

Analysis of Window-Observation Recurrence Data

Jianying Zuo

William Q. Meeker

Huaiqing Wu

Statistics Department

Iowa State University

Ames, IA 50011

June 5, 2007

Overview

- Introduction to recurrence data.
- Examples.
- Parametric and nonparametric estimation for window-observation recurrence data.
- Hybrid estimator for the MCF.
- Concluding remarks.

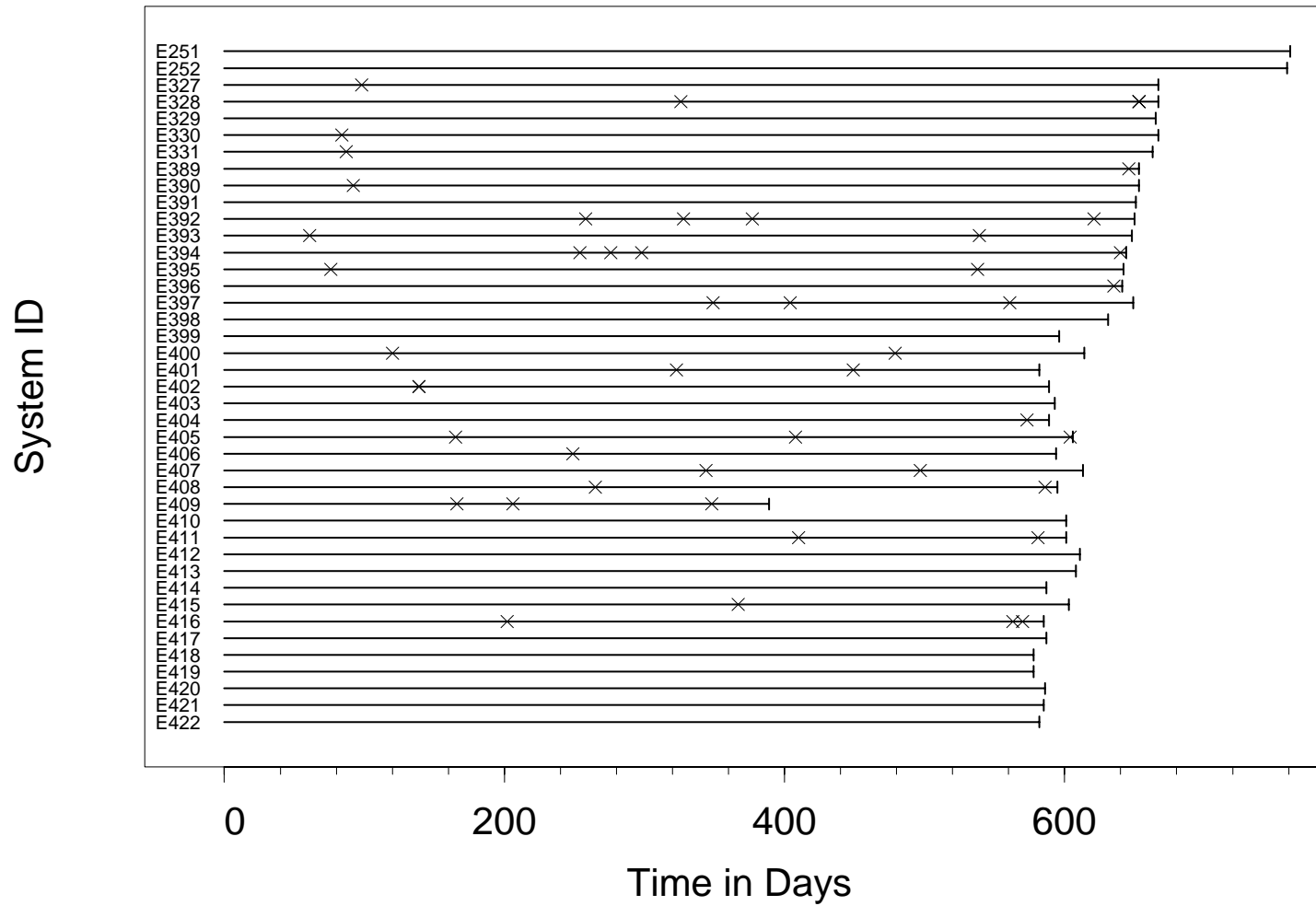
Introduction on Recurrence Data

- Recurrence data record events and related covariates such as cost over time.
- Statistics of interest include:
 - ▶ The expected cumulative number or cost of events per unit over the period $[0, t]$: $\mu(t)$.
 - ▶ The recurrence rate as a function of t : $\nu(t)$. If $\mu(t)$ is differentiable, then $\nu(t) = \frac{d\mu(t)}{dt}$.

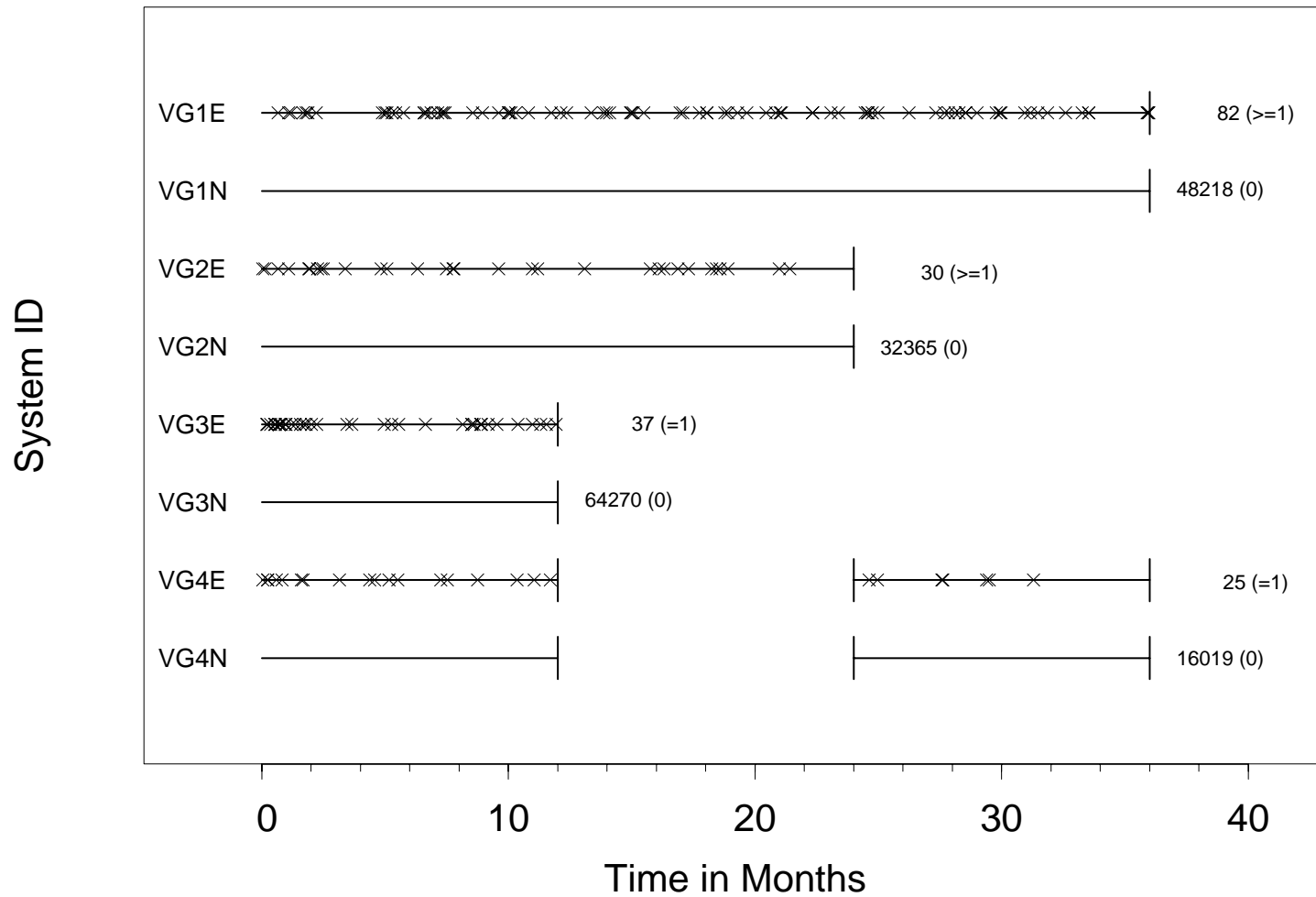
Example 1. ValveSeat Data (Nelson 1995)

Event Plot

Valve-Seat Replacement Data

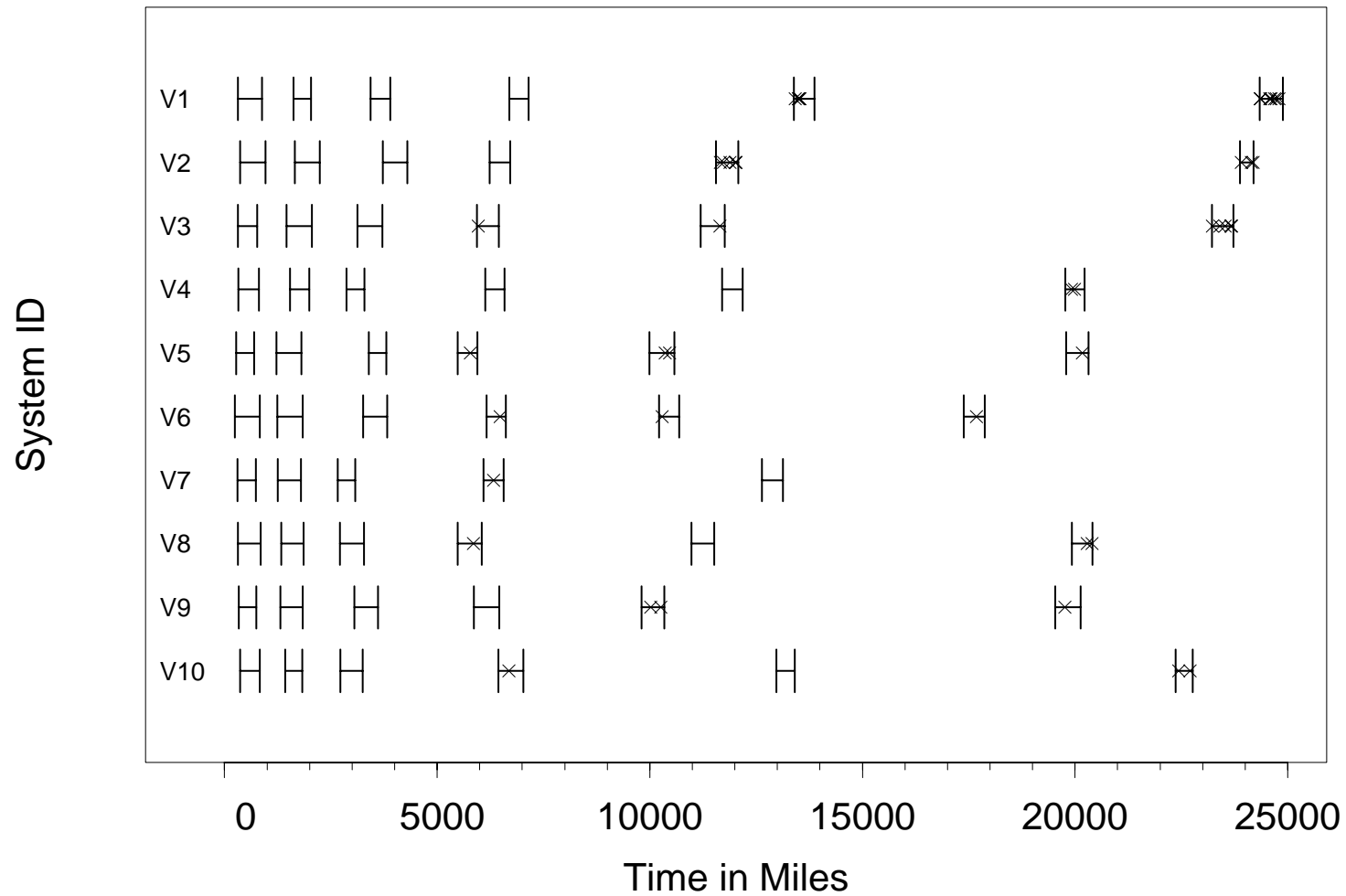


Example 2. Extended Warranty Data Grouped Event Plot for Labor Code C6050



Example 3. AMSAA Vehicle Fleet Data

Event Plot of Non-random Selection Window Data



Recurrence Data Structure Examples

- *Simple Recurrence Data:*

|—x—x—x—————x—x—|

- *Interval-censored Recurrence Data:*

|—x—x—x—————x—x—|

(1](2](0](2]

- *Window-observation Recurrence Data:*

|—x—] (—-] (-x—]

Related Literature

- **Nonparametric methods:**

Nelson(1988), Nelson(1995), Nelson(2003), Lawless and Nadeau (1995), and Chapter 16 of Meeker and Escobar (1998).

- **Parametric methods:**

Cox and Lewis (1966), and Basu and Rigdon (2000).

- **Nonparametric, parametric, covariates, random effects:**

Cook and Lawless (2007)

Extension of Nonparametric Estimation Method to Window-observation Recurrence Data

Let t_1, \dots, t_m be the unique report times,

$$\hat{\mu}(t_j) = \sum_{k=1}^j \left[\frac{\sum_{i=1}^n \delta_i(t_k) \times d_i(t_k)}{\sum_{i=1}^n \delta_i(t_k)} \right] = \sum_{k=1}^j \frac{d_{\cdot}(t_k)}{\delta_{\cdot}(t_k)} = \sum_{k=1}^j \bar{d}(t_k),$$

$j = 1, \dots, m, (1)$

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

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

Note that computation of variance estimator is more complicated, and formula are omitted here.

Model Assumptions for Nonparametric Methods

- The units in the sample data are a simple random sample from a well-defined target population.
- The population MCF is zero at time zero and exists (i.e., is finite) at any age t of interest up to the greatest censoring age.
- The stochastic process that generates the observation windows is independent from the stochastic process that generates the recurrences.
- The size of the risk set must be positive from time zero up to the maximum time that population MCF would be estimated.

Extension of NHPP Model to Window-observation Recurrence Data – Likelihood Function Comparison for Unit i

- For the usual case for a unit i observed from 0 to t_{a_i} ,

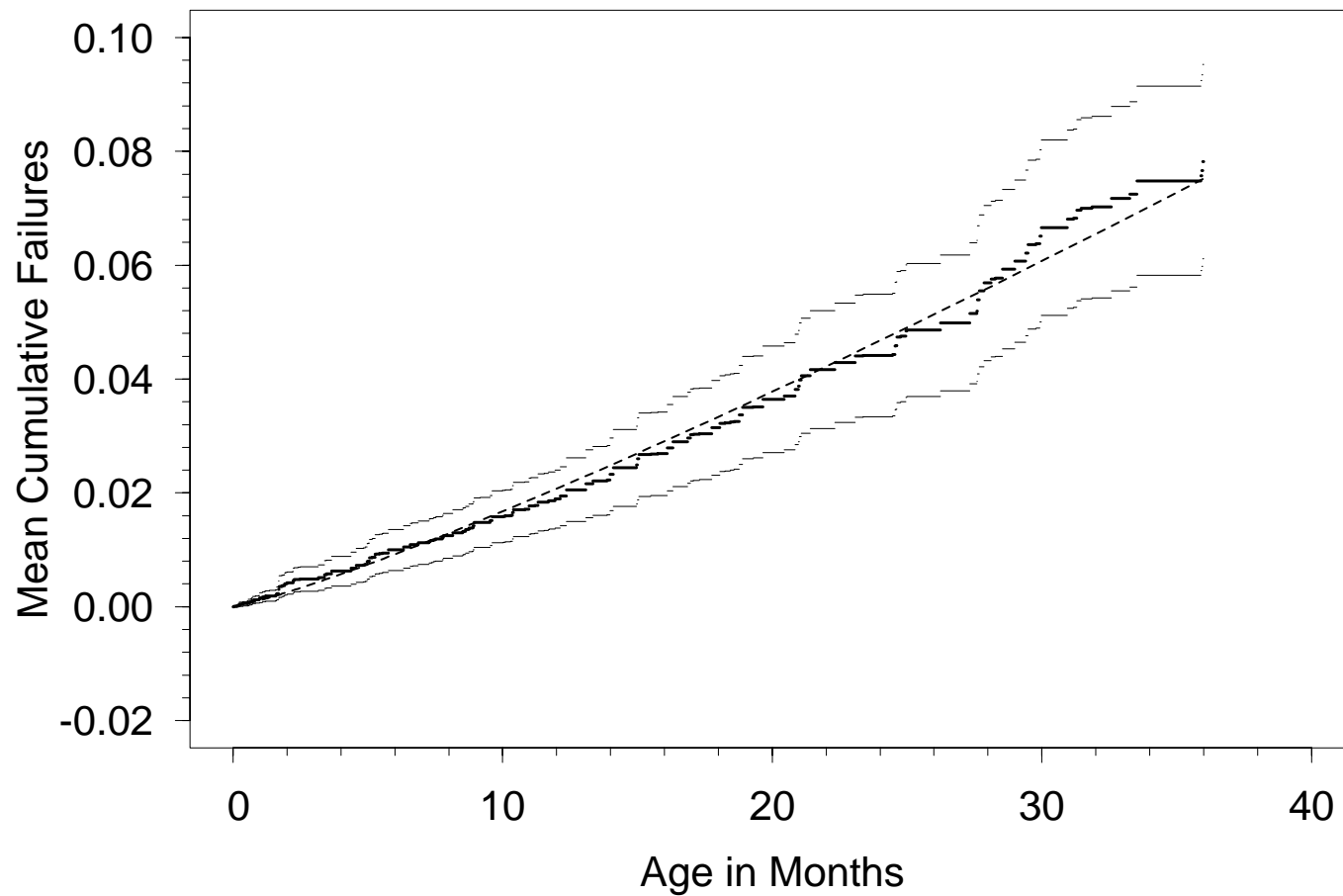
$$L_i(\boldsymbol{\theta}) = \prod_{j=1}^{r_i} \nu(t_{ij}; \boldsymbol{\theta}) \times \exp[-\mu(0, t_{a_i}; \boldsymbol{\theta})]$$

- For the general window-observation recurrence data,

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

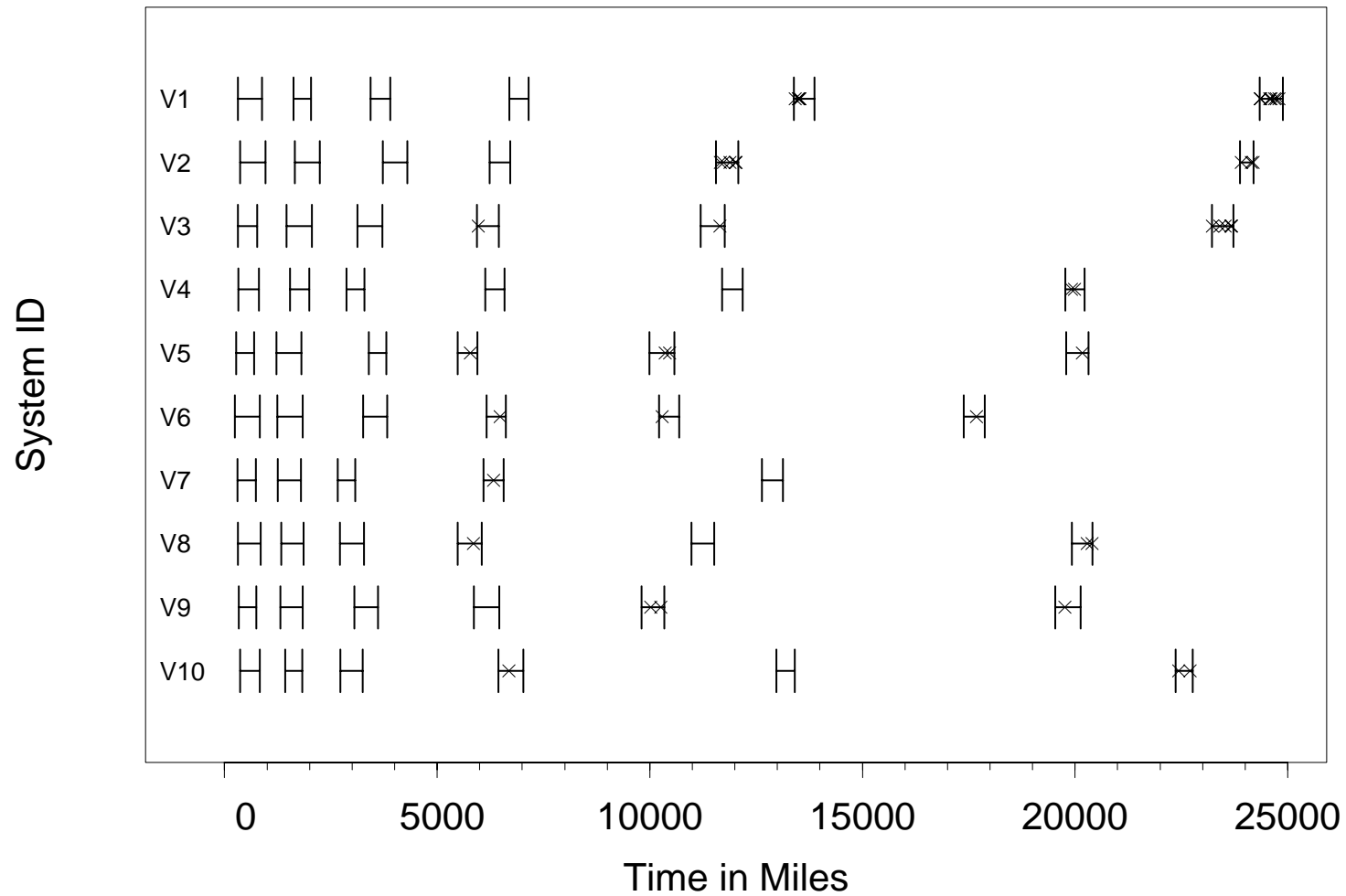
where $(t_{i1L}, t_{i1U}], (t_{i2L}, t_{i2U}], \dots, (t_{ip_iL}, t_{ip_iU}]$ (with $t_{i1L} \geq 0, t_{i(k-1)U} \leq t_{ikL}, t_{ip_iU} < t_{a_i}$) are the non-overlapping windows of observation for unit i .

**Nonparametric MCF estimates, with 95% CI,
and NHPP MCF estimates,
for Extended Warranty Data for C6050**



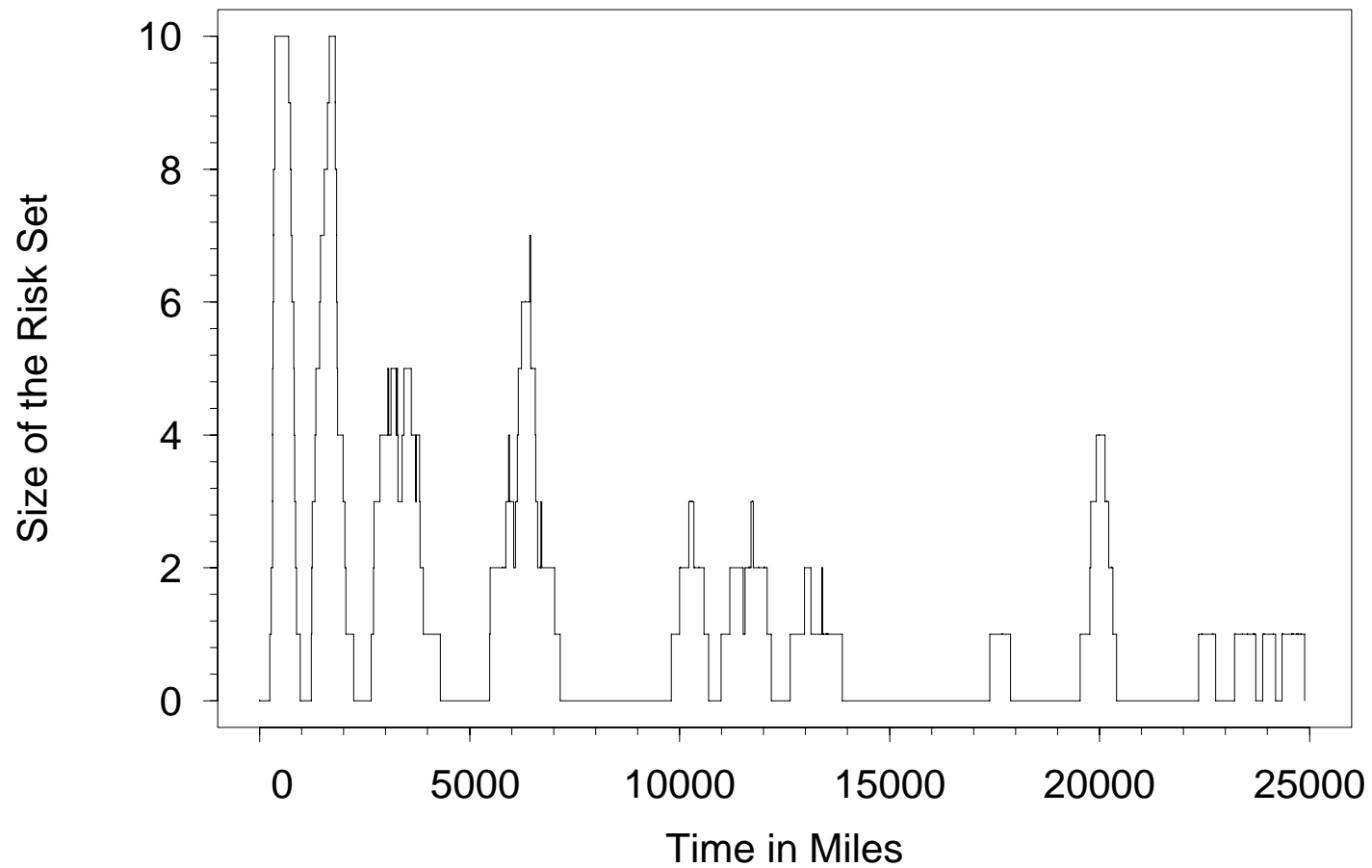
Example 3. AMSAA Vehicle Fleet Data

Event Plot of Non-random Selection Window Data



Example 3. AMSAA Vehicle Fleet Data

Risk Set Plot of Non-random Selection Window Data

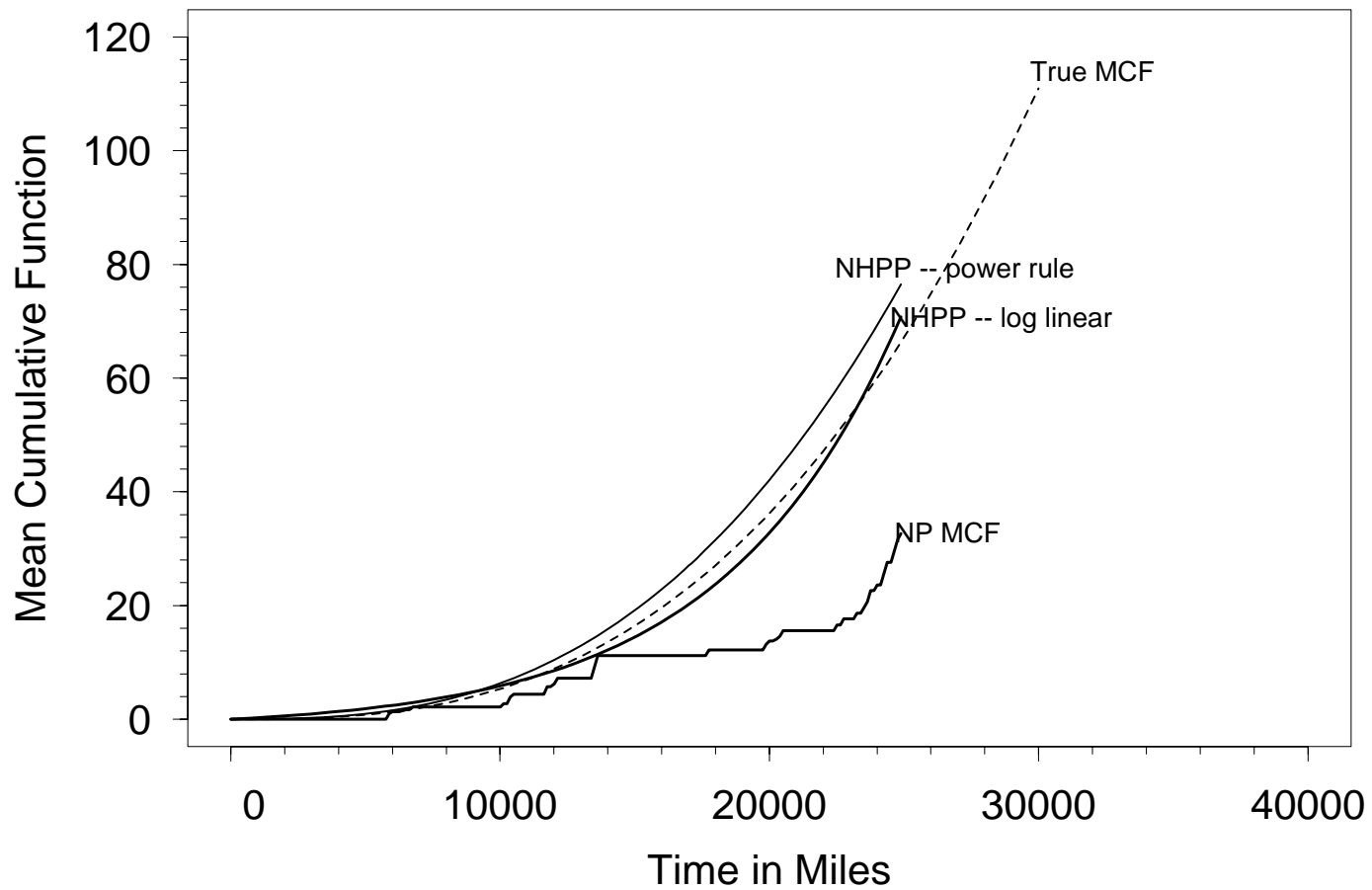


What if the size of the risk set is not always positive?

- No problems for the NHPP model.
- Nonparametrically, the change in MCF is not estimable over the risk-set-size-zero (RSSZ) intervals.
- Naive nonparametric estimator ignoring the riskset-size-zero (RSSZ) intervals can be severely biased.
- Nonparametrically, the variance of the change in the MCF is not estimable when the size of the risk set is less than two.

Example 3. AMSAA Vehicle Fleet Data

Comparing True Model and the MCF Estimates



Hybrid Estimator of the MCF

- For the risk-set-size-positive (RSSP) intervals, apply the nonparametric method with (1) to estimate the increases in the MCF.
- For the RSSZ intervals, borrow strength from the RSSP intervals, and two alternatives are proposed here:
- Hybrid estimator is obtained by summing over time the estimated increases in the MCF from both the RSSP intervals and RSSZ intervals.
- Two alternatives: Parametric using NHPP, and Local, using estimates from the surrounding RSSP intervals.

NHPP Hybrid Estimator

NHPP hybrid estimator assume an NHPP model for the recurrence process over the RSSZ intervals, and estimate the model parameters with all the data in the RSSP intervals.

Then the estimated increase in the MCF for the RSSZ interval i is

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

Local Hybrid Estimator

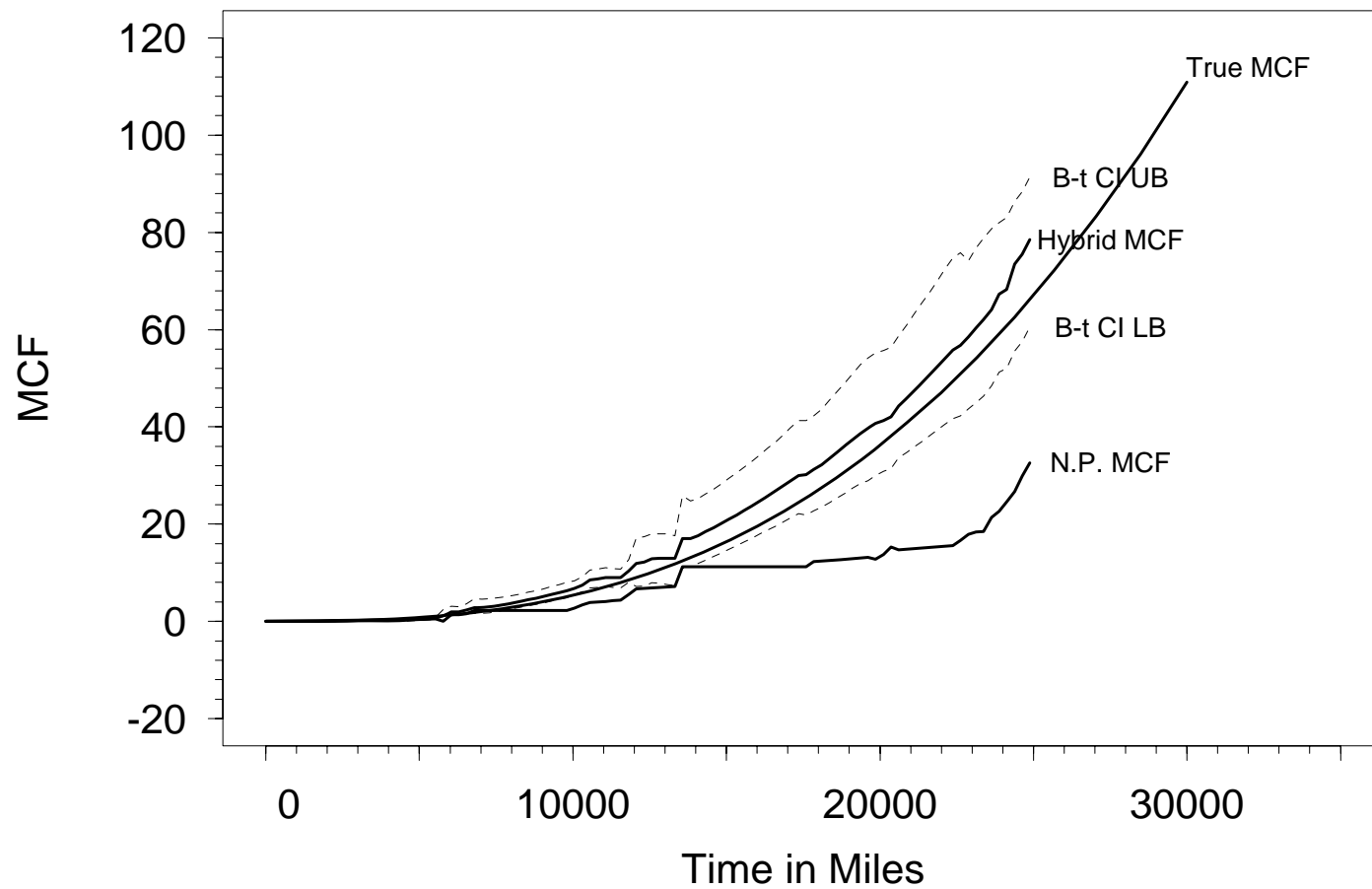
Local hybrid estimator is a purely nonparametric approach, using an average of the recurrence rate of the neighboring RSSP intervals. The estimated increase in the MCF for RSSZ interval i is

$$d_i^\dagger = (t_{iU} - t_{iL}) \times \nu_i,$$

$$\nu_i = \begin{cases} \frac{\bar{d}.(t_{1U}, t_{2L})}{(t_{2L} - t_{1U})} & \text{if } i = 1 \text{ and RSSZ at start} \\ \frac{\bar{d}.(0, t_{1L}) + \bar{d}.(t_{1U}, t_{2L})}{(t_{1L} - 0) + (t_{2L} - t_{1U})} & \text{if } i = 1 \text{ and RSSP at start} \\ \frac{\bar{d}.(t_{(i-1)U}, t_{iL}) + \bar{d}.(t_{iU}, t_{(i+1)L})}{(t_{iL} - t_{(i-1)U}) + (t_{(i+1)L} - t_{iU})} & \text{for } i = 2, 3, \dots, q-1 \\ \frac{\bar{d}.(t_{(q-1)U}, t_{qL}) + \bar{d}.(t_{qU}, t_{max})}{(t_{qL} - t_{(q-1)U}) + (t_{max} - t_{qU})} & \text{for } i = q. \end{cases}$$

Example 3. AMSAA Vehicle Fleet Data

Comparing True Model and the MCF Estimates



Concluding Remarks and Areas for Further Research

- Existing methods, both nonparametric and parametric, can be extended to analyze window observation recurrence data.
- Choice among the NHPP estimator, the local hybrid estimator and the NHPP hybrid estimator involves a tradeoff between variance and model-error bias.
- The local hybrid is better than the NHPP hybrid when the assumed NHPP model is far from the truth.
- Further work
 - ▶ Alternative CI methods.
 - ▶ Properties of alternative CI methods.
 - ▶ Covariate adjustment