Analysis of Window-Observation Recurrence Data

Jianying Zuo
William Q. Meeker
Huaiqing Wu
Statistics Department
Iowa State University
Ames, IA 50011

June 30, 2005
13h 52min

Abstract

Many systems experience recurrent events. Recurrence data are collected to analyze quantities of interest, such as the mean cumulative number of events or the mean cumulative cost of events. Methods of analysis are available for recurrence data with left and/or right censoring. Due to practical constraints, however, recurrence data are sometimes recorded only in windows with gaps between the windows. This paper extends existing methods, both nonparametric and parametric, to window-observation recurrence data. The nonparametric estimator requires minimum assumptions, but will be biased if the size of the risk set is not positive over the entire period of interest. There is no such difficulty when using a parametric model for the recurrence data. For cases in which the size of the risk set is zero for some periods of time, we propose a simple method that uses a parametric adjustment to the nonparametric estimator. The methods are illustrated with two numerical examples.

KEY WORDS: Forecast; Mean cumulative function; Nonhomogeneous Poisson process; Nonparametric estimation; Repairable system data.

1 INTRODUCTION

1.1 Background

Recurrent events in systems are of interest in many applications. Here we use a broad definition of a system. For example, machines or automobiles in a company's fleet break down and are repaired; people become ill and visit a doctor; customers encounter financial need and apply for loans from banks. For these systems, quantities of interest might include the expected cumulative number (or cost) of events over a specific time range of system operation. Recurrence data are collected to analyze and estimate these and other quantities of interest.

Recurrence data usually record the type and number of events over time. Often, covariates related to the events, such as cost, are also recorded. An event could be recorded as having occurred at an exact time or within a particular time interval. Nelson (2003) describes many examples and data analysis methods for such data.

In some applications, recurrence data are recorded only in observation windows with gaps between the windows, even though the underlying process is continuous in time. Nelson (2003, page 75) describes an example in which window-observation recurrence data arise when "patients may enter and leave a medical study of a disease any number of times." Section 2 describes two other applications that we have encountered.

Note that observation windows can have random length, and the length of the gaps between windows can also be random. Furthermore, there is no requirement to have the same beginning or ending time points of windows for different observational units. Note that window-observation data are different from the "interval-grouped recurrent-event data" (Lawless and Zhan, 1998), which record the number of recurrent events in time intervals, with no gaps between the intervals (i.e., the number of events is known but their exact times are not known).

1.2 Previous Work on Analysis of Recurrence Data

Much work has been done on the development of methods for the analysis of recurrence data. Nelson (1988) presents a nonparametric estimator of the mean cumulative function (MCF), and shows how to use the estimator to make predictions. Nelson (1995) provides an unbiased variance estimator for the MCF estimator and the confidence limits for the MCF. Nelson (2003) gives a comprehensive treatment of the most important nonparametric methods for analyzing recurrence data. This book also presents many examples, for a wide range of application areas. These methods do not require strong model assumptions, and are easy to apply in practice.

Lawless and Nadeau (1995) provide an alternative variance estimator for the nonparametric

MCF estimator. Although their variance estimator is not unbiased, it is always nonnegative. The same paper also presents a flexible semi-parametric regression model that allows for covariates in the analysis of recurrence data. Basu and Rigdon (2000) is a good source for parametric statistical models and methods for repairable systems, from which recurrence data are often recorded. The purpose of this paper is to extend the use of these nonparametric and parametric methods to allow analysis of window-observation recurrence data.

1.3 Overview

The remainder of this paper is organized as follows. Section 2 describes two applications that have recurrence data with observation windows: Extended Warranty Data and AMSAA Vehicle Fleet Data. Section 3 reviews the nonparametric estimation method for the MCF and shows how it can be extended to handle window-observation recurrence data, and illustrates the method with the Extended Warranty Data. Section 4 reviews a parametric estimation method for the MCF with window-observation recurrence data. Section 5 shows how to apply the methods to the AMSAA Vehicle Fleet Data in which the gaps between observation windows result in some periods of time with risk set that is empty, resulting in downward bias. Section 6 proposes a simple method to reduce the downward bias by applying a parametric adjustment to the nonparametric MCF estimator for periods over which the risk set has size zero in window-observation recurrence data. Section 7 contains concluding remarks.

2 EXAMPLES

This section describes two examples that we encountered with window-observation recurrence data.

2.1 Extended Warranty Data

Extended warranties are often available to businesses or individuals for high-cost items, such as automobiles, large appliances, and computers. Only recurrence records within the period under extended warranty are available. Warranty data for a particular unit can have disconnected warranty periods. For example, a customer might purchase an extended warranty at a point in time after the expiration of the initial warranty. This results in window-observation data.

Actual extended warranty data that we have seen are not available for publication. To help illustrate the extension of existing estimation methods to window-observation recurrence data, we used automobile warranty data to which we do have access to create an extended warranty data set with simulated windows.

Table 1: Extended Warranty Data

VIN Group:	VG1	VG2	VG3	VG4	Total
Observation Window(s):	[0, 36]	[0, 24]	[0, 12]	[0, 12] & [24, 36]	
Number of Automobiles:	48,300	32,395	64,307	16,044	161,046
Percent in Group(%):	30	20	40	10	100
For Labor Code C6050, the number of automobiles					
With Events	82	30	37	25	174
Without Events	48,218	32,365	64,270	16,019	160,872

The original warranty data are for 161,046 automobiles from model year 1995 that were sold between August 1994 and November 1995. For each automobile, the following information is available: VIN number, build date, sale date, and the complete warranty-report history up to three years of service or until the end of November 1998 (the cutoff date of the data), whichever came first. For these automobiles, there had been 586,750 repair events with 1,745 distinct labor codes. Some of these automobiles never had a warranty report. Many had more than one.

We used the complete warranty data to generate the extended warranty data for these 161,046 automobiles, by randomly assigning these automobiles to one of the four warranty plans shown in Table 1, according to the percentages shown there. That is, all automobiles are covered in their first year of service; 30% elect for two extra years, 20% for one additional year, 40% do not extend beyond the first year, while 10% skip year 2, but return to warranty coverage in year 3.

The time scale is months of service (months since an automobile had been sold). We focus on events for labor code C6050. Figure 1 shows the event plot of the window-observation warranty data, with two lines for each of the four groups of automobiles. The line with "E" after group number is for all the automobiles with events in the group, while the line with "N" after group number is for all the automobiles that have no events in the group. For example, there are 48,300 automobiles in group VG1, 82 of which have at least one event with labor code C6050, and all the events are plotted on the line corresponding to System ID VG1E. The other 48,218 automobiles do not have any event with labor code C6050, and thus the line corresponding to System ID VG1N does not have any event along the line. The same explanation applies to the other three groups, and the corresponding number of automobiles for each line is shown in Table 1.

Because of the observation windows imposed on the data, the number of automobiles that is under observation changes during the period of 0 to 36 months. This is reflected in Figure 2, which shows the changing size of the risk set along the time line. The risk set plot of the original warranty data over the period from 0 to 36 months is a horizontal line at the value of 161,046 (not shown).

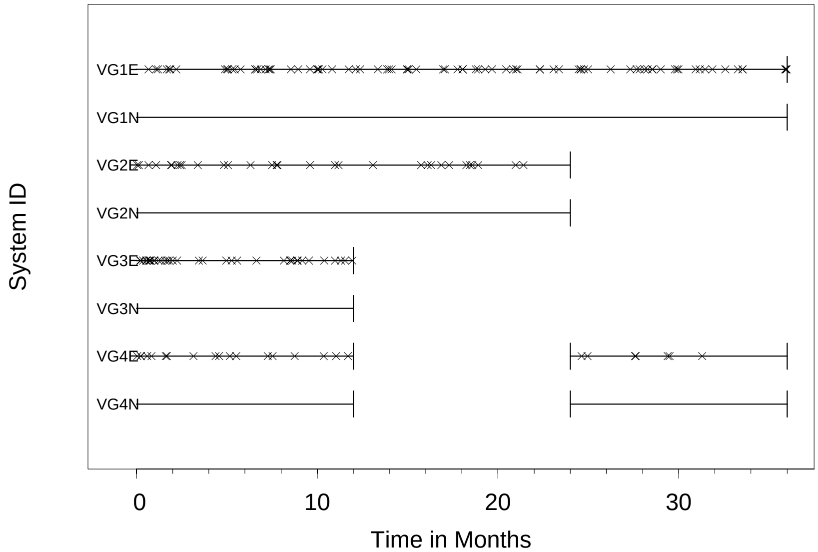

Figure 1: Event Plot for Labor Code C6050 Extended Warranty Data

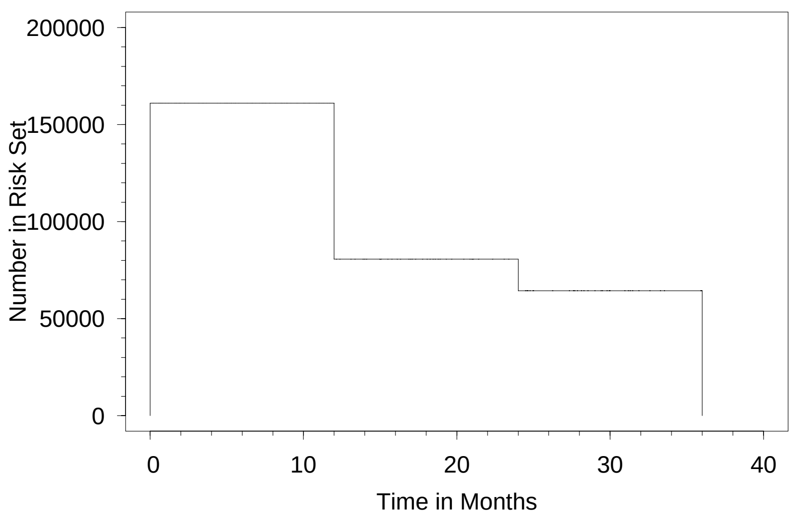

Figure 2: Risk Set Plot for Labor Code C6050 Extended Warranty Data

2.2 AMSAA Vehicle Fleet Data

U.S. Army Materiel Systems Analysis Activity (AMSAA) oversees the Army's Field Exercise Data Collection (FEDC) Program. This program maintains a database of part replacement rates for mission-essential weapon systems (e.g., various types of vehicles) that are used during intensive field training exercises. FEDC data are collected from a number of sites around the world and are used to answer such questions as whether a fleet is aging or not, how fast a fleet is aging, and when units should be overhauled or replaced. During an FEDC field exercise, a group of vehicles is used and careful records are taken for all maintenance and repair actions. An exercise generally lasts for approximately 500 miles. Fleet vehicles appear in a number of such exercises but with considerable gaps in between. Not all vehicles in a fleet are used in all FEDC exercises. Vehicles accrue mileage during the gaps between FEDC exercise uses. That is, gaps may contain non-exercise use and other unobserved field exercises. Recurrences occur but are not observed in those gaps.

Because actual FEDC data are sensitive, we were asked to analyze data that had been simulated by an analyst at AMSAA, according to a process that is similar to the process that generates actual FEDC data. In particular,

- The recurrences in the AMSAA Vehicle Fleet *Complete Data* were simulated from a nonhomogeneous Poisson process (NHPP) with a power recurrence rate $\beta = 2.76$ and $\eta = 5447$ miles, using the method described in Section 16.7 of Meeker and Escobar (1998). For each of the ten vehicles in the fleet, the simulation was run until the end of observation period had been reached. For each vehicle, the end of observation period was simulated from a uniform distribution between 20 and 30 thousand miles.

- The AMSAA Vehicle Fleet *Random Selection Window Data* were obtained by starting with the Complete AMSAA Vehicle Fleet Data simulating the lengths of the FEDC exercises and the gaps in time between these exercises. The length of each exercise (observation window) was simulated from a uniform distribution between 400 to 600 miles. The gaps between exercises (observation windows) were simulated from a uniform distribution between 600 and 1400 miles.

- The AMSAA Vehicle Fleet *Non-random Selection Window Data* used a model in which vehicles with more miles of service are less likely to be chosen for an exercise (in effect increasing the length of the gaps for vehicles as the number of miles of service gets larger). The data also started with the Complete AMSAA Vehicle Fleet Data. The length of each exercise was also simulated from a uniform distribution between 400 to 600 miles. However, the ith gap between exercises was simulated from a uniform distribution between 200×2^i and 400×2^i miles, while the length of time to the first exercise was simulated from a uniform distribution between

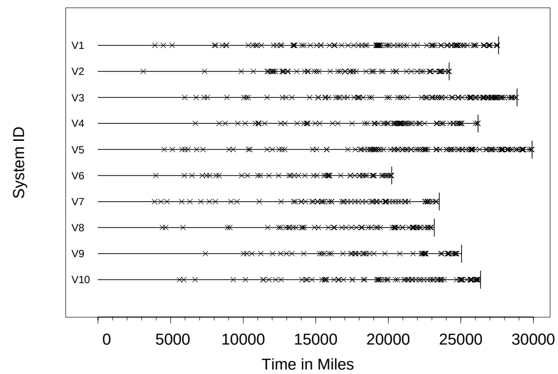

Figure 3: Event Plot for AMSAA Vehicle Fleet Complete Data

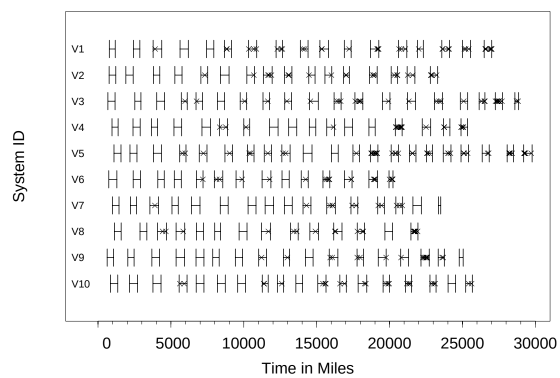

Figure 4: Event Plot for AMSAA Vehicle Fleet Random Selection Window Data

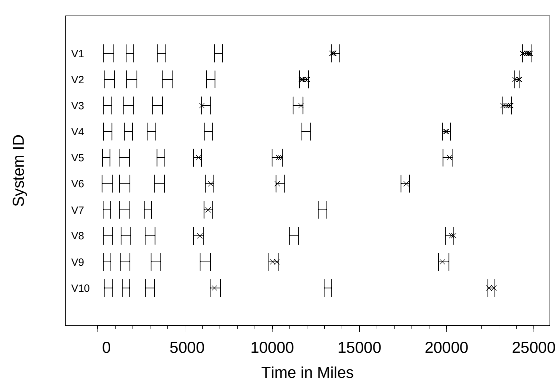

Figure 5: Event Plot for AMSAA Vehicle Fleet Non-random Selection Window Data

200 and 400 miles. Such data arise, for example, when commanders are permitted to choose lower-mileage vehicles that they might perceive as being less likely to fail during the exercise.

Figures 3 to 5 are the event plots for these three data sets, showing the recurrent events and the observation windows.

3 NONPARAMETRIC ESTIMATION METHODS FOR WINDOW-OBSERVATION RECURRENCE DATA

3.1 Notation

We use the following notation. Let $N(t)$ denote the cumulative number of events for a single unit under analysis for the period $[0, t]$. The population mean cumulative function is denoted by $\mu(t) = \mathrm{E}[N(t)]$. If $\mu(t)$ is differentiable, then

$$\nu(t) = \frac{d\mu(t)}{dt} \tag{1}$$

is the recurrence rate, and $\nu(t) \times \Delta t$ can be interpreted as the approximate expected number of events to occur during the next short time interval $(t,\ t + \Delta t)$.

3.2 Nonparametric Estimation of the Mean Cumulative Function

First we review the nonparametric method for estimating the population MCF, with some changes in presentation that extend the method to allow for window-observation recurrence data. Nonparametric MCF estimation methods are described, for example, in Nelson (1988), Lawless and Nadeau (1995), Chapter 16 of Meeker and Escobar (1998), and Chapters 3 to 5 of Nelson (2003).

Let n denote the number of observed units and let m denote the number of unique event times. Also, let $t_1, ..., t_m$ be the unique event times. Then the nonparametric estimator of the population MCF $\mu(t_j)$ is

$$\hat{\mu}(t_j) = \sum_{k=1}^{j} \left[\frac{\sum_{i=1}^{n} \delta_i(t_k) \times d_i(t_k)}{\sum_{i=1}^{n} \delta_i(t_k)} \right] = \sum_{k=1}^{j} \frac{d.(t_k)}{\delta.(t_k)} = \sum_{k=1}^{j} \bar{d}(t_k), \quad j = 1, ..., m, \tag{2}$$

where $d_i(t_k)$ is the number of events recorded at time t_k for unit i, and

$$\delta_i(t_k) = \begin{cases} 1 & \text{if unit } i \text{ is under observation in a time window at time } t_k, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $d.(t_k) = \sum_{i=1}^{n} \delta_i(t_k) \times d_i(t_k)$ is the total number of events reported at time t_k, $\delta.(t_k) = \sum_{i=1}^{n} \delta_i(t_k)$ is the size of the risk set (the number of systems at risk to have an event) at time t_k, taking account of gaps between observation windows and censoring, and $\bar{d}(t_k)$ is the sample mean number of events at time t_k. Thus $\hat{\mu}(t_j)$ can be viewed as an estimate of the cumulative mean number of events at time t_j, where the mean at the time of each event is computed with respect to the risk set at the time of the event.

The MCF estimator can be extended to estimate the mean cumulative cost (or other covariate of interest) for recurrent events. Section 3.3 of Nelson (2003) describes the "MCF for cost from exact age data with right censoring." In this more general cost model, $\hat{\mu}(t_j)$ is the cumulative average cost per unit by time t_j, while $d_i(t_k)$ is the total cost recorded at time t_k for unit i, and $\sum_{i=1}^{n} \delta_i(t_k) \times d_i(t_k)$ is the total cost at that time for all units under observation.

Because the estimate $\hat{\mu}(t_j)$ is not continuous (there are jumps in the estimate at each event time), we cannot use (1) to estimate the recurrence rate. We can, however, obtain an estimate of the recurrence rate over any interval $[t_a, t_b]$ as

$$\hat{\nu}(t_a, t_b) = \frac{\hat{\mu}(t_b) - \hat{\mu}(t_a)}{t_b - t_a}. \tag{3}$$

3.3 Assumptions

The nonparametric methods described here require no assumptions on the form of the recurrence process that produces the recurrent events. Therefore there is no risk of choosing an incorrect functional form for the MCF and there is no need for the assumption of independent increments. Refer to Nelson (2003, pages 51-54) for a detailed explanation of what is assumed and what is not assumed for the nonparametric estimator of the MCF when exact recurrence times are recorded. Under these assumptions, $\hat{\mu}(t_j)$ has the desirable properties of unbiasedness for recurrence data without gaps. For $\hat{\mu}(t_j)$ to be unbiased for window-observation recurrence data, an additional assumption, which is not stated explicitly in Nelson (2003) or Lawless and Nadeau (1995), is that the size of the risk set is always positive for the entire period of estimation. This is because the incremental changes in the process MCF are not estimable over the intervals corresponding to gaps where the size of the risk set is zero. However, in a naive implementation of the nonparametric estimation method, this incremental change is estimated to be 0.

Analyses of window-observation data also require the assumption that the recurrence process is not affected by whether a unit is being observed or not. For the extended warranty example, this assumption would be violated if customers tended to purchase an extended warranty because they have higher recurrence rates than those who do not. For the FEDC example, the assumption would be violated if intensive use during exercises changes the rate of failures with respect to miles of service.

3.4 Variance of $\hat{\mu}(t)$ and Its Estimator

In addition to the point estimate of $\mu(t_j)$ in (2), there is usually need to provide a standard error for $\hat{\mu}(t_j)$. The variance of $\hat{\mu}(t_j)$ is

$$\text{Var}[\hat{\mu}(t_j)] = \sum_{k=1}^{j} \text{Var}[\bar{d}(t_k)] + 2\sum_{k=1}^{j-1}\sum_{v=k+1}^{j} \text{Cov}[\bar{d}(t_k), \bar{d}(t_v)]. \quad (4)$$

Nelson (1995) provides an unbiased estimator for $\text{Var}[\hat{\mu}(t_j)]$, which could (generally with small probability) be negative. Lawless and Nadeau (1995) provide an alternative estimator that is not unbiased, but is always nonnegative. To estimate the variance of $\hat{\mu}(t_j)$ with window-observation data, we use a modification of the variance estimator given by Lawless and Nadeau (1995). The method and formula described below make it possible to handle large recurrence data sets. Because the method is valid for the more general cost-accumulation model, we will use the term "cost," instead of "number of events," in describing the data and estimation method.

The variance estimate $\widehat{\text{Var}}[\widehat{\mu}(t_j)]$ can be computed recursively as follows.

$$\widehat{\text{Var}}[\widehat{\mu}(t_1)] = \widehat{\text{Var}}(\bar{d}_1), \quad (5)$$

$$\widehat{\text{Var}}[\widehat{\mu}(t_j)] = \widehat{\text{Var}}[\widehat{\mu}(t_{j-1})] + \widehat{\text{Var}}(\bar{d}_j) + 2\sum_{k=1}^{j-1} \widehat{\text{Cov}}(\bar{d}_k, \bar{d}_j), \text{ for } j = 2, \ldots, m, \quad (6)$$

where $\bar{d}_j = \bar{d}(t_j)$ is the average cost of events per unit at $t_j, j = 1, \ldots, m$.

To present the details of the calculations, we use the following notation, defined for $j, k = 1, \ldots, m$.

- A_j is the total number of events in $(0, t_j]$.
- $\delta_j = \delta.(t_j)$ is the size of the risk set at time t_j- (i.e., just before t_j).
- $\delta_{j,k} = \sum_{i=1}^{n} \delta_i(t_j)\delta_i(t_k)$ is the number of units that were under observation at both t_j- and t_k-.
- Also,

$$\bar{d}_j^{\,k} = \bar{d}^k(t_j) = \frac{\sum_{i=1}^{n} \delta_i(t_k)\delta_i(t_j)d_i(t_j)}{\delta_{j,k}} \quad (7)$$

  is the average increase in cost at time t_j for all of those units that were under observation at both t_j- and t_k-.

Then, we have

$$\widehat{\mathrm{Var}}(\bar{d}_j) = \frac{1}{\delta_j^2}\sum_{u=1}^{n}\delta_u(t_j)[d_u(t_j)-\bar{d}_j]^2$$
$$= \frac{1}{\delta_j^2}\left[\sum_{i=A_{j-1}+1}^{A_j}(d_i-\bar{d}_j)^2+(\delta_j-A_j+A_{j-1})\bar{d}_j^{\,2}\right]. \qquad (8)$$

In (8), the term $(\delta_j - A_j + A_{j-1})\bar{d}_j^{\,2}$ accounts for units that were under observation at time t_j- but had no events at that time. Also,

$$\widehat{\mathrm{Cov}}(\bar{d}_k,\bar{d}_j) = \frac{1}{\delta_k\delta_j}\sum_{u=1}^{n}\delta_u(t_k)d_u(t_k)\delta_u(t_j)[d_u(t_j)-\bar{d}_j^{\,k}]$$
$$= \frac{1}{\delta_k\delta_j}\sum_{l=A_{k-1}+1}^{A_k}\left\{d_l I(t_j\in O_l)\left[\sum_{i=A_{j-1}+1}^{A_j}I(K_i=K_l)d_i-\bar{d}_j^{\,k}\right]\right\}, \qquad (9)$$

where for event l, K_l is the unit ID (identification number) for the system having the event, d_l is the corresponding cost of the event, and O_l is the set of observation intervals for the system. In addition, $I(t_j \in O_l)$ is the indicator function for whether the time t_j is in the set of observation intervals O_l while $I(K_i = K_l)$ is the indicator function for whether event i and event l are for the same system. By (7), $\bar{d}_j^{\,k} = \frac{1}{\delta_{j,k}}\sum_{i=A_{j-1}+1}^{A_j} I(t_k \in O_i)d_i$ is the sample mean cost of recurrent events at time t_j for units that are under observation at time t_k.

When there are no gaps in the observation period for any units, our variance estimator is equivalent to that of Lawless and Nadeau (1995). The main difference, required for the window-observation data, is in the computation of the sample mean in (7) and the covariances in (9), which must account for which units with an event at time t_j were under observation at event times $t_k < t_j$.

3.5 Confidence Intervals for $\mu(t)$

Using $\hat{\mu}(t)$ and $\widehat{\mathrm{Var}}[\hat{\mu}(t)]$, pointwise normal-approximation confidence intervals are easy to compute from the following equations (Meeker and Escobar, 1998, Chapter 16, Page 400).

- Based on $Z_{\hat{\mu}(t)} = [\hat{\mu}(t) - \mu(t)]/\widehat{\mathrm{se}}_{\hat{\mu}(t)} \mathrel{\dot\sim} \mathrm{NOR}(0, 1)$, an approximate $100(1 - \alpha)\%$ confidence interval for $\mu(t)$ is

$$\hat{\mu}(t) \pm z_{(1-\alpha/2)}\widehat{\mathrm{se}}_{\hat{\mu}(t)}, \qquad (10)$$

  where $\widehat{\mathrm{se}}_{\hat{\mu}(t)} = \sqrt{\widehat{\mathrm{Var}}[\hat{\mu}(t)]}$, and $z_{(1-\alpha/2)}$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution.

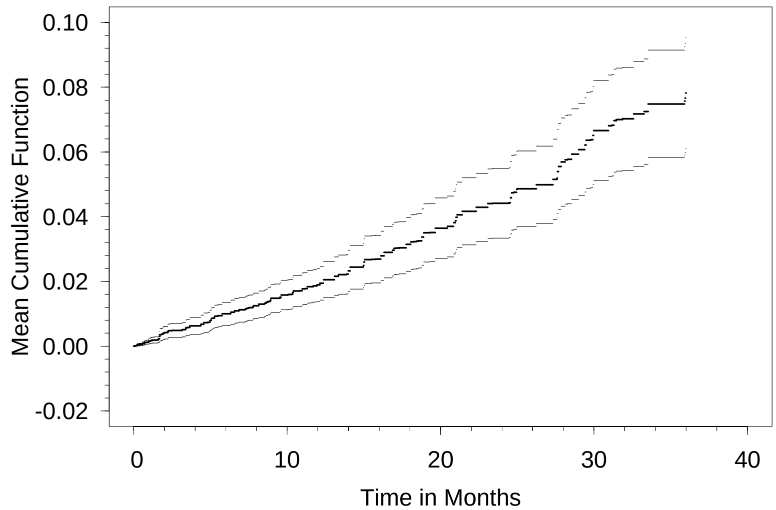

Figure 6: MCF plot for Labor Code C6050 Extended Warranty Data

- Based on $Z_{\log[\hat{\mu}(t)]} = \{\log[\hat{\mu}(t)] - \log[\mu(t)]\} / \widehat{\text{se}}_{\log[\hat{\mu}(t)]} \dot{\sim} \text{NOR}(0,1)$, an approximate $100(1-\alpha)\%$ confidence interval for $\mu(t)$ is

$$[\hat{\mu}(t)/w, \quad \hat{\mu}(t) \times w], \tag{11}$$

  where $w = \exp[z_{(1-\alpha/2)} \widehat{\text{se}}_{\hat{\mu}(t)} / \hat{\mu}(t)]$.

Equation (11) might provide a better approximate confidence interval procedure in applications where $\mu(t)$ is strictly positive (and the interval endpoints will always be positive). Meeker and Escobar (1998, Chapter 16, Page 402) point out that when the size of the risk set is small (say, less than 30), using $t_{(p;\nu)}$ (the p quantile of a t distribution with ν degrees of freedom) instead of $z_{(p)}$ in (10) and (11) can provide a confidence interval procedure with a coverage probability that is closer to the nominal value.

3.6 Nonparametric Estimation of the MCF for the Extended Warranty Data

Figure 6 shows, for the Extended Warranty Data, the nonparametric MCF estimate of labor code C6050, and the corresponding approximate 95% pointwise confidence intervals. The discontinuities in the estimates reflect the jumps at the time of events. The confidence intervals were computed using (10).

4 MAXIMUM LIKELIHOOD ESTIMATION OF A NON-HOMOGENEOUS POISSON PROCESS MODEL USING WINDOW-OBSERVATION RECURRENCE DATA

This section describes how to fit a parametric model to the recurrence data. This is, in effect, fitting a curve through the nonparametric MCF estimate. Compared to nonparametric methods, fitting a parametric model has advantages such as a more concise model with just a few parameters and the ability to extrapolate outside the range of the data. Also, the maximum likelihood (ML) estimator is generally consistent for window-observation recurrence data, even if the size of the risk set is sometimes equal to zero.

The Poisson process is a widely used parametric model for point process data. Detailed descriptions of this model and corresponding statistical methods are available in many books, such as Cox and Lewis (1966) and Basu and Rigdon (2000). In this section, we give an expression for the likelihood and illustrate fitting the nonhomogeneous Poisson process (NHPP) model to window-observation recurrence data.

4.1 NHPP Model

A particular NHPP model is specified by its recurrence rate $\nu(t)$. Let $N(a,b]$ be the number of events in the time range $(a,b]$ from an NHPP with recurrence rate $\nu(t)$. The expectation of $N(a,b]$ over this interval is $\mu(a,b) = \int_a^b \nu(t)dt$. The most commonly used NHPP recurrence rate functions are:

- NHPP model with power recurrence rate:

$$\nu(t;\beta,\eta) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}, \quad \beta > 0, \eta > 0.$$

- NHPP model with loglinear recurrence rate:

$$\nu(t;\gamma_0,\gamma_1) = \exp(\gamma_0 \ + \ \gamma_1 \times t).$$

Of course an important special case of the NHPP model is the homogeneous Poisson process (HPP), in which $\nu(t)$ is a constant (does not depend on t).

4.2 NHPP Likelihood for the Window-Observation Data

For unit i with exact event times for the observation period $(0, t_{a_i}]$ and no gaps in observation, the NHPP likelihood is

$$L_i(\boldsymbol{\theta}) = \left\{ \prod_{j=1}^{r_i} \nu(t_{ij}; \boldsymbol{\theta}) \right\} \left\{ \exp\left[-\mu(0, t_{a_i}; \boldsymbol{\theta})\right] \right\}, \tag{12}$$

where $\boldsymbol{\theta}$ is the unknown parameter vector (e.g. $\boldsymbol{\theta} = (\beta, \eta)'$ for the NHPP model with power recurrence rate), r_i is the total number of events being observed for unit i, and $t_{i1}, ..., t_{ir_i}$ are the corresponding unique event times.

For window-observation recurrence data, denote the non-overlapping windows of observation for unit i as $(t_{i1L}, t_{i1U}]$, $(t_{i2L}, t_{i2U}]$, $\ldots$, $(t_{ip_iL}, t_{ip_iU}]$ (with $t_{i1L} \geq 0, t_{i(k-1)U} < t_{ikL}, t_{ip_iU} \leq t_{a_i}$). The NHPP likelihood for unit i with events in observation windows reported at exact times is

$$L_i(\boldsymbol{\theta}) = \left\{ \prod_{j=1}^{r_i} \nu(t_{ij}; \boldsymbol{\theta}) \right\} \left\{ \prod_{k=1}^{p_i} \exp\left[-\mu(t_{ikL}, t_{ikU}; \boldsymbol{\theta})\right] \right\}. \tag{13}$$

It is easy to establish (12) and (13) by writing the probability of the data for interval censored data (event times are known to be in specific intervals) and then allowing the width of the intervals to approach 0.

For a sample of n independent NHPP systems with the same intensity function, the overall likelihood is simply the product of the likelihoods for the individual units,

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} L_i(\boldsymbol{\theta}), \tag{14}$$

and $\hat{\boldsymbol{\theta}}$, the ML estimator of $\boldsymbol{\theta}$, is obtained by maximizing (14) or its logarithm. Generally, this must be done numerically.

Given $\hat{\boldsymbol{\theta}}$, the NHPP MCF ML estimator is

$$\hat{\mu}(t) = \int_0^t \nu(x; \hat{\boldsymbol{\theta}}) dx. \tag{15}$$

4.3 Variance of the NHPP $\hat{\mu}(t)$ and Its Estimator

Using delta method, an approximate variance of the NHPP $\hat{\mu}(t)$ is

$$\text{Var}[\hat{\mu}(t)] \doteq \left[\frac{\partial \mu(t)}{\partial \boldsymbol{\theta}} \right]' \Sigma_{\hat{\boldsymbol{\theta}}} \left[\frac{\partial \mu(t)}{\partial \boldsymbol{\theta}} \right], \tag{16}$$

where $\Sigma_{\hat{\boldsymbol{\theta}}}$ is the variance-covariance matrix for $\hat{\boldsymbol{\theta}}$. An estimator $\widehat{\Sigma}_{\hat{\boldsymbol{\theta}}}$ of $\Sigma_{\hat{\boldsymbol{\theta}}}$ can be obtained by evaluating the inverse of the negative Hessian matrix,

$$-\left[\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]^{-1}, \tag{17}$$

at $\hat{\boldsymbol{\theta}}$.

An estimator for the variance in (16) is

$$\widehat{\text{Var}}[\hat{\mu}(t)] = \left[\frac{\partial\mu(t)}{\partial\boldsymbol{\theta}}\right]'_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \widehat{\Sigma}_{\hat{\boldsymbol{\theta}}} \left[\frac{\partial\mu(t)}{\partial\boldsymbol{\theta}}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \tag{18}$$

For example, using the NHPP model with the power recurrence rate,

$$\mu(t) = \int_0^t \nu(x;\beta,\eta)dx = \left(\frac{t}{\eta}\right)^{\beta}$$

and

$$\frac{\partial\mu(t)}{\partial\boldsymbol{\theta}} = \left[\left(\frac{t}{\eta}\right)^{\beta}\ln\left(\frac{t}{\eta}\right),\ -\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta}\right]', \quad \text{where } \boldsymbol{\theta} = [\beta,\ \eta]'.$$

Thus an estimator of the variance of $\hat{\mu}(t)$ for the NHPP power recurrence rate model is

$$\begin{aligned}
\widehat{\text{Var}}[\hat{\mu}(t)] &= \left[\left(\frac{t}{\hat{\eta}}\right)^{\hat{\beta}}\ln\left(\frac{t}{\hat{\eta}}\right)\right]^2 \widehat{\text{Var}}(\hat{\beta}) \\
&+ 2\left[\left(\frac{t}{\hat{\eta}}\right)^{\hat{\beta}}\ln\left(\frac{t}{\hat{\eta}}\right)\right]\left[-\frac{\hat{\beta}}{\hat{\eta}}\left(\frac{t}{\hat{\eta}}\right)^{\hat{\beta}}\right]\widehat{\text{Cov}}(\hat{\beta},\ \hat{\eta}) \\
&+ \left[-\frac{\hat{\beta}}{\hat{\eta}}\left(\frac{t}{\hat{\eta}}\right)^{\hat{\beta}}\right]^2 \widehat{\text{Var}}(\hat{\eta}).
\end{aligned} \tag{19}$$

5 ANALYSIS OF THE AMSAA VEHICLE FLEET DATA

In this section, we apply both the nonparametric MCF estimation method and the NHPP power recurrence rate model to the three simulated AMSAA Vehicle Fleet data sets. An important distinguishing feature of these window-observation data (not present in the Extended Warranty Data) is that there are periods of time over which the size of the risk set is zero.

5.1 Nonparametric and Parametric Estimation

Figures 7 to 9 compare the MCF estimate using the nonparametric method, the NHPP model with the power recurrence rate, and the “true” model that generated the simulated data. For the NHPP

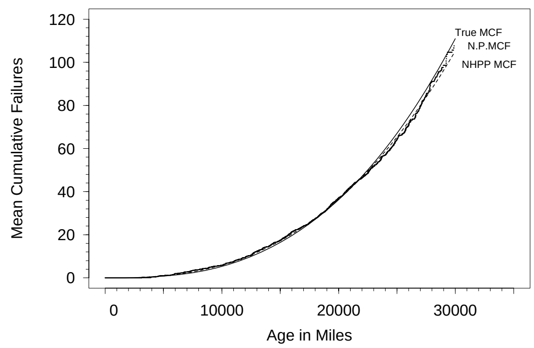

Figure 7: AMSAA Vehicle Fleet Complete Data: Nonparametric MCF Estimate, NHPP MCF Estimate, and the True MCF

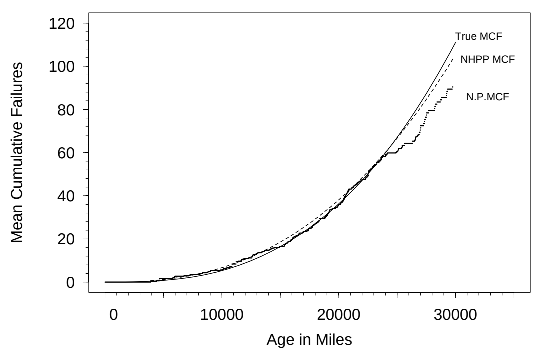

Figure 8: AMSAA Vehicle Fleet Random Selection Window Data: Nonparametric MCF Estimate, NHPP MCF Estimate, and the True MCF

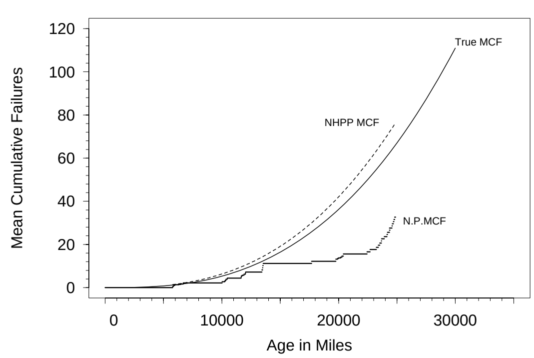

Figure 9: AMSAA Vehicle Fleet Non-random Selection Window Data: Nonparametric MCF Estimate, NHPP MCF Estimate, and the True MCF

Table 2: AMSAA Vehicle Fleet Data – Comparisons of NHPP Power Recurrence Rate Model Estimates

Parameter		*Complete Data*	*Random Selection Window Data*	*Non-random Selection Window Data*
	MLE	5063.070	4686.747	5098.635
η	Std.Err.	310.798	515.508	801.295
(true value: 5447)	95% Lower Bound	4453.920	3676.370	3528.126
	95% Upper Bound	5672.223	5697.123	6669.144
	MLE	2.617	2.509	2.736
β	Std.Err.	0.095	0.156	0.276
(true value: 2.76)	95% Lower Bound	2.430	2.202	2.195
	95% Upper Bound	2.804	2.815	3.276
Maximum Log Likelihood		-4606	-1564	-298

model, we fit both the power recurrence rate model and the loglinear recurrence rate model. Tables 2 and 3 summarize the results for the three data sets. We note the following:

- The NHPP estimates all have good agreement with the true MCF.
- For all data sets, the values of the maximum log likelihood are somewhat larger for the power recurrence rate model (from which the data were simulated), when compared with the loglinear recurrence rate model.
- For the Complete Data, the nonparametric estimate is very close to the true MCF.
- For the Random Selection Window Data, the nonparametric estimate is somewhat smaller than the true MCF, especially after 24,000 miles.
- For the Non-random Selection Window Data, the nonparametric estimate is importantly below the true MCF.

We will discuss the downward bias of the nonparametric estimator with window-observation data in Section 5.2.

5.2 Problems Caused by Periods with Zero-Size Risk Set

In Section 3.3 we stated that an additional assumption needed for the MCF estimator in (2) to be unbiased is that the size of the risk set is always positive for the entire period of estimation. This

Table 3: AMSAA Vehicle Fleet Data – Comparisons of NHPP Loglinear Recurrence Rate Model Estimates

Parameter		Complete Data	Random Selection Window Data	Non-random Selection Window Data
	MLE	-7.728	-7.558	-8.559
γ_0	Std.Err.	0.114	0.190	0.395
	95% Lower Bound	-7.952	-7.931	-9.333
	95% Upper Bound	-7.504	-7.185	-7.786
	MLE	0.0001140	0.0001060	0.0001591
γ_1	Std.Err.	0.0000057	0.0000095	0.0000209
	95% Lower Bound	0.0001030	0.0000870	0.0001182
	95% Upper Bound	0.0001250	0.0001250	0.0002001
Maximum Log Likelihood		-4624	-1570	-303

Table 4: Simulated Vehicle Data – Comparisons of Risk Set Sizes

Risk Set	Complete Data		Random Selection Window Data		Non-random Selection Window Data	
size	time	pct (%)	time	pct (%)	time	pct (%)
0	0	0.00	3949	13.26	13371	53.73
1	1042	3.48	5349	17.96	5235	21.04
2	1271	4.25	5444	18.28	2493	10.02
> 2	27593	92.27	15037	50.49	3786	15.21
Total	29906	100	29779	100	24885	100

assumption is satisfied by the Complete Data, but not by the Random Selection Window Data or the Non-random Selection Window Data, because for some time intervals, the size of the risk set is zero.

Table 4 shows the amount of total observation time with risk set size at the values of 0, 1, 2 and greater than 2, for the Complete Data, the Random Selection Window Data, and the Non-random Selection Window Data. Note that for the Non-random Selection Window Data, the size of the risk set is zero for 53.73% of the total observation time, while for the Random Selection Window Data, the figure is 13.26%. This difference can be visualized by looking at the risk set plots in Figures 10 to 12.

Equation (2) would produce an unbiased estimate for the MCF, conditional on the assumption

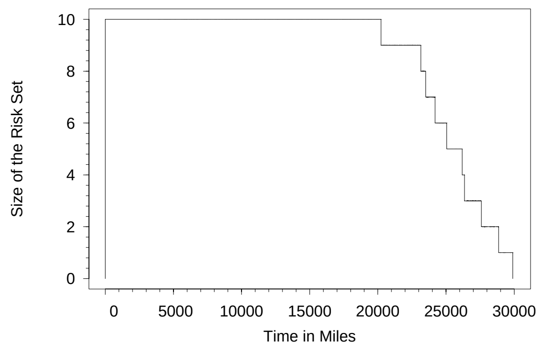

Figure 10: Risk Set Plot for AMSAA Vehicle Fleet Complete Data

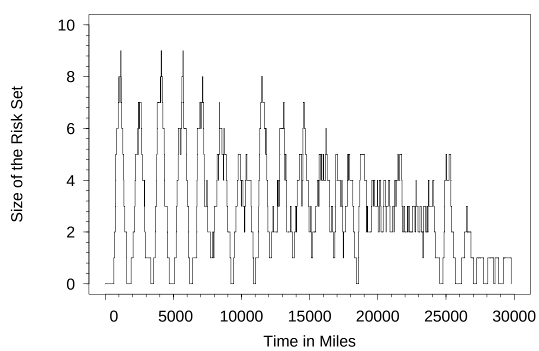

Figure 11: Risk Set Plot for AMSAA Vehicle Fleet Random Selection Window Data

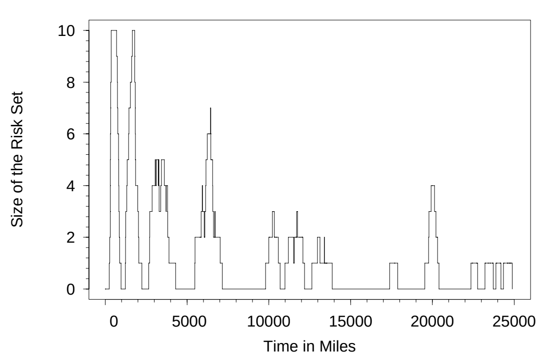

Figure 12: Risk Set Plot for AMSAA Vehicle Fleet Non-random Selection Window Data

that the true MCF does not increase when the size of the sample risk set is zero. Usually, this assumption is not realistic. In Section 6, we will present a hybrid estimator that uses a parametric model to estimate the expected number of events in the gaps with zero-size risk set.

When estimating the variance of the nonparametric MCF estimator, there can be estimation problems even when the size of the risk set is positive over the period of interest. In particular, when the size of the risk set is 1 at event time t_k, then $\mathrm{Var}[\bar{d}(t_k)]$ is not estimable. Also, unless at least two units are being observed at both event times t_k and t_v, $\mathrm{Cov}[\bar{d}(t_k), \bar{d}(t_v)]$ is not estimable. The problem here is similar to trying to estimate a variance or a covariance from a sample of size one.

6 PARAMETRIC ADJUSTMENTS TO THE NONPARAMETRIC ESTIMATOR

6.1 Hybrid Estimator

Here we propose a simple adjustment method for the nonparametric estimator that uses a parametric model to estimate the increase in the MCF for the intervals with zero-size risk set. This method helps reduce the downward bias induced by intervals with an empty risk set, yet retaining, as much as possible, the desirable property of keeping the risks of using inappropriate model assumptions to a minimum. The key steps needed to obtain this hybrid estimator $\hat{\mu}_{CNP}(t)$ are outlined below.

1. For the time intervals with a nonempty risk set, apply (2) to obtain the estimate of increase in the MCF as before. More specifically, at each unique event time t_k, the nonparametric estimate of the increase in the MCF is

$$\bar{d}(t_k) = \frac{\sum_{i=1}^{n} \delta_i(t_k) \times d_i(t_k)}{\sum_{i=1}^{n} \delta_i(t_k)}.$$

This is contribution due to events in the observation windows from the nonparametric model.

2. Use all the data available to fit a parametric model for the recurrence events, and obtain the parametric estimate for $\nu(t)$. In our example, we use the power recurrence rate NHPP model.

3. For the time intervals with a zero-size risk set, use the estimated increase in the MCF from the parametric model as the estimate of the increase in the MCF in the interval. Assuming that there are q gaps with zero-size risk set between the observation windows, $(t_{1L}, t_{1U}]$, $(t_{2L}, t_{2U}]$, $\ldots$, $(t_{qL}, t_{qU}]$ (with $t_{1L} \geq 0, t_{(i-1)U} < t_{iL}, t_{qU} \leq t_a$), the estimated increase in MCF for gap i is

$$d_i^{\dagger} = \int_{t_{iL}}^{t_{iU}} \hat{\nu}(t) dt$$

where $\hat{\nu}(t)$ is the estimated recurrence rate, evaluated at the ML estimate of the NHPP model parameters.

The $d_i^\dagger$ values provide estimates for the contribution (number of events or cost) for the time intervals with zero-size risk set.

4. Calculate the estimate for the MCF by summing over time the estimated increase in the MCF obtained in Steps 1 to 3. That is,

$$\hat{\mu}_{CNP}(t) = \bar{d}.(t) + d_{\cdot}^{\dagger}(t) \tag{20}$$

where $\bar{d}.(t) = \sum_{k:t_k \le t} \bar{d}(t_k)$

and $d_{\cdot}^{\dagger}(t) = \begin{cases} \sum_{i:t_{iU} \le t} d_i^\dagger & \text{if } t \text{ is not in a gap with zero-size risk set} \\ \sum_{i:t_{iU} \le t} d_i^\dagger + \int_{t_{jL}}^{t} \hat{\nu}(x)dx & \text{if } t \text{ is in the } j^{th} \text{ gap with zero-size risk set } (t_{jL} < t \le t_{jU}). \end{cases}$

The resulting hybrid estimator $\hat{\mu}_{CNP}(t)$ consists of two parts: the contribution from the nonparametric model $\bar{d}.(t)$ and the contribution from the parametric adjustment $d_{\cdot}^{\dagger}(t)$. If the proportion of time with zero-size risk set is relatively small, the hybrid estimator will be dominated by the nonparametric model; otherwise, it will be more strongly affected by the parametric model.

6.2 Approximate Variance of the Hybrid Estimator $\hat{\mu}_{CNP}(t)$ and Its Estimator

Direct computation from (20) provides the following equation for variance of $\hat{\mu}_{CNP}(t)$:

$$\text{Var}[\hat{\mu}_{CNP}(t)] = \text{Var}[\bar{d}.(t)] + \text{Var}[d_{\cdot}^{\dagger}(t)] + 2\text{Cov}[\bar{d}.(t),\ d_{\cdot}^{\dagger}(t)], \tag{21}$$

where $\text{Var}[\bar{d}.(t)]$ can be estimated by using the method described in Section 3.4, if the size of the risk set is greater than one at each recorded event time. However, $\text{Var}[\bar{d}(t_k)]$ is not estimable if the size of the risk set is one at the time t_k. For such event times, Appendix B suggests a conservative approach for estimating $\text{Var}[\bar{d}(t_k)]$.

$\text{Var}[d_{\cdot}^{\dagger}(t)]$ can be estimated using the delta method described in Section 4.3. Appendix A gives more detail on the computation of $\widehat{\text{Var}}[d_{\cdot}^{\dagger}(t)]$.

The data, however, contain no information about the terms $\text{Cov}[\bar{d}.(t),\ d_{\cdot}^{\dagger}(t)]$. It is assumed that these terms are zero, which would be implied by the commonly used assumption of independent increments (e.g., in fitting the NHPP model).

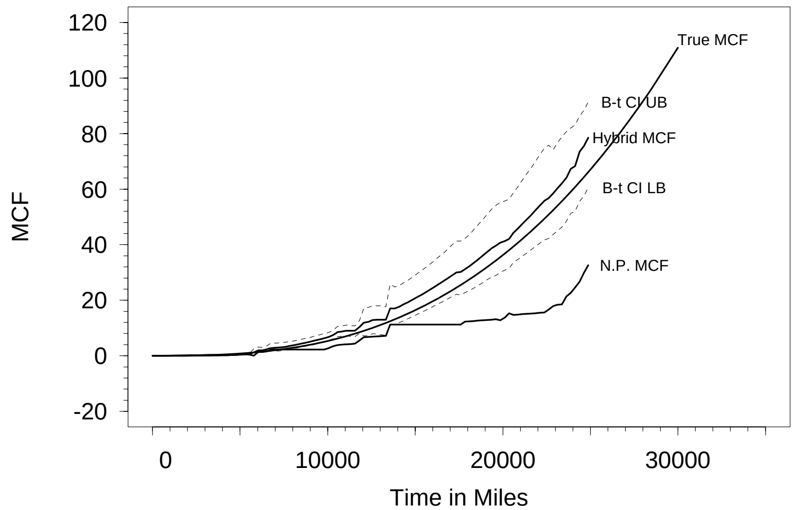

Figure 13: MCF Plots for AMSAA Vehicle Fleet Non-random Selection Window Data, Comparing True Model, Nonparametric MCF Estimates, Hybrid MCF Estimates, and Bootstrap-t CI Based on B=5000 Nonparametric Re-samplings

6.3 Confidence Intervals for $\mu(t)$

To compute confidence intervals (CI's) for $\mu(t)$ with the hybrid estimator $\hat{\mu}_{CNP}(t)$, one can use the normal approximation method outlined in Section 3.5, along with the estimate of the variance in (21), and use either (10) or (11). Obtaining CI's using Bootstrap methods is another option. For example, Efron and Tibshirani (1993) describe how to calculate Bootstrap-t CI on pages 160-161. They also describe other Bootstrap methods to calculate CI's.

6.4 Hybrid Estimation of the MCF for the AMSAA Vehicle Fleet Data

Figure 13 shows, for the Non-random Selection Window Data, the MCF estimates from the nonparametric model, the hybrid MCF estimates, the true MCF, and the 95% Bootstrap-t CI's. We use the Bootstrap method instead of the normal approximation method to calculate confidence intervals here for the hybrid estimation method, because our simulation studies have shown that the simple normal approximation method does not work well when the expected number of events is small. MCF estimates from the hybrid estimator show substantial improvement, compared to the nonparametric MCF estimates, and the true MCF is within the Bootstrap-t CI's.

7 CONCLUDING REMARKS AND AREAS FOR FURTHER RESEARCH

We have shown how to extend existing nonparametric and parametric methods for recurrence data to analyze window-observation recurrence data. No additional assumptions are needed for the parametric methods, but one must specify a particular form for the NHPP recurrence rate function.

In our paper, the example using a parametric adjustment to nonparametric estimator applies to the number of events. To use the same method for the more general cumulative cost model, one option of the parametric models is to fit compound Poisson processes. Descriptions of compound Poisson processes are available in many books, such as Cox and Isham (1980) or Parzen (1962).

ACKNOWLEDGMENTS

We would like to thank Dr. Michael J. Cushing for suggesting that we work on the development of statistical methods for window-observation data, for providing the background information on FEDC exercises, and for providing the simulated AMSAA Vehicle Fleet data. We would also like to thank Luis Escobar for helpful comments on an earlier version of this paper.

APPENDIX

A CALCULATION OF $\widehat{\mathrm{Var}}[d_{\cdot}^{\dagger}(t)]$

Recall that $d_{\cdot}^{\dagger}(t)$ is the estimator of the sum of the increases in the MCF $\mu(t)$ over all time intervals (gaps) from 0 to t, where the size of the risk set is zero. Here, using the NHPP power recurrence model, we illustrate how to use delta method to calculate an estimate of $\mathrm{Var}[d_{\cdot}^{\dagger}(t)]$.

We start from a simple case with only one zero-size risk set interval before time t, say (t_{a1}, t_{b1}). Then the parametric adjustment is

$$d_{\cdot}^{\dagger}(t) = d_1^{\dagger}(t_{a1}, t_{b1}) = \int_{t_{a1}}^{t_{b1}} \hat{\nu}(x)dx = \int_{t_{a1}}^{t_{b1}} \nu(x; \hat{\beta}, \hat{\eta})dx = \left(\frac{t_{b1}}{\hat{\eta}}\right)^{\hat{\beta}} - \left(\frac{t_{a1}}{\hat{\eta}}\right)^{\hat{\beta}}.$$

Let $g_1(t; \beta, \eta) = \int_{t_{a1}}^{t_{b1}} \nu(x; \beta, \eta)dx$. Then taking first derivative with respect to the model parameters $\boldsymbol{\theta} = [\beta,\ \eta]'$, we have

$$\frac{\partial g_1(t; \beta, \eta)}{\partial \boldsymbol{\theta}} = \left[\left(\frac{t_{b1}}{\eta}\right)^{\beta} \ln\left(\frac{t_{b1}}{\eta}\right) - \left(\frac{t_{a1}}{\eta}\right)^{\beta} \ln\left(\frac{t_{a1}}{\eta}\right),\ -\frac{\beta}{\eta}\left(\left(\frac{t_{b1}}{\eta}\right)^{\beta} - \left(\frac{t_{a1}}{\eta}\right)^{\beta}\right)\right]'. \tag{22}$$

Applying the delta method, one can calculate an estimate for $\mathrm{Var}[d_{\cdot}^{\dagger}(t)]$ as

$$\widehat{\mathrm{Var}}[d_1^{\dagger}(t_{a1}, t_{b1})] = \left[\frac{\partial g_1(t; \beta, \eta)}{\partial \boldsymbol{\theta}}\right]'_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \widehat{\Sigma}_{\hat{\boldsymbol{\theta}}} \left[\frac{\partial g_1(t; \beta, \eta)}{\partial \boldsymbol{\theta}}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \tag{23}$$

Now we extend to the more general cases, which allow more than one zero-size risk set intervals before time t as well as allow time t in one of the zero-size risk set intervals.

In the more general cases, we first identify all the intervals with zero-size risk set on or before time t, say $(t_{a1}, t_{b1}), (t_{a2}, t_{b2}), ..., (t_{ak}, t_{bk})$. If t is in the interval (t_{ak}, t_{bk}), then the last interval is replaced by (t_{ak}, t). Without loss of generality, we take the last interval as (t_{ak}, t_{bk}). Then

$$d_{\cdot}^{\dagger}(t) = \sum_{i=1}^{k} d_i^{\dagger}(t_{ai}, t_{bi}),$$

where

$$d_i^{\dagger}(t_{ai}, t_{bi}) = \int_{t_{ai}}^{t_{bi}} \hat{\nu}(x)dx = \int_{t_{ai}}^{t_{bi}} \nu(x; \hat{\beta}, \hat{\eta})dx = \left(\frac{t_{bi}}{\hat{\eta}}\right)^{\hat{\beta}} - \left(\frac{t_{ai}}{\hat{\eta}}\right)^{\hat{\beta}}.$$

Let $g(t; \beta, \eta)$ be the row vector with element i given by $g_i(t; \beta, \eta) = \int_{t_{ai}}^{t_{bi}} \nu(x; \beta, \eta)dx$. Then by calculating

$$\left[\frac{\partial g(t;\beta,\eta)}{\partial \boldsymbol{\theta}}\right]'_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \widehat{\Sigma}_{\hat{\boldsymbol{\theta}}} \left[\frac{\partial g(t;\beta,\eta)}{\partial \boldsymbol{\theta}}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

we obtain a k by k symmetric matrix, with the diagonal elements $\widehat{\text{Var}}[d_i^\dagger(t_{ai}, t_{bi})]$ and the off-diagonal elements $\widehat{\text{Cov}}[d_i^\dagger(t_{ai}, t_{bi}), d_j^\dagger(t_{aj}, t_{bj})]$, $i, j = 1, 2, ..., k$. It can be shown that the sum of all the elements in this matrix is an estimator of $\text{Var}[d_\cdot^\dagger(t)]$.

B A CONSERVATIVE ESTIMATOR OF $\text{Var}[\bar{d}(t_k)]$ WHEN THE SIZE OF THE RISK SET IS 1

When the size of the risk set at t_k, $\delta.(t_k)$, is 1, variance of $\bar{d}(t_k)$ in (2) is not estimable. The simple moment estimator with the following formula generates an estimate at the value of zero (as in estimating the variance with a sample size of one):

$$\widehat{\text{Var}}[\bar{d}(t_k)] = \frac{\widehat{\text{Var}}[d_1(t_k)]}{\delta.(t_k)} \tag{24}$$

$$= \frac{\left\{\left[d_i(t_k) - \frac{d_i(t_k)}{\delta.(t_k)}\right]^2 + [\delta.(t_k) - 1]\left[0 - \frac{d_i(t_k)}{\delta.(t_k)}\right]^2\right\} / \delta.(t_k)}{\delta.(t_k)}. \tag{25}$$

Note that an assumption for (25) is that no two units have events at the same time. Therefore, for a time point with an event, t_k, only "cost" of one unit (say unit i) $d_i(t_k)$ is recorded, and all other units are with "cost" at the value of zero.

It can be shown that $\widehat{\text{Var}}[d_1(t_k)]$ in (24) is maximized when $\delta.(t_k)$ equals 2, independent of the value of $d_i(t_k)$. Therefore, for events recorded in the intervals with the size of the risk set at 1, we use (25) with $\delta.(t_k)$ at the value of 2. Then simplifying (25), we obtain a conservative estimator for $\text{Var}[\bar{d}(t_k)]$ as $[d_i(t_k)]^2/8$, in intervals with risk set size having a value of 1.

REFERENCES

Basu, A. P., and Rigdon, S. E. (2000), *Statistical Methods for the Reliability of Repairable Systems*, New York: Wiley.

Cox, D. R., and Isham, V. (1980), *Point Processes*, New York: Chapman and Hall.

Cox, D. R., and Lewis, P. A. W. (1966), *The Statistical Analysis of Series of Events*, New York: Wiley.

Efron, B., and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, New York: Chapman & Hall.

Lawless, J. F., and Nadeau, C. (1995), "Some Simple Robust Methods for the Analysis of Recurrent Events," *Technometrics*, 37, 158-168.

Lawless, J. F., and Zhan, M., (1998), "Analysis of Interval-grouped Recurrent-event Data Using Piecewise Constant Rate Functions," *The Canadian Journal of Statistics*, 26, 549-565.

Meeker, W. Q., and Escobar, L. A. (1998), *Statistical Methods for Reliability Data*, New York: Wiley.

Nelson, W. (1988), "Graphical Analysis of System Repair Data," *Journal of Quality Technology*, 20, 24-35.

Nelson, W. (1995), "Confidence Limits for Recurrence Data – Applied to Cost or Number of Product Repairs," *Technometrics*, 37, 147-157.

Nelson, W. (2003), *Recurrent Events Data Analysis for Product Repairs, Disease Recurrences, and Other Applications*, Philadelphia: ASA-SIAM.

Parzen, E. (1962), *Stochastic Processes*, San Francisco: Holden-Day.