On Dynamic Recurrent Event Modeling and Analysis

Edsel A. Peña

E-Mail: pena@stat.sc.edu

Dept of Statistics, Univ of South Carolina

Portions joint with E. Slate (MUSC) and J. Gonzalez (Spain)

Research support from NIH Grants (R01, COBRE)

Joint Statistical Meeting, August 2006

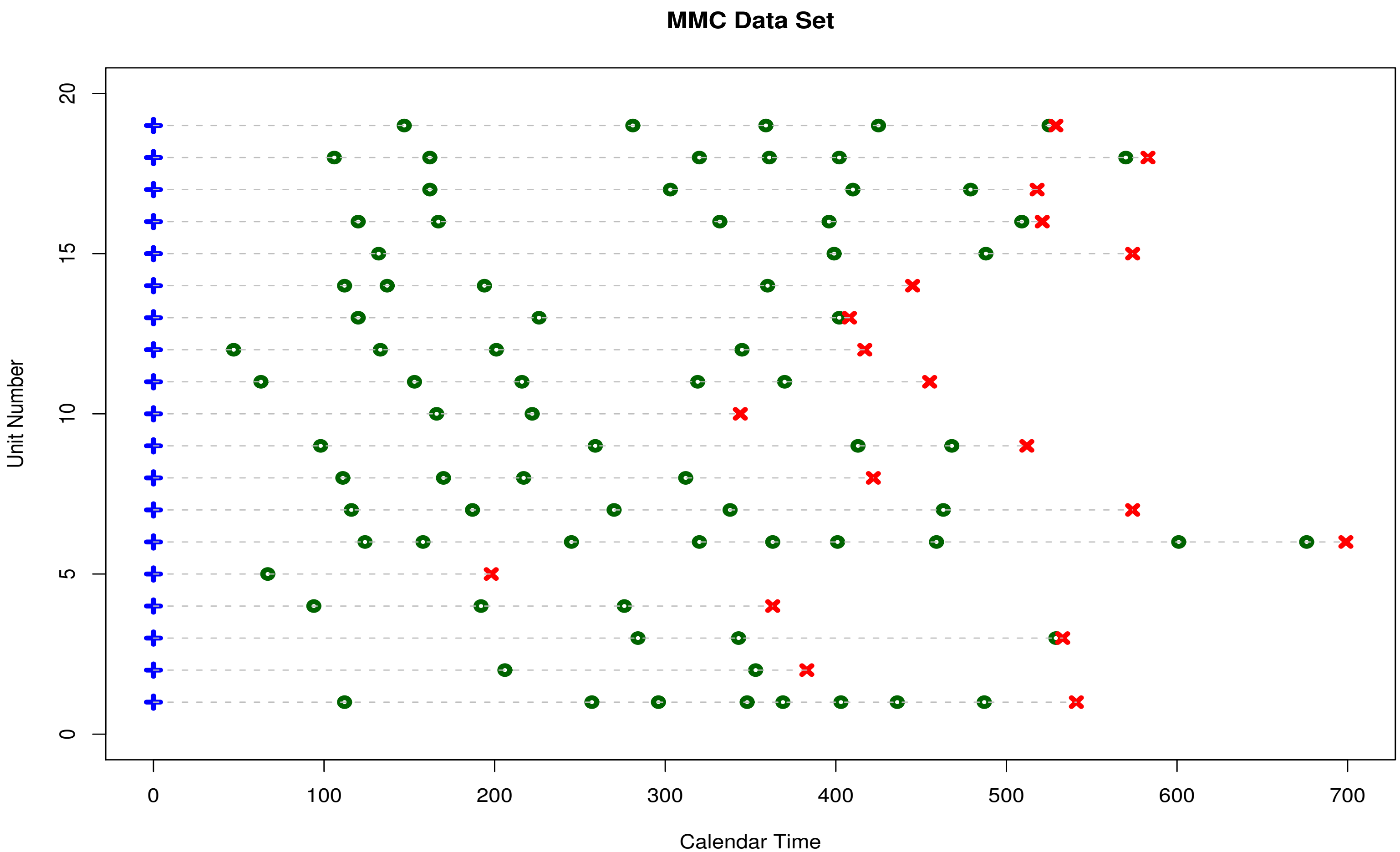
Seattle, Washington

Recurrent Phenomena

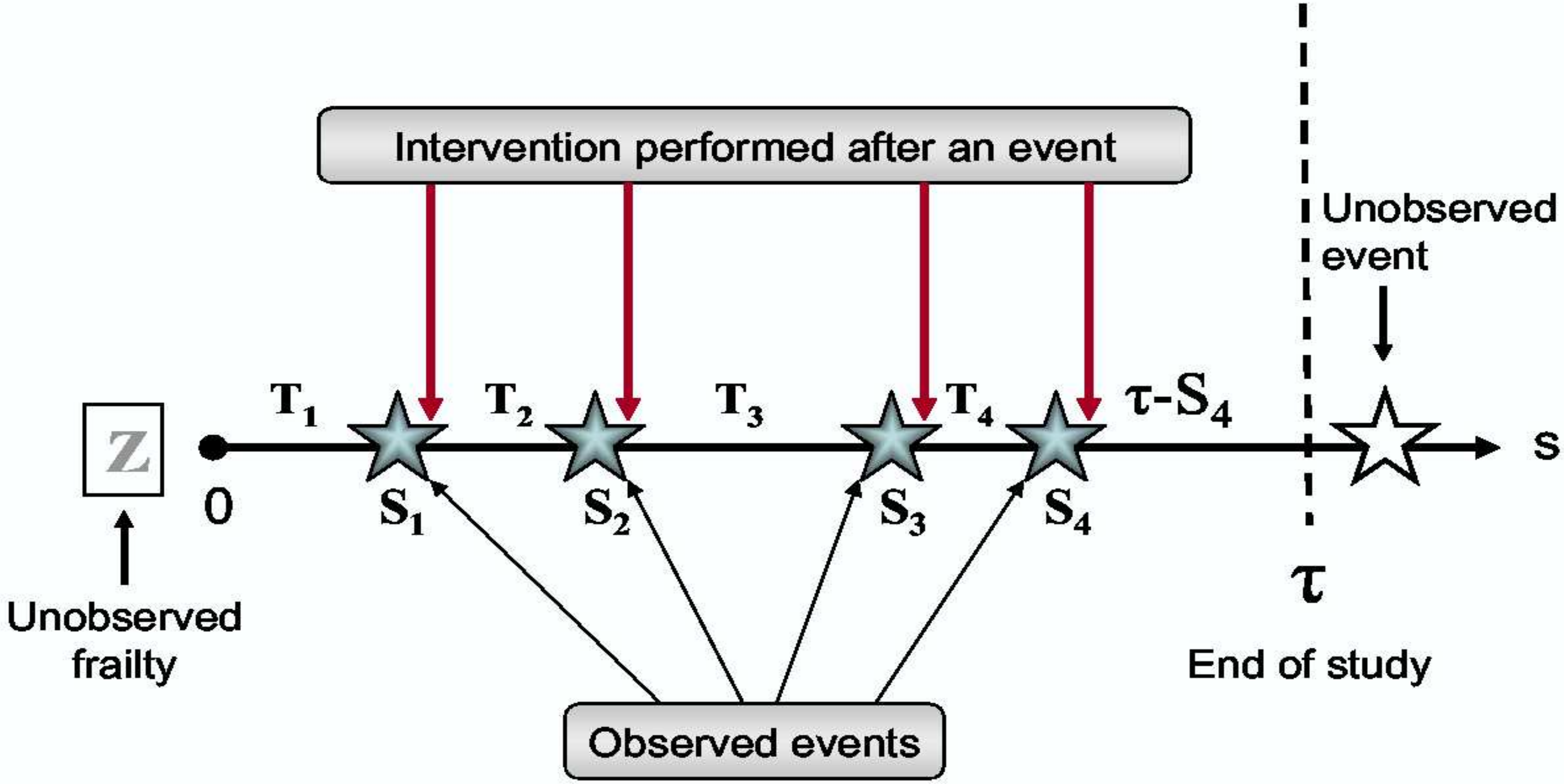
- Hospitalization due to a chronic disease.
- Drug/alcohol abuse
- Occurrence of migraine headaches.
- Onset of depression.
- Episodes of epileptic seizures.
- Non-fatal heart attacks.
- Software crashes and subsystem failures.
- Non-life insurance claims.
- In sociology: serious marital disagreements.
- Publication of a research paper or book.

Motivating Data Set: MMC Data Set

Migratory Motor Complex (MMC) Times for 19 Subjects
(Aalen and Husebye, 1991)



Representation: One Subject



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

Observables: One Subject

- $\mathbf{X}(s)$ = covariate vector, possibly time-dependent
- $T_1, T_2, T_3, \ldots$ = inter-event or gap times
- $S_1, S_2, S_3, \ldots$ = calendar times of event occurrences
- τ = end of observation period.
- $K = \max\{k : S_k \leq \tau\}$ = number of events in $[0, \tau]$
- Z = unobserved frailty variable
- $N^\dagger(s)$ = number of events in $[0, s]$
- $Y^\dagger(s) = I\{\tau \geq s\}$ = at-risk indicator at time s
- $\mathbf{F}^\dagger = \{\mathcal{F}_s^\dagger : s \geq 0\}$ = filtration: information that includes interventions, covariates, etc.

Aspect of Sum-Quota Accrual

Remark: A unique feature of recurrent event modeling is the sum-quota constraint that arises due to a fixed or random observation window. Failure to recognize this in the statistical analysis leads to erroneous conclusions.

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

$(T_1, T_2, \ldots, T_K)$ satisfies $\displaystyle\sum_{j=1}^{K} T_j \leq \tau < \sum_{j=1}^{K+1} T_j$.

General Class of Dynamic Models

- Peña and Hollander (2004) model.

$$N^\dagger(s) = A^\dagger(s|Z) + M^\dagger(s|Z)$$

$$M^\dagger(s|Z) \in \mathcal{M}_0^2 = \text{square-integrable martingales}$$

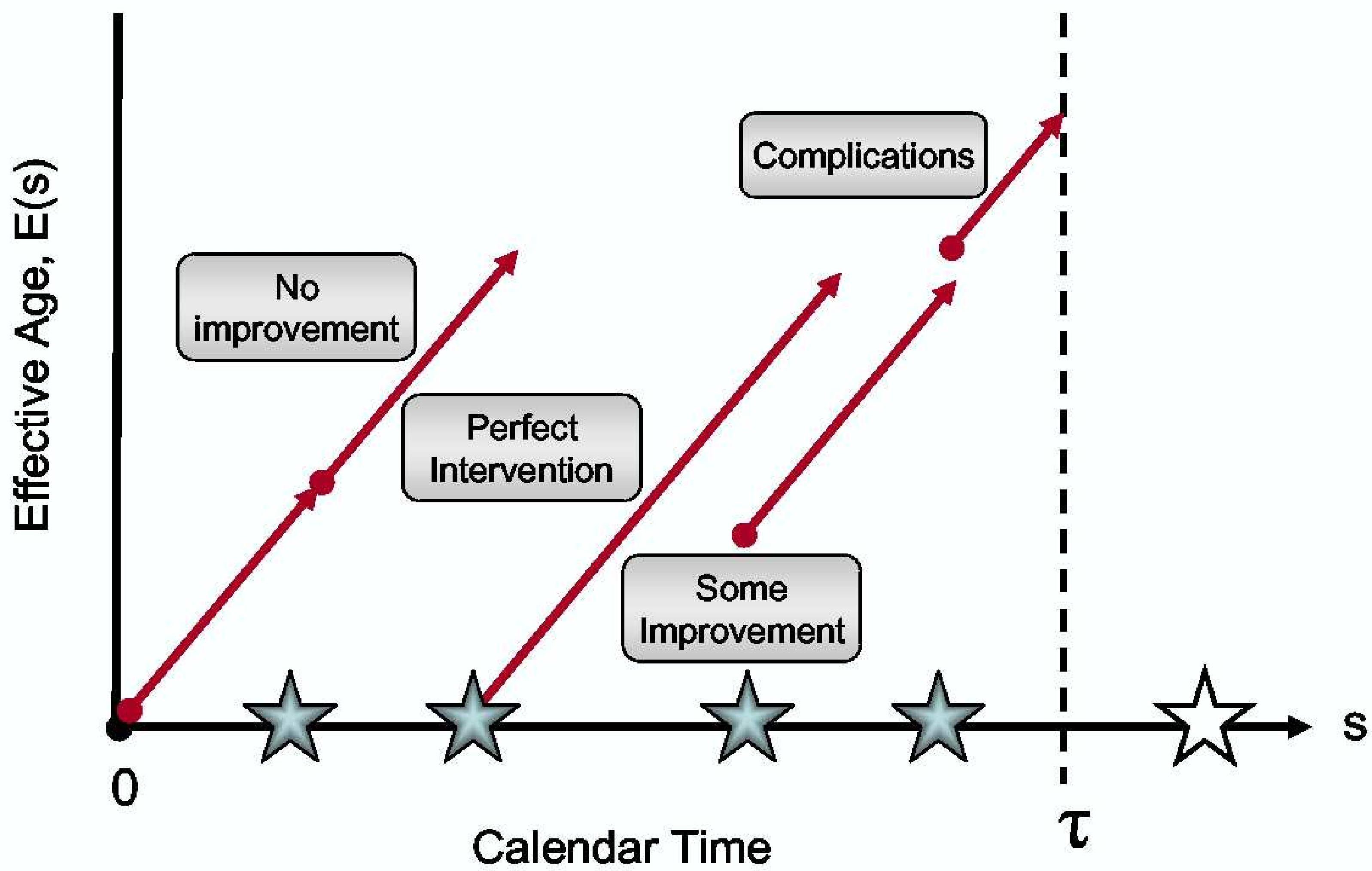
$$A^\dagger(s|Z) = \int_0^s Y^\dagger(w)\lambda(w|Z)dw$$

- Intensity Rate Process:

$$\lambda(s|Z) = Z\,\lambda_0[\mathcal{E}(s)]\,\rho[N^\dagger(s-);\alpha]\,\psi[\beta^{\mathsf{t}} X(s)]$$

- Class includes as special cases many models in reliability and survival analysis.

Effective Age Process



Some Effective Age Processes

- Perfect Intervention: $\mathcal{E}(s) = s - S_{N^\dagger(s-)}$.
- Imperfect Intervention: $\mathcal{E}(s) = s$.
- Minimal Intervention (Brown & Proschan, '83; Block, Borges & Savits, '85):

$$\mathcal{E}(s) = s - S_{\Gamma_{\eta(s-)}}$$

where, with $I_1, I_2, \ldots$ IID BER(p),

$$\eta(s) = \sum_{i=1}^{N^\dagger(s)} I_i \quad \text{and} \quad \Gamma_k = \min\{j > \Gamma_{k-1} : I_j = 1\}.$$

Semi-Parametric Estimation: No Frailty

Observed Data for n Subjects:

$$\{(\mathbf{X}_i(s), N_i^\dagger(s), Y_i^\dagger(s), \mathcal{E}_i(s)) :\ 0 \le s \le s^*\}, i = 1, \ldots, n$$

$$N_i^\dagger(s) = \text{\# of events in } [0, s] \text{ for } i\text{th unit}$$

$$Y_i^\dagger(s) = \text{at-risk indicator at } s \text{ for } i\text{th unit}$$

$$A_i^\dagger(s) = \int_0^s Y_i^\dagger(v)\, \rho[N_i^\dagger(v-); \alpha]\, \psi[\beta^{\mathsf{t}} \mathbf{X}_i(v)]\, \lambda_0[\mathcal{E}_i(v)] dv$$

Baseline gap-time distribution associated with $\lambda_0(\cdot)$ will be denoted by $\bar{F}_0$.

Processes and Notations

Calendar/Gap Time Processes:

$$N_i(s,t) = \int_0^s I\{\mathcal{E}_i(v) \le t\} N_i^\dagger(dv)$$

$$A_i(s,t) = \int_0^s I\{\mathcal{E}_i(v) \le t\} A_i^\dagger(dv)$$

Notational Reductions:

$$\mathcal{E}_{ij-1}(v) \equiv \mathcal{E}_i(v) I_{(S_{ij-1}, S_{ij}]}(v) I\{Y_i^\dagger(v) > 0\}$$

$$\varphi_{ij-1}(w|\alpha,\beta) \equiv \frac{\rho(j-1;\alpha)\psi\{\beta^{\mathsf{t}} \mathbf{X}_i[\mathcal{E}_{ij-1}^{-1}(w)]\}}{\mathcal{E}'_{ij-1}[\mathcal{E}_{ij-1}^{-1}(w)]}$$

Change-of-Variable Transformations

$$\int_0^s H(s,\mathcal{E}(v))I\{\mathcal{E}_i(v)\le t\}N^\dagger(dv) = \int_0^t H(s,w)N(s,dw);$$

$$\int_0^s H(s,\mathcal{E}(v))I\{\mathcal{E}_i(v)\le t\}A^\dagger(dv) = \int_0^t H(s,w)Y(s,w)\Lambda_0(dw);$$

$$Y(s,w) = \sum_{j=1}^{N^\dagger(s-)} I_{(\mathcal{E}_{j-1}(S_{j-1}),\ \mathcal{E}_{j-1}(S_j)]}(w)\ \varphi_{j-1}(w) +$$

$$I_{(\mathcal{E}_{N^\dagger(s-)}(S_{N^\dagger(s-)}),\ \mathcal{E}_{N^\dagger(s-)}((s\wedge\tau))]}(w)\ \varphi_{N_i^\dagger(s-)}(w|\alpha,\beta);$$

$$\int_0^s H(s,\mathcal{E}(v))I\{\mathcal{E}_i(v)\le t\}M^\dagger(dv) = \int_0^t H(s,w)M(s,dw).$$

Generalized At-Risk Processes

$$Y_i(s, w|\alpha, \beta) = \sum_{j=1}^{N_i^\dagger(s-)} I_{(\mathcal{E}_{ij-1}(S_{ij-1}),\, \mathcal{E}_{ij-1}(S_{ij})]}(w)\ \varphi_{ij-1}(w|\alpha, \beta) +$$

$$I_{(\mathcal{E}_{iN_i^\dagger(s-)}(S_{iN_i^\dagger(s-)}),\, \mathcal{E}_{iN_i^\dagger(s-)}((s\wedge\tau_i))]}(w)\ \varphi_{iN_i^\dagger(s-)}(w|\alpha, \beta)$$

For IID Renewal Model (PSH, 01) this simplifies to:

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

Estimation of Λ_0

$$A_i(s,t|\alpha,\beta) = \int_0^t Y_i(s,w|\alpha,\beta)\Lambda_0(dw)$$

$$S_0(s,t|\alpha,\beta) = \sum_{i=1}^n Y_i(s,t|\alpha,\beta)$$

$$J(s,t|\alpha,\beta) = I\{S_0(s,t|\alpha,\beta) > 0\}$$

Generalized Nelson-Aalen ‘Estimator’:

$$\hat{\Lambda}_0(s,t|\alpha,\beta) = \int_0^t \left\{\frac{J(s,w|\alpha,\beta)}{S_0(s,w|\alpha,\beta)}\right\}\left\{\sum_{i=1}^n N_i(s,dw)\right\}$$

Estimation of α and β

- Partial Likelihood (PL) Process:

$$L_P(s^*|\alpha,\beta) = \prod_{i=1}^{n} \prod_{j=1}^{N_i^\dagger(s^*)} \left[\frac{\rho(j-1;\alpha)\psi[\beta^{\mathtt{t}}\mathbf{X}_i(S_{ij})]}{S_0[s^*,\mathcal{E}_i(S_{ij})|\alpha,\beta]} \right]^{\Delta N_i^\dagger(S_{ij})}$$

- PL-MLE: $\hat{\alpha}$ and $\hat{\beta}$ are maximizers of the mapping

$$(\alpha,\beta) \mapsto L_P(s^*|\alpha,\beta)$$

- Iterative procedures. Implemented in an `R` package called `gcmrec` (Gonzaléz, Slate, Peña '04).

Estimation of $\bar{F}_0$

- G-NAE of $\Lambda_0(\cdot)$: $\hat{\Lambda}_0(s^*, t) \equiv \hat{\Lambda}_0(s^*, t|\hat{\alpha}, \hat{\beta})$

- G-PLE of $\bar{F}_0(t)$:

$$\hat{\bar{F}}_0(s^*, t) = \prod_{w=0}^{t} \left[1 - \frac{\sum_{i=1}^{n} N_i(s^*, dw)}{S_0(s^*, w|\hat{\alpha}, \hat{\beta})} \right]$$

- For IID renewal model with $\mathcal{E}_i(s) = s - S_{iN_i^\dagger(s-)}$, $\rho(k;\alpha) = 1$, and $\psi(w) = 1$, the estimator in PSH (2001) obtains.

Semi-Parametric Estimation: With Frailty

- Recall the intensity rate:

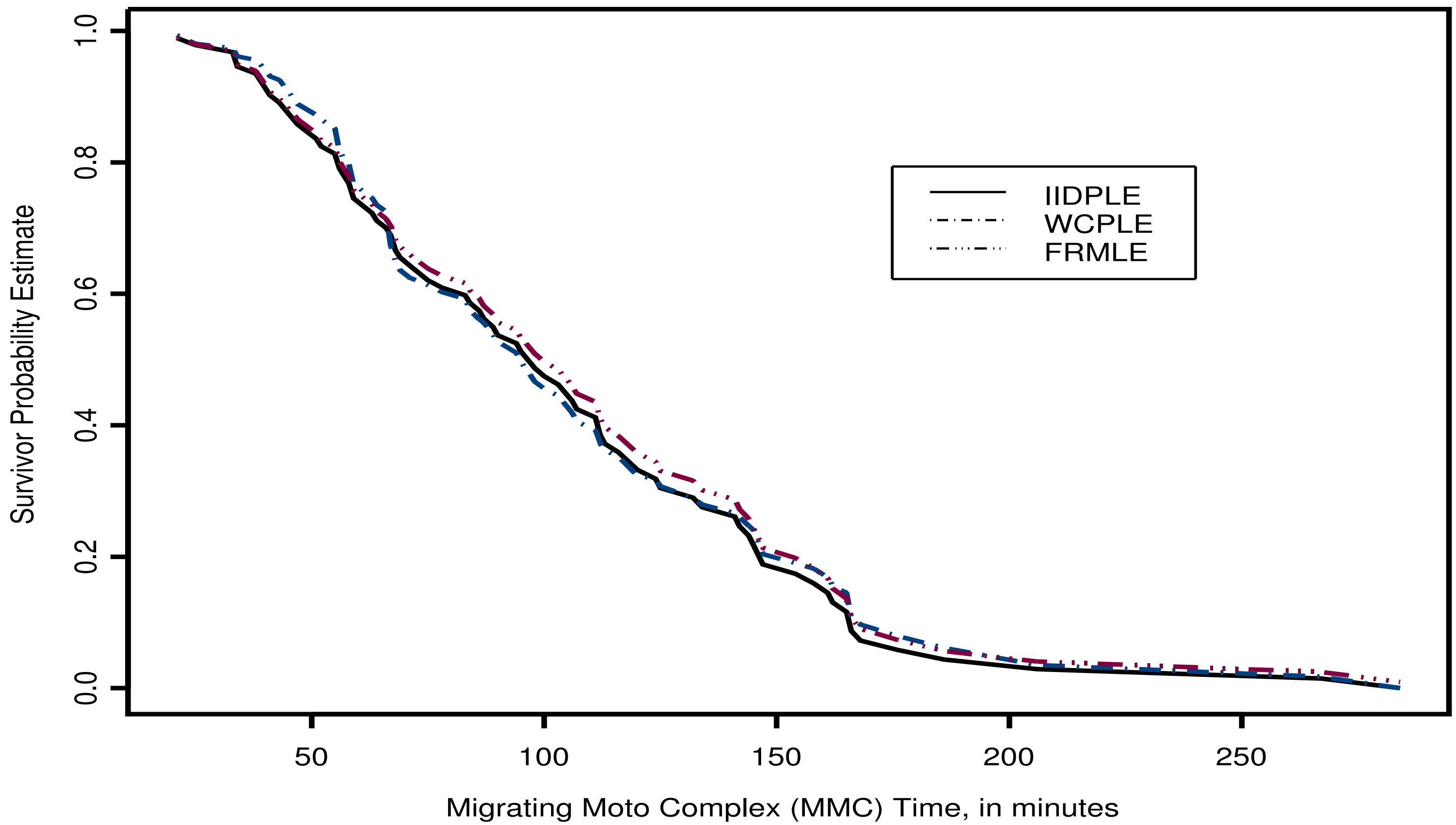
$$\lambda_i(s|Z_i, \mathbf{X}_i) = Z_i\,\lambda_0[\mathcal{E}_i(s)]\,\rho[N_i^\dagger(s-);\alpha]\,\psi(\beta^{\mathtt{t}}\mathbf{X}_i(s))$$

- Frailties $Z_1, Z_2, \ldots, Z_n$ are unobserved and assumed to be IID Gamma(ξ, ξ)
- Unknown parameters: $(\xi, \alpha, \beta, \lambda_0(\cdot))$
- Use of the EM algorithm (Dempster, et al; Nielsen, et al), with frailties as missing observations.
- Estimator of baseline hazard function under no-frailty model plays an important role.
- Details in Peña, Slate & Gonzalez (*JSPI*, to appear).

An Application: MMC Data Set

Aalen and Husebye (1991) Data

Estimates of distribution of MMC period



On Asymptotic Properties

- Asymptotics under the no-frailty models.
- Difficulty: $\lambda_0(\cdot)$ has $\mathcal{E}(s)$ as argument in the model; whereas, interest is usually on $\Lambda_0(t)$.
- No martingale structure in gap-time axis. MCLT not directly applicable.
- Under regularity conditions: consistency and joint weak convergence to Gaussian processes of standardized $(\hat{\alpha}, \hat{\beta})$ and $\hat{\Lambda}_0(s^*, \cdot)$.
- Results extend those in Andersen and Gill (AoS 82) regarding Cox PHM, though proofs different.
- Research on the asymptotics for the model with frailty *in progress.*

Asymptotics: Master Theorem

- $\{\mathbf{H}_i\}$ a sequence defined on $[0, s^*] \times [0, t^*]$.
- $M_i(s,t) = \int_0^s I\{\mathcal{E}_i(v) \leq t\} M_i^\dagger(dv)$.
- $Y_i(s,t)$ - generalized at-risk process.
- Under some regularity conditions, and if

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{H}_i^{\otimes 2}(s^*, \cdot) Y_i(s^*, \cdot) \xrightarrow{upr} \mathbf{v}(s^*, \cdot),$$

- then, with $\Sigma(s^*, t) = \int_0^t \mathbf{v}(s^*, w)\Lambda_0(dw)$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\cdot} \mathbf{H}_i(s^*, w) M_i(s^*, dw) \Longrightarrow \mathbf{GP}(0, \Sigma(s^*, \cdot)).$$

Relevant Empirical Measures

- Simplified model (one unit):

$$\Pr\{dN_i^\dagger(v) = 1 | \mathcal{F}_{s-}\} = Y_i^\dagger(v)\lambda_0[\mathcal{E}_i(v)]\Xi_i(v;\eta)dv.$$

- *Conditional* PM $Q(s^*, w; \eta)$ on $\{1, 2, \ldots, N^\dagger(s-) + 1\}$:

$$Q(\{j\}; s^*, w; \eta) = \frac{\varphi_{j-1}(w;\eta) I\{\mathcal{E}(S_{j-1}) < w \leq \mathcal{E}(S_j)\}}{Y(s^*, w)}$$

with $S_{N^\dagger(s-)+1} = \min(s, \tau)$.

- *Conditional* PM $P(s^*, w; \eta)$ on $\{1, 2, \ldots, n\}$:

$$P(\{i\}; s^*, w; \eta) = \frac{Y_i(s^*, w; \eta)}{\mathbb{P}Y(s^*, w; \eta)}.$$

Empirical Means & Variances

$$\mathbb{P}f(\mathbf{D}) = \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{D}_i)$$

$$\mathbb{E}_{Q(s^*,w;\eta)}g(J) = \sum_{j=1}^{N^\dagger(s^*-)+1} g(j)Q(\{j\}; s^*, w; \eta)$$

$$\mathbb{V}_{Q(s^*,w;\eta)}g(J) = \mathbb{E}_{Q(s^*,w;\eta)}[g^2(J)] - (\mathbb{E}_{Q(s^*,w;\eta)}g(J))^2$$

$$\mathbb{E}_{P(s^*,w;\eta)}g(I) = \sum_{i=1}^{n} g(i)P(\{i\}; s^*, w; \eta)$$

$$\mathbb{V}_{P(s^*,w;\eta)}g(I) = \mathbb{E}_{Q(s^*,w;\eta)}[g^2(I)] - (\mathbb{E}_{Q(s^*,w;\eta)}g(I))^2$$

Relevant Limit Functions

- $s_0(s^*, w; \eta, \Lambda_0) =$ plim $\mathbb{P}Y(s^*, w; \eta)$.
- Partial Likelihood Information Limit:

$$\mathcal{I}_p(s^*, t; \eta, \Lambda_0) = \text{plim}$$
$$\int_0^t \left\{ \left[\mathbb{E}_{P(s^*,w;\eta)} \mathbb{V}_{Q(s^*,w;\eta)} \left(\nabla_\eta \log \Xi_I(\mathcal{E}_{IJ-1}^{-1}(w); \eta) \right) + \right.\right.$$
$$\left.\left. \mathbb{V}_{P(s^*,w;\eta)} \mathbb{E}_{Q(s^*,w;\eta)} \left(\nabla_\eta \log \Xi_I(\mathcal{E}_{IJ-1}^{-1}(w); \eta) \right) \right] \right\} \times$$
$$s_0(s^*, w; \eta, \Lambda_0)\ \Lambda_0(dw).$$

- With $\mathbf{e}(s^*, w; \eta, \Lambda_0) =$ plim $\frac{\mathbb{P}\nabla_\eta Y(s^*,w;\eta)}{\mathbb{P}Y(s^*,w;\eta)}$, let

$$A(s^*, t; \eta, \Lambda_0) = \int_0^t \mathbf{e}(s^*, w; \eta, \Lambda_0)\Lambda_0(dw).$$

Weak Convergence Results

As $n \to \infty$ and under certain regularity conditions:

$$\sqrt{n}(\hat{\eta}(s^*, t^*) - \eta) \Rightarrow N(0, \mathcal{I}_p(s^*, t^*; \eta, \Lambda_0)^{-1})$$

$$\sqrt{n}(\hat{\Lambda}_0(s^*, \cdot) - \Lambda_0(\cdot)) \Rightarrow GP(0, \Gamma(s^*, \cdot; \eta, \Lambda_0))$$

where the limiting variance function is given by

$$\begin{aligned}\Gamma(s^*, t; \eta, \Lambda_0) &= \int_0^t \frac{\Lambda_0(dw)}{s_0(s^*, w; \eta)} \\ &+ \; A(s^*, t; \eta, \Lambda_0)\mathcal{I}_p(s^*, t^*; \eta, \Lambda_0)^{-1} A(s^*, t; \eta, \Lambda_0)^{\mathtt{t}}.\end{aligned}$$

Concluding Remarks

- Many aspects of the general dynamic recurrent event model still under investigation.
- Asymptotics for the model with frailty.
- Testing hypothesis procedures.
- Goodness-of-fit and residual analysis.
- Its practical relevance still needs exploring, e.g., could the effective age process be determined appropriately in practice.
- Comparisons with marginal-based models (PWP, WLW).
- *Dynamic recurrent event modeling* remains a challenge and is a fertile area for research.