Lambda_0;X)-m_2^{ps}(\theta_0,\Lambda_0;X)[\mathbf{h}^*]\\
&=\sum_{j=1}^{K}\left\{Z\left(\mathbb{N}_{Kj}-e^{\theta_0'Z}\Lambda_{0Kj}\right)-\left(\frac{\mathbb{N}_{Kj}}{\Lambda_{0Kj}}-e^{\theta_0'Z}\right)\Lambda_{0Kj}\frac{E\left(Ze^{\theta_0'Z}|K,T_{K,j}\right)}{E\left(e^{\theta_0'Z}|K,T_{K,j}\right)}\right\}\\
&=\sum_{j=1}^{K}\left(\mathbb{N}_{Kj}-e^{\theta_0'Z}\Lambda_{0Kj}\right)\left[Z-\frac{E\left(Ze^{\theta_0'Z}|K,T_{K,j}\right)}{E\left(e^{\theta_0'Z}|K,T_{K,j}\right)}\right].
\end{aligned}
$$

By Theorem 2.3.5 of [35], page 32, the asymptotic variance will be $(A^{ps})^{-1}B^{ps}\left((A^{ps})^{-1}\right)'$, where

$$
\begin{aligned}
B^{ps}&=Em^{*ps}(\theta_0,\Lambda_0;X)^{\otimes 2}\\
&=E_{(K,T_K,Z)}\left\{\sum_{j,j'=1}^{K}C^{ps}(T_{K,j},T_{K,j'};Z)\left[Z-\frac{E\left(Ze^{\theta_0'Z}|K,T_{K,j}\right)}{E\left(e^{\theta_0'Z}|K,T_{K,j}\right)}\right]\right.\\
&\qquad\left.\left[Z-\frac{E\left(Ze^{\theta_0'Z}|K,T_{K,j'}\right)}{E\left(e^{\theta_0'Z}|K,T_{K,j'}\right)}\right]'\right\},
\end{aligned}
$$

$$
\begin{aligned}
A^{ps}&=-\,\dot S_{11}^{ps}(\theta_0,\Lambda_0)+\dot S_{21}^{ps}(\theta_0,\Lambda_0)[\mathbf{h}^*]\\
&=E\left\{\sum_{j=1}^{K}\left[\Lambda_{0Kj}e^{\theta_0'Z}ZZ'-e^{\theta_0'Z}\Lambda_{0Kj}\frac{E\left(Ze^{\theta_0'Z}|K,T_{K,j'}\right)}{E\left(e^{\theta_0'Z}|K,T_{K,j'}\right)}Z'\right]\right\}\\
&=E_{(K,T_K,Z)}\left\{\sum_{j=1}^{K}\Lambda_{0Kj}e^{\theta_0'Z}\left[Z-\frac{E\left(Ze^{\theta_0'Z}|K,T_{K,j}\right)}{E\left(e^{\theta_0'Z}|K,T_{K,j}\right)}\right]Z'\right\}
\end{aligned}
$$

$$= E_{(K,\underline{T}_K,Z)} \left\{ \sum_{j=1}^{K} \Lambda_{0Kj} e^{\theta_0' Z} \left[Z - \frac{E\left(Ze^{\theta_0' Z}|K, T_{K,j}\right)}{E\left(e^{\theta_0' Z}|K, T_{K,j}\right)} \right]^{\otimes 2} \right\},$$

and

$$\begin{aligned} & C^{ps}(T_{K,j}, T_{K,j'}; Z) \\ &= E\left[\left(\mathbb{N}_{Kj} - e^{\theta_0' Z}\Lambda_{0Kj}\right) \left(\mathbb{N}_{Kj'} - e^{\theta_0' Z}\Lambda_{0Kj'}\right) |Z, K, T_{K,j}, T_{K,j'} \right]. \end{aligned}$$

Note that if the counting process is indeed a conditional Poisson process with the mean function given as specified,

$$C^{ps}(T_{K,j}, T_{K,j'}; Z) = e^{\theta_0' Z}\Lambda_{0K(j\wedge j')} .$$

This yields

$$\begin{aligned} B^{ps} = A^{ps} + 2E_{(K,\underline{T}_K,Z)} & \left\{ \sum_{j<j'} e^{\theta_0 Z}\Lambda_{0Kj} \left[Z - \frac{E\left(Ze^{\theta_0' Z}|K, T_{K,j}\right)}{E\left(e^{\theta_0' Z}|K, T_{K,j}\right)} \right] \right. \\ & \left. \left[Z - \frac{E\left(Ze^{\theta_0' Z}|K, T_{K,j'}\right)}{E\left(e^{\theta_0' Z}|K, T_{K,j'}\right)} \right]' \right\} \\ \neq A^{ps} . \end{aligned}$$

5 Comparisons: MLE versus pseudo-MLE.

Scenario 1. We first suppose that the underlying counting process is in fact a standard homogeneous Poisson process conditionally given Z, with baseline mean function $\Lambda_0(t) = \lambda t$, We will also assume that the distribution of $(K, \underline{T}_K)$ is independent of Z. As a consequence, Z is independent of $(K, \underline{T}_K)$, and the formulas in the preceding section simplify considerably. Because of the Poisson process assumption, $A = B = I(\theta_0)$, and this matrix is given by

$$\begin{aligned} I(\theta_0) &= E_{(K,\underline{T}_K,Z)} \left\{ \sum_{j=1}^{K} \Delta\Lambda_{0Kj} e^{\theta_0' Z} \left[Z - \frac{E\left(Ze^{\theta_0' Z}|K, T_{K,j-1}, T_{K,j}\right)}{E\left(e^{\theta_0' Z}|K, T_{K,j-1}, T_{K,j}\right)} \right]^{\otimes 2} \right\} \\ &= E_{(K,\underline{T}_K)} \left\{ \Lambda_0(T_{K,K}) \right\} E_Z \left\{ e^{\theta_0' Z} \left[Z - \frac{E\left(Ze^{\theta_0' Z}\right)}{E\left(e^{\theta_0' Z}\right)} \right]^{\otimes 2} \right\} \\ &\equiv E_{(K,\underline{T}_K)} \left\{ \Lambda_0(T_{K,K}) \right\} C , \end{aligned}$$

so that if C is nonsingular,

$$I(\theta_0)^{-1} = C^{-1} \frac{1}{E_{(K,T_K)}\left\{\Lambda_0(T_{K,K})\right\}} .$$

On the other hand,

$$\begin{aligned} A^{ps} &= E_{(K,T_K,Z)} \left\{ \sum_{j=1}^{K} \Lambda_{0Kj} e^{\theta_0' Z} \left[Z - \frac{E\left(Z e^{\theta_0' Z} | K, T_{K,j}\right)}{E\left(e^{\theta_0' Z} | K, T_{K,j}\right)} \right]^{\otimes 2} \right\} \\ &= E_{(K,T_K)} \left\{ \sum_{j=1}^{K} \Lambda_0(T_{Kj}) \right\} E \left\{ e^{\theta_0' Z} \left[Z - \frac{E\left(Z e^{\theta_0' Z}\right)}{E\left(e^{\theta_0' Z}\right)} \right]^{\otimes 2} \right\} , \end{aligned}$$

while

$$B^{ps} = E_{(K,T_K,Z)} \left\{ \sum_{j,j'=1}^{K} \Lambda_0(T_{K,j\wedge j'}) \right\} E_Z \left\{ e^{\theta_0' Z} \left[Z - \frac{E\left(Z e^{\theta_0' Z}\right)}{E\left(e^{\theta_0' Z}\right)} \right]^{\otimes 2} \right\} ,$$

so that

$$(A^{ps})^{-1} B^{ps} \left((A^{ps})^{-1}\right)' = C^{-1} \frac{E_{(K,T_K,Z)}\left\{\sum_{j,j'=1}^{K} \Lambda_0(T_{K,j\wedge j'})\right\}}{\left(E_{(K,T_K)}\left\{\sum_{j=1}^{K} \Lambda_0(T_{Kj})\right\}\right)^2} .$$

Thus it follows that the ARE of the pseudo-MLE of θ_0 relative to the MLE of θ_0 under the above scenario is given by

$$\begin{aligned} ARE(pseudo, mle) &= \frac{A^{-1}B(A^{-1})^T}{(A^{ps})^{-1}B^{ps}((A^{ps})^{-1})^T} \\ &= \frac{\left\{E\left\{\sum_{j=1}^{K} \Lambda_0(T_{K,j})\right\}\right\}^2}{E\left\{\Lambda_0(T_{K,K})\right\} E\left\{\sum_{j,j'=1}^{K} \Lambda_0(T_{K,j\wedge j'})\right\}} . \end{aligned}$$

Note that this equals 1 if $P(K=1)=1$. Actually, we have not yet used the assumption about Λ_0. If we assume that $\Lambda_0(t) = \lambda t$, then

$$ARE(pseudo, mle) = \frac{\left\{E\left\{\sum_{j=1}^{K} T_{K,j}\right\}\right\}^2}{E\left\{T_{K,K}\right\} E\left\{\sum_{j,j'=1}^{K} T_{K,j\wedge j'}\right\}}$$

Scenario 1A. If we assume, further, that $P(K=k)=1$ for a fixed integer $k \geq 2$, and $\underline{T}_K = (T_{K,1}, \ldots, T_{K,K})$ are the order statistics of a sample of k uniformly distributed random variables on an interval $[0, M]$, then

$$E\left\{\sum_{j=1}^{K} T_{K,j}\right\} = \sum_{j=1}^{k} \frac{j}{k+1} M = \frac{k}{2} M ,$$

$$E\{T_{K,K}\} = \frac{k}{k+1} M\,,$$

and

$$E\left\{\sum_{j,j'=1}^{K} T_{K,j\wedge j'}\right\} = \sum_{j,j'=1}^{k} \frac{j\wedge j'}{k+1} M = \frac{k(2k+1)}{6} M\,.$$

Hence in this case

$$ARE(pseudo, mle) = \frac{(k/2)^2}{\frac{k}{k+1}\frac{k(2k+1)}{6}} = \frac{3(k+1)}{2(2k+1)} \to \frac{3}{4}$$

as $k \to \infty$.

k	1	2	3	4	5	6	7	8	9	10
$I(\theta)^{-1}(k)M\lambda$	2	3/2	4/3	5/4	6/5	7/6	8/7	9/8	10/9	11/10
$V^{ps}(k)M\lambda$	2	5/3	14/9	3/2	22/15	13/9	10/7	17/12	38/27	7/5
ARE(k)	1.00	.900	0.857	0.833	0.818	0.808	0.800	0.794	0.789	0.786

Scenario 1B. If we assume instead that K is random and for some $0 \le c < 1/2$

$$T_{K,j} \sim \text{Uniform}[j-c, j+c], \qquad j = 1,\ldots,K,$$

and are independent (conditionally on K), then we calculate

$$E(T_{K,K}) = E(K)\,,$$
$$E(\sum_{j=1}^{K} T_{K,j}) = E\left(\frac{K(K+1)}{2}\right)\,,$$
$$E(\sum_{j,j'=1}^{K} T_{K,j\wedge j'}) = E\left(\frac{K(K+1)(2K+1)}{6}\right)\,,$$

and hence

$$ARE(pseudo, mle) = \frac{\left\{E\left(\frac{K(K+1)}{2}\right)\right\}^2}{E(K)E\left(\frac{K(K+1)(2K+1)}{6}\right)}\,.$$

Note that this depends only on the first three moments of K, with the third moment appearing in the denominator. It does not depend on c, but only on $E(T_{K,j}|K) = j$, $j = 1,\ldots,K$. When $P(K = k) = 1$ we find that under this scenario

$$ARE(pseudo, mle) = \frac{3(k+1)}{2(2k+1)} \ge \frac{3}{4}\,.$$

It is clear that for random K the ARE in this scenario can be arbitrarily small if the distribution of K is heavy-tailed. This will be shown in more detail in Scenario 2.

Scenario 2. A variant on these calculations is to repeat all the assumptions about Z and $(K, \underline{T}_K)$, assume, conditionally on K, that $\underline{T}_K = (T_{K,1}, \ldots, T_{K,K})$ are the order statistics of a sample of K uniformly distributed random variables on an interval $[0, M]$, but now allow a distribution for K. Then

$$E\left\{\sum_{j=1}^{K} T_{K,j} \Big| K\right\} = \sum_{j=1}^{K} \frac{j}{K+1} M = \frac{K}{2} M \,,$$

$$E\left\{T_{K,K} \Big| K\right\} = \frac{K}{K+1} M \,,$$

and

$$E\left\{\sum_{j,j'=1}^{K} T_{K,j\wedge j'} \Big| K\right\} = \sum_{j,j'=1}^{K} \frac{j\wedge j'}{K+1} M = \frac{K(2K+1)}{6} M \,.$$

If we let K be distributed according to 1+Poisson(μ), then the ARE will be asymptotically $3/4$ again as $\mu \to \infty$. A more interesting choice of the distribution of K is the Zeta(α) distribution given as follows: for $\alpha > 1$,

$$P(K = k) = \frac{1/k^{\alpha}}{\zeta(\alpha)}, \qquad k = 1, 2, \ldots,$$

where $\zeta(\alpha) = \sum_{j=1}^{\infty} j^{-\alpha}$ is the Riemann Zeta function. Then we can compute

$$E\left\{\sum_{j=1}^{K} T_{K,j}\right\} = E\left\{\sum_{j=1}^{K} \frac{j}{K+1}\right\} M = \frac{E_{\alpha}(K)}{2} M = \frac{M}{2} \frac{\zeta(\alpha-1)}{\zeta(\alpha)} \,,$$

$$E\left\{T_{K,K}\right\} = E_{\alpha}\left(\frac{K}{K+1}\right) M \,,$$

and

$$\begin{aligned} E\left\{\sum_{j,j'=1}^{K} T_{K,j\wedge j'}\right\} &= E_{\alpha}\left(\sum_{j,j'=1}^{K} \frac{j\wedge j'}{K+1}\right) M \\ &= \frac{M}{6} E_{\alpha}(K(2K+1)) \\ &= \frac{M}{6}\left\{2E_{\alpha}(K^2) + E_{\alpha}(K)\right\} \\ &= \frac{M}{6}\left\{\frac{2\zeta(\alpha-2) + \zeta(\alpha-1)}{\zeta(\alpha)}\right\} \end{aligned}$$

Hence in this case

$$I(\theta)^{-1}(\alpha) = C^{-1} \frac{1}{M\lambda E_{\alpha}\left(\frac{K}{K+1}\right)} \,,$$

$$V^{ps}(\alpha) = C^{-1} \frac{\left\{\frac{2\zeta(\alpha-2)+\zeta(\alpha-1)}{\zeta(\alpha)}\right\}}{6M\lambda\left(\frac{\zeta(\alpha-1)}{2\zeta(\alpha)}\right)^2},$$

and

$$\begin{aligned} ARE(pseudo, mle)(\alpha) &\equiv ARE(\alpha) \\ &= \frac{(E_\alpha(K)/2)^2}{E_\alpha\{\frac{K}{K+1}\}E_\alpha\frac{K(2K+1)}{6}} \\ &= \frac{3}{2}\frac{\zeta(\alpha-1)/\zeta(\alpha)}{E_\alpha\{\frac{K}{K+1}\}\frac{2\zeta(\alpha-2)+\zeta(\alpha-1)}{\zeta(\alpha)}} \\ &= \frac{3}{2}\frac{\zeta(\alpha-1)}{\{2\zeta(\alpha-2)+\zeta(\alpha-1)\}\,E_\alpha\{\frac{K}{K+1}\}}, \end{aligned}$$

and this varies between 0 and 1 as α varies from 3 to ∞; note that for $\alpha = 3$, $E(K^2) = \infty = \zeta(1)$, while $E(K) = \zeta(2)/\zeta(3) \doteq 10.5844\ldots$. See Figure 1.

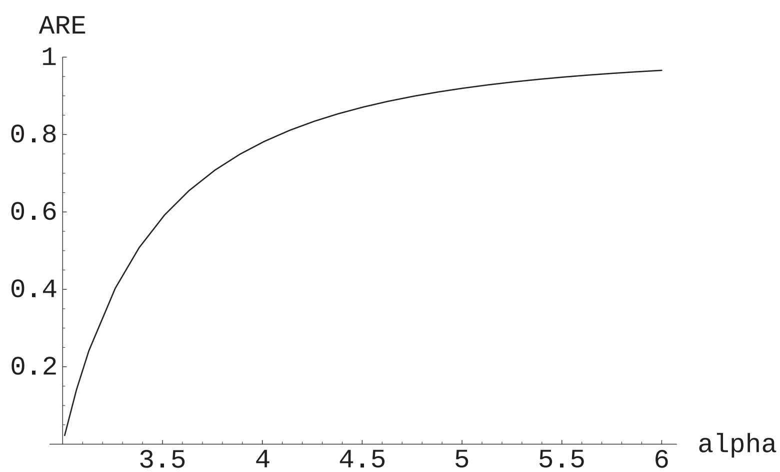

Fig. 1. ARE(α)

If we change the distribution of K to $K \sim \text{Unif}\{1, \ldots, k_0\}$, then

$$E\left\{\sum_{j=1}^{K} T_{K,j}\right\} = E\left\{\sum_{j=1}^{K} \frac{j}{K+1}\right\} M = \frac{E(K)}{2} M = \frac{k_0+1}{4} M,$$

$$\begin{aligned}
E\left\{\sum_{j,j'=1}^{K} T_{K,j\wedge j'}\right\} &= E_\alpha\left(\sum_{j,j'=1}^{K}\frac{j\wedge j'}{K+1}\right)M\\
&= \frac{M}{6}E(K(2K+1))\\
&= \frac{M}{6}\left\{2E(K^2)+E(K)\right\}\\
&= \frac{M}{6}\left\{2\frac{(k_0+1)(2k_0+1)}{6}+\frac{k_0+1}{2}\right\}
\end{aligned}$$

while

$$E\left\{T_{K,K}\right\} = E\left(\frac{K}{K+1}\right)M\,,$$

Hence in this case

$$I(\theta)^{-1}(\alpha) = C^{-1}\frac{1}{M\lambda E\left(\frac{K}{K+1}\right)}\,,$$

$$V^{ps}(k_0) = C^{-1}\frac{\frac{(k_0+1)(2k_0+1)/3+(k_0+1)/2}{6}}{M\lambda(k_0+1)^2/16}\,,$$

and

$$\begin{aligned}
ARE(pseudo, mle)(k_0) &\equiv ARE(k_0)\\
&= \frac{(E(K)/2)^2}{E\{\frac{K}{K+1}\}E\frac{K(2K+1)}{6}}\\
&= \frac{(k_0+1)^2/16}{E\{\frac{K}{K+1}\}\frac{(k_0+1)(2k_0+1)/3+(k_0+1)/2}{6}}\,,
\end{aligned}$$

and this varies between 1 and 9/16 as k_0 varies from 1 to ∞.

Scenario 3. We now suppose that the underlying counting process is *not* a Poisson process, conditionally given Z, but is, instead, defined as in terms of the "negative-binomialization" of an empirical counting process, as follows: suppose that $X_1, X_2, \ldots$ are i.i.d. with distribution function F on R, and define

$$\mathbb{N}_n(t) = \sum_{i=1}^{n} 1_{[X_i\le t]} \qquad \text{for} \quad t\ge 0\,.$$

Suppose that $(N|Z) \sim$ Negative Binomial$(r(Z,\gamma,\theta_0),p)$ where $r(Z,\gamma,\theta_0) = \gamma e^{\theta_0' Z}$, and define $\mathbb{N}$ by

$$\mathbb{N}(t) \equiv \mathbb{N}_N(t) = \sum_{i=1}^{N} 1_{[X_i\le t]}\,.$$

Then, since

$$E(N|Z) = r(Z, \gamma, \theta_0)\frac{q}{p}\,, \qquad Var(N|Z) = r(Z, \gamma, \theta_0)\frac{q}{p^2}\,,$$

it follows that

$$\begin{aligned} E\{\mathbb{N}(t)|Z\} &= E\{E[\mathbb{N}(t)|Z, N]|Z\} = E\{NF(t)|Z\} \\ &= \gamma e^{\theta_0' Z} F(t)\frac{q}{p} = e^{\theta_0' Z}\Lambda_0(t)\,, \end{aligned}$$

with $\Lambda_0(t) \equiv \gamma F(t)(q/p)$. Alternatively,

$$(\mathbb{N}(t)|Z) \sim \text{Negative Binomial}(r, \frac{p}{p + qF(t)})$$

by straightforward computation, and hence it follows that

$$E\{\mathbb{N}(t)|Z\} = r\frac{qF(t)/(p + qF(t))}{p/(p + qF(t))} = rF(t)\frac{q}{p} = \gamma e^{\theta_0' Z}\frac{q}{p}F(t) = e^{\theta_0' Z}\Lambda_0(t)\,,$$

in agreement with the above calculation. Moreover

$$\begin{aligned} Var\{\mathbb{N}(t)|Z\} &= r\frac{qF(t)/(p + qF(t))}{(p/(p + qF(t)))^2} = r\frac{q}{p}F(t)(1 + \frac{q}{p}F(t)) \\ &= r\frac{q}{p}F(t) + r\left(\frac{q}{p}\right)^2 F(t)^2\,. \end{aligned}$$

Remark. It should be noted that this is somewhat different than the model obtained by supposing that the underlying counting process is a nonhomogeneous Poisson process conditionally given Z and an unobserved frailty $Y \sim \text{Gamma}(\gamma, \gamma)$ and conditional (on Y and Z) mean function

$$E\{\mathbb{N}(t)|Z, Y\} = Ye^{\theta_0' Z}\Lambda_0(t)\,.$$

In this model we have

$$E\{\mathbb{N}(t)|Z\} = e^{\theta_0' Z}\Lambda_0(t)\,,$$

but

$$\begin{aligned} Var\{\mathbb{N}(t)|Z\} &= E\{Var[\mathbb{N}(t)|Y, Z]|Z\} + Var\{E[\mathbb{N}(t)|Y, Z]|Z\} \\ &= e^{\theta_0' Z}\Lambda_0(t) + \gamma^{-1}e^{2\theta_0' Z}\Lambda_0^2(t)\,. \end{aligned}$$

Now we want to calculate

$$
\begin{aligned}
&C(T_{K,j}, T_{K,j'}, T_{K,j-1}, T_{K,j'-1}; Z) \\
&= E\Big[\Big(\Delta\mathbb{N}_{Kj} - e^{\theta_0' Z}\Delta\Lambda_{0Kj}\Big) \\
&\qquad \Big(\Delta\mathbb{N}_{Kj'} - e^{\theta_0' Z}\Delta\Lambda_{0Kj'}\Big)\,|Z, K, T_{K,j-1}, T_{K,j}, T_{K,j'-1}, T_{K,j'}\Big]\,.
\end{aligned}
$$

By computing conditionally on N, and using the fact that conditionally on $Z, K, T_{K,j-1}, T_{K,j}, T_{K,j'-1}, T_{K,j'}$ and N, $\mathbb{N}$ has increments with a joint multinomial distribution, we find that

$$
\begin{aligned}
&C(T_{K,j}, T_{K,j'}, T_{K,j-1}, T_{K,j'-1}; Z) \\
&= E\Big[\Big(\Delta\mathbb{N}_{Kj} - e^{\theta_0' Z}\Delta\Lambda_{0Kj}\Big) \\
&\qquad \Big(\Delta\mathbb{N}_{Kj'} - e^{\theta_0' Z}\Delta\Lambda_{0Kj'}\Big)\Big|Z, K, T_{K,j-1}, T_{K,j}, T_{K,j'-1}, T_{K,j'}\Big] \\
&= E\Big\{E\Big[\Big(\Delta\mathbb{N}_{Kj} - E(\Delta\mathbb{N}_{Kj}|N) + E(\Delta\mathbb{N}_{Kj}|N) - e^{\theta_0' Z}\Delta\Lambda_{0Kj}\Big) \\
&\qquad \cdot\ \Big(\Delta\mathbb{N}_{Kj'} - E(\Delta\mathbb{N}_{Kj'}|N) + E(\Delta\mathbb{N}_{Kj'}|N) - e^{\theta_0' Z}\Delta\Lambda_{0Kj'}\Big) \\
&\qquad\qquad \Big|N, Z, K, T_{K,j-1}, T_{K,j}, T_{K,j'-1}, T_{K,j'}\Big] \\
&\qquad\qquad \Big|Z, K, T_{K,j-1}, T_{K,j}, T_{K,j'-1}, T_{K,j'}\Big\} \\
&= \begin{cases} E\left\{N\Delta F_{Kj}(1-\Delta F_{Kj})|Z, K, \underline{T}_K\right\} & \\ \quad + E\left\{(N - rq/p)^2(\Delta F_{Kj})^2|Z, K, \underline{T}_K\right\}, & \text{if } j = j' \\ -E\left\{N\Delta F_{Kj}\Delta F_{Kj'}|Z, K, \underline{T}_K\right\} & \\ \quad + E\left\{(N - rq/p)^2\Delta F_{Kj}\Delta F_{Kj'}|Z, K, \underline{T}_K\right\}, & \text{if } j \neq j' \end{cases} \\
&= \begin{cases} r\frac{q}{p}\left(\Delta F_{K,j} - (\Delta F_{K,j})^2\right) + r\frac{q}{p^2}\Delta F_{K,j}^2, & \text{if } j = j' \\ \left\{r\frac{q}{p^2} - r\frac{q}{p}\right\}\Delta F_{Kj}\Delta F_{Kj'}, & \text{if } j \neq j' \end{cases} \\
&= \begin{cases} r\frac{q}{p}\Delta F_{K,j} + r\frac{q^2}{p^2}(\Delta F_{K,j})^2, & \text{if } j = j' \\ r\frac{q^2}{p^2}\Delta F_{Kj}\Delta F_{Kj'}, & \text{if } j < j' \end{cases} \\
&= \begin{cases} e^{\theta_0' Z}\Delta\Lambda_{0,K,j} + \gamma^{-1}e^{\theta_0' Z}\Delta(\Lambda_{0,K,j})^2, & \text{if } j = j'\,, \\ \gamma e^{\theta_0' Z}\frac{q^2}{p^2}\Delta F_{Kj}\Delta F_{Kj'}, & \text{if } j \neq j'\,, \end{cases} \\
&= \begin{cases} e^{\theta_0' Z}\Delta\Lambda_{0,K,j} + \gamma^{-1}e^{\theta_0' Z}(\Delta\Lambda_{0,K,j})^2, & \text{if } j = j'\,, \\ \gamma^{-1}e^{\theta_0' Z}\Delta\Lambda_{0Kj}\Delta\Lambda_{0Kj'}, & \text{if } j \neq j'\,. \end{cases}
\end{aligned}
$$

Remark. Note that if $(N|Z) \sim \text{Poisson}(r(Z, \gamma, \theta_0))$, then the process $\mathbb{N} \equiv \mathbb{N}_N$ is conditionally, given Z, a non-homogeneous Poisson process with conditional mean function

$$E\{\mathbb{N}(t)|Z\} = \gamma e^{\theta_0' Z}F(t) = e^{\theta_0' Z}\Lambda_0(t)\,,$$

and conditional variance function

$$Var\{\mathbb{N}(t)|Z\} = \gamma e^{\theta_0' Z}F(t) = e^{\theta_0' Z}\Lambda_0(t)$$

where $\Lambda_0(t) = \gamma F(t)$.

We will also assume, as in Scenarios 1 and 2, that the distribution of $(K, \underline{T}_K)$ is independent of Z. As a consequence, Z is independent of $(K, \underline{T}_K)$, and the formulas in the preceding section again simplify. By taking F uniform on $[0, M]$, $\Lambda_0(t) = \lambda t$ where $\lambda = (\gamma/M)(q/p)$, and we compute, using moments and covariances of uniform spacings, as found on page 721 of [28],

$$
\begin{aligned}
B &= Em^*(\theta_0, \Lambda_0; X)^{\otimes 2} \\
&= E_{(K,T_K,Z)} \left\{ \sum_{j,j'=1}^{K} C(T_{K,j}, T_{K,j'}, T_{K,j-1}, T_{K,j'-1}; Z) \right. \\
&\qquad \left[Z - \frac{E\left(Ze^{\theta_0' Z} | K, T_{K,j-1}, T_{K,j}\right)}{E\left(e^{\theta_0' Z} | K, T_{K,j-1}, T_{K,j}\right)} \right] \\
&\qquad \left. \left[Z - \frac{E\left(Ze^{\theta_0' Z} | K, T_{K,j'-1}, T_{K,j'}\right)}{E\left(e^{\theta_0' Z} | K, T_{K,j'-1}, T_{K,j'}\right)} \right]' \right\} \\
&= E_{(K,T_K,Z)} \left\{ \sum_{j=1}^{K} (e^{\theta_0' Z} \Delta\Lambda_{0Kj} \right. \\
&\qquad \left. + \gamma^{-1} e^{\theta_0' Z} (\Delta\Lambda_{0Kj})^2) \left[Z - \frac{E\left(Ze^{\theta_0' Z}\right)}{E\left(e^{\theta_0' Z}\right)} \right]^{\otimes 2} \right\} \\
&\qquad + E_{(K,T_K,Z)} \left\{ \sum_{j \neq j'}^{K} \gamma^{-1} e^{\theta_0' Z} \Delta\Lambda_{0Kj} \Delta\Lambda_{0Kj'} \left[Z - \frac{E\left(Ze^{\theta_0' Z}\right)}{E\left(e^{\theta_0' Z}\right)} \right]^{\otimes 2} \right\} \\
&= C \left\{ E_{K,T_K} \left\{ \sum_{j=1}^{K} (\Delta\Lambda_{0Kj} + \gamma^{-1} (\Delta\Lambda_{0Kj})^2) \right\} \right. \\
&\qquad \left. + \gamma^{-1} E_{K,T_K} \left\{ \sum_{j \neq j'}^{K} \Delta\Lambda_{0Kj} \Delta\Lambda_{0Kj'} \right\} \right\} \\
&= C \left\{ \lambda M E\left(\frac{K}{K+1}\right) + \gamma^{-1} \lambda^2 M^2 E\left(\frac{K}{K+2}\right) \right\} \\
&= \lambda M C \left\{ E\left(\frac{K}{K+1}\right) + \gamma^{-1} \lambda M E\left(\frac{K}{K+2}\right) \right\}
\end{aligned}
$$

where, as in scenario 1,

$$C \equiv E_Z \left\{ e^{\theta_0' Z} \left[Z - \frac{E\left(Ze^{\theta_0' Z}\right)}{E\left(e^{\theta_0' Z}\right)} \right]^{\otimes 2} \right\} .$$

On the other hand, we find that

$$\begin{aligned} A &= E_{(K,T_K,Z)} \left\{ \sum_{j=1}^{K} \Delta\Lambda_{0Kj} e^{\theta_0' Z} \left[Z - \frac{E\left(Ze^{\theta_0' Z} | K, T_{K,j-1}, T_{K,j}\right)}{E\left(e^{\theta_0' Z} | K, T_{K,j-1}, T_{K,j}\right)} \right]^{\otimes 2} \right\} \\ &= C\lambda M E \left\{ \frac{K}{K+1} \right\} . \end{aligned}$$

Thus the asymptotic variance of the MLE for this scenario is

$$A^{-1} B (A^{-1})' = (\lambda M C)^{-1} \frac{E\left\{\frac{K}{K+1}\right\} + \frac{\lambda M}{\gamma} E\left\{\frac{K}{K+2}\right\}}{\left\{E\left(\frac{K}{K+1}\right)\right\}^2} .$$

Now for the asymptotic variance of the pseudo-MLE under scenario 3. To calculate B^{ps} we first need to calculate

$$\begin{aligned} & C^{ps}(T_{K,j}, T_{K,j'}; Z) \\ &= E\left[\left(\mathbb{N}_{Kj} - e^{\theta_0' Z} \Lambda_{0Kj} \right) \left(\mathbb{N}_{Kj'} - e^{\theta_0' Z} \Lambda_{0Kj'} \right) | Z, K, T_{K,j}, T_{K,j'} \right] \\ &= E\left\{ E\left[\left(\mathbb{N}_{Kj} - e^{\theta_0' Z} \Lambda_{0Kj} \right) \left(\mathbb{N}_{Kj'} - e^{\theta_0' Z} \Lambda_{0Kj'} \right) \right.\right. \\ &\qquad \left.\left. | N, Z, K, T_{K,j}, T_{K,j'} \right] \Big| Z, K, T_{K,j}, T_{K,j'} \right\} \\ &= E\left\{ N (F(T_{K,j} \wedge T_{K,j'}) - F(T_{K,j}) F(T_{K,j'})) \right. \\ &\qquad \left. + (N - rq/p)^2 F(T_{K,j}) F(T_{K,j'}) \Big| Z, K, T_{K,j}, T_{K,j'} \right\} \\ &= r\frac{q}{p} \{ F(T_{K,j} \wedge T_{K,j'}) - F(T_{K,j}) F(T_{K,j}) \} + r\frac{q}{p^2} F(T_{K,j}) F(T_{K,j'}) \\ &= e^{\theta_0' Z} \Lambda_0 (T_{K,j} \wedge T_{K,j'}) + \gamma^{-1} e^{\theta_0' Z} \Lambda_0(T_{K,j}) \Lambda_0(T_{K,j}) \\ &= e^{\theta_0' Z} \Lambda_0(T_{K,j}) \left(1 + \gamma^{-1} \Lambda_0(T_{K,j'}) \right) \qquad \text{if} \;\; j \le j' . \end{aligned}$$

We can then calculate

$$\begin{aligned} B^{ps} &= E m^{*ps}(\theta_0, \Lambda_0; X)^{\otimes 2} \\ &= E_{(K,T_K,Z)} \left\{ \sum_{j,j'=1}^{K} C^{ps}(T_{K,j}, T_{K,j'}; Z) \left[Z - \frac{E\left(Ze^{\theta_0' Z} | K, T_{K,j}\right)}{E\left(e^{\theta_0' Z} | K, T_{K,j}\right)} \right] \right. \end{aligned}$$

$$\left[Z - \frac{E\left(Ze^{\theta_0' Z}|K, T_{K,j'}\right)}{E\left(e^{\theta_0' Z}|K, T_{K,j'}\right)} \right]' \Bigg\}$$

$$= E_{(K,T_K,Z)} \Bigg\{ \sum_{j,j'=1}^{K} \left(\Lambda_0(T_{K,j}) \left(1 + \gamma^{-1}\Lambda_0(T_{K,j'})\right)\right) e^{\theta_0' Z}$$

$$\left[Z - \frac{E\left(Ze^{\theta_0' Z}\right)}{E\left(e^{\theta_0' Z}\right)} \right]^{\otimes 2} \Bigg\}$$

$$= C \left\{ \lambda M E \left(\sum_{j,j'=1}^{K} \frac{j \wedge j'}{k+1} \right) + \frac{\lambda^2 M^2}{\gamma} E \left(\sum_{j,j'}^{K} U_{K,j} U_{K,j'} \right) \right\}$$

$$= C \left\{ \lambda M E \left(\frac{K(2K+1)}{6} \right) + \frac{\lambda^2 M^2}{\gamma} E \left(\sum_{j,j'}^{K} U_{K,j} U_{K,j'} \right) \right\}$$

$$= C \left\{ \lambda M E \left(\frac{K(2K+1)}{6} \right) + \frac{\lambda^2 M^2}{\gamma} E \left(\frac{K(3K+1)}{12} \right) \right\}$$

$$= \lambda M C \left\{ E \left(\frac{K(2K+1)}{6} \right) + \frac{\lambda M}{\gamma} E \left(\frac{K(3K+1)}{12} \right) \right\}$$

$$A^{ps} = E_{(K,T_K,Z)} \left\{ \sum_{j=1}^{K} \Lambda_{0Kj} e^{\theta_0' Z} \left[Z - \frac{E\left(Ze^{\theta_0' Z}|K, T_{K,j}\right)}{E\left(e^{\theta_0' Z}|K, T_{K,j}\right)} \right]^{\otimes 2} \right\}$$

$$= \lambda M C E \left(\frac{K}{2} \right).$$

Thus we find that the asymptotic variance of the pseudo-MLE $\widehat{\theta}_n^{ps}$ is given by

$$(A^{ps})^{-1} B^{ps} ((A^{ps})^{-1})' = (\lambda M C)^{-1} \frac{E\left(\frac{K(2K+1)}{6}\right) + \frac{\lambda M}{\gamma} E\left(\frac{K(3K+1)}{12}\right)}{\left\{E\left(\frac{K}{2}\right)\right\}^2},$$

and the asymptotic relative efficiency of the pseudo-mle to the mle is, under scenario 3,

$$ARE(pseudo, mle)(NegBin)$$

$$= \frac{\frac{E\left\{\frac{K}{K+1}\right\} + \frac{\lambda M}{\gamma} E\left\{\frac{K}{K+2}\right\}}{\left\{E\left(\frac{K}{K+1}\right)\right\}^2}}{\frac{E\left(\frac{K(2K+1)}{6}\right) + \frac{\lambda M}{\gamma} E\left(\frac{K(3K+1)}{12}\right)}{\left\{E\left(\frac{K}{2}\right)\right\}^2}}$$

$$= \frac{E\left\{\frac{K}{K+1}\right\} + \frac{\lambda M}{\gamma} E\left\{\frac{K}{K+2}\right\}}{E\left(\frac{K(2K+1)}{6}\right) + \frac{\lambda M}{\gamma} E\left(\frac{K(3K+1)}{12}\right)} \cdot \left(\frac{E\left(K/2\right))}{E\left(\frac{K}{K+1}\right)}\right)^2$$

$$= \frac{E\left\{\frac{K}{K+1}\right\} + \frac{\lambda M}{\gamma} E\left\{\frac{K}{K+2}\right\}}{E\left(\frac{K}{K+1}\right)} \cdot \frac{E\left(\frac{K(2K+1)}{6}\right)}{E\left(\frac{K(2K+1)}{6}\right) + \frac{\lambda M}{\gamma} E\left(\frac{K(3K+1)}{12}\right)}$$

$$\cdot ARE(pseudo, mle)(Poisson)$$

$$= \frac{\left(1 + \frac{\lambda M}{\gamma} \frac{E\left(\frac{K}{K+2}\right)}{E\left(\frac{K}{K+1}\right)}\right)}{\left(1 + \frac{\lambda M}{\gamma} \frac{E\left(\frac{K(3K+1)}{12}\right)}{E\left(\frac{K(2K+1)}{6}\right)}\right)} \cdot ARE(pseudo, mle)(Poisson)\,.$$

Note that when factor $\lambda M/\gamma = q/p \to 0$, is zero then we recover our earlier result for the Poisson case. This is to be expected since $Poisson(\Lambda_0(t)\exp(\theta_0' Z))$ becomes the limiting distribution of Negative Binomial$(r, p/(p + qF(t)))$ as $p \to 1$.

6 Conclusions and Further Problems

Conclusions

As in the case of panel count data without covariates studied in [29] and [33], the pseudo likelihood estimation method for the semiparametric proportional mean model with panel count data proposed and studied in [35] and [36] also has advantages in terms of computational simplicity. The results of section 5 above show that the maximum pseudo-likelihood estimator of the regression parameters can be very inefficient relative to the maximum likelihood estimator, especially when the distribution of K is heavy - tailed. In such cases it is clear that we will want to avoid the pseudo-likelihood estimator, and the computational effort required by the "full" maximum likelihood estimators can be justified by the consequent gain in efficiency.

Our derivation of the asymptotic normality of the maximum likelihood estimator of the regression parameters results in a relatively complicated expression for the asymptotic variance which may be difficult to estimate directly. Hence it becomes important to develop efficient algorithms for computation of the maximum likelihood estimator in order to allow implementation, for example, of bootstrap inference procedures. Alternatively, profile likelihood inference may be quite feasible in this model; see e.g. [24], [25], [26] for likelihood ratio procedures in some related interval censoring models.

Further problems

The asymptotic normality results stated in section 4 will be developed and given in detail in [34]. There are quite a large number of interesting problems still open concerning the semiparametric model for panel count data which has been studied here. Here is a short list:

- Find a fast and reliable *algorithm* for computation of the MLE $\widehat{\theta}$ of θ. Although reasonable algorithms for computation of the maximum pseudo-likelihood estimators have been proposed in [35] and [36] based on the earlier work of [29], good algorithms for computation of the maximum likelihood estimators have yet to be developed and implemented.
- Show that the natural semiparametric profile likelihood ratio procedures are valid for inference about the regression parameter θ via the theorems of [24], [25], and [26].
- Do the non-standard likelihood ratio procdures and methods of [1] extend to the present model to give tests and confidence intervals for $\Lambda_0(t)$?
- Are there compromise or hybrid estimators between the maximum pseudo-likelihood estimators and the full maximum likelihood estimators which have the computational advantages of the former and the efficiency advantages of the latter?
- Do similar results continue to hold for panel count data with covariates, but with other models for the mean function replacing the proportional mean model given by (1)?

• Are there computational or efficiency advantages to using the MLE's for one of the class of Mixed Poisson Process $(\mathbb{N}|Z)$, for example the Negative-Binomial model? Further comparisons with the work of [6], [14], and [20], [21] would be useful.

References

[1] Banerjee, M. and Wellner, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29**, 1699 - 1731.

[2] Begun, J. M., Hall, W. J., Huang, W.M., and Wellner, J. A. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.* **11**, 432 - 452.

[3] Betensky, R.A., Rabinowitz, D., and Tsiatis, A. A. (2001). Computationally simple accelerated failure time regression for interval censored data. *Biometrika* *88*, 703-711.

[4] Bickel, P. J., Klaassen, C.A.J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins University Press, Baltimore.

[5] Cook, R. J., Lawless, J. F., and Nadeau, C. (1996). Robust tests for treatment comparisons based on recurrent event responses. *Biometrics* **52**, 557 - 571.

[6] Dean, C. B. and Balshaw, R. (1997). Efficiency lost by analyzing counts rather than event times in Poisson and overdispersed Poisson regression models. *J. Amer. Statist. Assoc.* **92**, 1387-1398.

[7] Gaver, D. P., and O'Muircheartaigh, I. G. (1987). Robust Empirical Bayes analysis of event rates, *Technometrics,* **29**, 1-15.

[8] Geskus, R. and Groeneboom, P. (1996). Asymptotically optimal estimation of smooth functionals for interval censoring, part 1. *Statist. Neerlandica* **50**, 69 - 88.

[9] Geskus, R. and Groeneboom, P. (1997). Asymptotically optimal estimation of smooth functionals for interval censoring, part 2. *Statist. Neerlandica* **51**, 201-219.

[10] Geskus, R. and Groeneboom, P. (1999). Asymptotically optimal estimation of smooth functionals for interval censoring case 2. *Ann. Statist.* **27**, 626 - 674.

[11] Groeneboom, P. (1991). Nonparametric maximum likelihood estimators for interval censoring and deconvolution. *Technical Report 378*, Department of Statistics, Stanford University.

[12] Groeneboom, P. (1996). Inverse problems in statistics. Proceedings of the St. Flour Summer School in Probability, 1994. *Lecture Notes in Math.* **1648**, 67 - 164. Springer Verlag, Berlin.

[13] Groeneboom, P. and Wellner, J. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation.* Birkhäuser, Basel.

[14] Hougaard, P., Lee, M.T., and Whitmore, G. A. (1997). Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes. *Biometrics* **53**, 1225 - 1238.

[15] Huang, J. (1996). Efficient estimation for the Cox model with interval censoring, *Annals of Statistics,* **24**, 540-568.

[16] Huang, J., and Wellner, J. A. (1995). Efficient estimation for the Cox model with case 2 interval censoring, *Technical Report* **290**, University of Washington Department of Statistics, 35 pages.

[17] Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *Journal of Computation and Graphical Statistics* **7**, 310-321.

[18] Kalbfleisch, J. D. and Lawless, J. F. (1981). Statistical inference for observational plans arising in the study of life history processes. In *Symposium on Statistical Inference and Applications In Honour of George Barnard's 65th Birthday.* University of Waterloo, August 5-8, 1981.

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

[20] Lawless, J. F. (1987a). Regression methods for Poisson process data. *J. Amer. Statist. Assoc.* **82**, 808-815.

[21] Lawless, J. F. (1987b). Negative binomial and mixed Poisson regression. *Canad. J. Statist.* **15**, 209-225.

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

[23] 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. B*, 711-730.

[24] Murphy, S. and Van der Vaart, A. W. (1997). Semiparametric likelihood ratio inference. *Ann. Statist.* **25**, 1471 - 1509.

[25] Murphy, S. and Van der Vaart, A. W. (1999). Observed information in semiparametric models. *Bernoulli* **5**, 381–412.

[26] Murphy, S. and Van der Vaart, A. W. (2000). On profile likelihood. *J. Amer. Statist. Assoc.* **95**, 449 - 485.

[27] Rabinowitz, D., Betensky, R. A., and Tsiatis, A. A. (2000). Using conditional logistic regression to fit proportional odds models to interval censored data. *Biometrics* **56**, 511-518.

[28] Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

[29] Sun, J. and Kalbfleisch, J. D. (1995). Estimation of the mean function of point processes based on panel count data. *Statistica Sinica* **5**, 279 - 290.

[30] Sun, J. and Wei, L. J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *J. R. Stat. Soc. Ser. B* **62**, 293 - 302.

[31] Thall, P. F., and Lachin, J. M. (1988). Analysis of Recurrent Events: Nonparametric Methods for Random-Interval Count Data. *J. Amer. Statist. Assoc.* **83**, 339-347.

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

[33] 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.

[34] Wellner, J. A., Zhang, Y., and Liu (2002). Large sample theory for two estimators in a semiparametric model for panel count data. Manuscript in progress.

[35] Zhang, Y. (1998). Estimation for Counting Processes Based on Incomplete Data. Unpublished Ph.D. dissertation, University of Washington.

[36] Zhang, Y. (2002). A semiparametric pseudo likelihood estimation method for panel count data. *Biometrika* **89**, 39 - 48.

University of Washington
Statistics
Box 354322
Seattle, Washington 98195-4322
U.S.A.
e-mail: jaw@stat.washington.edu

Department of Statistics
University of Central Florida
P.O. Box 162370
Orlando, Florida 32816-2370
e-mail: zhang@pegasus.cc.ucf.edu

University of Washington
Biostatistics
Box 357232
Seattle, Washington 98195-7232
U.S.A.
e-mail: hliu@u.washington.edu