Regression Parameter Estimation from Panel Counts
Author(s): X. Joan Hu, Jianguo Sun, Lee-Jen Wei
Source: *Scandinavian Journal of Statistics*, Vol. 30, No. 1 (Mar., 2003), pp. 25-43
Published by: Blackwell Publishing on behalf of Board of the Foundation of the Scandinavian Journal of Statistics
Stable URL: http://www.jstor.org/stable/4616747

 Vol 30: 25–43, 2003

Regression Parameter Estimation from Panel Counts

X. JOAN HU

University of Memphis

JIANGUO SUN

University of Missouri-Columbia

LEE-JEN WEI

Harvard School of Public Health

ABSTRACT. This paper considers a study where each subject may experience multiple occurrences of an event and the rate of the event occurrences is of primary interest. Specifically, we are concerned with the situations where, for each subject, there are only records of the accumulated counts for the event occurrences at a finite number of time points over the study period. Sets of observation times may vary from subject to subject and differ between groups. We model the mean of the event occurrence number over time semiparametrically, and estimate the regression parameter. The proposed estimation procedures are illustrated with data from a bladder cancer study (Byar, 1980). Both asymptotics and simulation studies on the estimators are presented.

Key words: estimation equation, recurrent events, semiparametric regression model

1. Introduction

We consider a longitudinal study, such as a clinical trial or an industrial experiment, where the primary endpoint is occurrence of a specific event and each subject may experience the event repeatedly over time. There are well-developed methods for analysing recurrent event data when time to each event occurrence is observed, such as those reported in Prentice *et al.* (1981), Andersen & Gill (1982), Wei *et al.* (1989), Pepe & Cai (1993), Lawless & Nadeau (1995). This paper is concerned with the situations where observations on each subject are taken at several distinct time points, and thus only at the time points are accumulated counts of the event occurrences, as well as covariates, recorded. Data of this type, often referred to as panel counts, arise frequently in practice. For example, there may only exist records on the cumulative numbers of asthma attacks and other related information of a patient at his clinical visits.

There have been some notable studies on panel counts data. Thall & Lachin (1988), for example, considered the two-sample testing problem when subjects have the same set of observation times. Sometimes sets of observation time points may vary from subject to subject and may also depend on covariates. For example, in a bladder cancer study conducted by the Veterans Administration Cooperative Urological Research Group of USA, all subjects had superficial bladder tumors when they entered the trial (Byar, 1980). These tumors were removed transurethrally and then subjects were randomly allocated to one of the three treatments, placebo, thiotepa and pyridoxine. Many subjects had multiple new tumors during the study. The new tumors were removed at the clinical visits of the subjects. Tables 1 and 2 present the numbers of new tumors discovered at those visits for the subjects in the placebo

Table 1. *Number of new tumors at each visit for subjects in placebo group*

Patient ID	Initial number	Initial size	Month since entry: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48
1	1	3	0 . . . .
2	2	1	0 . . 0 . . . .
3	1	1	. . . . . . 0 . . . .
4	5	1	. . 0 . . . . . 0 0 . . . .
5	4	1	0 . . 0 . 1 0 . . 0 . . . .
6	1	1	. . 0 . . . . . . 0 . . . 0 . . . .
7	1	1	. 0 . . . . 0 . 2 . . . 3 . 0 . . . .
8	1	1	. . 0 . . . . 0 . . . 0 . . . .
9	1	3	. . . . 2 . . . . 0 . 0 . . . . . 0 . . . .
10	1	3	. . 0 . . . 0 . . 6 . . . . 3 . . . 0 . . . 0 . . . .
11	1	1	0 . 8 . . . . 0 . 0 . . 0 . . 8 . . 0 . . . 8 . . . .
12	3	1	. . 1 . . 0 . . 1 0 . . 0 . 0 . 0 . . 0 8 . 0 . . . .
13	3	3	. . 0 . . 0 . . 0 . 0 . . 0 . . . 0 . . . . 0 . . . .
14	2	3	. . 0 . . . 8 . . 7 . . 0 . . 5 . . . . 7 . . . .
15	1	1	. . 1 . . 0 . . 0 . . 0 . . 1 . 0 . . 0 . 0 . . 3 . . . .
16	8	1	8 . . 0 . . 0 . . . 0 . . 0 . . 0 . . 0 . . . . 0 0 . . . .
17	1	4	. 4 . . . 0 . . . . 8 . . . .
18	1	2	. . 0 . . 0 . . . . 0 . . 0 . . . . . 0 0 . . 0 . . . .
19	1	2	. . . . . 0 . . . . 3 . . 0 . . . .
20	1	4	. . . . 0 . . 0 . . . . 0 . . . . 0 . . . .
21	1	2	. . 0 . . 0 . . . . 0 . . . . 0 . . . .
22	4	1	. 0 . . 0 . . 0 . . 0 . . . . 0 . . . . . . 0 . . . . 0 . . . .
23	1	5	. 4 . . . . 0 . . . . 2 . . . . 4 0 . . . . 0 . 0 . . . .
24	2	1	. . 1 . . 3 . 3 . . . 3 0 . 0 0 0 0 . . . 0 0 . . 3 . . . 0 . . . .
25	1	6	. 0 . . 0 0 . . 0 . . . 0 . . . . 0 . . . . . 2 . 1 . . . .
26	1	3	0 . 0 0 . 0 . . 0 . . 2 . . 3 . . 0 . . . . . . 1 . . . 0 . . 0 . . . .
27	1	2	. . 0 . . 0 . . 0 . . 0 . 0 . . 0 . . 0 . . . . . . 0 . . . . . . 0 . . . .
28	2	1	. . . 0 . . . 0 . 0 . . . 0 . . . . 0 . . . . 0 0 . 0 0 . . . .
29	2	1	. . 0 . . 0 . . 0 . 0 . . 0 . . 0 . . 0 . . 0 . . . . 0 . . . . . . 0 . . . .
30	3	1	. . . . 0 . . . . 0 . . . . 8 . . . . 0 . . . .
31	1	2	0 . . 0 . . . . . . 0 . . . . 0 . . . . 0 . . . . 0 . . . . 0 . . . .
32	4	1	. 0 . . . . . . 8 . . . . 0 . . 2 . . . . 5 . 1 . . . 0 . . 0 . . . . 0 . . . 0 . . . .
33	5	1	. 0 . 0 0 . . 0 . . 0 . . . 0 1 . . 8 . . . 1 . . 0 . . 2 . . 0 . 1 . . 0 . . 3 . . . .
34	1	2	. . 0 . . 0 . . 0 . 0 . . 0 . . 0 . . 0 . . 0 . . 0 . . 0 . . . . . 0 0 . . . . 0 . . . .
35	1	1	. . 3 . 0 . 0 . . 0 . . . 0 . . . . . 0 . . . . 0 . . 0 . . . 0 . . 0 . . . . . . 0 . . . . . .
36	2	6	. . 0 . . 1 0 0 . . 0 . . 0 . 0 0 . . 0 . . 0 . . . . 0 . . 0 . . . . 0 . . . . .
37	2	1	. . 5 . 0 3 . . 4 . . . . 0 . . 0 . . 0 . . 0 . . . 0 . . 0 . . 0 . . 0 . . 0 . 0 . . 0 . . . .
38	1	1	. 0 . . . 0 . . 1 . 3 . 0 . . . 0 . . 1 . . 0 . . 4 . . . 3 . 0 . . 0 . . . 0 . . 0 . . 0 . . .
39	1	1	. . 0 . . 0 . . 0 . 0 . . . 0 . . 1 . . . 0 . . 0 . . 0 . . 0 . . 0 . . 0 . 0 . . . . . 0 . . 0
40	1	3	. . . . . 0 . 0 . . . 0 . . . . 0 . . . . . 0 . . . . . . 0 . . . . . 0 . . . . . 0 . . . 0 . . . .
41	3	1	. 0 . . 0 . . . . . . 0 . . . . . . 0 . . . . . . 0 . . . . . . 0 . . . . . . 1 . . . . 0 . . . 0 .
42	1	7	. . 0 . . . . . . 0 . . 0 0 . . 1 . . . 0 . . 0 . . 0 . . 0 . . 0 . 0 . 0 . . . 0 . . 0 . . 0 .
43	3	1	. . 7 . . . . 0 . . . 2 . . . . 0 . . . . 0 . . . 0 . . . . . 0 . . . . 3 . . .
44	1	1	0 . . 0 . . 0 . . 0 . . 0 . . 0 . . 0 . . 0 . . . . . . 0 . . . . 0 . . . . . . 0 . 0 . . . 0 . .
45	3	2	. 1 . . 0 . 0 . . 0 . . 0 . 3 . . . 0 . . . . 4 . . 0 . . 3 . . . 4 . . 0 . 1 . . . 1 . . 0 . .
46	1	3	. 0 . . 3 . . 0 . . 0 . . 4 . . . 0 2 0 . . 0 . . . 5 . 0 . . . 0 . . 0 . . . . 9 . . 0 0 0 . .
47	2	3	. 1 . . 0 . . 3 . . . 6 2 . . . 2 . . . 1 . 0 0 . . 0 . . 0 . . 1 . . . 0 . . . 0 . . . 0 . . .

and thiotepa groups, while for each subject time points without a visit are marked by dots. We envision that we would observe a counting process for each subject's tumor accumulation if we could monitor closely and record the tumor occurrences in real time. Yet, in reality only the

Table 2. *Number of new tumors at each visit for subjects in thiotepa group*

			Month since entry
Patient ID	Initial number	Initial size	1 2 3 4 5 6 7 8 9 10 11 12 13 14 16 18 20 21 22 24 26 28 30 32 34 36 38 40 42 44 46 48
48	1	3	0..
49	1	1	0..
50	8	1	....8...
51	1	2	000000000..
52	1	1	.....0...0..
53	1	1	..0.........0...
54	2	6	..1.0...0.0..0...
55	5	3	5.2.5.2..2..0..00.....................................
56	1	3	....0.....0.....20....................................
57	5	1	..................0.....................................
58	5	1	020................0....................................
59	1	1	000000000...00001.1.0................................
60	1	1	.0..0....0....0..0...0.................................
61	1	3	000....0...0.....0..0...0............................
62	1	5	0.......0.0.......0.....0..............................
63	1	1	0000000000.0.00..00..00.0...........................
64	1	1	000002...0031......0..0..0..........................
65	1	1	0.0..1...0..0..0....0.....0..........................
66	2	1	.2............................0.........................
67	8	3	..0..0..0..........0.00..3..0..0..30................
68	1	1	00000000000000.0000.000.00............0............
69	6	1	0001..00000000.3.00.003.003.0...8.098.0...........
70	1	1	0000000000000.000.000.2100.20...3.0....0...........
71	3	1	....0........0..0......3.2..1..........2.............
72	3	2	00000000.00000000000000.0000000000.00...0.......
73	1	1	0000..0000000.0000000000000000.00000.....0.......
74	1	1	1.0...0.....0...000.0000.01.00.00..00.....0.......
75	1	1	.......0................0....................0.........
76	6	1	02.0.00.0000000000001..2...10.0.......8..0..0......
77	1	2	..0..0..0..0..0.....0.0...0.....0.....0.....0.......
78	1	4	010..0000.0000000000000000..0.....0.....0.....0....
79	1	4	000000000000000000000000000000.00000........0..
80	3	3	..........0...0........0....0............0............
81	4	1	...1...0....0.......0..1......0.0..0......0...1.....
82	1	1	00..................0....0......0.....0.....0..0.....
83	2	1	0.0000.0..0.00..0..0..0..0..0..0.....2..0...0..0...
84	3	4	00000000000.0000000000000000000000.0..0..0..0..0
85	1	3	0000000000000.0.00.000.000000000000.....0.......

counting process is observed at a finite number of time points. Two potentially important covariates, the number of initial tumors and the size of the largest initial tumor of each subject, are also shown in the tables. It is noticeable, by comparing the two tables, that the subjects in the thiotepa group tended to visit the clinics more often than those in the placebo group. This is probably due to the fact that thiotepa had to be instilled in the bladder and, therefore, the subjects in this group needed to visit the clinics more. As indicated in Byar (1980), one of the study objectives was to evaluate the effectiveness of thiotepa based on comparison of the placebo and the thiotepa groups in tumor accumulation. The different observation patterns should be taken into account in the treatment evaluation. The number of observation times is large, compared to the sample size, and subjects within each group seem to have various observation times. This creates concerns about efficiency and implementation of the commonly used methods, such as the generalized estimation equation approach (GEE, cf. Diggle *et al.* 1994).

Motivated by the bladder cancer data, we consider a semiparametric regression model and estimate the regression parameter with the baseline function unspecified. Learning from the structure of the estimation equation based on the Cox partial score function for Poisson processes (see Andersen & Gill, 1982; Lawless & Nadeau, 1995), we construct two estimation equations. Our approaches allow for imbalanced and different observation mechanisms among subjects and between groups. This work provides an alternative to the existing approaches as discussed in Lawless & Zhan (1998) and Sun & Wei (2000). Assuming piecewise constant rate functions, Lawless & Zhan (1998) considered a likelihood based approach and an extension of GEE. The approach given by Sun & Wei (2000) may be viewed as a special case of our second approach, which is presented in section 2.3. The methods and associated discussions in this paper apply quite generally. With little modification, for example, our methods can be used in the situations where the response process is not necessarily a counting process. To focus, we concentrate on the panel counts data that motivated the study. The remainder of this paper is organized as follows. Section 2 proposes two estimation procedures for the regression parameter and presents the asymptotics properties. In section 3, the proposed methods are illustrated with the bladder tumor data. Section 4 presents a simulation study on the finite sample performances of the two estimators. Some remarks are given in section 5.

2. Regression estimation

We begin by introducing the notation and describing the general setting. Two estimation equation based approaches to estimating the regression parameter along with their asymptotics properties are then presented.

2.1. Notation and setting

Suppose that a counting process $\{N_i(t), t \geqslant 0\}$ associated with subject i from a group of independent individuals, $i = 1, \ldots, n$, is potentially observed only at finite time points ξ_{il}: $0 < \xi_{i1} < \xi_{i2} < \cdots < \xi_{iK_i} < \infty$, where ξ_{il} and K_i are random. Here $N_i(t)$ represents the number of accumulated occurrences of a specific event that subject i has experienced up to time t since his entry to a study at $t = 0$. Without loss of generality, assume that $N_i(0) = 0$. In the bladder cancer study described in the previous section, for example, $N_i(t)$ is the cumulative number of new tumors of subject i at study time t, and ξ_{il} is the time of his lth clinical visit, $l = 1, \ldots, K_i$. The objective is to assess the effects of covariates on the event occurrences. To avoid strong assumptions about the counting process, we consider a "marginal" model for the means or the rates of the event occurrences as follows.

Suppose that conditional on a p-dimensional covariate Z_i, the expectation of $N_i(t)$ is

$$\mathrm{E}\{N_i(t)|Z_i = z_i\} = \Lambda_0(t)\exp(\beta' z_i), \quad t \geqslant 0, \tag{1}$$

where the baseline $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ is unspecified, and the parameter β is p-dimensional. Notice that if $\{N_i(t), t \geqslant 0\}$ is a Poisson process, its intensity function conditional on the covariate is the same as the conditional rate function corresponding to (1), that is, $\mathrm{E}\{dN_i(t)|Z_i = z_i\} = \lambda_0(t)\exp(\beta' z_i)dt$, $t \geqslant 0$. See Lawless (1995) and Lawless & Nadeau (1995) for more discussion on this model.

It is of primary interest in this paper to estimate the regression parameter β from the panel counts data along with observed covariates:

$$\{[(\xi_{il}, N_i(\xi_{il})) : l = 1, \ldots, K_i; z_i] : i = 1, \ldots, n\}. \tag{2}$$

We denote the accumulated number of observations on subject i up to time t by $O_i(t)$, that is, $O_i(t)$ is the counting process

$$O_i(t) = \sum_{l=1}^{K_i} \mathrm{I}(\xi_{il} \leqslant t), \quad t \geqslant 0,$$

where $\mathrm{I}(A)$ is the indicator of set A. Let $o_i(t) = O_i(t) - O_i(t^-)$ so that $o_i(t)$ indicates whether subject i has an observation at time t. Suppose that c_i, $i = 1, \ldots, n$, are censoring times associated with the subjects, and that the c_is are observed values of random variables $C_1, \ldots, C_n$ generated from a distribution $G(\cdot)$ on $[0, \tau]$ with $\lim_{t \uparrow \tau} G(t) < 1$, where $0 < \tau < \infty$. We assume that the censoring distribution has a mass at the maximum point τ. Note that the condition on $G(\cdot)$ is not essential but without it we have to spend a substantial portion in the study of asymptotics to control the so-called tail instability, see the related discussion in Tsiatis (1981) and Bilias *et al.* (1997). In practice, in order to satisfy the constraint, we may choose, artificially, a finite value which is close to but smaller than the maximum of observed censoring times, and use the value to approximate the censoring times beyond it. Using this value as τ will not result in much efficiency loss in analysing the data.

This paper supposes that $\{N_i(\cdot), O_i(\cdot), Z_i, C_i\}$, $i = 1, \ldots, n$, are i.i.d. Suppose further that $N_i(\tau)$ and $O_i(\tau)$, along with $\mathrm{E}\{Z_i^{\otimes j} \exp(\beta' Z_i)\}$, for $j = 0, 1, 2$, are bounded, where $Z_i^{\otimes 0} = 1$, $Z_i^{\otimes 1} = Z_i$, and $Z_i^{\otimes 2} = Z_i Z_i'$. A crucial assumption we make in this paper is that $N_i(\cdot)$, $O_i(\cdot)$, and the censoring time C_i are independent, given the covariate Z_i. This assumption is plausible in the bladder cancer example. Firstly, the censoring was mainly due to the subjects' staggered entries and there were no informative drop-outs. Secondly, the model assumes that the cumulative new tumor numbers and the clinical visits are independent for subjects within each treatment and with the same initial condition, when the covariate includes the treatment indicator and the initial condition. Tumor accumulation, especially with small amount in the case of the bladder cancer example, unlikely affects the pattern of subjects' clinical visits. There is a discussion in section 5 on the situations where the independence assumption is violated and on a diagnosis for the assumption.

2.2. *Estimation conditional on observation process*

To begin with, we form a new process associated with subject i based on the data of (2):

$$\tilde{N}_i(t) = \int_0^t N_i(s) dO_i(s), \quad t \geqslant 0.$$

This newly defined process only has possible jumps at the observation time points $\{\xi_{il} : l = 1, \ldots, K_i\}$ with respective jump sizes $N_i(\xi_{il})$. It is available under the current observation mechanism up to censoring. The process satisfies

$$\mathrm{E}\{d\tilde{N}_i(t) | O_i(s), 0 < s \leqslant t; Z_i = z_i\} = \Lambda_0(t) \exp(\beta' z_i) dO_i(t). \tag{3}$$

This leads to the following approach.

In the following we assume $\mathrm{E}\{o_i(t)\} = p(t) > 0$ for $t \in \mathscr{T} \subseteq [0, \tau]$, where $\mathscr{T}$ is not empty. This ensures that there is more than one subject having observation at some time points when the study size n is large enough. Note that the set $\mathscr{T}$ is finite in the situations considered in this paper. Define

$$S_C^{(j)}(\beta; t) = \frac{\sum_{i=1}^{n} \mathrm{I}(c_i \geqslant t) z_i^{\otimes j} \exp(\beta' z_i) o_i(t)}{\sum_{i=1}^{n} o_i(t)}$$

for t with $\sum_{i=1}^{n} o_i(t) > 0$ and $j = 0, 1, 2$. If $\{\tilde{N}_i(t), t \geqslant 0\}$ were Poisson, we would consider the corresponding Cox partial score function to estimate β (see Andersen & Gill, 1982; Lawless & Nadeau, 1995). Borrowing the structure of the Cox partial score function, we construct an estimating function of β in the form of

$$U_n^C(\beta; \tilde{N}, w) = \sum_{i=1}^{n} \int_0^{\tau} w(t) \mathrm{I}(c_i \geqslant t)\{z_i - \bar{Z}_C(t; \beta)\} d\tilde{N}_i(t) \tag{4}$$

with $w(\cdot)$ being a weight function and $\bar{Z}_C(t;\beta) = S_C^{(1)}(\beta; t)/S_C^{(0)}(\beta; t)$. Here $\bar{Z}_C(t;\beta)$ is well-defined at $t \in [0, \tau]$ when $\sum_{i=1}^{n} o_i(t) > 0$, which is sufficient for (4) and other integrals involving $\bar{Z}_C(t;\beta)$ in the remaining paper. Since $\mathscr{T}$ is a finite set, the integral in (4) and similar integrals below are finite sums.

The estimating function is evaluable from the data in (2). It is easy to verify that, for a general process satisfying (3), the estimation equation $U_n^C(\beta; \tilde{N}, w) = 0$ is still unbiased for β. Solving this estimation equation provides us with an estimator of β, denoted by $\hat{\beta}_n^C$. Notice that, in the simple situation where every subject only has observation at time $T_0 < \tau$ and is censored at τ, with a constant weight function, the estimating function (4) reduces to

$$U_n^C(\beta; \tilde{N}, 1) = \sum_{i=1}^{n} z_i N_i(T_0) - \left\{ \sum_{i=1}^{n} \int_0^{T_0} \frac{1}{\sum_{j=1}^{n} \exp(\beta' z_j)} dN_i(t) \right\} \times \left\{ \sum_{i=1}^{n} z_i \exp(\beta' z_i) \right\},$$

which may be viewed as the estimating function obtained through approximating $\sum_{i=1}^{n} z_i N_i(T_0)$ by its estimated expectation conditional on $Z_i = z_i$s, $\hat{\Lambda}_0(T_0) \sum_{i=1}^{n} z_i \exp(\beta' z_i)$ with $\hat{\Lambda}_0(\cdot)$ being the Breslow estimator of $\Lambda_0(\cdot)$ (see Fleming & Harrington, 1991, ch. 4).

The following discussion focuses on $w(\cdot) = 1$ for $t \in [0, \tau]$. Results presented below can be straightforwardly extended to situations with any deterministic weight function.

In appendix A, we show that the estimator $\hat{\beta}_n^C$ is consistent and asymptotically normal. The asymptotical covariance matrix of $\sqrt{n}(\hat{\beta}_n^C - \beta)$ is $\Sigma_C = A_C(\beta)^{-1} B_C(\beta) A_C(\beta)^{-1}$, where

$$A_C(\beta) = -\mathrm{E}\left\{ \int_0^{\tau} \mathrm{I}(C_1 \geqslant t) \left[\frac{s_C^{(2)}(\beta; t)}{s_C^{(0)}(\beta; t)} - \bar{z}_C(t; \beta)^{\otimes 2} \right] d\tilde{N}_1(t) \right\} \tag{5}$$

and

$$B_C(\beta) = \mathrm{E}\left\{ \left[\int_0^{\tau} [Z_1 - \bar{z}_C(t; \beta)] dM_1^C(t; \beta) \right]^{\otimes 2} \right\} = \mathrm{var}\left\{ \int_0^{\tau} [Z_1 - \bar{z}_C(t; \beta)] dM_1^C(t; \beta) \right\} \tag{6}$$

with

$$s_C^{(j)}(\beta; t) = \frac{1}{p(t)} \mathrm{E}\left\{ \mathrm{I}(C_1 \geqslant t) Z_1^{\otimes j} \exp(\beta' Z_1) o_1(t) \right\}$$

for $j = 0, 1, 2$,

$$\bar{z}_C(t; \beta) = \lim_{n \to \infty} \bar{Z}_C(t; \beta) = \frac{s_C^{(1)}(\beta; t)}{s_C^{(0)}(\beta; t)}$$

almost surely, and

$$M_i^C(t; \beta) = \int_0^t \mathrm{I}(c_i \geqslant s)[d\tilde{N}_i(s) - \Lambda_0(s) \exp(\beta' z_i)\, dO_i(s)]$$

for $t \in [0, \tau]$. In the above, we assign zero to $s_C^{(2)}(\beta; t)/s_C^{(0)}(\beta; t)$ and $\bar{z}_C(t; \beta)$ when $\mathrm{E}\big[\mathrm{I}(C_1 \geqslant t) \exp(\beta' Z_1) o_1(t)\big] = 0$.

The covariance matrix Σ_C is consistently estimated by

$$\hat{\Sigma}_C = \hat{A}_C(\beta)^{-1}\hat{B}_C(\beta)\hat{A}_C(\beta)^{-1} \tag{7}$$

with the parameter β substituted by its estimate $\hat{\beta}_n^C$, where

$$\hat{A}_C(\beta) = \frac{1}{n}\frac{\partial U_n^C(\beta; \tilde{N}, 1)}{\partial \beta} = -\frac{1}{n}\sum_{i=1}^n \int_0^\tau \mathrm{I}(c_i \geqslant t)\left[\frac{S_C^{(2)}(\beta; t)}{S_C^{(0)}(\beta; t)} - \bar{Z}_C(t; \beta)^{\otimes 2}\right] d\tilde{N}_i(t) \tag{8}$$

and

$$\hat{B}_C(\beta) = \frac{1}{n}\left[\sum_{i=1}^n \int_0^\tau (z_i - \bar{Z}_C(t; \beta))\, d\hat{M}_i^C(t; \beta)\right]^{\otimes 2} \tag{9}$$

with

$$\hat{\Lambda}_0^C(t; \beta) = \frac{\sum_{i=1}^n \mathrm{I}(c_i \geqslant t) N_i(t) o_i(t)}{\sum_{i=1}^n \mathrm{I}(c_i \geqslant t)\exp(\beta' z_i) o_i(t)} \tag{10}$$

and

$$\hat{M}_i^C(t; \beta) = \int_0^t \mathrm{I}(c_i \geqslant s)[N_i(s) - \hat{\Lambda}_0^C(s; \beta)\exp(\beta' z_i)]\, dO_i(s)$$

for $t \in [0, \tau]$. The previous comment on the definition of $\bar{Z}_C(t; \beta)$ applies to $\hat{\Lambda}_0^C(t; \beta)$.

We remark that, to employ this approach in the continuous time situations, we may need to discretize the time scale to meet the requirement of $\mathrm{E}[o_i(t)] = p(t) > 0$ at some time points in $[0, \tau]$.

2.3. Estimation by modelling observation process

An alternative to the above approach is obtained by modelling the observation process. Suppose that the observation process follows the model

$$\mathrm{E}\{dO_i(t) | Z_i = z_i\} = \exp(\alpha' z_i) d\mu_0(t), \tag{11}$$

where $\mu_0(\cdot)$ is either discrete or absolutely continuous with respect to Lebesgue measure. Combining models (1) and (11), we have

$$\mathrm{E}\{d\tilde{N}_i(t) | Z_i = z_i\} = \exp(\tilde{\beta}' z_i) d\tilde{\Lambda}_0(t), \quad t \geqslant 0 \tag{12}$$

with $\tilde{\Lambda}_0(t) = \int_0^t \Lambda_0(s) d\mu_0(s)$ and $\tilde{\beta} = \beta + \alpha$.

The Cox partial score function of $\tilde{\beta}$, if $\{\tilde{N}_i(t), t \geqslant 0\}$ were Poisson, would be

$$\sum_{i=1}^n \int_0^\tau \mathrm{I}(c_i \geqslant t)\left\{z_i - \bar{Z}_M(t; \tilde{\beta})\right\} d\tilde{N}_i(t),$$

where

$$S_M^{(j)}(\tilde{\beta}; t) = \frac{1}{n}\sum_{i=1}^n \mathrm{I}(c_i \geqslant t) z_i^{\otimes j} \exp(\tilde{\beta}' z_i)$$

for $j = 0, 1, 2$, and $\bar{Z}_M(t; \tilde{\beta}) = S_M^{(1)}(\tilde{\beta}; t)/S_M^{(0)}(\tilde{\beta}; t)$. This suggests an unbiased estimation equation for $\tilde{\beta}$, $U_n^M(\tilde{\beta}; \tilde{N}, w) = 0$, where

$$U_n^M(\tilde{\beta}; \tilde{N}, w) = \sum_{i=1}^n \int_0^\tau w(t) \mathrm{I}(c_i \geqslant t)\left\{z_i - \bar{Z}_M(t; \tilde{\beta})\right\} d\tilde{N}_i(t) \tag{13}$$

with $w(\cdot)$ a weight function. Solving the estimation equation gives an estimator for $\tilde{\beta}$, denoted by $\hat{\tilde{\beta}}_n$. An estimator for the regression parameter β is then obtained as given by $\hat{\beta}_n^M = \hat{\tilde{\beta}}_n - \alpha$. In practice, α is usually unknown. We can consider to substitute α by the estimator $\hat{\alpha}_n$ obtained by solving an unbiased estimation equation, $U_n^M(\alpha; O, w_O) = 0$, where

$$U_n^M(\alpha; O, w_O) = \sum_{i=1}^n \int_0^\tau w_O(t)\mathrm{I}(c_i \geqslant t)\{z_i - \bar{Z}_M(t; \alpha)\} dO_i(t).$$

The weight function $w_O(\cdot)$ could be different from the weight in $U_n^M(\tilde{\beta}; \tilde{N}, w)$ of (13).

The estimating function (13) could be viewed as a generalization of a reasonable estimating function

$$\sum_{i=1}^n z_i \int_0^\tau N_i(t)dt - \Big\{\sum_{i=1}^n z_i \exp(\beta' z_i)\Big\} \times \int_0^\tau \frac{\sum_{l=1}^n N_l(t)}{\sum_{j=1}^n \exp(\beta' z_j)} dt$$

for the situation where every subject is censored at τ, the weight function is constant, and $N_i(\cdot)$ is observed over $[0, \tau]$.

We remark that, in contrast with the approach in section 2.2, this approach requires knowledge of the censoring times c_is. In practice, we may use the last observation times to approximate c_is when the values of c_is are not available. This will not have much effect on the inference, since c_is are used only for evaluating $\bar{Z}_M(t; \gamma)$ in the estimating functions $U_n^M(\tilde{\beta}; \tilde{N}, w)$ and $U_n^M(\alpha; O, w_O)$. Again we focus on the constant weight function $w(\cdot) = 1$ and $w_O(\cdot) = 1$, for $t \in [0, \tau]$. Results presented below can also be straightforwardly extended to situations with general deterministic weight functions.

Appendix B shows that the estimator $\hat{\beta}_n^M$ is also consistent and asymptotically normal. The asymptotical covariance matrix of $\sqrt{n}(\hat{\beta}_n^M - \beta)$ is

$$\Sigma_M = (\mathbf{1}_{1\times p}, -\mathbf{1}_{1\times p}) A_M(\tilde{\beta}, \alpha)^{-1} B_M(\tilde{\beta}, \alpha) A_M(\tilde{\beta}, \alpha)^{-1} (\mathbf{1}_{1\times p}, -\mathbf{1}_{1\times p})' \tag{14}$$

with $\mathbf{1}_{1\times p}$ the $1 \times p$ matrix consisting of all 1s, where

$$A_M(\tilde{\beta}, \alpha) = -\mathrm{E}\left\{ \mathrm{diag}\left(A_M^{11}(\tilde{\beta}, \alpha), A_M^{22}(\tilde{\beta}, \alpha) \right) \right\} \tag{15}$$

with

$$A_M^{11}(\tilde{\beta}, \alpha) = \int_0^\tau \mathrm{I}(C_1 \geqslant t) \left[\frac{s_M^{(2)}(\tilde{\beta}; t)}{s_M^{(0)}(\tilde{\beta}; t)} - \bar{z}_M(t; \tilde{\beta})^{\otimes 2} \right] d\tilde{N}_1(t)$$

and

$$A_M^{22}(\tilde{\beta}, \alpha) = \int_0^\tau \mathrm{I}(c_1 \geqslant t) \left[\frac{s_M^{(2)}(\alpha; t)}{s_M^{(0)}(\alpha; t)} - \bar{z}_M(t; \alpha)^{\otimes 2} \right] dO_1(t),$$

and

$$B_M(\tilde{\beta}, \alpha) = \mathrm{E}\left\{ \begin{bmatrix} \int_0^\tau [Z_1 - \bar{z}_M(t; \tilde{\beta})] dM_1^M(t; \tilde{\beta}) \\ \int_0^\tau [Z_1 - \bar{z}_M(t; \alpha)] dM_1^O(t; \alpha) \end{bmatrix}^{\otimes 2} \right\} \tag{16}$$

with

$$s_M^{(j)}(\gamma; t) = \mathrm{E}\{\mathrm{I}(C_1 \geqslant t) Z_1^{\otimes j} \exp(\gamma' Z_1)\},$$

for $j = 0, 1, 2$,

$$\bar{z}_M(t;\gamma) = \lim_{n\to\infty} \bar{Z}_M(t;\gamma) = \frac{s_M^{(1)}(\gamma;t)}{s_M^{(0)}(\gamma;t)}$$

almost surely, and

$$M_i^M(t;\tilde{\beta}) = \int_0^t \mathrm{I}(c_i \geqslant s)[d\tilde{N}_i(s) - \Lambda_0(s)\exp(\tilde{\beta}' z_i) d\mu_0(s)]$$

and

$$M_i^O(t;\alpha) = \int_0^t \mathrm{I}(c_i \geqslant s)[dO_i(s) - \exp(\alpha' z_i) d\mu_0(s)],$$

for $t \in [0, \tau]$.

We may obtain a consistent estimator for Σ_M, denoted by $\hat{\Sigma}_M$, by substituting $A_M(\tilde{\beta}, \alpha)$ and $B_M(\tilde{\beta}, \alpha)$ in (14) with $\hat{A}_M(\hat{\tilde{\beta}}_n, \hat{\alpha}_n)$ and $\hat{B}_M(\hat{\tilde{\beta}}_n, \hat{\alpha}_n)$, respectively, where

$$\hat{A}_M(\tilde{\beta}, \alpha) = \frac{1}{n} \frac{\partial\left(U_n^M(\tilde{\beta}; \tilde{N}, 1), U_n^M(\alpha; O, 1)\right)}{\partial(\tilde{\beta}, \alpha)}, \tag{17}$$

that is,

$$-\frac{1}{n}\sum_{i=1}^{n} \mathrm{diag}\left(a_{M,i}^{11}(\tilde{\beta}, \alpha), a_{M,i}^{22}(\tilde{\beta}, \alpha)\right)$$

with

$$a_{M,i}^{11}(\tilde{\beta}, \alpha) = \int_0^\tau \mathrm{I}(c_i \geqslant t)\left[\frac{S_M^{(2)}(\tilde{\beta};t)}{S_M^{(0)}(\tilde{\beta};t)} - \bar{Z}_M(t;\tilde{\beta})^{\otimes 2}\right] d\tilde{N}_i(t)$$

and

$$a_{M,i}^{22}(\tilde{\beta}, \alpha) = \int_0^\tau \mathrm{I}(c_i \geqslant t)\left[\frac{S_M^{(2)}(\alpha;t)}{S_M^{(0)}(\alpha;t)} - \bar{Z}_M(t;\alpha)^{\otimes 2}\right] dO_i(t),$$

and

$$\hat{B}_M(\tilde{\beta}, \alpha) = \frac{1}{n}\sum_{i=1}^{n} \begin{bmatrix} \int_0^\tau [z_i - \bar{Z}_M(t;\tilde{\beta})] d\hat{M}_i^M(t;\tilde{\beta}) \\ \int_0^\tau [z_i - \bar{Z}_M(t;\alpha)] d\hat{M}_i^O(t;\tilde{\alpha}) \end{bmatrix}^{\otimes 2} \tag{18}$$

with

$$\hat{M}_i^M(t;\tilde{\beta}) = \int_0^t \mathrm{I}(c_i \geqslant s)[d\tilde{N}_i(s) - \exp(\tilde{\beta}' z_i) d\hat{\tilde{\Lambda}}_0(s;\tilde{\beta})],$$

and

$$\hat{M}_i^O(t;\alpha) = \int_0^t \mathrm{I}(c_i \geqslant s)[dO_i(s) - \exp(\alpha' z_i) d\hat{\mu}_0(s;\alpha)].$$

Here $\hat{\tilde{\Lambda}}_0(\cdot;\tilde{\beta})$ and $\hat{\mu}_0(\cdot;\alpha)$ are the Breslow estimators for $\tilde{\Lambda}_0(\cdot)$ and $\mu_0(\cdot)$, respectively.

Notice that the modelling of the observation process links information obtained at different observation times. The method presented above is applicable to the situations where $\mathrm{E}\{o_i(t)\} = 0$, for all $t \in [0, \tau]$, such as where $O_i(\cdot)$ is a Poisson process.

3. Example

This section presents an analysis of the bladder cancer data by applying the two proposed methods. The bladder cancer study consisted of three treatment groups: placebo, thiotepa and pyridoxine. There were 121 subjects with superficial bladder tumors randomized into one of the three groups at study entry. Identical tablets were given daily by mouth to the subjects in the placebo and pyridoxine groups, while for the subjects in the thiotepa group, thiotepa was instilled into the bladder for two hours once a week for four weeks and once a month thereafter. In the following we focus on the placebo and thiotepa groups with respective sizes of 47 and 38.

The total number of subjects of interest is $n = 85$, and the time period under consideration is over four years since the entry of the first subject (i.e. $\tau = 48$ months). Let $N_i(\cdot)$, a counting process, represent the accumulated new tumor numbers of subject i over study period. Records on $N_i(\cdot)$ available are only at the subject's finite number of clinical visits. See Tables 1 and 2. The data are in months, a discrete time scale. As mentioned in section 1, Tables 1 and 2 indicate that the visiting patterns were different from subject to subject, and the thiotepa group tended to visit the clinics more frequently than the placebo group. Let z_i be a 3-dimensional vector with the first component indicating whether the subject was in the thiotepa group, and the second and third components being the number of tumors observed at the beginning of the study and the size of the largest initial tumors, respectively. The censoring time c_i of subject i was his internal time at the database closing. Since there was no record of c_i on hand, we took the last visit time of the subject to approximate c_i in the analysis when the approach described in section 2.3 was used. We considered the semiparametric model (1) and estimated the regression parameter $\beta = (\beta_1, \beta_2, \beta_3)'$ from the cancer data.

Using weight function $w(\cdot) = 1$ and solving the estimation equation $U_n^C(\beta; \tilde{N}, w) = 0$ given in (4), we obtained an estimate for the regression parameter, $\hat{\beta}_n^C = (-1.36, 0.28, -0.07)'$. The estimated standard deviations were $0.45, 0.09, 0.12$ for each component of β, respectively. The second approach, taking $w(\cdot) = w_O(\cdot) = 1$ and the model of (11) for the observation process, gave another estimate for β as $\hat{\beta}_n^M = (-1.48, 0.28, -0.08)'$ with respective estimated standard deviations $0.35, 0.07, 0.12$. Both estimates for β_1, the coefficient of the indicator for the thiotepa group, are negative. This indicates that thiotepa reduced the tumor recurrences. In fact, based on the asymptotic results of the two estimators and the estimated standard deviations, we see the effect of thiotepa was statistically significant. The results of both approaches suggest the number of initial tumors was significantly a useful prognostic factor to tumor growth, and the initial size of tumors was not. This is consistent with the analysis results presented in both Lawless & Zhan (1998) and Sun & Wei (2000).

As a side-product, the second approach provided an estimate for the regression parameter of the observation process, $\hat{\alpha}_n = (0.51, -0.01, -0.03)'$, with estimated standard deviations $0.28, 0.08, 0.09$, respectively for each of the components. Note that an approximately 95% confidence interval for the coefficient of the indicator of the thiotepa group was $(-0.04, 1.05)$. This suggests the thiotepa group visited clinics rather more frequently than the placebo group, while visiting pattern did not appear to be associated with the initial tumor number and size. It, from another perspective, indicates that inferences on the effect of thiotepa without adjusting for observation mechanism could be biased. Specifically in this situation, the effect could be underestimated heavily. Approaches taking into account observation patterns of the different treatment groups, such as the two methods presented in section 2, are in need for analysing data of this type.

To check whether the modelling (11) is appropriate for the study, we stratified the subjects into eight groups according to their covariates: in the placebo or thiotepa group, with number

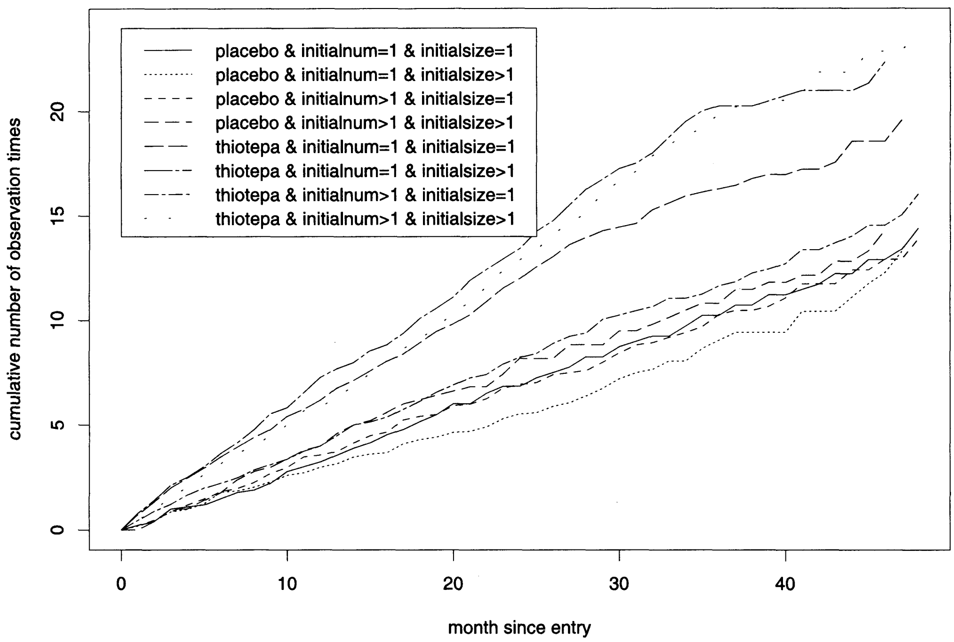

Fig. 1. Estimated means of observation times for different groups.

of initial tumors equal to 1 or larger, and with the initial tumor size equal to 1 unit or larger. Figure 1 shows non-parametrical estimates of the mean numbers of observation times for the groups, obtained using

$$\hat{\mu}_k(t) = \sum_{i \in \mathcal{G}_k} \int_0^t \frac{\mathrm{I}(c_i \geqslant s)}{\sum_{j \in \mathcal{G}_k} \mathrm{I}(c_j \geqslant s)} dO_i(s),$$

where $\mathcal{G}_k$ represents group k. It appears that the means over time are quite proportional to each other. This indicates that in the current situation it is feasible to apply the second approach.

The use of weight functions allows us to place greater emphasis on certain time periods. Table 3 presents estimates for $\beta = (\beta_1, \beta_2, \beta_3)'$, as well as their estimated standard deviations (in brackets), obtained by applying the two proposed approaches with different deterministic weight functions. Conclusions on the effect of thiotepa and on the influence of the baseline measures drawn by using different weight functions are consistent except for the case with weight function of $w(t) = 1/t^2$. The weight function heavily emphasizes the response at early stage of subjects, where there were few observations. This may explain the apparently larger standard deviations in this case than in the others. The result in this case, combined with the results in the other cases, may suggest that it takes time for the thiotepa treatment to take effect.

Wei *et al.* (1989) analysed the same cancer data. They focused on times to the clinical visits when new tumors were discovered. By using a counting process structure, the analysis presented here takes into account every visit and numbers of new tumors discovered at the visits.

4. Simulation

Finite sample behaviours of the proposed estimators were studied through a simulation. Motivated by the tumor data, we chose the following two simulation settings.

Table 3. *Estimates and their estimated standard deviations by different weight functions*

Weight function	Approach I[1] β_1	β_2	β_3	Approach II[2] β_1	β_2	β_3
t^2	−1.333 (0.51)	0.317 (0.11)	−0.101 (0.14)	−1.615 (0.41)	0.348 (0.09)	−0.103 (0.13)
t	−1.349 (0.49)	0.291 (0.10)	−0.095 (0.13)	−1.551 (0.38)	0.311 (0.08)	−0.102 (0.12)
$\sqrt{t}$	−1.362 (0.47)	0.280 (0.10)	−0.086 (0.12)	−1.524 (0.34)	0.295 (0.07)	−0.095 (0.12)
1	−1.364 (0.45)	0.275 (0.09)	−0.070 (0.12)	−1.478 (0.35)	0.284 (0.07)	−0.083 (0.12)
$1/\sqrt{t}$	−1.338 (0.44)	0.285 (0.09)	−0.039 (0.12)	−1.412 (0.35)	0.288 (0.08)	−0.058 (0.12)
$1/t$	−1.288 (0.48)	0.327 (0.10)	0.018 (0.13)	−1.336 (0.43)	0.325 (0.10)	−0.011 (0.13)
$1/t^2$	−1.073 (0.73)	0.494 (0.13)	0.245 (0.20)	−1.277 (0.70)	0.500 (0.13)	0.126 (0.16)

[1] Estimation conditional on observation process, presented in section 2.2.
[2] Estimation by modelling observation process, presented in section 2.3.

4.1. Simulation setting A

Simulate a trial over two years (i.e. 24 months) with two treatment arms, A and B; suppose that $n = 100$ subjects are enrolled staggeringly, and randomly distributed to each of the arms at entry. We generated independent mixed-Poisson processes $N_i(t), t \in [0, 24]$, with a rate of $\varsigma_i/6$ for subjects $i = 1, \ldots, 50$ on arm A, and with a rate of $\exp(-3/5)\varsigma_i/6$ for subjects $i = 51, \ldots, 100$ on arm B, where ς_i were generated independently from the Gamma distribution with mean 1 and variance 5/6. We simulated observation processes $O_i(t), t \in [0, 24]$ as follows: (i) generate independent Poisson processes over $[0, 24]$ with a rate of μ for $i = 1, \ldots, 50$ and a rate of $\mu\exp(2/5)$ for $i = 51, \ldots, 100$, where μ was chosen to be (a) 1/2, (b) 1/6, (c) 1/12 to reflect observations from dense to sparse; (ii) round up event times to the integers next to them. A sample of 100 independent random variables was generated from the uniform distribution on $[0, 24]$ to play the role of censoring times.

Taking the indicator of arm B as the covariate, from the simulated data we evaluated the estimators $\hat{\beta}_n^C$ and $\hat{\beta}_n^M$ along with their variance estimates proposed in section 2. We approximated the censoring times by the last observation times for evaluating $\bar{Z}_M(t; \gamma)$ in $U_n^M(\tilde{\beta}; \tilde{N}, 1)$ and $U_n^M(\alpha; O, 1)$ to get $\hat{\beta}_n^M$. Table 4 shows results from 1000 repetitions of the estimate evaluation. The $\hat{\beta}$s in Table 4 are sample means of the estimates, the "mse/mv"s are the ratios of sample mean square errors vs sample variances, and the $\hat{\sigma}$s and $\tilde{\sigma}$s are sample means of the standard deviation estimates and the sample standard deviations of the estimates, respectively.

Both approaches provided estimates for the regression parameter β in the three situations (a), (b) and (c) with almost no bias relative to the standard deviations. See the corresponding ratios of sample mean square errors vs sample variance ("mse/mv"). Notice the standard deviations of the two approaches increase as μ decreases, i.e. as the observations become sparser. From the comparison between the values of $\hat{\sigma}$ and $\tilde{\sigma}$, we see the variance estimates are close to the corresponding sample variances when μ is 1/2 or 1/6. In the situations with sparse observation, where $\mu = 1/12$, the difference is getting larger. This is rather obvious in the second approach, which partially resulted from the approximation of censoring times by the last observation times.

To study the asymptotic normal approximations, which one would use to obtain confidence intervals or tests on the regression parameter, we examined the Q–Q plots of the approximate pivotals $(\hat{\beta} - \beta_0)/\hat{\sigma}$ vs the standard normal random variables. The plots suggest that the asymptotic approximations are sufficiently accurate for practical purposes.

Table 4. *Summary on 1000 simulation repetitions in settings A & B*

	Setting A						Setting B					
Estimates	Approach I[1]			Approach II[2]			Approach I			Approach II		
($\beta_0 = -0.6$)	$\mu = 1/2$	$\mu = 1/6$	$\mu = 1/12$	$\mu = 1/2$	$\mu = 1/6$	$\mu = 1/12$	$\mu_0 = 12$	$\mu_0 = 4$	$\mu_0 = 2$	$\mu_0 = 12$	$\mu_0 = 4$	$\mu_0 = 2$
$\hat{\beta}^a$	−0.609	−0.599	−0.594	−0.612	−0.605	−0.594	−0.617	−0.588	−0.603	−0.630	−0.681	−0.727
mse/sv^b	1.001	1.000	1.000	1.001	1.000	1.000	1.003	1.001	1.000	1.009	1.059	1.119
$\hat{\sigma}^c$	0.310	0.346	0.405	0.317	0.344	0.384	0.393	0.408	0.341	0.308	0.324	0.351
$\tilde{\sigma}^d$	0.330	0.365	0.439	0.332	0.367	0.435	0.315	0.332	0.366	0.316	0.332	0.367

[1] Estimation conditional on observation process, presented in section 2.2.
[2] Estimation by modelling observation process, presented in section 2.3.
[a] $\hat{\beta}$ = sample mean of parameter estimator evaluations.
[b] mse/sv = sample mean square error : sample variance of parameter estimator evaluations.
[c] $\hat{\sigma}$ = sample mean of standard deviation estimator evaluations.
[d] $\tilde{\sigma}$ = sample standard deviation of parameter estimator evaluations.

4.2. Simulation setting B

We generated data in the same way as in setting A except that the observation processes were simulated based on independent Poisson processes $O_i(t), t \in [0, 24]$, with intensity $\mu(t) = \mu_0/(6+t)$ for the first 50 subjects and intensity $\mu(t) = \mu_0/(2+t)$ for the last 50 subjects, where μ_0 was chosen to be (a) 12, (b) 4, (c) 2 to match the previous setting with the same expected number of observations. Thus the model (11) used for the observation processes is inappropriate for the simulated data.

Table 4 also summarizes results from 1000 repetitions of evaluations of the proposed estimators from the simulated data in this setting. Estimates for the parameter by the approach conditional on observation time points are close to the true value, and the two estimates for the standard deviations are similar. However, estimates for the parameter by the second approach depart from the true value as μ_0 decreases, that is, as observation is getting sparser. The bias in the case of $\mu_0 = 2$ is not ignorable, relative to the corresponding standard deviations. This suggests that we should apply the second approach with caution. Checking for the model (11) of observation process, like the one performed in section 3, is necessary in practice.

To address the asymptotic normal approximations, we also studied the Q–Q plots of the approximate pivotals $(\hat{\beta} - \beta_0)/\hat{\sigma}$ vs standard normal random variables. The plots associated with the first approach indicate that the asymptotic approximations are, for the cases considered, sufficiently accurate for practical purposes. The plots for the approach modelling observation process show that the mean of the approximate pivotals differs from 0.

5. Discussion

This paper considers estimation of the regression parameter β in the semiparametric model for the conditional mean function of a counting process, $\mathrm{E}\{N_i(t)|Z_i = z_i\} = \Lambda_0(t)\exp(\beta z_i)$, where only panel counts along with covariates for each subject are available and observation patterns possibly differ from subject to subject. Two estimation equation based approaches are proposed. They are easy to implement. The first approach is readily applicable to situations with time-dependent covariates and does not require modelling of the observation process, provided there are some time points where more than one subject has observations. In practice, this restriction may be met by discretizing the time scale. The second approach is based on modelling the observation process. It can be applied in situations not meeting the requirement on the observation. The second approach requires the values of the censoring times c_i, and, if extended to situations with time-dependent covariates, it also needs the values on the covariate processes $z_i(t)$ at all the study observation points. This is in general not realistic, although there are situations where the state space of the covariate process contains only finite number of values and they may be observed. We remark that the proposed methods and discussions, with little modification, apply to the situations under the general models

$$\mathrm{E}\{N_i(t)|z_i(t)\} = \Lambda_0(t)\phi(\beta; z_i(t)), \quad t \geqslant 0$$

and

$$\mathrm{E}\{dO_i(t)|z_i(t)\} = \psi(\alpha; z_i(t))\, d\mu_0(t), \quad t \geqslant 0,$$

where the functions $\phi(\cdot)$ and $\psi(\cdot)$ are known, positive-valued and differentiable with respect to the parameters.

Another problem, which is of theoretical and practical interest, is estimation for the baseline $\Lambda_0(t)$ of the mean function. Notice that both $\tilde{\Lambda}_0^C(t;\beta)$ in (10) and $d\hat{\tilde{\Lambda}}_0(t;\tilde{\beta})/d\hat{\mu}_0(t;\alpha)$ in section 2.3 are natural estimators, where $\hat{\tilde{\Lambda}}_0(t;\tilde{\beta})$ and $\hat{\mu}_0(t;\alpha)$ are the Breslow estimators for $\tilde{\Lambda}_0(t)$ and $\mu_0(t)$, respectively. Neither of them, however, ensures its evaluations are non-decreasing over time.

This paper considers situations where $N_i(\cdot), O_i(\cdot)$, and the censoring time C_i are independent conditional on covariate Z_i. There are situations where the assumption is violated. For example, the clinical visits of asthma patients may be driven by the number of the asthma attacks previously to the visits. Then, the presented approaches may lead to biased estimates. If the dependence is specified, however, the approaches can be adjusted accordingly. For example, suppose that

$$\mathrm{E}\{N_i(t)|O_i(t), z_i(t)\} = \Lambda_0(t)\exp(\gamma O_i(t) + \beta' z_i(t)). \tag{19}$$

We may apply the approach conditional on observation process to obtain an estimator for the parameter $(\gamma,\beta)'$, and thus to obtain an estimator of β, the regression parameter of interest. Notice that the estimate of γ could reveal the dependence between the response and observation processes. We may conduct a model diagnosis for the independence assumption based on a hypothesis testing on γ. In some practical situations, as pointed out by a referee, it would be more natural to model how observation pattern depends on the history information of the counting process. We need to search for a reasonable model on $O_i(t)$ conditional on $\{N_i(s) : 0 \leqslant s < t\}$, and then consider inferences accordingly.

There are several other issues worthwhile further investigation. One interesting problem, for example, is how to choose the optimal weight function or an almost optimal weight function, not necessarily deterministic, in each of the two approaches.

Acknowledgements

The authors thank Professors R. J. Gray and S. W. Lagakos, and two referees and the Associate Editor for their insightful comments and valuable suggestions, which led to significant improvements in the paper. The research was partially supported by the New Faculty Research Initiation Award of the University of Memphis to the first author.

References

Andersen, P. K. & Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100–1120.

Bilias, Y., Gu, M. & Ying, Z. (1997). Towards a general asymptotic theory for Cox model with staggered entry. *Ann. Statist.* **25**, 662–682.

Byar, D. P. (1980). The veterans administration study of chemoprophylaxis for recurrent stage I bladder tumors: comparison of placebo, pyridoxine, and topical thiotepa. In *Bladder tumors and other topics in urological oncology* (eds M. Pavone-Macaluso, P. H. Smith & F. Edsmyn) 363–370, Plenum, New York.

Diggle, P. J., Liang, K. Y. & Zeger, S. L. (1994). *Analysis of longitudinal data.* Clarendon Press, Oxford.

Fleming, T. R. & Harrington, D. P. (1991). *Counting processes and survival analysis.* Wiley, New York.

Lawless, J. F. (1995). The analysis of recurrent events for multiple subjects. *Appl. Statist.* **44**, 487–498.

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

Lawless, J. F. & Zhan, M. (1998). Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *Canad. J. Statist.* **26**, 549–565.

Pollard, D. (1990). *Empirical processes: theory and applications.* Regional conference series in probability and statistics 2. Institute of Mathematical Statistics, Hayward, CA.

Pepe, M. S. & 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.

Prentice, R. L., Williams, B. J. & Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **68**, 373–379.
Shorack, G. R. & Wellner, J. A. (1986). *Empirical processes with applications to statistics*, Wiley Series in Probability and Mathematical Statistics. Wiley, New York.
Sun, J. & Wei, L. J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *J. Roy. Statist. Soc. Ser. B* **62**, 293–302.
Thall, P. F. & Lachin, J. M. (1988). Analysis of recurrent events: nonparametric methods for random-interval count data. *J. Amer. Statist. Assoc.* **83**, 339–347.
Tsiatis, A. A. (1981). A large sample study of Cox's regression model. *Ann. Statist.* **9**, 93–108.
Wei, L. J., Lin, D. Y. & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Amer. Statist. Assoc.* **84**, 1065–1073.

Received March 1999, in final form November 2001

X. Joan Hu, Department of Mathematical Sciences, University of Memphis, Memphis, TN 38152-3240, USA.
E-mail: joanhu@msci.memphis.edu

Appendix A. Asymptotics of $\hat{\beta}_n^C$

In the following we establish the consistency and weak convergence of the estimator $\hat{\beta}_n^C$ and the consistency of the covariance estimator $\hat{\Sigma}_C$. We are concerned with the case of $w(\cdot) = 1$, for $t \in [0, \tau]$. Results given below may be straightforwardly extended to situations with general deterministic weight functions. We denote the true value of β by β_0. The proofs below utilize approaches similar to those of Bilias *et al.* (1997).

A.1. Consistency of $\hat{\beta}_n^C$

For $j = 0, 1, 2$ and $\beta \in \mathscr{B}_\epsilon = \{\beta : ||\beta - \beta_0|| \leqslant \epsilon\}$ with $\epsilon > 0$, by the strong law of large numbers, $S_C^{(j)}(\beta; t)$ converges almost surely to $s_C^{(j)}(\beta; t)$. Define a pseudo partial log-likelihood as

$$X_n^C(\beta) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \mathrm{I}(c_i \geqslant t)\left[(\beta - \beta_0)' z_i - \log\frac{S_C^{(0)}(\beta; t)}{S_C^{(0)}(\beta_0; t)}\right] d\tilde{N}_i(t).$$

We see that $X_n^C(\beta)$ converges almost surely to

$$\mathscr{X}^C(\beta) = \mathrm{E}\left\{\int_0^\tau \mathrm{I}(C_1 \geqslant t)\left[(\beta - \beta_0)' Z_1 - \log\frac{s_C^{(0)}(\beta; t)}{s_C^{(0)}(\beta_0; t)}\right] d\tilde{N}_1(t)\right\}.$$

Notice that $\partial^2 X_n^C(\beta)/\partial\beta^2 = \hat{A}_C(\beta)$ in (8) is negative semi-definite. Therefore we have

$$\sup_{\beta \in \mathscr{B}_\epsilon} ||X_n^C(\beta) - \mathscr{X}^C(\beta)|| \to 0 \tag{20}$$

almost surely. Since $\partial\mathscr{X}^C(\beta_0)/\partial\beta = 0$ and $\partial^2\mathscr{X}^C(\beta_0)/\partial\beta^2 = A_C(\beta_0)$ given in (5) is negative definite, $\mathscr{X}(\beta)$ has an unique maximizer β_0. This fact, together with (20), implies that there must exist a maximizer of $X_n^C(\beta)$ in the interior of $\mathscr{B}_\epsilon$. Therefore,

$$\frac{\partial X_n^C(\beta)}{\partial\beta} = \frac{1}{n} U_n^C(\beta; \tilde{N}, 1) = 0$$

has a solution, say, $\hat{\beta}_n^C$, within $\mathscr{B}_\epsilon$.

On the other hand, $\partial^2 X_n^C(\beta_0)/\partial\beta^2$ converges almost surely to $\partial^2\mathscr{X}^C(\beta_0)/\partial\beta^2$. This along with the fact that $\partial^3 X_n^C(\beta)/\partial\beta^3$ is bounded ensures existence of ϵ such that $\partial^2 X_n^C(\beta)/\partial\beta^2$ is negative definite for $\beta \in \mathscr{B}_\epsilon$ when n is large enough. Thus $\partial^2 X_n^C(\hat{\beta}_n^C)/\partial\beta^2$ is negative definite, which implies that $\hat{\beta}_n^C$ is the unique global maximizer of $X(\beta)$ in $\mathscr{B}_\epsilon$, i.e. the unique solution of

$U_n^C(\beta; \tilde{N}, 1) = 0$. Finally, since ϵ can be chosen arbitrarily small, $\hat{\beta}_n^C$ must converge to β_0 almost surely, as $n \to \infty$.

A.2. Weak convergence of $\hat{\beta}_n^C$

Notice that

$$U_n^C(\beta_0; \tilde{N}, 1) = \sum_{i=1}^{n} \int_0^{\tau} [z_i - \bar{Z}_C(t; \beta_0)]\, dM_i^C(t, \beta_0).$$

Following the arguments in Bilias *et al.* (1997) we can show both $\{[M_i^C(t; \beta_0), t \in \mathscr{T}] : i = 1, \ldots, n\}$ and $\{[\int_0^t z_i dM_i^C(s; \beta_0), t \in \mathscr{T}] : i = 1, \ldots, n\}$ are manageable (Pollard, 1990, pa. 38) and hence $\{(n^{-1/2} \sum_{i=1}^n M_i^C(t; \beta_0),\ n^{-1/2} \sum_{i=1}^n \int_0^t z_i dM_i^C(s; \beta_0))'\}_{t \in \mathscr{T}}$ converges weakly to a zero-mean Gaussian process with index in $\mathscr{T}$, say $(W_M^C, W_{M_z}^C)'$.

By the strong embedding theorem (cf: Shorack & Wellner, 1986, th. 2.3.4), we can construct a new probability space and have $\{(n^{-1/2} \sum_{i=1}^n M_i^C(t; \beta_0), n^{-1/2} \sum_{i=1}^n z_i \int_0^t dM_i^C(s; \beta_0), S_C^{(0)}(\beta_0; t), S_C^{(1)}(\beta_0; t))\}_{t \in \mathscr{T}}$ almost surely convergence to $\{W_M^C, W_{M_z}^C, s_C^{(0)}(\beta_0; \cdot), s_C^{(1)}(\beta_0; \cdot)\}$. Then we may show that almost surely, as $n \to \infty$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{\tau} [\bar{z}_C(t; \beta) - \bar{Z}_C(t; \beta)]\, dM_i^C(t, \beta_0) \to 0.$$

Thus $n^{-1/2} U_n^C(\beta_0; \tilde{N}, 1)$ is asymptotically normal in the original space.

By the Taylor expansion,

$$\sqrt{n}(\hat{\beta}_n^C - \beta_0) = -\hat{A}_C(\beta^*)^{-1} \frac{1}{\sqrt{n}} U_n^C(\beta_0; \tilde{N}, 1)$$

with $\hat{A}_C(\beta) = n^{-1} \partial U_n^C(\beta; \tilde{N}, 1)/\partial \beta$ in (8) and β^* on the line segment between β_0 and $\hat{\beta}_n^C$. The consistency of $\hat{\beta}_n^C$ to β_0 and $\hat{A}_C(\beta_0)$ to $\partial^2 \mathscr{X}^C(\beta_0)/\partial \beta^2 = A_C(\beta_0)$ together with the weak convergence of $n^{-1/2} U_n^C(\beta_0; \tilde{N}, 1)$ yields the asymptotical normality of $\sqrt{n}(\hat{\beta}_n^C - \beta_0)$ with mean zero and covariance matrix Σ_C. Notice that the weak convergence of the estimator may be derived by using a multivariate approach since $\mathscr{T}$ is finite in the current situation.

A.3. Consistency of $\hat{\Sigma}_C$

Notice that $\hat{B}_C(\beta_0)$ in (9) is asymptotically equivalent to

$$\frac{1}{n} \sum_{i=1}^{n} \left[\int_0^{\tau} (z_i - \bar{z}_C(t; \beta_0)) dM_i^C(t; \beta_0) \right]^{\otimes 2},$$

which is consistent to the asymptotical covariance of $n^{-1/2} U_n^C(\beta_0; \tilde{N}, 1)$, $B_C(\beta_0)$ given in (6). This along with the consistency of $\hat{\beta}_n^C$ and $\hat{A}_C(\beta_0)$ and the continuity of $A_C(\cdot)$ and $B_C(\cdot)$ leads $\hat{\Sigma}_C$ in (7) a consistent estimate of Σ_C.

Appendix B. Asymptotics of $\hat{\beta}_n^M$

In the following we establish the consistency and weak convergence of the estimator $\hat{\beta}_n^M$ and the consistency of the covariance estimator $\hat{\Sigma}_M$. We are again concerned only with the case of $w(\cdot) = 1$ and $w_O(\cdot) = 1$, for $t \in [0, \tau]$. Results given below may be straightforwardly extended to situations with general weight functions. We denote the true values of $\tilde{\beta}$, β and α by $\tilde{\beta}_0$, β_0 and α_0, respectively. We again utilize approaches similar to those of Bilias *et al.* (1997) in the proofs below.

B.1. Consistency of $\hat{\beta}_n^M$

For $j = 0, 1, 2$ and either $\gamma = \tilde{\beta} = \beta + \alpha \in \tilde{\mathscr{B}}_\epsilon = \{\tilde{\beta} : ||\tilde{\beta} - \tilde{\beta}_0|| \leqslant \epsilon\}$ or $\gamma = \alpha \in \mathscr{A}_\epsilon = \{\alpha : ||\alpha - \alpha_0|| \leqslant \epsilon\}$ with $\epsilon > 0$, by the strong law of large numbers, $S_M^{(j)}(\gamma; t)$ converges almost surely to $s_M^{(j)}(\gamma; t)$. Define

$$X_n^M(\tilde{\beta}, \alpha) = \frac{1}{n}\sum_{i=1}^n \begin{bmatrix} \int_0^\tau \mathrm{I}(c_i \geqslant t)\left\{(\tilde{\beta} - \tilde{\beta}_0)' z_i - \log \dfrac{S_M^{(0)}(\tilde{\beta}; t)}{S_M^{(0)}(\tilde{\beta}_0; t)}\right\} d\tilde{N}_i(t) \\ \int_0^\tau \mathrm{I}(c_i \geqslant t)\left\{(\alpha - \alpha_0)' z_i - \log \dfrac{S_M^{(0)}(\alpha; t)}{S_M^{(0)}(\alpha_0; t)}\right\} dO_i(t) \end{bmatrix},$$

and

$$\mathscr{X}^M(\tilde{\beta}, \alpha) = \mathrm{E}\left\{\begin{bmatrix} \int_0^\tau \mathrm{I}(C_1 \geqslant t)\left[(\tilde{\beta} - \tilde{\beta}_0)' Z_1 - \log \dfrac{s_M^{(0)}(\tilde{\beta}; t)}{s_M^{(0)}(\tilde{\beta}_0; t)}\right] d\tilde{N}_1(t) \\ \int_0^\tau \mathrm{I}(c_1 \geqslant t)\left[(\alpha - \alpha_0)' Z_1 - \log \dfrac{s_M^{(0)}(\alpha; t)}{s_M^{(0)}(\alpha_0; t)}\right] dO_1(t) \end{bmatrix}\right\}.$$

Taking arguments similarly to appendix A, we have

$$\sup_{(\tilde{\beta}, \alpha) \in \tilde{\mathscr{B}}_\epsilon \times \mathscr{A}_\epsilon} ||X_n^M(\tilde{\beta}, \alpha) - \mathscr{X}^M(\tilde{\beta}, \alpha)|| \to 0, \quad \text{as } n \to \infty, \tag{21}$$

almost surely. For $H = (H_1, H_2)'$ and $a = (a_1, a_2)'$, let $\partial^* H / \partial^* a = \left(\partial H_1 / \partial a_1, \partial H_2 / \partial a_2\right)'$. Notice that $\partial^* \mathscr{X}^M(\tilde{\beta}_0, \alpha_0) / \partial^*(\tilde{\beta}, \alpha) = 0$ and $\partial\left[\partial^* \mathscr{X}^M(\tilde{\beta}_0, \alpha_0) / \partial^*(\tilde{\beta}, \alpha)\right] / \partial(\tilde{\beta}, \alpha) = A_M(\tilde{\beta}_0, \alpha_0)$ given in (15) is negative definite. This with (21) implies that there must exist a maximizer of $X_n^M(\tilde{\beta}, \alpha)$ in the interior of $\mathscr{B}_\epsilon \times \mathscr{A}_\epsilon$. Therefore,

$$\frac{\partial^* X_n^M(\tilde{\beta}, \alpha)}{\partial^*(\tilde{\beta}, \alpha)} = \frac{1}{n}\begin{bmatrix} U_n^M(\tilde{\beta}; \tilde{N}, 1) \\ U_n^M(\alpha; O, 1) \end{bmatrix} = 0 \tag{22}$$

has a solution, say, $(\hat{\tilde{\beta}}_n, \hat{\alpha}_n)$, within $\tilde{\mathscr{B}}_\epsilon \times \mathscr{A}_\epsilon$. We can show that $(\hat{\tilde{\beta}}_n, \hat{\alpha}_n)$ is the unique solution of (22) and converges to $(\tilde{\beta}_0, \alpha_0)$ almost surely. Therefore $\hat{\beta}_n^M = \hat{\tilde{\beta}}_n - \hat{\alpha}_n$ is consistent.

B.2. Weak convergence of $\hat{\beta}_n^M$

Notice that

$$U_n^M(\tilde{\beta}_0; \tilde{N}, 1) = \sum_{i=1}^n \int_0^\tau \{z_i - \bar{Z}_M(t; \tilde{\beta}_0)\}\, dM_i^M(t, \tilde{\beta}_0)$$

and

$$U_n^M(\alpha_0; O, 1) = \sum_{i=1}^n \int_0^\tau \{z_i - \bar{Z}_M(t; \alpha_0)\}\, dM_i^O(t, \alpha_0).$$

Similarly to appendix A but with $(0, \tau]$ instead of the set $\mathscr{T}$, we may show almost surely

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{bmatrix} \int_0^\tau \{\bar{z}_M(t; \tilde{\beta}_0) - \bar{Z}_M(t; \tilde{\beta}_0)\}\, dM_i^M(t, \tilde{\beta}_0) \\ \int_0^\tau \{\bar{z}_M(t; \alpha_0) - \bar{Z}_M(t; \alpha_0)\}\, dM_i^O(t, \alpha_0) \end{bmatrix} \to 0, \quad \text{as } n \to \infty.$$

Thus $\left(n^{-1/2} U_n^M(\tilde{\beta}_0; \tilde{N}, 1), n^{-1/2} U_n^M(\alpha_0; O, 1)\right)$ is asymptotically normal. By the Taylor expansion,

$$\sqrt{n}\begin{bmatrix} \hat{\tilde{\beta}}_n - \tilde{\beta}_0 \\ \hat{\alpha}_n - \alpha_0 \end{bmatrix} = -\hat{A}_M^{-1}(\tilde{\beta}^*, \alpha^*)\frac{1}{\sqrt{n}}\begin{bmatrix} U_n^M(\tilde{\beta}_0; \tilde{N}, 1) \\ U_n^M(\alpha_0; O, 1) \end{bmatrix},$$

where $\hat{A}_M(\tilde{\beta},\alpha) = n^{-1}\partial(U_n^M(\tilde{\beta};\tilde{N},1), U_n^M(\alpha;O,1))'/\partial(\tilde{\beta},\alpha)$ given in (17) and $\tilde{\beta}^*$ and α^* are on the line segments between $\tilde{\beta}_0$ and $\hat{\tilde{\beta}}_n$ and between α_0 and $\hat{\alpha}_n$, respectively. The consistency of $(\hat{\tilde{\beta}}_n,\hat{\alpha})$ to $(\tilde{\beta}_0,\alpha_0)$ and $\hat{A}_M(\tilde{\beta}_0,\alpha_0)$ to $A_M(\tilde{\beta}_0,\alpha_0)$ together with the weak convergence of $(n^{-1/2}U_n^M(\tilde{\beta}_0;\tilde{N},1), n^{-1/2}U_n^M(\alpha_0;O,1))'$ ensures the asymptotical normality of $\sqrt{n}(\hat{\tilde{\beta}}_n - \tilde{\beta}_0, \hat{\alpha} - \alpha_0)'$ with mean zero and covariance matrix

$$\Gamma_M = A_M(\tilde{\beta}_0,\alpha_0)^{-1}B_M(\tilde{\beta}_0,\alpha_0)A_M(\tilde{\beta}_0,\alpha_0)^{-1}.$$

Therefore $\sqrt{n}(\hat{\beta}_n^M - \beta_0)$ weakly converges to a normal distribution with mean zero and covariance matrix Σ_M.

B.3. Consistency of $\hat{\Sigma}_M$

Notice that $\hat{B}_M(\tilde{\beta},\alpha)$ in (18) is asymptotically equivalent to

$$\frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} \int_0^\tau [z_i - \bar{z}_M(t;\tilde{\beta})]dM_i^M(t;\tilde{\beta}) \\ \int_0^\tau [z_i - \bar{z}_M(t;\alpha)]dM_i^O(t;\alpha) \end{bmatrix}^{\otimes 2},$$

which is consistent to $B_M(\tilde{\beta},\alpha)$ (16). Recalling the consistency of $(\hat{\tilde{\beta}}_n,\hat{\alpha}_n)$, $\hat{A}^M(\tilde{\beta},\alpha)$, $\hat{\tilde{\Lambda}}_0(\cdot;\tilde{\beta})$, and $\hat{\mu}_0(\cdot;\alpha)$ and the continuity of $A_M(\cdot)$, $B_M(\cdot)$, $\hat{\tilde{\Lambda}}_0(\cdot;\tilde{\beta})$ with respect to $\tilde{\beta}$, and $\hat{\mu}_0(\cdot;\alpha)$ with respect to α, we see that $\hat{\Sigma}_M$ is consistent to Σ_M.