Frailty Modeling for clustered survival data: a simulation study

IAA Oslo 2015

Souad ROMDHANE

LaREMFiQ - IHEC
University of Sousse (Tunisia)

souad_romdhane@yahoo.fr

Lotfi BELKACEM

LaREMFiQ - IHEC
University of Sousse (Tunisia)

Monday, June 8, 2015

Motivation

- The current regulatory context of **Solvency II** –> partial internal model –> the construction of experience table as an alternative of the national life table.

- Risks within an insured portfolio are heterogeneous –> difference in mortality rates and life expectancy between :

  Men / women

  Smokers / Non-smokers

  Chronically-diseases dependant (diabetes, obesity..) / Insurers with good health (etc.)

Motivation

First approach

$(+)$ Modeling the behavior of each specific group in an independent way.
$(-)$ Problems of insufficient data, problems of choice of optimal segmentation and raise the risk estimation.

Motivation

First approach

$(+)$ Modeling the behavior of each specific group in an independent way.
$(-)$ Problems of insufficient data, problems of choice of optimal segmentation and raise the risk estimation.

Second approach

$(+)$ Turning to models integrating the observable factors of heterogeneity starting from explanatory variables :

Motivation

First approach

(+) Modeling the behavior of each specific group in an independent way.
(−) Problems of insufficient data, problems of choice of optimal segmentation and raise the risk estimation.

Second approach

(+) Turning to models integrating the observable factors of heterogeneity starting from explanatory variables : **"Proportional Hazards Cox model"**

Motivation

First approach

$(+)$ Modeling the behavior of each specific group in an independent way.
$(-)$ Problems of insufficient data, problems of choice of optimal segmentation and raise the risk estimation.

Second approach

$(+)$ Turning to models integrating the observable factors of heterogeneity starting from explanatory variables : **"Proportional Hazards Cox model"**
$(-)$Unobservable risk factors have been disregarded.

In such case of human life expectancy studies : No matter how many covariates we want to add, it will never be the complete one.

Motivation

First approach

(+) Modeling the behavior of each specific group in an independent way.
(−) Problems of insufficient data, problems of choice of optimal segmentation and raise the risk estimation.

Second approach

(+) Turning to models integrating the observable factors of heterogeneity starting from explanatory variables : **"Proportional Hazards Cox model"**
(−)Unobservable risk factors have been disregarded.

In such case of human life expectancy studies : No matter how many covariates we want to add, it will never be the complete one.

Third approach

(+) Integrating the unobservable factors of heterogeneity :**"Frailty models"**

Objective

We are interested in the models of the 3^{rd} approach which integrates observable and non observable factors of heterogeneity as an explanatory variables.

- Investigate, by means of simulation studies in a comprehensive way, the particular situations where this shared frailty model becomes applicable and preferable to classical Cox proportional hazards model and when it becomes inaccurate to use such model.
- Explore in more details the link between fixed factors estimates and some parameters combinations such as percent censoring, group size, number of groups, magnitude of the variance parameters associated with the distribution of the random effects etc.
- Compare several frequentist estimation approaches of frailty models and studying their limitations for different parameters combinations (e.g EM, PPL, REML, marginal likelihood approach, MCEM).

Survival Analysis

- Survival analysis is concerned with studying the time between entry to a study (birth date) and a subsequent event (death date).
- A Cox model is a statistical technique for exploring the relationship between the survival of an insured and several explanotory variable.
- A significant feature :
  - **censored survival times :** The period of observation was cut off before the event of interest occurred.

Survival Analysis

- Survival analysis is concerned with studying the time between entry to a study (birth date) and a subsequent event (death date).
- A Cox model is a statistical technique for exploring the relationship between the survival of an insured and several explanotory variable.
- A significant feature :
  - **censored survival times :** The period of observation was cut off before the event of interest occurred.
  - **Proportional Hazards assumption :** This means that the hazard functions for any two individuals at any point in time are proportional and time independent.

Survival Analysis

- Survival analysis is concerned with studying the time between entry to a study (birth date) and a subsequent event (death date).
- A Cox model is a statistical technique for exploring the relationship between the survival of an insured and several explanotory variable.
- A significant feature :
  - **censored survival times :** The period of observation was cut off before the event of interest occurred.
  - **Proportional Hazards assumption :** This means that the hazard functions for any two individuals at any point in time are proportional and time independent.
  - **Unknown baseline hazard function.**

Survival models : Cox model

Heterogeneity **only** due to observable risk factors

$$\lambda(t_i \left|\beta\right.) = \lambda_0(t)\, exp\left(x_i^t \beta\right)$$

- **Assumption :** The hazard function for each individual is proportional to the basline hazard $\lambda_0(t)$. This assumption implies that the hazard function is fully determined by the covariate vector.
- **Problem :** There may be unobserved covariates that cause this assumption to be violated.

Survival models :**Shared Frailty model**

Heterogeneity due to observable and **unobservable** risk factors.

- $$\begin{aligned}\lambda(t_{ij}\,|\beta_i, b_i) &= \lambda_0(t_{ij})\, w_{ij}\, exp\left(x_{ij}^t\beta_i\right)\\ &= \lambda_0(t_{ij})\, exp\left(x_{ij}^t\beta_i + z_{ij}^t b_i\right)\end{aligned}$$

-
  - $\delta_{ij} = \mathbb{1}_{\{T_{ij} \leq D_{ij}\}}$ i.e δ_{ij} equal to 1 if $T_{ij} = Y_{ij}$, and 0 if $T_{ij} = C_{ij}$.
  - $W_{ij} = e^{z_{ij}^t b_i}$ with $b_i \sim N(0, \theta)$
  - If $b_i = 0$ i.e. $W_{ij} = 1$, Shared frailty model ==> Cox's proportional hazards model.

Generation survival times

- Bender(2005)

$$T = H_0^{-1}\left[\frac{-log(U)}{w_{ij}\ exp(\beta_i^t x_{ij})}\right]$$

where :

  - $\mathbf{U} \sim Uni[0,1]$,
  - $\mathbf{H_0^{-1}}$ is the inverse of a cumulative baseline hazard function.
  - $\mathbf{H_0}$ can take Exponential, Weibull or Gompertz distribution.

Data simulation

- Each simulated dataset is composed of $(x = 1000)$ *i.i.d.* replications of the quartet $(T_{ij}, \delta_{ij}, X_{ij}, Z_{ij})$:
  where :
  - $\mathbf{T_{ij}} = min(Y_{ij}, C_{ij})$ with
    - Y_{ij} is the failure-time variable corresponding to individual j from cluster i.
    - C_{ij} is non-informative right-censoring time, independent of Y_{ij}.
  - $\boldsymbol{\delta_{ij}}$ is the censoring indicator (The indicator of failure) with δ_{ij} equal to 1 if $T_{ij} = Y_{ij}$, and 0 if $T_{ij} = C_{ij}$.
  - $\mathbf{X_{ij}}$ is a vector of covariates associated with the fixed-effect parameters.
  - $\mathbf{Z_{ij}}$ is a vector of covariates associated with the random-effect parameter b_i.

Data simulation

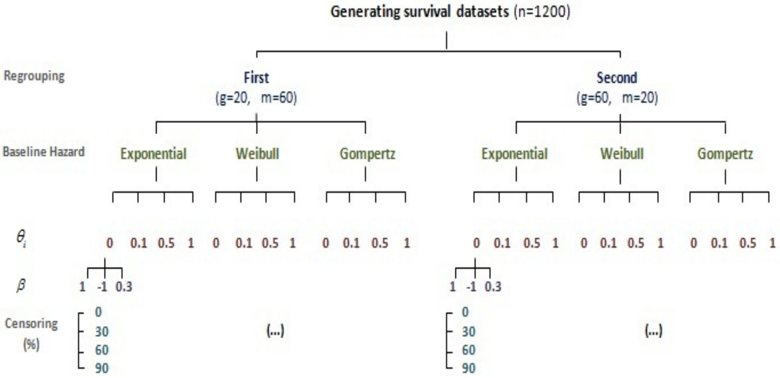


FIGURE: Mechanism of the simulation study for different datasets combination.

Data simulation

- Cluster size :20 clusters of 60 observations - 60 clusters of 20 observations.
- Fixed censoring effects : 0%, 30%, 60% and 90%
- Variance of frailty :
  - Classical Cox model ($\theta = 0$)
  - 2 Shared frailty models with various levels of heterogeneity ($\theta = 0.1$, $\theta = 0.5$, $\theta = 1$).
- 3 Covariates :

Covariates	Distribution	regression coefficients
X_1	Binomial $B[n, p = 0.5]$	$B_1 = 1$
X_2	Normal $N[0, 1]$	$B_2 = -1$
X_3	Uniform $U[0, 2]$	$B_3 = 0.3$

TABLE: Simulation with an example of 3 covariates

Estimation methods

- To compare the performance of different inference procedures for shared frailty :
  - **Fit1** : method PPL
  - **Fit2** : method PPL
  - **Fit3** : method EM
  - **Fit4** : method REML
  - **Fit5** : method PPL
  - **Fit6** : method MCEM

Illustration Results :

- $\sigma^2(\theta) = 0$
  Censoring : 30%(left colun), 90%(right colun)

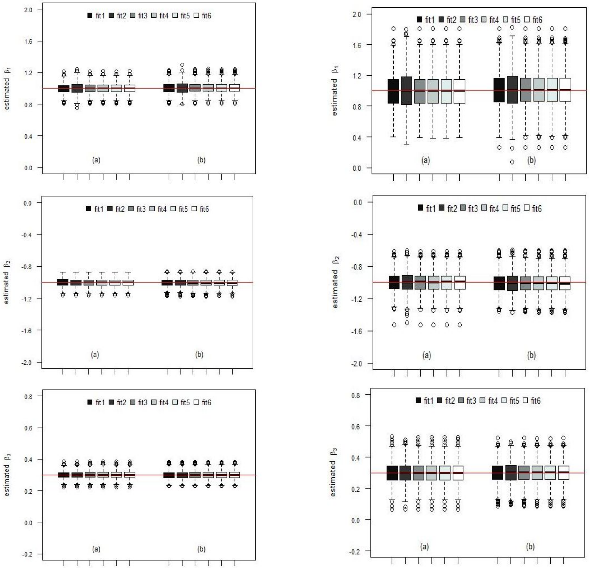


Illustration Results :

- Gamma frailty with $\theta = 0.5$ (left colun)
  log-normal frailty with $\theta = 0.5$ (right column)

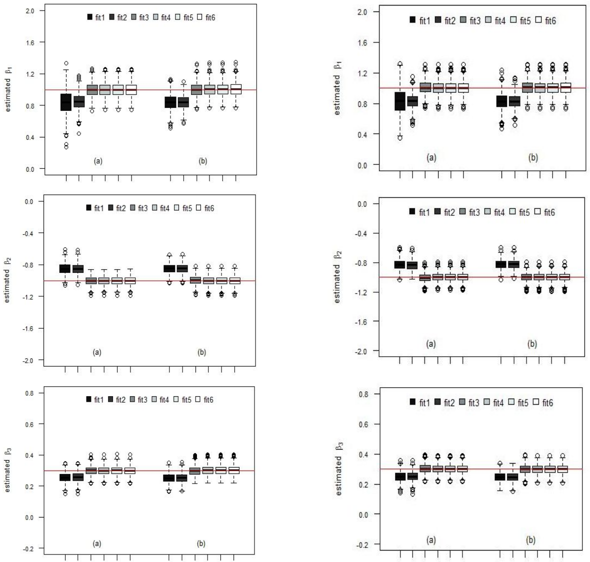
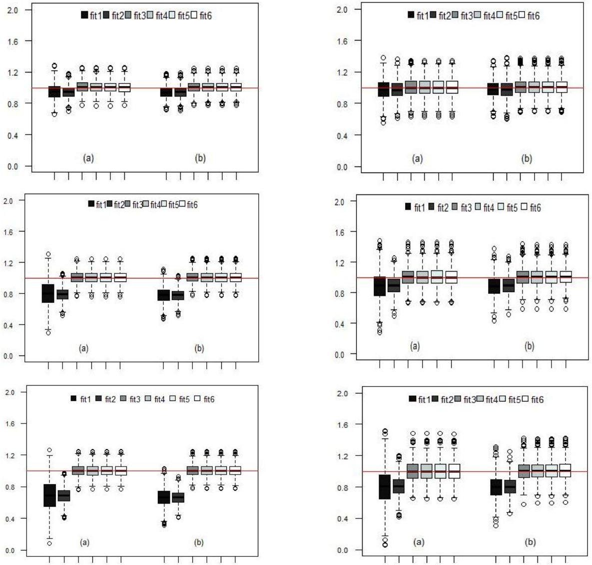


Illustration Results :

- lognormal frailty with $\theta = 0.1$ (1st line),$\theta = 0.5$ (2nd line) ; $\theta = 1$ (3rd line)
  Censoring : 30%(left colun), 90%(right colun)



Conclusions

- The empirical variability of the parameter estimates, by means of EM, REML, PPL or MCEM methods, was in general similar for both gamma and lognormal frailty survival data.
- A misspecification of the distribution of the frailties has no significant impact on the estimates of β.
- The bias increases with increasing
  - the censoring rates
  - the variance of random effects θ, if we do not take it into account through an appropriate frailty model.
-
  - for survival data without frailty effect ($\theta = 0$), Cox's model provides better results in terms of bias and MAE than the estimates of any frailty approaches.
  - for further increase in heterogeneity effects, the bias in the estimates obtained from Cox's model increases. However, those from the frailty model decrease and converge towards true parameter.

Conclusions

- when survival data includes a some level of heterogeneity whatever its distribution, the frailty model approach to estimate β coefficients performs better than the standard Cox's proportional hazard model.
  - –> Cox approach has lost its usefulness when heterogeneity parameter increases more than 0.1.
  - –> Cox model is still better than any frailty models when the heterogeneity parameter is close to zero.
- for all models, the bias for the fixed-effect estimates seems to reduce with the increase in cluster size and for the random-effect estimates with the increasing of the number of clusters.

Thank You