

# Semiparametric Estimation for a Generalized KG Model with Recurrent Event Data

Edsel A. Peña  
(pena@stat.sc.edu)

Department of Statistics  
University of South Carolina  
Columbia, SC 29208

MMR Conference  
Beijing, China  
June 2011

# Historical Perspective: Random Censorship Model (RCM)

$$T_1, T_2, \dots, T_n \stackrel{IID}{\sim} F$$

$$C_1, C_2, \dots, C_n \stackrel{IID}{\sim} G$$

$F$  and  $G$  not related

$$\{T_i\} \perp \{C_i\}$$

## Random Observables:

$$(Z_1, \delta_1), (Z_2, \delta_2), \dots, (Z_n, \delta_n)$$

$$Z_i = T_i \wedge C_i \quad \text{and} \quad \delta_i = I\{T_i \leq C_i\}$$

**Goal:** To make inference on the distribution  $F$  or the hazard  
 $\Lambda = \int dF/\bar{F}_-$ , or functionals (mean, median, etc.).

# Nonparametric Inference

$F \in \mathcal{F}$  = space of continuous distributions

$$\Lambda = -\log \bar{F}; \bar{F} = 1 - F = \exp(-\Lambda); \lambda = \Lambda'$$

$$N(s) = \sum_i I\{Z_i \leq s; \delta_i = 1\} \quad \text{and} \quad Y(s) = \sum_i I\{Z_i \geq s\}$$

$$\text{NAE: } \hat{\Lambda} = \int \frac{dN}{Y} \quad \text{and} \quad \text{PLE: } \hat{\bar{F}} = \prod \left[ 1 - \frac{dN}{Y} \right]$$

Properties of  $\hat{\Lambda}$  and  $\hat{\bar{F}}$  well-known, e.g., biased; consistent; and when normalized, weakly convergent to Gaussian processes.

$$\text{Avar}(\hat{\bar{F}}(t)) = \frac{1}{n} \bar{F}(t)^2 \int_0^t \frac{dF(s)}{\bar{F}(s)^2 \bar{G}(s)}$$

# An Informative RCM

- ▶ Koziol-Green (KG) Model (1976):

$$\exists \beta \geq 0, \quad \bar{G}(t) = \bar{F}(t)^\beta$$

- ▶ Lehmann-type alternatives
- ▶ Proportional hazards:

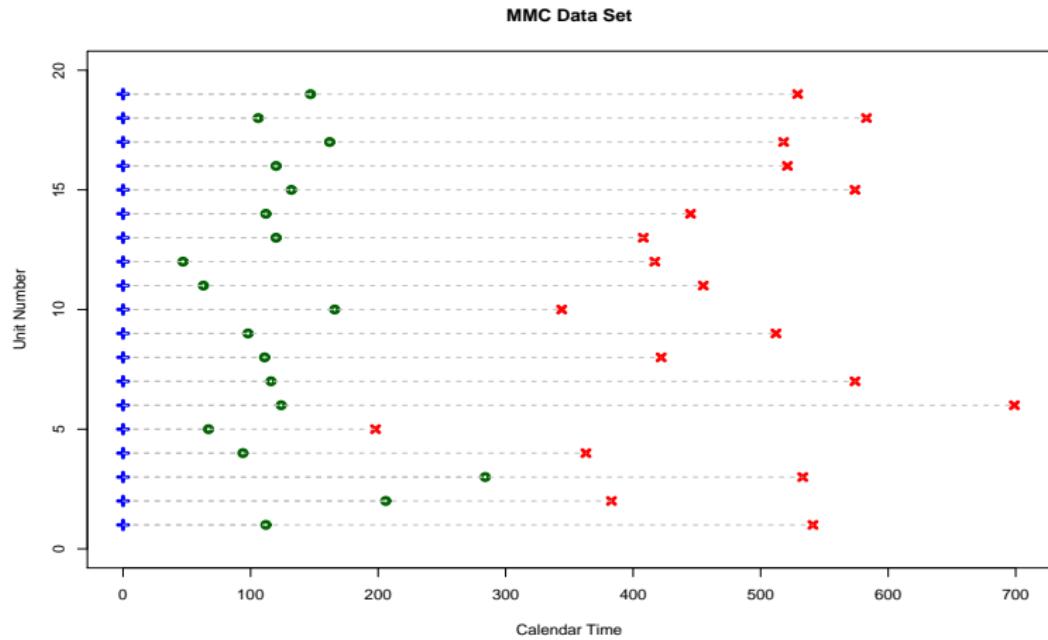
$$\Lambda_G = \beta \Lambda_F$$

- ▶  $Z_i = \min(T_i, C_i) \sim \bar{F}^{\beta+1}$
- ▶  $\delta_i = I\{T_i \leq C_i\} \sim \text{BER}(1/(\beta + 1))$
- ▶ An important characterizing property:

$$Z_i \perp \delta_i$$

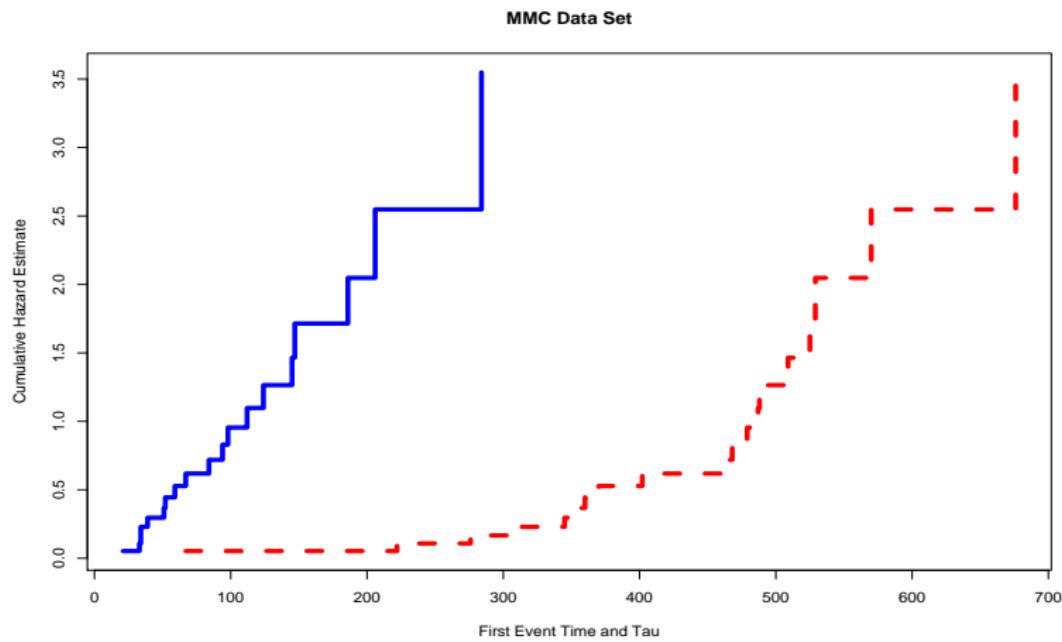
# MMC Data: First Event Times and Tau's

Migratory Motor Complex (MMC) data set from Aalen and Husebye, Stat Med, 1991.



# MMC Data: Hazard Estimates

KG model holds if and only if  $\Lambda_\tau \propto \Lambda_{T_1}$



## KG Model's Utility

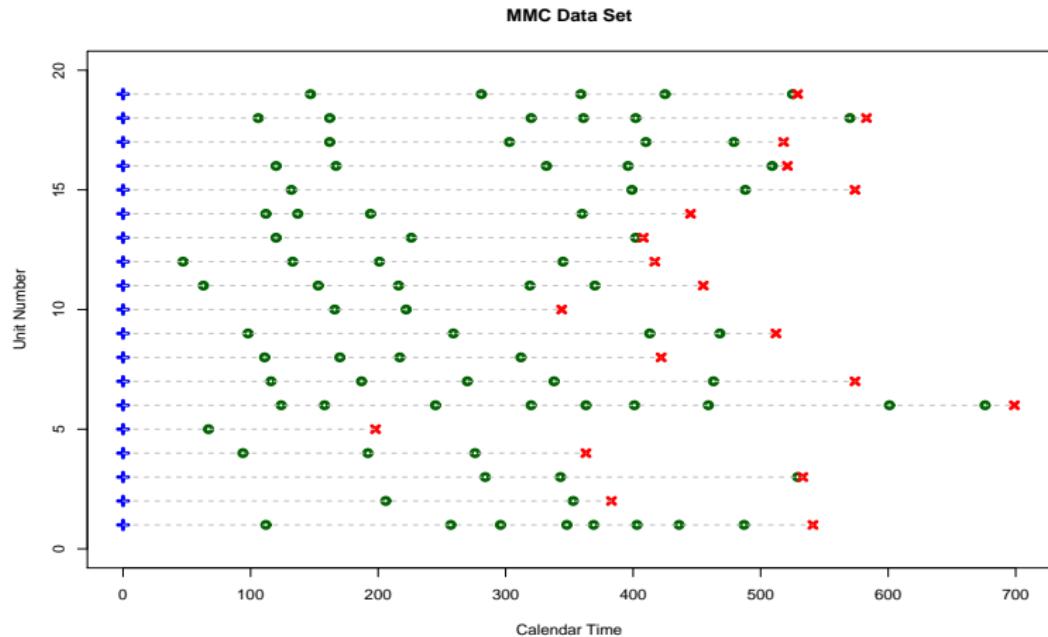
- ▶ Chen, Hollander and Langberg (JASA, 1982): exploited independence between  $Z_i$  and  $\delta_i$  to obtain the exact bias, variance, and MSE functions of PLE. Comparisons with asymptotic results.
- ▶ Cheng and Lin (1987): exploited semiparametric nature of KG model to obtain a more efficient estimator of  $F$  compared to the PLE.
- ▶ Hollander and Peña (1988): obtained better confidence bands for  $F$  under KG model.
- ▶ Csorgo & Faraway (1998): KG model **not** practically viable, **but** serves as a mathematical specimen (a la yeast) for examining exact properties of procedures and for providing pinpoint assessment of efficiency losses/gains.

## Recurrent Events: Some Examples

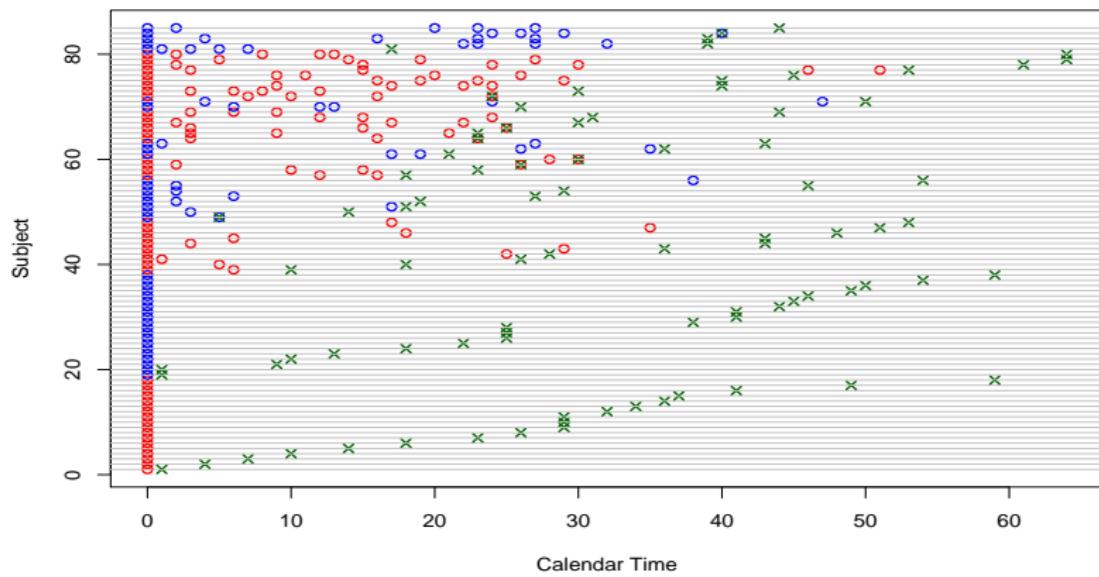
- ▶ admission to hospital due to chronic disease
- ▶ tumor re-occurrence
- ▶ migraine attacks
- ▶ alcohol or drug (eg cocaine) addiction
- ▶ machine failure or discovery of a bug in a software
- ▶ commission of a criminal act by a delinquent minor!
- ▶ major disagreements between a couple
- ▶ non-life insurance claim
- ▶ drop of  $\geq 200$  points in DJIA during trading day
- ▶ publication of a research paper by a professor

# Full MMC Data: With Recurrences

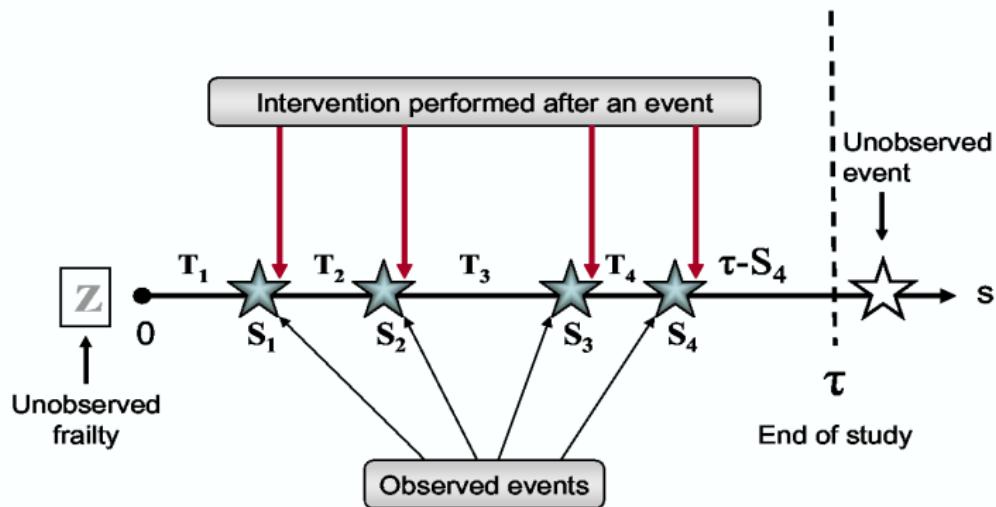
$n = 19$  subjects; event = end of migratory motor complex cycle;  
random length of monitoring period per subject.



# Bladder Cancer Data Set: Groups [control (red), thiotepa (blue)]; Covariates; in WLW (89)



# Data Accrual: One Subject



Covariate vector:  $\mathbf{X}(s) = (X_1(s), \dots, X_q(s))$

# Some Aspects in Recurrent Data

- ▶ random monitoring length ( $\tau$ ).
- ▶ random # of events ( $K$ ) and sum-quota constraint:

$$K = \max \left\{ k : \sum_{j=1}^k T_j \leq \tau \right\} \text{ with } \sum_{j=1}^K T_j \leq \tau < \sum_{j=1}^{K+1} T_j$$

- ▶ **Basic Observable:**  $(K, \tau, T_1, T_2, \dots, T_K, \tau - S_K)$
- ▶ always a right-censored observation.
- ▶ dependent and informative censoring.
- ▶ effects of covariates, frailties, interventions after each event, and accumulation of events.

## Approaches: Recurrent Event Analysis

- ▶ First-event analysis. Inefficient. How much do we gain by utilizing the recurrences?
- ▶ Full modeling approach: AG (82) and in PH (04), PSG (07), book by ABG (08).
- ▶ Full modeling: harder to implement and requires intensity process specification.
- ▶ Marginal modeling approach: Wei, Lin, and Weissfeld (89).
- ▶ Conditional modeling approach: Prentice, Williams, and Peterson (81).
- ▶ Marginal and conditional approaches: quite popular, but are there foundational issues?

# Simplest Model: One Subject

- ▶  $T_1, T_2, \dots \stackrel{IID}{\sim} F$ : (renewal model)
- ▶ 'perfect interventions' after each event
- ▶  $\tau \sim G$
- ▶  $F$  and  $G$  not related
- ▶ no covariates ( $X$ )
- ▶ no frailties ( $Z$ )
- ▶  $F$  could be parametric or nonparametric
- ▶ Peña, Strawderman and Hollander (JASA, 01): nonparametric estimation of  $F$ .

# Nonparametric Estimation of $F$

$$N(t) = \sum_{i=1}^n \sum_{j=1}^{K_i} I\{T_{ij} \leq t\}$$

$$Y(t) = \sum_{i=1}^n \left\{ \sum_{j=1}^{K_i} I\{T_{ij} \geq t\} + I\{\tau_i - S_{iK_i} \geq t\} \right\}$$

$$\textbf{GNAE : } \tilde{\Lambda}(t) = \int_0^t \frac{dN(w)}{Y(w)}$$

$$\textbf{GPLE : } \tilde{\tilde{F}}(t) = \prod_0^t \left[ 1 - \frac{dN(w)}{Y(w)} \right]$$

# Main Asymptotic Result

**$k$ th Convolution:**  $F^{\star(k)}(t) = \Pr \left\{ \sum_{j=1}^k T_j \leq t \right\}$

**Renewal Function:**  $\rho(t) = \sum_{k=1}^{\infty} F^{\star(k)}(t)$

$$\nu(t) = \frac{1}{\bar{G}(t)} \int_t^{\infty} \rho(w-t) dG(w)$$

$$\sigma^2(t) = \bar{F}(t)^2 \int_0^t \frac{dF(w)}{\bar{F}(w)^2 \bar{G}(w)[1 + \nu(w)]}$$

**Theorem (JASA, 01):**  $\sqrt{n}(\tilde{\bar{F}}(t) - \bar{F}(t)) \Rightarrow \text{GP}(0, \sigma^2(t))$

## Extending KG Model: Recurrent Setting

- ▶ Wanted: a **tractable** model with monitoring time **informative** about  $F$ .
- ▶ Potential to refine analysis of efficiency gains/losses.
- ▶ Idea: Why not simply generalize the KG model for the RCM.
- ▶ **Generalized KG Model** (GKG) for Recurrent Events:

$$\exists \beta > 0, \quad \bar{G}(t) = \bar{F}(t)^\beta$$

with  $\beta$  unknown, and  $F$  the common inter-event time distribution function.

- ▶ **Remark:**  $\tau$  may also represent system failure/death, while the recurrent event could be shocks to the system.
- ▶ **Remark:** Association (within unit) could be modeled through a frailty.

# Estimation Issues and Some Questions

- ▶ How to semiparametrically estimate  $\beta$ ,  $\Lambda$ , and  $\bar{F}$ ?
- ▶ Parametric estimation in Adekpedjou, Peña, and Quiton (2010, JSPI).
- ▶ How much **efficiency loss** is incurred when the informative monitoring model structure is ignored?
- ▶ How much **penalty** is incurred with Single-event analysis relative to Recurrent-event analysis?
- ▶ In particular, what is the **efficiency loss** for estimating  $F$  when **using the nonparametric estimator** in PSH (2001) relative to the semiparametric estimator that exploits the informative monitoring structure?

# Basic Processes

$$S_{ij} = \sum_{k=1}^j T_{ik}$$

$$N_i^\dagger(s) = \sum_{j=1}^{\infty} I\{S_{ij} \leq s\}$$

$$Y_i^\dagger(s) = I\{\tau_i \geq s\}$$

$R_i(s) = s - S_{iN_i^\dagger(s-)} =$  backward recurrence time

$$A_i^\dagger(s) = \int_0^s Y_i^\dagger(v) \lambda[R_i(v)] dv$$

$$N_i^\tau(s) = I\{\tau_i \leq s\}$$

$$Y_i^\tau(s) = I\{\tau_i \geq s\}$$

# Transformed Processes

$$Z_i(s, t) = I\{R_i(s) \leq t\}$$

$$N_i(s, t) = \int_0^s Z_i(v, t) N_i^\dagger(dv) = \sum_{j=1}^{N_i^\dagger(s))} I\{\tau_{ij} \leq t\}$$

$$Y_i(s, t) = \sum_{j=1}^{N_i^\dagger(s-)} I\{\tau_{ij} \geq t\} + I\{(s \wedge \tau_i) - S_{iN_i^\dagger(s-)} \geq t\}$$

$$A_i(s, t) = \int_0^s Z_i(v, t) A_i^\dagger(dv) = \int_0^t Y_i(s, w) \lambda(w) dw$$

# Aggregated Processes

$$N(s, t) = \sum_{i=1}^n N_i(s, t)$$

$$Y(s, t) = \sum_{i=1}^n Y_i(s, t)$$

$$A(s, t) = \sum_{i=1}^n A_i(s, t)$$

$$N^\tau(s) = \sum_{i=1}^n N_i^\tau(s)$$

$$Y^\tau(s) = \sum_{i=1}^n Y_i^\tau(s)$$

## First, Assume $\beta$ Known

Via Method-of-Moments Approach, 'estimator' of  $\Lambda$ :

$$\hat{\Lambda}(s, t|\beta) = \int_0^t \left\{ \frac{N(s, dw) + N^\tau(dw)}{Y(s, w) + \beta Y^\tau(w)} \right\}$$

Using product-integral representation of  $\bar{F}$  in terms of  $\Lambda$ ,  
'estimator' of  $\bar{F}$ :

$$\hat{\bar{F}}(s, t|\beta) = \prod_{w=0}^t \left\{ 1 - \frac{N(s, dw) + N^\tau(dw)}{Y(s, w) + \beta Y^\tau(w)} \right\}$$

# Estimating $\beta$ : Profile Likelihood MLE

Profile Likelihood:

$$L_P(s^*; \beta) = \beta^{N^\tau(s^*)} \times \\ \prod_{i=1}^n \left\{ \left[ \prod_{v=0}^{s^*} \left\{ \frac{1}{Y(s^*, v) + \beta Y^\tau(v)} \right\}^{N_i^\tau(\Delta v)} \right] \times \right. \\ \left. \left[ \prod_{v=0}^{s^*} \left\{ \frac{1}{Y(s^*, v) + \beta Y^\tau(v)} \right\}^{N_i(s^*, \Delta v)} \right] \right\}$$

Estimator of  $\beta$ :

$$\hat{\beta} = \arg \max_{\beta} L_P(s^*; \beta)$$

**Computational Aspect:** in R, we used `optimize` to get **good** seed for the Newton-Raphson iteration.

# Estimators of $\Lambda$ and $\bar{F}$

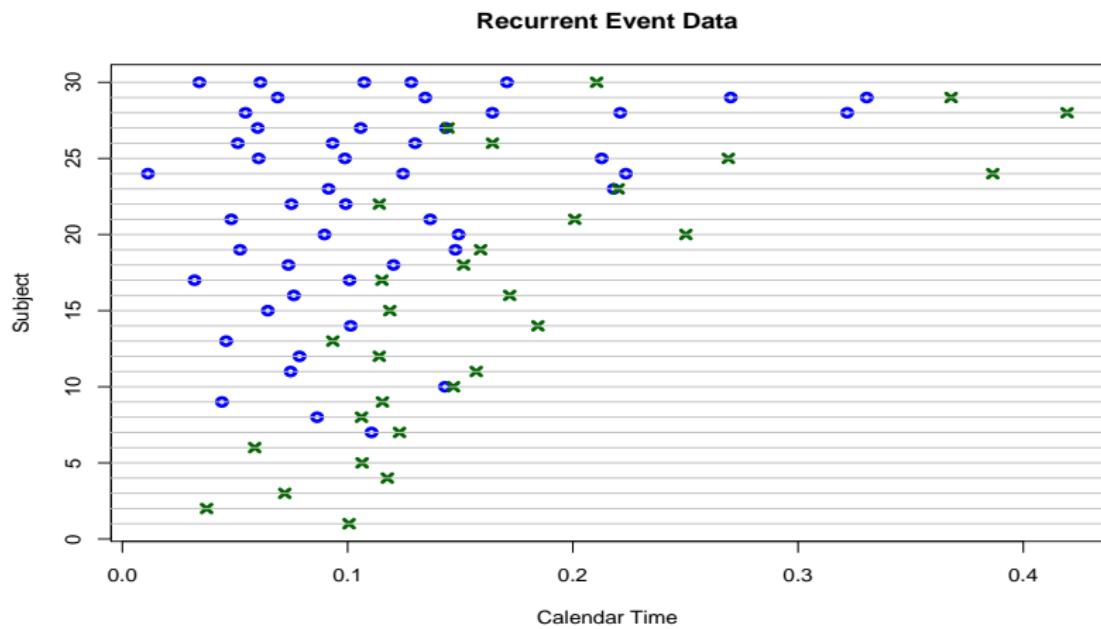
Estimator of  $\Lambda$ :

$$\hat{\Lambda}(s^*, t) = \hat{\Lambda}(s^*, t|\hat{\beta}) = \int_0^t \left\{ \frac{N(s^*, dw) + N^\tau(dw)}{Y(s^*, w) + \hat{\beta} Y^\tau(w)} \right\}$$

Estimator of  $\bar{F}$ :

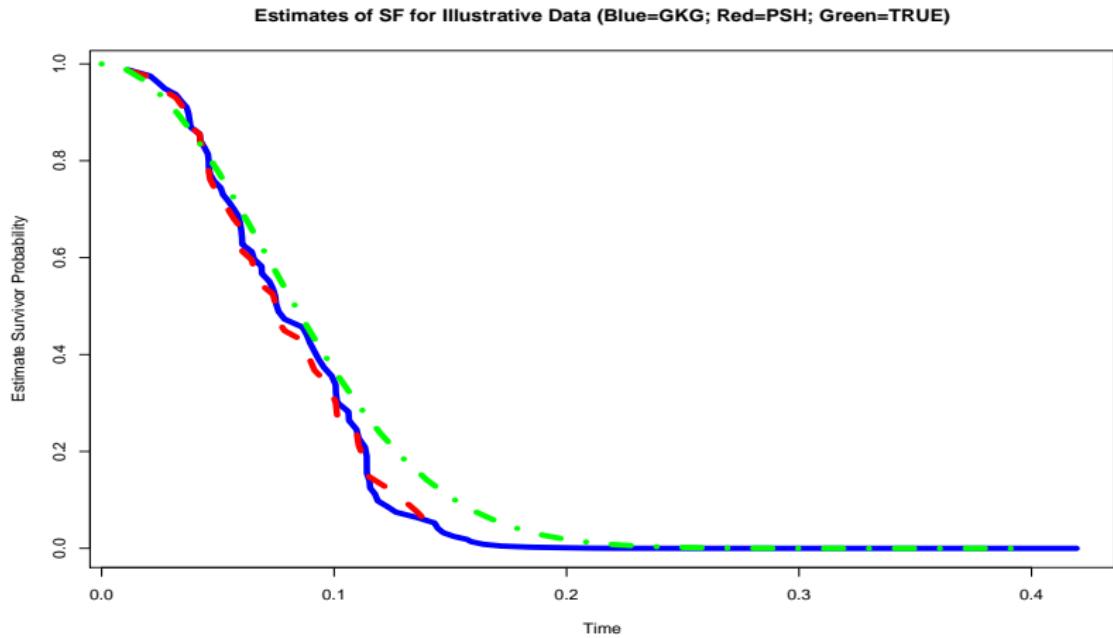
$$\hat{\bar{F}}(s^*, t) = \hat{\bar{F}}(s^*, t|\hat{\beta}) = \prod_{w=0}^t \left\{ 1 - \frac{N(s^*, dw) + N^\tau(dw)}{Y(s^*, w) + \hat{\beta} Y^\tau(w)} \right\}$$

# Illustrative Data ( $n = 30$ ): GKG[Wei(2,.1), $\beta = .2$ ]



# Estimates of $\beta$ and $\bar{F}$

$$\hat{\beta} = .2331$$



# Properties of Estimators

$$G_s(w) = G(w)I\{w < s\} + I\{w \geq s\}$$

$$\mathbb{E}\{Y_1(s, t)\} \equiv y(s, t) = \bar{F}(t)\bar{G}_s(t) + \bar{F}(t) \int_t^{\infty} \rho(w - t)dG_s(w)$$

$$\mathbb{E}\{Y_1^\tau(t)\} \equiv y^\tau(t) = \bar{F}(t)^\beta$$

True Values =  $(F_0, \Lambda_0, \beta_0)$

$$y_0(s, t) = y(s, t; \Lambda_0, \beta_0)$$

$$y_0^\tau(s) = y^\tau(s; \Lambda_0, \beta_0)$$

# Existence, Consistency, Normality

## Theorem

*There is a sequence of  $\hat{\beta}$  that is consistent, and  $\hat{\Lambda}(s^*, \cdot)$  and  $\hat{F}(s^*, \cdot)$  are both uniformly strongly consistent.*

## Theorem

*As  $n \rightarrow \infty$ , we have*

$$\sqrt{n}(\hat{\beta} - \beta_0) \Rightarrow N(0, [\mathcal{I}_P(s^*; \Lambda_0, \beta_0)]^{-1})$$

*with*

$$\mathcal{I}_P(s^*; \Lambda_0, \beta_0) = \frac{1}{\beta_0} \int_0^{s^*} \frac{y_0^\tau(v)y_0(s^*, v)}{y_0(s^*, v) + \beta_0 y_0^\tau(v)} \lambda_0(v) dv.$$

# Weak Convergence of $\hat{\Lambda}(s^*, \cdot)$

## Theorem

As  $n \rightarrow \infty$ ,  $\{\sqrt{n}[\hat{\Lambda}(s^*, t) - \Lambda_0(t)] : t \in [0, t^*]\}$  converges weakly to a zero-mean Gaussian process with variance function

$$\begin{aligned}\sigma_{\hat{\Lambda}}^2(s^*, t) &= \int_0^t \frac{\Lambda_0(dv)}{y_0(s^*, v) + \beta_0 y_0^\tau(v)} + \\ &\left[ \int_0^{s^*} \frac{y_0(s^*, v)y_0^\tau(v)}{\beta_0[y_0(s^*, v) + \beta_0 y_0^\tau(v)]} \Lambda_0(dv) \right]^{-1} \times \\ &\left[ \int_0^t \frac{y_0^\tau(v)}{y_0(s^*, v) + \beta_0 y_0^\tau(v)} \Lambda_0(dv) \right]^2.\end{aligned}$$

**Remark:** The last product term is the effect of estimating  $\beta$ . It inflates the asymptotic variance.

# Weak Convergence of $\hat{F}(s^*, \cdot)$ and $\tilde{F}(s^*, \cdot)$

## Corollary

As  $n \rightarrow \infty$ ,  $\{\sqrt{n}[\hat{F}(s^*, t) - \bar{F}_0(t)] : t \in [0, t^*]\}$  converges weakly to a zero-mean Gaussian process whose variance function is

$$\sigma_{\hat{F}}^2(s^*, t) = \bar{F}_0(t)^2 \sigma_{\hat{\Lambda}}^2(s^*, t) \equiv \bar{F}_0(t)^2 \sigma_{\tilde{\Lambda}}^2(s^*, t).$$

Recall/Compare!

## Theorem (PSH, 2001)

As  $n \rightarrow \infty$ ,  $\{\sqrt{n}[\tilde{F}(s^*, t) - \bar{F}_0(t)] : t \in [0, t^*]\}$  converges weakly to a zero-mean Gaussian process whose variance function is

$$\sigma_{\tilde{F}}^2(s^*, t) = \bar{F}_0(t)^2 \int_0^t \frac{\Lambda_0(dv)}{y_0(s^*, v)}.$$

## Asymptotic Relative Efficiency: $\beta_0$ Known

If we **know**  $\beta_0$ :

$$\begin{aligned} ARE\{\tilde{F}(s^*, t) : \hat{F}(s^*, t|\beta_0)\} &= \\ &\left\{ \int_0^t \frac{\Lambda_0(dw)}{y_0(s^*, w)} \right\}^{-1} \times \\ &\left\{ \int_0^t \frac{\Lambda_0(dw)}{y_0(s^*, w) + \beta_0 y_0^\tau(w)} \right\} \end{aligned}$$

Clearly, this could not exceed unity, as is to be expected.

# Case of Exponential $F$ : $\beta_0$ Known

## Theorem

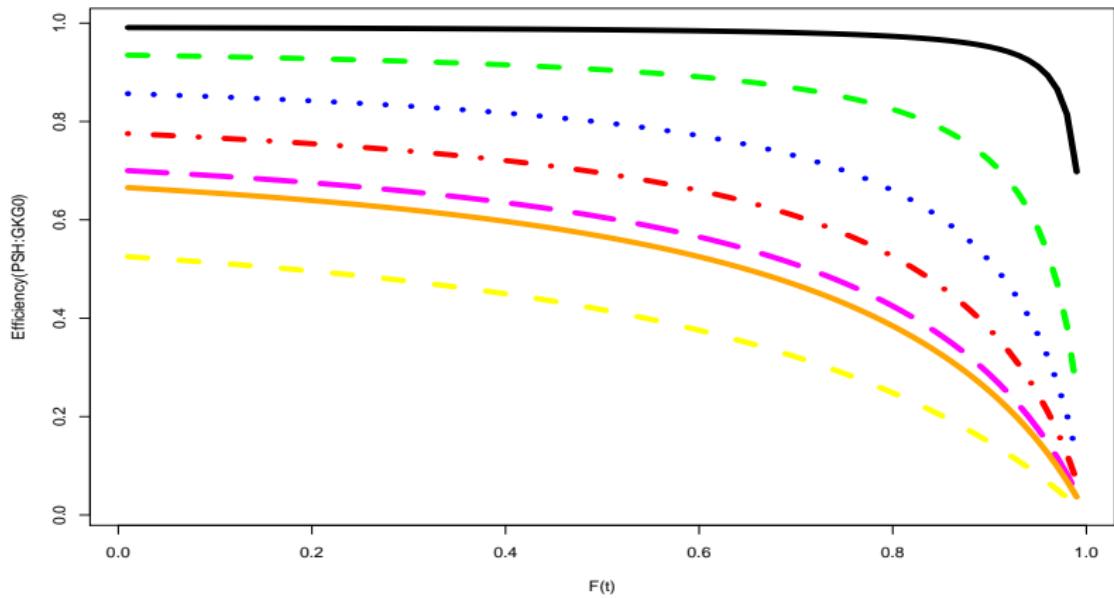
If  $\bar{F}_0(t) = \exp\{-\theta_0 t\}$  for  $t \geq 0$  and  $s^* \rightarrow \infty$ , then

$$\begin{aligned} ARE\{\tilde{\bar{F}}(\infty, t) : \hat{\bar{F}}(\infty, t|\beta_0)\} &= \\ &\left\{ \int_{\bar{F}_0(t)}^1 \frac{du}{(1 + \beta_0)u^{2+\beta_0}} \right\}^{-1} \times \\ &\left\{ \int_{\bar{F}_0(t)}^1 \frac{du}{(1 + \beta_0)u^{2+\beta_0} + \beta_0^2 u^{1+\beta_0}} \right\}. \end{aligned}$$

Also,  $\forall t \geq 0$ ,

$$ARE\{\tilde{\bar{F}}(\infty, t) : \hat{\bar{F}}(\infty, t; \beta_0)\} \leq \frac{1 + \beta_0}{1 + \beta_0 + \beta_0^2}.$$

ARE-Plots;  $\beta_0 \in \{.1, .3, .5, .7, .9, 1.0, 1.5\}$  Known;  
 $F = Exponential$



## Case of $\beta_0$ Unknown

- ▶ As to be expected, if  $\beta_0$  is known, then the estimator exploiting the GKG structure is more efficient.
- ▶ **Question:** Does this dominance hold true still if  $\beta_0$  is now estimated?

### Theorem

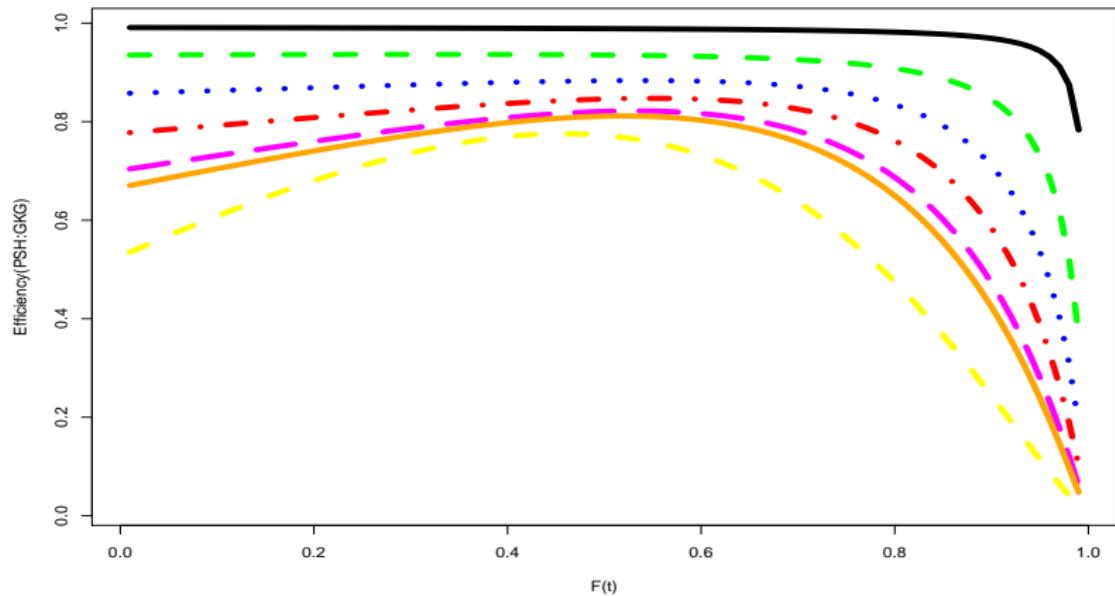
*Under the GKG model, for all  $(\bar{F}_0, \beta_0)$  with  $\beta_0 > 0$ ,  $\tilde{\bar{F}}(s^*, t)$  is asymptotically dominated by  $\hat{\bar{F}}(s^*, t)$  in the sense that*

$$ARE(\tilde{\bar{F}}(s^*, t) : \hat{\bar{F}}(s^*, t)) \leq 1.$$

### Proof.

Neat application of Cauchy-Schwartz Inequality. □

ARE-Plots;  $\beta_0 \in \{.1, .3, .5, .7, .9, 1.0, 1.5\}$  Unknown;  
 $F = Exponential$



# Assessing Asymptotic Approximations under Exponential

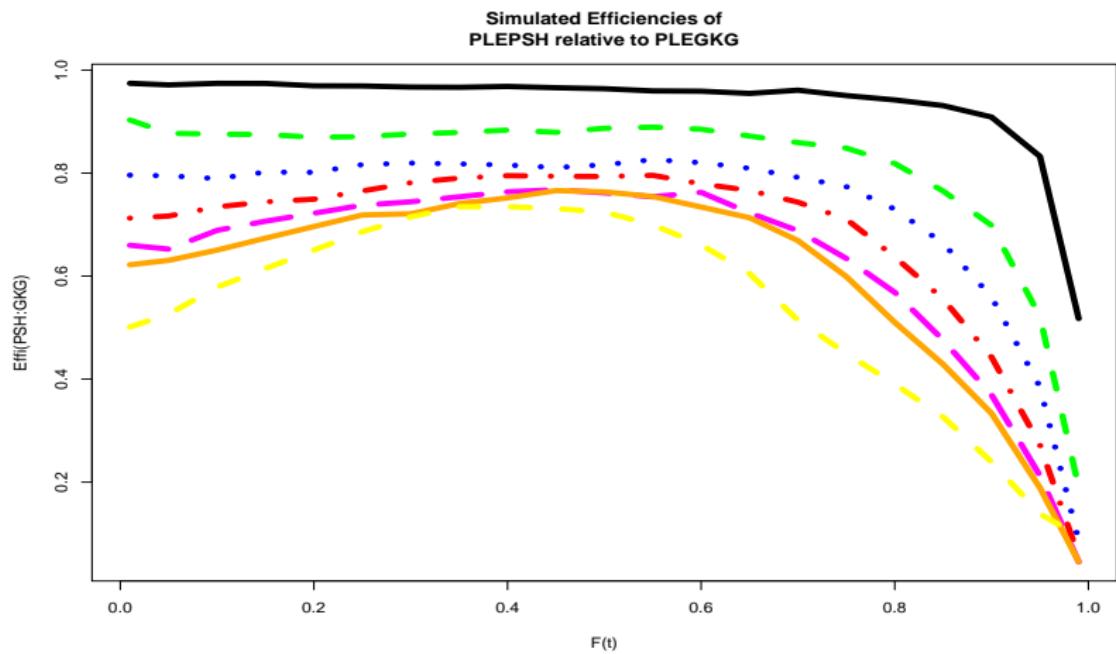
$\beta_0$	$n = 50$			$n = 100$		
	Mean	SE	ASE	Mean	SE	ASE
0.1	.1014	.0245	.0234	.1005	.0169	.0165
0.3	.3048	.0624	.0596	.3018	.0430	.0422
0.5	.5119	.1038	.0983	.5045	.0708	.0695
0.7	.7176	.1518	.1406	.7075	.1032	.0994
0.9	.9213	.1964	.1864	.9093	.1356	.1318
1.0	1.0294	.2298	.2107	1.0172	.1579	.1490
1.5	1.5615	.3963	.3445	1.5316	.2572	.2436

- ▶ Mean: mean of the simulated values of  $\hat{\beta}$ .
- ▶ SE: standard error of the simulated values of  $\hat{\beta}$ .
- ▶ ASE: asymptotic standard error obtained from earlier formula.
- ▶ 10000 replications were performed.

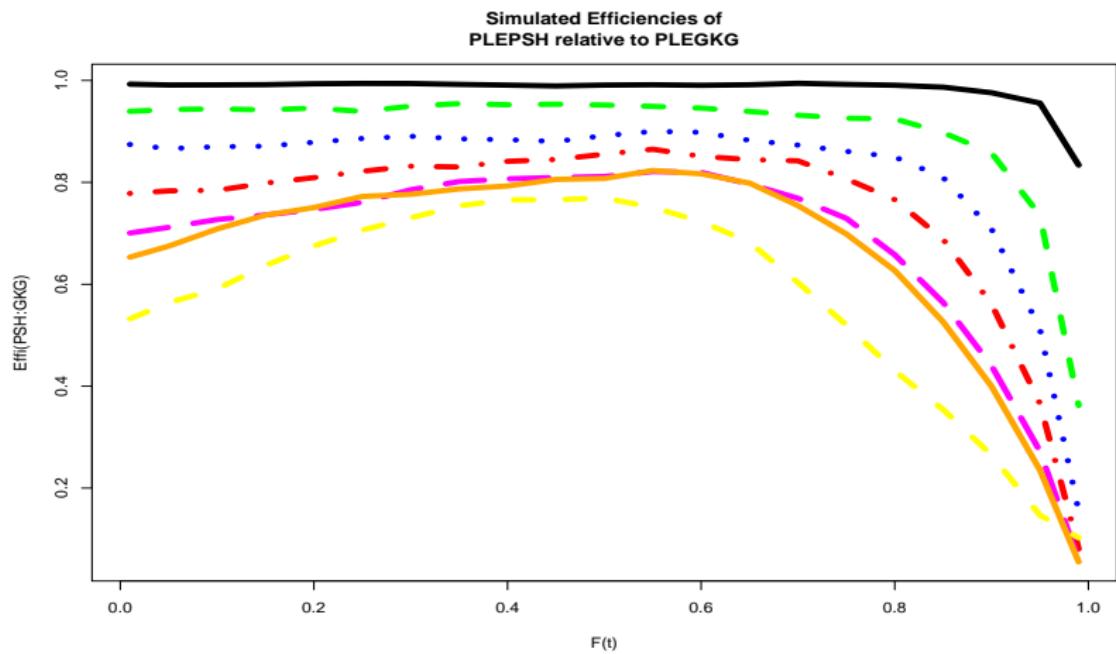
# Simulations under Weibull $F$ : $\hat{\beta}$ Results

$\beta_0$	$n = 30, \alpha = 2$		$n = 50, \alpha = 2$		$n = 30, \alpha = .9$		$n = 50, \alpha = .9$	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
0.1	.10	.03	.10	.02	.10	.03	.10	.02
0.3	.31	.09	.30	.07	.30	.07	.30	.06
0.5	.52	.16	.51	.12	.51	.13	.50	.10
0.7	.73	.23	.72	.17	.72	.20	.71	.14
0.9	.95	.32	.92	.23	.94	.26	.92	.19
1.0	1.06	.36	1.03	.26	1.04	.30	1.03	.22
1.5	1.64	.63	1.58	.43	1.60	.55	1.56	.38

# Simulated Rel Eff of $\tilde{F} : \hat{F}$ under a Weibull $F$ with $\alpha = 2$



# Simulated Rel Eff of $\tilde{F} : \hat{F}$ under a Weibull $F$ with $\alpha = 0.9$



## Some Concluding Thoughts

- ▶ Revisited the random censorship model with one event and the Koziol-Green model.
- ▶ Revisited nonparametric estimation of inter-event distribution in the presence of recurrent event data.
- ▶ Extended the Koziol-Green model to the situation with recurrent event.
- ▶ Considered the semiparametric estimation of the inter-event distribution under this generalized recurrent event data.
- ▶ Examined finite- and asymptotic properties of the estimators under this GKG structure.
- ▶ Assessed efficiency losses of the fully nonparametric estimator of  $F$  relative to that exploiting GKG structure.

# Acknowledgements

- ▶ Joint work with Akim Adekpedjou (Missouri University of Science and Technology).
- ▶ Research support from the National Science Foundation and National Institutes of Health.
- ▶ Thanks to Bo Lindqvist for inviting me to participate in this MMR and to Lirong Cui and his colleagues for organizing this MMR.