Frailty Modeling for clustered survival data: a simulation study

Souad ROMDHANE *

Lotfi BELKACEM†

Abstract

When referring to actuarial analysis of lifetime, only models accounting for observable risk factors have been developed. Within this context, Cox proportional hazards model (CPH model) is commonly used to assess the effects of observable covariates as gender, age, smoking habits, on the hazard rates. These covariates may fail to fully account for the true lifetime interval. This may be due to the existence of another random variable (frailty) that is still being ignored.
The aim of this paper is to examine the shared frailty issue in the Cox proportional hazard model by including two different parametric forms of frailty into the hazard function. Four estimated methods are used to fit them. The performance of the parameter estimates of observed covariates is assessed and compared to the classical Cox model and then to these frailty models through a general simulation study. This performance is investigated in terms of the bias of point estimates and their empirical standard errors in both fixed and random effect parts. This simulation study showed differences between classical Cox model and shared frailty model.

JEL classification : C02; C13; C52; C63; G22.

Keywords : Life insurance-Pension plan, Survival analysis, Risk factors, Cox proportional hazards model, Multivariate failure-time data, shared frailty, simulations study.

*Ph.D student, Research Laboratory for Economy, Management and Quantitative Finance (LaREMFiQ), IHEC - University of Sousse - Tunisia. E-mail: souad_romdhane@yahoo.fr. Tel: (216)96117703

†Professor and Director of LaREMFiQ, IHEC - University of Sousse, B.P. 40 - Route de la ceinture - Sahloul III, 4054 Sousse - Tunisia. Email: lotfi.belkacem@yahoo.fr. Tel: (216)98643785, Fax: (216)73368350

1 Introduction

In actuarial science and demography, survival data, or time to-event-data, measure the time elapsed from a given date of birth to the occurrence of an event of interest (the death). These are used extensively to derive different types of life tables.

The major problem to get these life tables is due to the heterogeneity of population which can be defined as the blend of individuals with dissimilar hazards and then with substantially different mortality rates or lifetimes. Several aspect of this heterogeneity are observable, as gender, smoking status, socioeconomic class, occupation etc. whilst others like the individual's attitude towards health, environmental factors, and some genetic characteristics are unobservable. Actuarial practice and also literature for life insurance have not included all relevant covariates related to lifetime. Sometimes because they don't know the values of the factor for each individual or sometimes it's due to difficulties inherent in their modeling.

Most commonly, survival data are handled by means of the proportional hazards model, which was first introduced by Cox (1972) and was widely known as the Cox regression model. The central objective of this model is to assess the effects on time to event of only observable covariates by estimating their coefficients. Thus, the conventional Cox model does not always provide an adequate fit to the data and then can generates biases and affects variances of the parameter estimates. One of the reasons is due to the omission of relevant covariates representing information that cannot be observed or have not been observed (univariate case) (Gail and al., 1984; Chamberlain ,1985; Trussel and Rodrigues, 1990). Another reason can be explained by the violation of the traditional assumption that event times are statistically independent and identically distributed when observed covariates are included. In fact, certain individual are linked by criteria that may share several of the above common unobserved factors (multivariate case)(Abrahantes and al, 2005).

For practical and simplicity reasons, such unobservable covariates are often ignored by considering them as a part of the error component and not controlled in conventional survival analysis. Then, heterogeneity due to unobserved covariates is added to the measurement error of observed covariates, and this increases the total variability of the hazard function and produces biased regression coefficient estimates. In such case of human life expectancy studies, no matter how many covariates we want to add, it will never be the complete one (Karim and Latif, 2008).

Therefore, during recent years, several extensions of the Cox proportional hazards model have been developed and lead to what are called "hidden heterogeneity" or "frailty" models. The basic inherent idea is that the hazard function depends upon an unobservable random quantity that impacts multiplicatively on it. Originally, Beard (1959) and Vaupel et al. (1979) proposed a random effects model in order to account for unobserved heterogeneity due to unobserved susceptibility to death. In their studies, the concept of frailty has been introduced and applied in univariate survival models. Their purpose behind introducing the random frailty effect to a Biostatistics framework was to improve the fit of mortality models in a given population.Later,Vaupel and Yashin (1985) used frailties model to explain the deviant behavior of mortality rates at older ages.

Hence, several models have been developed to take into account different forms of frailty between individuals. Most of the literature distinguishes between univariate and multivariate survival models where frailties may be individual-specific or group-specific respectively. One important model in the multivariate set is the shared frailty model. From a modeling point of view, the univariate frailty model is a special case of the shared frailty model with cluster size one. Because of this, we shall chiefly refer to the second. However, the interpretation of the two models is different. In the former case, the frailty distribution variability is related to a measure of over

dispersion between individuals, whereas it is rather interpreted as a measure of dependence in the latter.

The shared frailty model was introduced to the literature by David Clayton (1978) and was considered in the bivariate case without exploiting the notion of frailty. Later, this model was extended to multivariate case and was discussed in details by Hougaard (2000), Therneau and Grambsch (2000), Duchateau et al. (2002, 2003), and Duchateau and Janssen (2004). It is considered as a conditional independence model in which, given the frailty, all event times in a cluster are independent. It is also known as a mixture model because the frailties in each cluster are assumed to be random.

The principal problem with the use of such models is parameters estimation. This explains why various estimation approaches have been proposed in literature, McGilchrist and Aisbett (1991), Nielsen et al. (1992), Xue and Brookmeyer (1996), Ripatti et al. (2000, 2002), Vaida and Xu (2000) and Cortinas and Burzykowski (2004, 2005). Each method presents its advantages and drawbacks.

One of the fundamental objectives of this work is to identify particular situations and specific portfolios of insurers where this shared frailty model is applicable and preferable to classical (conventional) Cox's proportional hazards model and when it becomes inaccurate to use such model. The second aim of this work is to compare several frequentist estimation approaches of frailty models and studying their limitations for different parameter combinations. Therefore we compare a conventional Cox model with four versions of the frailty models differentiated by their frequentist estimation procedures. Bayesian approach, that offers an alternative, will not be studied here. For the shared gamma frailty estimation, we used the EM algorithm approach introduced by Nielsen and al (1992) and later approximated by the Penalized Partial Likelihood solution, Therneau and Grambsch (2000). For the shared Gaussian frailty estimation, we used three estimation procedures: the REML estimation approach proposed byMcGilchrist and Aisbett (1991) and later approximated too by Penalized Partial Likelihood solution, Therneau and Grambsch (2000); the approximated marginal likelihood approach suggested by Ripatti and Palmgren (2000); and the Monte Carlo EM method (MCEM) proposed by Vaida and Xu (2000).The principal reason behind this choice was software availability and feasibility of conducting a simulation study. We implement these versions with available R-packages presented in section 4.3.

Such comparison through a simulation study is not new in the statistical litterature. There has been different studies to compare various shared frailty models. Lorino et al (2004) analyzed three models including fixed or random cluster effects with a conventional Cox model, Cortinas Abrahantes (2005) and Karim and Latif (2008) compared the performance of four estimation procedures available for proportional hazards models with random effects, Hirsch and Wienke (2012) conducted a large simulation study to compare the performance of different statistical software for the analysis of shared frailty mdels.

In contrast to the abundant literature applied to biostatistic context, financial and especially actuarial studies taking into account unobservable heterogeneity are still scare. For this purpose, the aim of this paper that makes the difference with the existing literature is to provide, by means of simulation studies in a comprehensive way, a guideline for actuaries on the performance of the Beta estimates from various characteristics of their database. New models are not proposed, rather, the effects of taking into account different heterogeneity models on the value of estimate covariates are discussed.

The paper is organized as follows. The shared frailty model, its conditional and marginal likelihood are shown in section 2. We outline briefly the estimation methods in section 3. Section 4, presents a general framework investigating by means of many simulations the bias and the efficiency of

the estimated regression coefficients for a variety of situations in particular for a classical Cox model and shared frailty model in a more general framework. Finally, section 5 draws remarks and conclusions based on the obtained results of simulations.

2 Theoretical consideration

2.1 Cox model with shared random Effects

In the following, we will consider clustered survival data from a total of N individuals that come from g different clusters. The failure-time variable corresponding to individual $j(j = 1, ..., n_i)$ from cluster $i(i = 1, ..., g)$ will be denoted by Y_{ij}. It is assumed that the individual j in the cluster i is observed until either an event time Y_{ij} or a non-informative right-censoring time C_{ij} independent of Y_{ij}. Thus, we observe $T_{ij} = min(Y_{ij}, C_{ij})$ and let $\delta_{ij} = \mathbb{1}_{\{T_{ij} \leq D_{ij}\}}$ be the censoring indicator with δ_{ij} equal to 1 if $T_{ij} = Y_{ij}$, and 0 if $T_{ij} = C_{ij}$.

The general mixed-effects proportional hazards model for such data T_{ij} is given by

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

where $\lambda(t_{ij}\,|.)$ is the conditional hazard function for the j^{th} individual from the i^{th} cluster at time t_{ij}, $\lambda_0(t_{ij})$ is the baseline hazard function at time t_{ij}, β_i is a vector of cluster-specific fixed-effects corresponding to a vector of covariates x_{ij}, and b_i is a vector of random effects associated with a vector of covariates z_{ij}.

In what follows, we must distinguish between "frailties" w_i and "random effects" b_i, $W_{ij} = e^{z_{ij}^t b_i}$. Shared frailty model reduces to Cox's proportional hazards model with $b_i = 0$ i.e. $W_{ij} = 1$. Here, this random effects b_i are assumed to be randomly distributed with mean 0 and variance-covariance matrix $\Sigma = \Sigma(\Theta)$, which depends on a d-dimensional vector of parameters $\Theta = (\theta_1, \theta_2, ..., \theta_d)$. This paper focus only on simplest case $d = 1$. Therefore, we consider, $\Theta = \theta$ and $\Sigma = \Sigma(\Theta) = \Sigma(\theta)$. The density function of the b_i which, except for θ, is assumed to be known, will be denoted by $f(b_i)$.

2.2 Likelihood Construction

Under assumptions of non-informative right-censoring and of independence between the censoring time and the survival times Y_{ij}, given the random effects b_i, observations within cluster i are assumed to be conditionally independent. Therefore, considering independence between clusters, **the conditional likelihood** function, denoted by $\mathbf{L^C}$, takes the form

$$L^C(\beta, \lambda_0, b) = \prod_{i=1}^{g} L_i^C(\beta_i, \lambda_0, b_i) = \prod_{i=1}^{g} exp\{l_i^C(\beta_i, \lambda_0, b_i)\} \tag{1}$$

where

$$l_i^C(\beta_i, \lambda_0, b_i) = \sum_{j=1}^{n_i} \left[\delta_{ij}\left(log\lambda_0(t_{ij}) + x_{ij}^t \beta_i + z_{ij}^t b_i\right) - \Lambda_0(t_{ij}) exp\left(x_{ij}^t \beta_i + z_{ij}^t b_i\right)\right] \tag{2}$$

is the conditional log-likelihood for the observed data in the ith cluster, and β and b denote the vectors resulting from "piling" vectors β_i and b_i for all clusters, respectively. $\Lambda_0(t_{ij})$ is the cumulative hazard function.
It follow that, **the marginal likelihood**,denoted by $\mathbf{L^M}$, of the observed data for all cluster can be expressed as

$$L^M(\beta, \theta, \lambda_0) = \prod_{i=1}^{g} \int L_i^A(\beta_i, \theta, \lambda_0, b_i) db_i \tag{3}$$

where

$$L_i^A(\beta_i, \theta, \lambda_0, b_i) = f(b_i) e^{l_i^C(\beta_i, \lambda_0, b_i)} \tag{4}$$

The last function (**??**) can be regarded as the likelihood of the "augmented" data for cluster i, treating b_i as additional observations. Accordingly, **the augmented likelihood**, denoted by $\mathbf{L^A}$ for all clusters is

$$L^A(\beta, \theta, \lambda_0, b) = \prod_{i=1}^{g} L_i^A(\beta_i, \theta, \lambda_0, b_i) \tag{5}$$

3 Estimation methods

3.1 Expectation Maximization algorithm (EM)

From a statistical point of view, frailties are usually viewed as an unobserved covariate. This has led to the application of the EM algorithm as a convenient estimation tool. As suggested by Gill (1985), Klein et al. (1992) and Nielsen et al. (1992) introduced a semi parametric inference for such frailty model by applying an EM algorithm to the Cox's partial likelihood function.
It consists of two major steps. In the first step and at each iteration, the expectation of the log-likelihood of the augmented-data (**??**), denoted by $Q(\beta, \theta, \lambda_0)$, given the observed data and given the currents estimates values from previous iteration $\widetilde{\Psi}(\widetilde{\beta}, \widetilde{\theta}, \widetilde{\lambda_0})$ of the parameters $\Psi(\beta, \theta, \lambda_0)$ respectively, is calculated ($\mathbf{E-step}$).

$$Q(\beta, \theta, \lambda_0) = Q_1(\beta, \lambda_0) + Q_2(\theta) \tag{6}$$

where

$$Q_1(\beta, \lambda_0) = \sum_{i=1}^{g} \sum_{j=1}^{n_i} \left[\delta_{ij} \left(log\lambda_0(t_{ij}) + x_{ij}^t \beta_i + z_{ij}^t E(b_i) \right) - \Lambda_0(t_{ij}) exp\left(x_{ij}^t \beta_i + logE(e^{z_{ij}^t b_i}) \right) \right] \tag{7}$$

and

$$Q_2(\theta) = \sum_{i=1}^{g} \left[log f(b_i) \right] \tag{8}$$

with $E(.)$ denoting conditional expected values given the observed values of T_{ij} and δ_{ij}. Then, maximizing the complete likelihood function $Q_1 + Q_2$ to provide the next parameter values $\widetilde{\beta}$, $\widetilde{\theta}$ as the non-observable frailties were observed ($\mathbf{M-step}$).

The algorithm iterates between the two steps until convergence is attained, Dempster and al.(1977).
However, this approach has some limitations. The EM algorithm has a slower convergence rate than the the Newton-Rapdson method. Thus, it is the most time consuming procedure. Furthermore, getting variance estimates for frailty parameter and regression coefficients require additional computation. Finally, the E-step in (**??**) requires an integral of the same dimension as in the observed data likelihood (Eq (0)). To remedy for integration problems, several solutions based on frequentest approach have been suggested in the literature, including penalized likelihood methods based on the Laplace approximation to the integral, REML and simulation based Monte Carlo EM.

3.2 Restricted maximum likelihood (REML)

McGilchrist and Aisbett (1991) and McGilchrist (1993) used the penalized likelihood approach to estimate the fixed effects and the residual maximum likelihood (REML) to estimate the variance components of a Gaussian random effects. In the first step, their methods consists of finding the best linear unbiased predictors (BLUP) of the fixed and random components by maximizing the sum of two components $l_r + l_f$. Assume that θ is a fixed random effect, the partial log-likelihood of the failure time, l_r is

$$l_r = \sum_{i=1}^{g} \sum_{j=1}^{n_i} \delta_{ij} \left[x_{ij}^t \beta_i + z_{ij}^t b_i - log \sum_{t_{kl} \geq t_{ij}} exp\left(x_{kl}^t \beta_k + z_{kl}^t b_k \right) \right] \tag{9}$$

and

$$l_f = -\frac{1}{2} \sum_{g=0}^{d} \left[g \ log 2\pi\theta_g + \sum_{i=1}^{g} \frac{b_i^2 g}{\theta_g} \right] \tag{10}$$

Then, in the second step, using these results to find both maximum likelihood (ML) and Residual maximum likelihood (REML) estimators. As described in McGilchrist (1993), the ML and REML estimators of β are the same as the BLUP estimator for any given estimate of θ. But the estimators for θ are different.

3.3 Approximate Marginal Likelihood approach

Ripatti and Palmgren (2000) used the derivation of a penalized likelihood solution obtained by Breslow and Clayton (1993) and proposed a parallel approximation for a lognormal frailty model. Their method consists of approximating inference based on the Laplace approximation of the marginal likelihood in (**??**). Assuming that the random effects are normally distributed with variance covariance matrix $D(\theta)$, they showed that, given the marginal likelihood

$$L^M(\beta, \lambda_0, b) = c|D(\theta)|^{-\frac{g}{2}} \int e^{-k(b)} db \tag{11}$$

where

$$k(b) = l^C(\beta, \lambda_0, b) - \frac{1}{2} b^t D(\theta)^{-1} b \tag{12}$$

with $l^C(\beta, \lambda_0, b)$ given by (**??**) and (**??**). Using the Laplace theorem, Ripatti and Palmgren (2000) showed that the logarithm of (**??**) can be approximated by

$$l^M(\beta, \theta, \lambda_0) \approx -\frac{g}{2}|log\, D(\theta)| - \frac{1}{2}\, log|k^{''}(\widetilde{b})| - k(\widetilde{b}) \tag{13}$$

where $k^{'}$ and $k^{''}$ denote, respectively, the g-vector and $g \prod g$ matrix of first and second order partial derivatives of k with respect to b, and $\widetilde{b} = \widetilde{b}(\beta, \theta)$ is the solution to $k^{'}(\widetilde{b}) = 0$. They further showed that, for fixed θ, the values $\widehat{\beta}(\theta)$ and $\widehat{b}(\theta)$, which maximize the penalized log likelihood (**??**) also maximize the penalized partial log likelihood

$$\sum_{i=1}^{g}\sum_{j=1}^{n_i}\delta_{ij}\left[x_{ij}^t\beta_i + z_{ij}^t b_i - log\sum_{t_{kl}\geq t_{ij}} exp\left(x_{kl}^t\beta_k + z_{kl}^t b_k\right)\right] - b^t D(\theta)^{-1} b \tag{14}$$

After calculating the $\widehat{\beta}(\theta)$ and $\widehat{b}(\theta)$ by maximising (**??**), θ can be apdated by maximizing the approximate profile likelihood as described in Cortinas et al. (2007)
.

3.4 Monte Carlo EM method (MCEM)

The MCEM algorithm is basically an EM algorithm where a Monte Carlo integration is required to calculate an expected value.
Klein et al. (1992) and Nielsen et al. (1992)applicate their EM algorithm only to a gamma frailty distribution. Unlike their case of gamma frailties, assuming normally distributed random effect, the computation of the conditional expectation in (**??**) and (**??**) is not trivial. The expectations of b_i and $exp(w_i^{'}b_i)$ are of the type $E\left[f(b_i)\right] = \int f(b_i).p(b_i \backslash y_i)\, db_i$ and then, are not available in closed form.

Thus, Vaida and Xu (2000)based the calculation of these integrals by an EM algorithm using Markov Chain Monte Carlo at the E step with the aim of obtaining the maximum likelihood solution rather than an approximation of it, Ripatti and Palmgren (2000). After, Donohue and Xu (2013) implemented their method with the phmm package available for R software.

4 Simulation study

In order to investigate in more details the particular situations where this shared frailty model becomes applicable and preferable to conventional Cox model and then to provide a practical guidelines for actuaries on the performance of the β estimates from different estimation methods and from various characteristics of their database, a simulation study is conducted using a setting similar to a real data set structure for insurance and pension sectors.

4.1 Generation survival times

As Cox and frailty models are formulated from the hazard function, appropriate survival times are not straightforward simulated. The effects of the regression coefficients have to be translated

from the hazards to the survival times, because the usual software packages for frailty models require the individual survival time date, not the hazard function. Bender (2005) derived a general formula relating between the hazard and the corresponding survival time of the usual Cox model, we follow their methods and we give the expression for a frailty model

$$T = H_0^{-1}\left[\frac{-log(U)}{w_{ij}\ exp(\beta_i^t x_{ij})}\right] \tag{15}$$

where U is a random variable with $U \sim Uni[0,1]$, H_0^{-1} is the inverse of a cumulative baseline hazard function.In this paper, we assume that the baseline hazard H_0 can take Exponential, Weibull and Gompertz distribution. This can be explained by the fact that among the known parametric distributions, only these three cited distributions have the property of proportional hazards.

Table 1: Simulating survival times with the Exponential, Weibull and Gompertz baseline hasard.

Characteristic	Exponential distribution	Weibull distribution	Gompertz distribution
Parameter	Scale parameter $\lambda > 0$	Scale parameter $\lambda > 0$ Shape parameter $\nu > 0$	Scale parameter $\lambda > 0$ Shape parameter $-\infty < \alpha < \infty$
Hazard function	$h_0(t) = \lambda$	$h_0(t) = \lambda\nu t^{\nu-1}$	$h_0(t) = \lambda exp(\alpha t)$
Survival times	$T = \frac{-log(u)}{W_{ij}\lambda exp(\beta^t x_{ij})}$	$T = \left(\frac{-log(u)}{W_{ij}\lambda exp(\beta^t x_{ij})}\right)^{1/\nu}$	$T = \frac{1}{\alpha} log\left(1 - \frac{\alpha log(u)}{W_{ij}\lambda exp(\beta^t x_{ij})}\right)$

4.2 Data simulation

The aim throughout the simulation study is to investigate in more detail 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. Moreover, we allow to compare the performance of different various inference procedures for shared frailty model in these different considered situations.

We simulate a dataset with an example of three covariates, X_1 from a Binomial $B[n, p = 0.5]$, X_2 from a Normal $N[0,1]$ and X_3 from a Normal distribution $U[0,2]$ with arbitrary parameter setting fixed throughout the entire study. The corresponding true regression coefficients are fixed as $B_1 = 1$, $B_2 = -1$ and $B_3 = 0.3$ respectively.

To investigate the importance of censoring, we used fixed and random censoring. First, we consider 4 fixed censoring rates : 0%, 30%, 60% and 90%. Second, we choose a random censure following a uniform distribution on the interval from 0 to 9 and then from 0 to 11 to create about 30% and 60% censoring respectively.

Also, we aim to study the effects of increasing the cluster size and the number of clusters distinctly on the parameter estimates. For each parameter combination, all simulated data contain the same number of observations ($n = 1200$) and was replicated 1000 times ($x = 1000$). Two simulation settings were considered to groupe data sets (20 clusters of 60 observations or 60 clusters of 20 observations), to analyze the effect of cluster size on the estimation performance. Referring to the literature, these amounts are optimal in terms of efficiency.

To generate appropriate survival times for simulation studies, the Exponential, the Weibull and

the Gompertz distributions are used as baseline hazard distributions where parameters are fixed according to the values resulting from the adjustment of a real insurance survival set. These survival times are generated under 2 main assumptions. In a first step, classical Cox model (without frailty or random effects, $\theta = 0$)is used to simulate the data. In a second step, data is generated from various shared frailty models. Different choices of frailty density function are possible. The most common are the one-parameter gamma and log-normal distributions. Therefore, we restrict our comparison to the shared gamma frailty model and the shared log-normal frailty model, respectively. These two frailty models are exhaustively discussed by Therneau and Grambsch (2000), Hougaard (2000) and Wienke (2010).

(a) Shared Gamma frailty model
For identifiability reasons, we restrict the expectation of the frailty term equal to one, $E(W_i) = 1$ and the variance to be finite $V(W_i) = \sigma^2$.Therefore, we consider one-parameter gamma distribution $\Gamma(1/\sigma^2, 1/\sigma^2)$ with $k = \lambda$. It follows that, to generate data with a true values of variance $\sigma^2 = 0.1, 0.5$ and 1, we must consider these gamma distributions $\Gamma(10, 10)$, $\Gamma(2, 2)$, $\Gamma(1, 1)$ respectively.

(b) Shared log-normal frailty model
Here, we suppose that the random effects b_i are normally distributed, $b_i \sim N(0, s^2)$. So therefore, the frailty is given by $W_i = e^{b_i}$. In this case, the expectation and the variance of the frailty W are function of the parameter s^2, $E(W) = e^{\frac{s^2}{2}}$ and $V(W) = e^{\frac{s^2}{2}}(e^{\frac{s^2}{2}} - 1)$. As precedent distribution, we generate data with equivalent true values of variance, so we consider $N(0, 0.1)$, $N(0, 0.5)$ and $N(0, 1)$ respectively.

It is important to note the difference between σ^2 and s^2. The first one denotes the variance of the frailty Z in the gamma frailty model, whereas the second one denotes the variance of the random effect $b = lnW$ in the log-normal model.

−**Fit1**: Conventional Cox Proportional Hazards Model
−**Fit2**: Cox Proportional Hazards Model including fixed cluster effect (method *EM* equivalent to *PPL*)
−**Fit3**: Cox Proportional Hazards Model with gamma shared frailty (method *EM*)
−**Fit4**: Cox Proportional Hazards Model with Gaussian shared frailty (method *REML*)
−**Fit5**: Cox Proportional Hazards Model with Gaussian shared frailty (method *PPL*)
−**Fit6**: Cox Proportional Hazards Model with Gaussian shared frailty (method *MCEM*)

For each fit, we calculate the average of the parameter estimates, their median, their lower and their upper quartiles and the corresponding average of their standard errors across the 1200 generated data sets. We also calculate the Wald and Score Test verifying the significance of each parameter. We use 95% coverage probability (CP) to specify how often the true parameter was covered by the 95% confidence interval based on the normal approximation. Furthermore, we compute the corresponding bias and mean absolute error (MAE) with, $Bias = \widehat{\beta_i} - \beta$, $i = 1, 2, \ldots, 1200$ and $MAE = \frac{|\widehat{\beta_i} - \beta|}{x}$, $x = 1000$. To illustrate these results, boxplots are provided.

4.3 Numerical implementation

A variety of recent estimation procedures becomes available in R statistical software.
The coxph () function from the survival package (Therneau 2014b) treats the semi-parametric model with gamma and lognormal frailties and allow analyses based on EM and approximate marginal likelihood estimates procedures . The coxme (Therneau 2014a) and the phmm packages (Donohue and Xu 2012) perform only semi parametric estimation in the log-normal frailties and implement the REML and the MCEM estimates approaches respectively. To increase the precision of the estimation procedure to an acceptable size, we adapt the convergence criteria (eps) and the maximum iteration number for each pakage as recommend Hirsch and al. (2012). In each replication, to obtain convergence, every estimation procedure requires different computer times to analyse a simulated sample size counting 1200 individuals. coxph and coxme take several minutes to run while phmm needs few hours to analyse one data set. This time can be extented or reduced by increasing or decreasing the size of dataset respectively.

4.4 Results of the simulations

In this simulation study, none of the methods showed convergence problems. While a variety of distributions for the baseline hazard have been explored in this simulation study, due to space limitations here, we present only the Weibull, which seems to represent a good tradeoff between simplicity and flexibility. In addition, all used approaches are semi-parametric estimation models that consider the baseline hazard as unknown. Then, generating the survival with exponential, Weibull or Gompertz baseline hazards don't considerably affect the estimation of fixed or random effect parameters.

In a first step, generated data without frailty/random effects (classical Cox models) was investigated to examine how the estimation approaches act in such condition. As the frailty variance is considered equal to zero, this data does not represent any clustering effects. In this case, the figure 1 shows clearly that the estimation procedures implementing clustering show no differences with those of the classical Cox model. Besides, for each beta estimate and for the six models, left Boxplots display the least Bias. So censoring level has a major effect on the estimation results than the cluter size.

In a second step, generated data from two shared frailty models was analyzed. The estimation results are presented according to the following four subsections.

a. Fixed-effects parameter estimates under Shared gamma frailty model

At the current setting, frailties are generated by gamma distribution. For the same levels of heterogeneity and censoring, all approaches (fit3, fit4, fit5 and fit6) produce on average similar estimates of the fixed parameter β. More difference is seen in the estimation with classical Cox model, fig.2.

Tables 2 and 4 (in the annex) highlight the effect of censoring and heterogeneity levels on the bias and the MAE of fixed-effects parameters respectively. It is shown that for a comparable setting, the bias increases greatly under very heavy and heavy censoring level (90% and 60% respectively). In the presence of moderate censored data (30%), the bias decreases slightly and continue further in the case without censoring. In addition, these two tables display the effect of heterogeneity

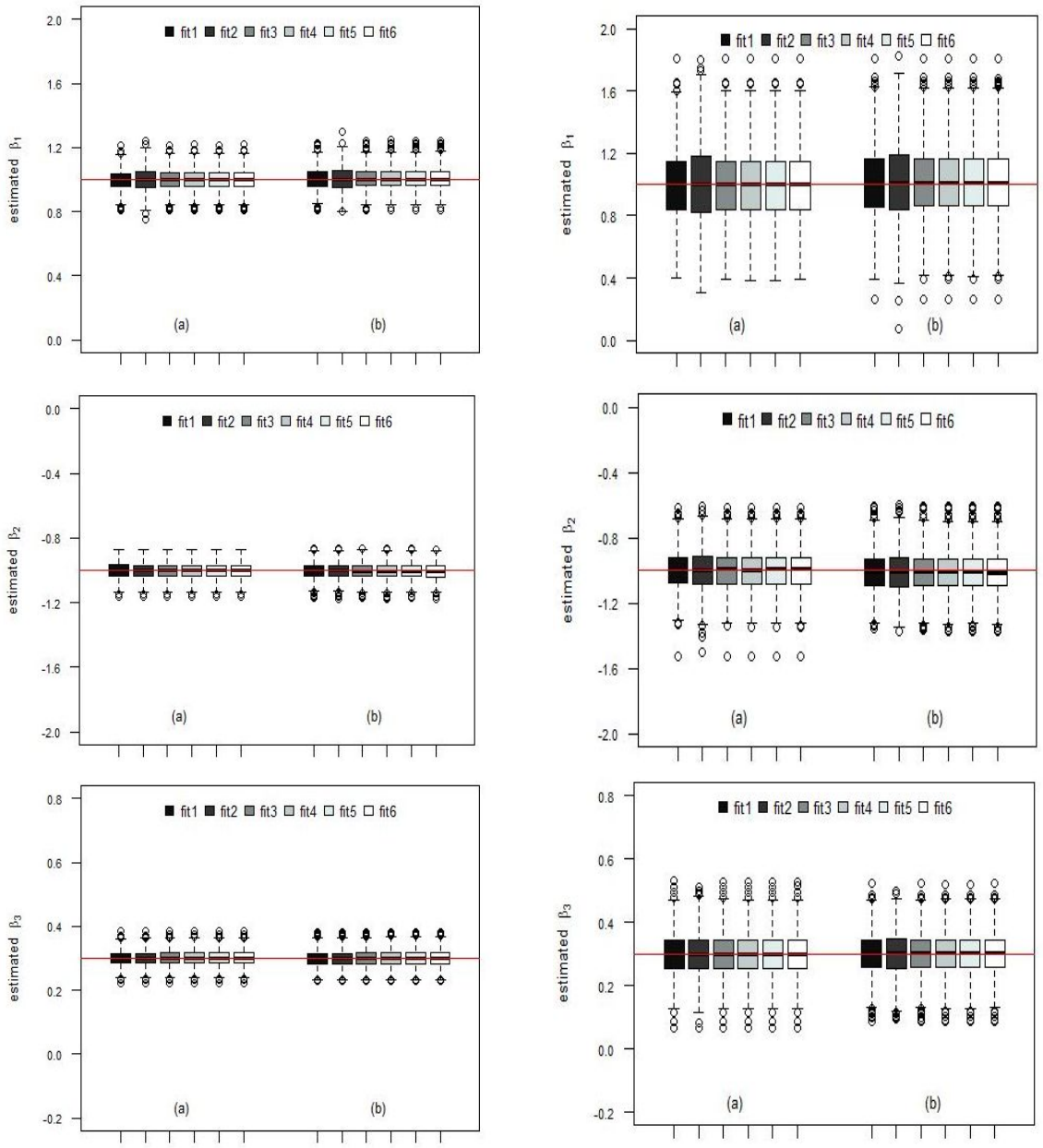

Figure 1: Fixed-effects parameter estimates in data generated by the Cox model ($\theta = 0$) and analyzed by the six fit models; cluster numbers of (a)20 and (b)60; left column 30% right column 90% censoring; true values are horizontal lines.

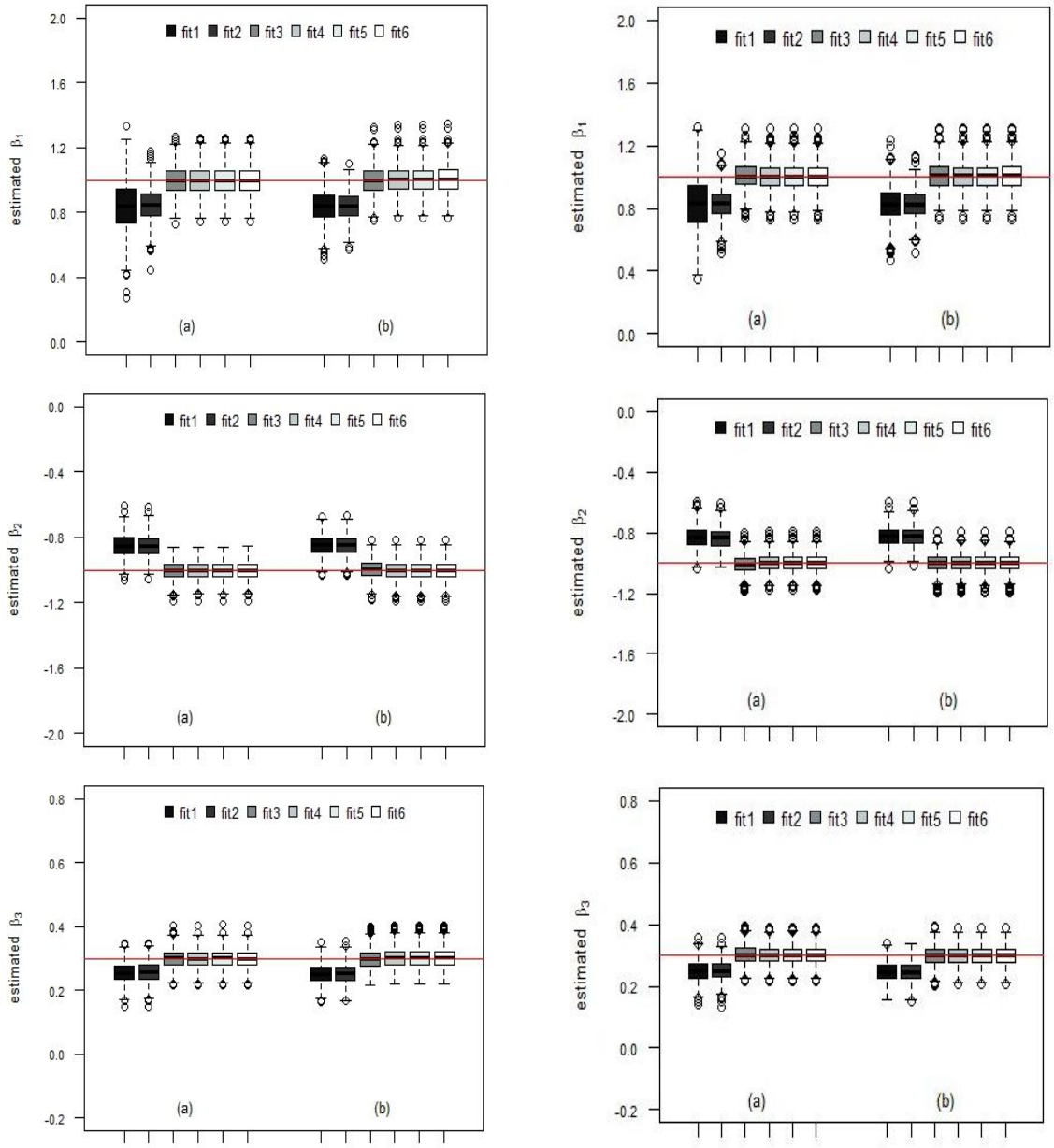

Figure 2: Fixed-effects parameter estimates in data generated by: the shared Gamma frailty model with $\theta = 0.5$(left column) and the shared log-normal frailty model with $\theta = 0.5$ (right column); cluster numbers of (a)20 and (b)60; 30% censoring level; true values are horizontal lines.

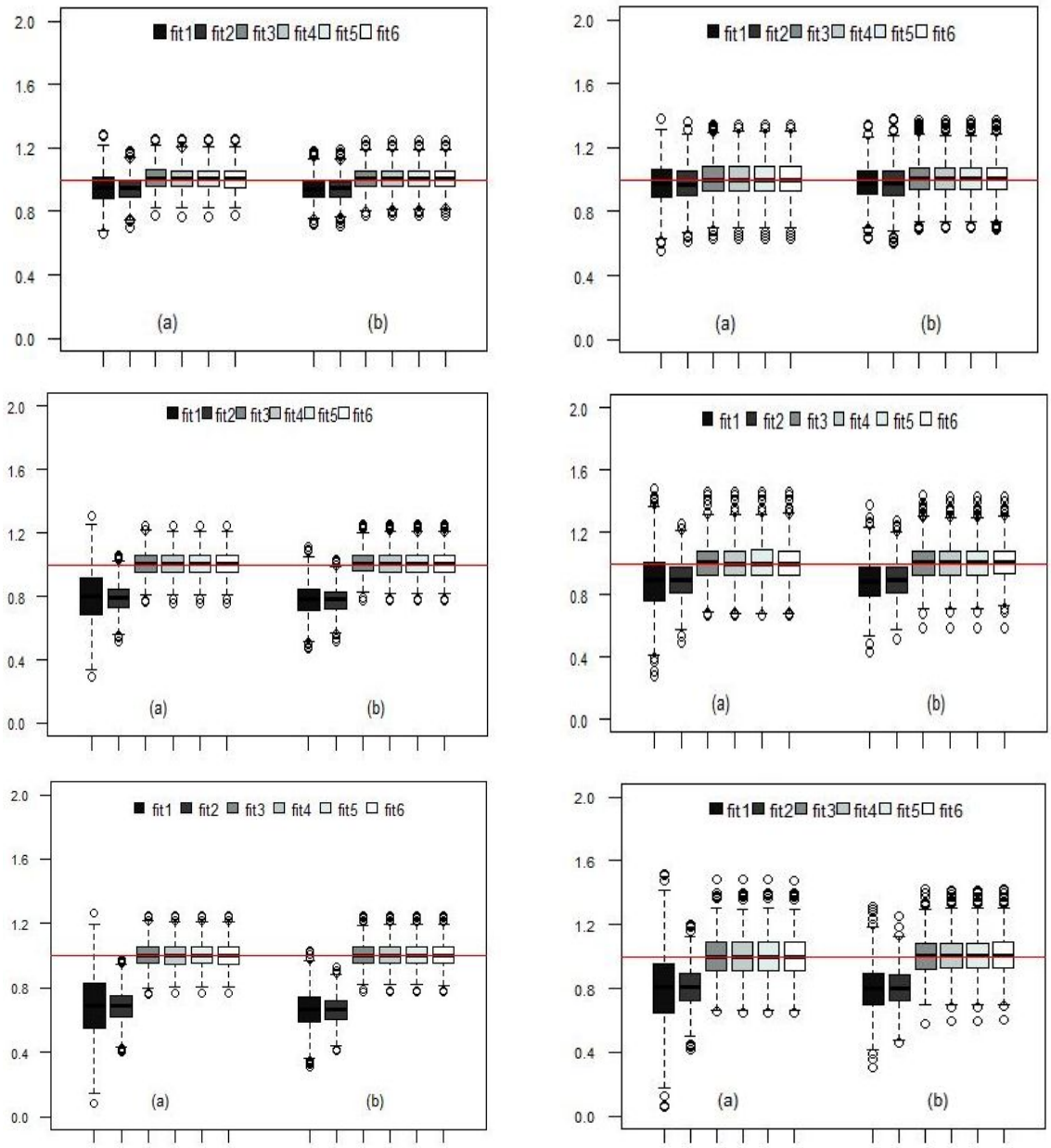

Figure 3: Fixed-effects parameter estimates in data generated by the shared log-normal frailty model with $\theta = 0.1$(1st line), $\theta = 0.5$(2nd line) and $\theta = 1$(3rd line); cluster numbers of (a)20 and (b)60; left column 30% right column 90% censoring; true values are horizontal lines.

parameter (variance of frailty) on the bias of fixed effects. It exhibits that for $\theta = 0$ (generating data without frailty effects), Cox's model provides better results in terms of bias and MAE than the estimates of any frailty approaches. When $\theta = 0.1$, the parameters of the fixed effect terms estimated by both Cox's and frailty modeling approaches are very small compared to the cases of $\theta = 0.5$. The right column of Fig 3 shows clearly the effect on the bias of various level of heterogeneity in gamma frailty survival data. referring on, for further increase in heterogeneity effects, it is obvious that the bias in the estimates obtained from Cox's model increases. Nonetheless, those from the frailty model decrease and converge towards true parameter.

On the other hand, one can remark that in the absence of heterogeneity and under heavy and very heavy censoring settings, the estimates of Cox model exhibit less bias and MAE than the shared frailty models.

b. Fixed-effects parameter estimates under shared log-normal frailty model

The distinctive feature here is that the frailties are assumed to be generated by log-normal distributions (the random effects are Gaussian distributed).

Keeping the same parameter setting as the precedent distribution, the six estimation approaches have shown similar results. In general, like the gamma frailty case, the bias increases with increasing the censoring rates or the variances of random effects θ, tables 5 and 6.

We conclude that 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 generated survaival data. Then, a misspecification of the distribution of the frailties has no significant impact on the estimates of . Even REML and MCEM approaches, proposed respectively by McGilchrist (1991) and Vaida et Xu (2000) to estimate Gaussian frailties, have shown good adjustment for the Gamma frailtty model. Conversely, Nielsen (1992) model which is specific to estimate Gamma farailty displayed good estimation results with Gaussian frailty. This finding coincides with the results obtained by Hirsch (2012). The detailed results of this simulation study are presented in table 4.

c. Random-effect parameter estimates

In some cases, the intention of the statistician or actuary is to obtain precise estimate of the heterogeneity parameter. For this reason, it is necessary in the present study to tell the difference that exists between the estimation of the frailty or random effects variance in the conventional Cox model (variance equal to zero) and in the shared frailty models (variance equal to 0,1; 0,5 and 1). The first model treats the data as it belongs to one group and then when the number of clusters increases, the variability of the variance rises too. The opposite effect is found in the second model where the variability of the variance estimates decreases considerably with increasing number of clusters.

The four shared frailty approaches produced on average similar estimates of the variance of Frailty/random effects at a comparable setting. Nevertheless, the variance estimates for the method of Ripatti and Palmgren (2000) showed smaller bias. More dissimilarity was seen between the samples composition (a) and (b).

d. Effect of size and sample composition

For moderate sample size ($n = 100$), the bias in the parameter estimates seemed to build up (not

shown). By increasing the data size (n = 1200), the bias in the fixed effect parameters, for all estimates approaches, seemed to decrease for the two samples composition (20 clusters of 60 members, 60 cluster of 20 members).

By altering the sample size composition without changing the total number n, the estimation results of each of the six frailty models are slightly different. The three previous tables give an overview of this difference. Then, we can conclude that there is a small significant effect of changing the sample size composition as long as the multiple of number group (g) and number of member (m) remains constant.
One can notice that, for all approaches, the bias in the fixed effects parameter estimates seemed to reduce with the increasing cluster size and decreasing the number of clusters. The situation is opposite in the random effect parameter estimates where the bias increase with reducing cluster number.

5 Discussion and Conclusions

Implementation of frailty models for the evaluation of lifetimes and then pension liabilities should be properly considered in actuarial mathematics and in insurance/pensions practice. In this paper, we compared frailty approaches in terms of the point estimates of the fixed effect regression coefficients. Through a simulation studies, many replications are carried with varied combinations and for each one, paramters are estimated via four commonly used estimation methods in the shared frailty models, all accessible in standard statistical software, for estimating fixed and random effects. These methods are from Nielsen et al. (1992), McGilchrist (1993) and Ripatti and Palmgren (2000) with the classical approach of not incorporating the frailty, that is, the approach of Cox (1972). The main reason for this choice was software availability and especially similarity between these frequentist estimation approaches as our purpose here isn't to test the performance of different estimation methods but to investigate the impact of taking into account a random effects in the Cox model.

It has been shown that, 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. In the current research, Cox approach has lost its usefulness when heterogeneity parameter increases more than 0.1. Conversely, the Cox model is still better than any frailty models when the heterogeneity parameter is close to zero.

Moreover, after evaluating the results obtained from the gamma frailty and Gaussian random effect, the fitted models (fit3, fit4, fit5, fit6) provide very similar values for the estimates of β even though the estimates of theta are slightly biased. On the one hand, we can say that each estimation approach has its own superiority as well as its very own drawbacks and then no unique approach is found as the superlative one. On the other hand, this is evident that the knowledge of the distribution of the frailty is not necessary as long as the sample size is reasonably well and the censoring level isn't heavy.

Another factor highly responsible for the increase of bias is the censoring rates of the data. This bias was, in almost all combinations, slightly higher as the amount of censoring increased in the

data. However, it can be noted that, in general, the frailty model behaved better in moderate or heavy censoring setting and was more flexible than the Cox model.

We also test whether there is any significant change if we model the composition of the sample size. We showed that the different allocations for sample affect slightly the estimation results. Then, we have to cofirm that for all models, the bias for the fixed-effect estimates seems to disappear with the increases in cluster size and for the random-effect estimates with the increasing of the number of clusters.

6 Annex

The four tables showed respectively the Bias and the MAE on the estimates results for a gamma shared frailty data.

Table 2: Simulation results: average value of bias over 1000 replications in data generated without and with shared Gamma frailty effect ($\theta = 0, 0.1$ and 0.5); cluster numbers of (a)20 and (b)60 with different censoring levels.

		Without censoring setting			Moderate censoring setting (30%)			Heavy censoring setting (90%)		
	Parameter	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
$\theta = 0$	**Cluster size = 1200, 20 clusters**									
	Fit 1	-1,16E-03	-1,51E-03	2,45E-04	-4,33E-03	-9,93E-04	3,56E-04	3,07E-04	2,20E-03	-1,27E-03
	Fit 2	2,84E-04	-2,10E-03	4,31E-04	-2,79E-03	-1,67E-03	5,29E-04	3,55E-03	-3,27E-04	-6,35E-05
	Fit 3	4,88E-05	-2,56E-03	5,36E-04	-2,86E-03	-2,22E-03	6,86E-04	1,39E-03	6,49E-04	-5,35E-04
	Fit 4	7,78E-04	-3,27E-03	7,37E-04	-2,28E-03	-2,74E-03	8,50E-04	1,55E-03	6,57E-05	-3,33E-04
	Fit 5	4,59E-04	-2,92E-03	6,31E-04	-2,54E-03	-2,48E-03	7,74E-04	1,38E-03	2,22E-04	-3,73E-04
	Fit 6	4,05E-04	-2,92E-03	6,38E-04	-2,53E-03	-2,49E-03	7,87E-04	2,00E-03	-8,14E-05	-3,13E-04
	Cluster size = 1200, 60 clusters									
	Fit 1	2,57E-03	-3,74E-03	-3,92E-04	2,16E-03	-4,48E-03	-2,64E-04	1,68E-02	-8,18E-03	9,33E-05
	Fit 2	2,81E-03	-4,38E-03	-2,13E-04	3,09E-03	-5,13E-03	-3,42E-04	1,62E-02	-1,03E-02	8,51E-04
	Fit 3	3,96E-03	-5,14E-03	2,50E-05	3,97E-03	-6,14E-03	2,73E-04	1,88E-02	-1,15E-02	1,17E-03
	Fit 4	5,51E-03	-6,64E-03	4,99E-04	5,42E-03	-7,38E-03	6,52E-04	1,91E-02	-1,25E-02	1,44E-03
	Fit 5	4,95E-03	-6,11E-03	3,36E-04	4,88E-03	-6,89E-03	5,00E-04	1,89E-02	-1,22E-02	1,35E-03
	Fit 6	5,92E-03	-7,05E-03	6,20E-04	6,08E-03	-7,96E-03	8,37E-04	2,09E-02	-1,42E-02	2,09E-03
$\theta = 0,1$	**Cluster size = 1200, 20 clusters**									
	Fit 1	-4,90E-02	5,16E-02	-1,56E-02	-4,03E-02	3,96E-02	-1,23E-02	2,76E-03	5,51E-03	-1,16E-03
	Fit 2	-5,14E-02	4,92E-02	-1,50E-02	-4,21E-02	3,79E-02	-1,19E-02	1,16E-02	5,75E-04	8,75E-04
	Fit 3	6,62E-03	-5,23E-03	5,17E-04	3,01E-03	-3,66E-03	3,26E-04	1,24E-02	-7,72E-03	2,82E-03
	Fit 4	3,70E-03	-2,35E-03	-3,03E-04	1,69E-03	-2,37E-03	-4,41E-05	1,16E-02	-8,55E-03	3,04E-03
	Fit 5	4,13E-03	-2,78E-03	-1,81E-04	2,11E-03	-2,80E-03	8,13E-05	1,17E-02	-8,81E-03	3,11E-03
	Fit 6	3,49E-03	-2,17E-03	-3,56E-04	1,72E-03	-2,21E-03	-8,21E-05	1,37E-02	-8,75E-03	3,17E-03
	Cluster size = 1200, 60 clusters									
	Fit 1	-5,55E-02	5,49E-02	-1,63E-02	2,16E-03	-4,48E-03	-2,64E-04	3,41E-03	1,50E-02	-7,80E-03
	Fit 2	-5,35E-03	5,38E-02	-1,20E-02	3,09E-03	-5,13E-03	-3,42E-05	1,22E-02	1,08E-02	-5,97E-03
	Fit 3	1,03E-03	-8,37E-04	-4,59E-04	3,97E-03	-6,14E-03	2,73E-04	1,27E-02	1,12E-03	-3,81E-03
	Fit 4	2,62E-03	-2,23E-03	-2,14E-05	5,42E-03	-7,38E-03	6,52E-04	1,10E-02	5,90E-04	-3,76E-03
	Fit 5	2,21E-03	-1,84E-03	-1,34E-04	4,88E-03	-6,89E-03	5,00E-04	1,08E-02	7,94E-04	-3,82E-03
	Fit 6	2,13E-03	-1,74E-03	-1,68E-04	6,08E-03	-7,96E-03	8,37E-04	1,45E-02	-1,26E-03	-3,08E-03
$\theta = 0,5$	**Cluster size = 1200, 20 clusters**									
	Fit 1	-1,70E-01	1,60E-01	-4,77E-02	-1,67E-01	1,54E-01	-4,58E-02	-7,34E-02	8,54E-02	-2,98E-02
	Fit 2	-1,60E-01	1,56E-01	-4,65E-02	-1,55E-01	1,49E-01	-4,43E-02	-5,02E-02	7,36E-02	-2,35E-02
	Fit 3	2,05E-03	-4,09E-03	8,38E-04	1,52E-04	-3,56E-03	8,64E-04	-4,40E-03	7,91E-03	-3,89E-03
	Fit 4	-1,19E-03	-1,22E-03	1,13E-06	-2,70E-03	-1,10E-03	1,43E-04	-4,23E-03	5,14E-03	-3,12E-03
	Fit 5	-6,90E-04	-1,67E-03	1,35E-04	-2,07E-03	-1,67E-03	3,15E-04	-2,41E-03	3,32E-03	-2,48E-03
	Fit 6	-1,19E-03	-1,17E-03	-4,63E-06	-2,74E-03	-1,05E-03	1,25E-04	1,59E-03	3,20E-03	-2,37E-03
	Cluster size = 1200, 60 clusters									
	Fit 1	-1,65E-01	1,65E-01	-5,06E-02	-1,62E-01	1,58E-01	-4,85E-02	-6,76E-02	8,53E-02	-2,88E-02
	Fit 2	-1,68E-01	1,63E-01	-5,02E-02	-1,63E-01	1,57E-01	-4,82E-02	-6,91E-02	8,24E-02	-2,76E-02
	Fit 3	1,67E-03	-4,06E-04	-9,03E-02	-7,72E-04	2,07E-03	-1,29E-03	-8,22E-03	9,58E-03	-4,32E-03
	Fit 4	5,11E-03	-3,94E-03	1,47E-04	3,96E-03	-3,08E-03	2,10E-04	-1,02E-02	6,00E-03	-4,18E-03
	Fit 5	4,90E-03	-3,73E-03	8,45E-05	4,06E-03	-3,18E-03	2,40E-04	-6,72E-03	2,40E-03	-3,03E-03
	Fit 6	4,90E-03	-3,80E-03	1,04E-04	4,26E-03	-3,21E-03	2,31E-04	6,32E-03	-1,99E-03	-1,34E-03

Table 3: Simulation results: average value of bias over 1000 replications in data generated with shared log-normal frailty effect ($\theta = 0, 1; 0.5$ and 1); cluster numbers of (a)20 and (b)60 with different censoring levels.

		Without censoring setting			Moderate censoring setting (30%)			Heavy censoring setting (90%)		
	Parameter	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
$\theta = 0,1$	**Cluster size = 1200, 20 clusters**									
	Fit 1	-5,31E-02	5,69E-02	-1,64E-02	-4,24E-02	4,11E-02	-1,13E-02	-1,22E-03	3,80E-03	1,29E-05
	Fit 2	-5,57E-02	5,40E-02	-1,56E-03	-4,26E-02	3,88E-02	-1,08E-02	-4,98E-03	2,07E-03	7,02E-04
	Fit 3	7,04E-03	-5,15E+00	2,12E-03	3,79E-03	-3,49E-03	2,20E-03	3,42E-03	-6,76E-03	3,66E-03
	Fit 4	3,29E-03	-1,48E-03	1,03E-03	1,91E-03	-1,84E-03	1,71E-03	1,61E-03	-7,15E-03	3,69E-03
	Fit 5	3,71E-03	-1,92E-03	1,16E-03	2,34E-03	-2,28E-03	1,84E-03	1,46E-03	-7,30E-03	3,75E-03
	Fit 6	3,11E-03	-1,30E-03	9,76E-04	1,83E-03	-1,68E-03	1,66E-03	3,66E-03	-7,22E-03	3,79E-03
	Cluster size = 1200, 60 clusters									
	Fit 1	-5,72E-02	6,01E-02	-1,87E-02	-4,27E-02	4,49E-02	-1,44E-02	1,84E-02	7,45E-03	-4,88E-03
	Fit 2	-5,91E-02	5,88E-02	-1,84E-02	-4,18E-02	4,38E-02	-1,41E-02	2,44E-02	3,92E-03	-3,22E-03
	Fit 3	4,70E-03	-1,21E-03	-5,58E-04	5,04E-03	-7,16E-04	-2,49E-04	2,57E-02	-3,25E-03	-1,21E-03
	Fit 4	4,94E-03	-1,65E-03	-4,22E-04	4,67E-03	-9,49E-04	-1,98E-04	2,45E-02	-3,92E-03	-1,09E-03
	Fit 5	4,55E-03	-1,27E-03	-5,34E-04	4,45E-03	-7,50E-04	-2,59E-04	2,43E-02	-3,67E-03	-1,18E-03
	Fit 6	4,49E-03	-1,18E-03	-5,67E-04	4,98E-03	-8,29E-04	-2,25E-04	2,77E-02	-5,59E-03	-4,96E-04
$\theta = 0,5$	**Cluster size = 1200, 20 clusters**									
	Fit 1	-2,04E-01	2,08E-01	-6,14E-02	-1,71E-01	1,68E-01	-4,99E-02	-2,10E-02	-1,20E-02	2,30E-01
	Fit 2	-2,07E-01	2,01E-01	-5,93E-02	-1,70E-01	1,62E-01	-4,84E-02	-2,83E-02	-1,09E-02	2,01E-01
	Fit 3	7,74E-03	-6,18E-03	2,51E-03	6,39E-03	-5,82E-03	2,59E-03	7,38E-03	5,73E-03	1,97E-01
	Fit 4	2,67E-03	-1,18E-03	1,04E-03	2,01E-03	-1,59E-03	1,33E-03	2,30E-03	5,30E-03	1,97E-01
	Fit 5	3,22E-03	-1,74E-03	1,21E-03	2,67E-03	-2,25E-03	1,52E-03	3,22E-03	5,83E-03	1,97E-01
	Fit 6	2,69E-03	-1,17E-03	1,04E-03	2,02E-03	-1,60E-03	1,34E-03	7,81E-03	6,00E-03	1,98E-01
	Cluster size = 1200, 60 clusters									
	Fit 1	-2,17E-01	2,19E-01	-6,57E-02	-1,76E-01	1,76E-01	-5,41E-02	-1,83E-02	5,39E-02	-1,92E-02
	Fit 2	-2,22E-02	2,16E-01	-6,50E-02	-1,73E-01	1,74E-01	-5,36E-02	-9,94E-03	4,92E-02	-1,72E-02
	Fit 3	9,31E-03	-6,04E-03	7,64E-04	6,92E-03	-1,10E-03	-1,77E-04	2,04E-02	-6,13E-04	-6,35E-04
	Fit 4	4,22E-03	-1,29E-03	-5,68E-04	5,25E-03	-2,64E-05	-5,21E-04	1,03E-02	3,37E-03	-2,43E-03
	Fit 5	4,06E-03	-1,13E-03	-6,15E-04	5,68E-03	-4,42E-04	-3,96E-04	1,17E-02	1,40E-03	-1,78E-03
	Fit 6	4,07E-03	-1,16E-03	-5,97E-04	6,05E-03	-4,17E-04	-4,01E-04	2,32E-02	-2,64E-03	-2,35E-04
$\theta = 1$	**Cluster size = 1200, 20 clusters**									
	Fit 1	-3,11E-01	3,16E-01	-9,36E-02	-2,74E-01	2,70E-01	-8,04E-02	-4,65E-02	9,61E-02	-2,65E-02
	Fit 2	-3,14E-01	3,06E-01	-9,09E-02	-2,72E-01	2,62E-01	-7,84E-02	5,58E-02	8,93E-02	-2,42E-02
	Fit 3	4,30E-03	-2,98E-03	1,54E-03	1,69E-03	-3,18E-03	1,72E-03	1,07E-02	-4,25E-03	6,13E-03
	Fit 4	2,26E-03	-1,12E-03	1,01E-03	3,92E-04	-2,05E-03	1,39E-03	6,65E-03	-4,14E-03	5,94E-03
	Fit 5	2,90E-03	-1,75E-03	1,20E-03	1,18E-03	-2,83E-03	1,62E-03	8,45E-03	-6,81E-03	6,83E-03
	Fit 6	2,32E-03	-1,12E-03	1,01E-03	4,65E-04	-2,10E-03	1,41E-03	1,27E-02	-6,61E-03	6,90E-03
	Cluster size = 1200, 60 clusters									
	Fit 1	-3,31E-01	3,31E-03	-9,93E-02	-2,80E-01	2,79E-01	-8,54E-02	-6,14E-02	1,03E-01	-3,65E-02
	Fit 2	-3,36E-01	3,28E-03	-9,84E-02	-2,78E-01	2,77E-01	-8,47E-02	-5,22E-02	9,75E-02	-3,42E-02
	Fit 3	5,53E-03	-2,42E-03	-3,91E-04	3,34E-03	1,46E-03	-1,24E-03	1,40E-02	9,94E-04	-2,65E-03
	Fit 4	3,71E-03	-1,05E-03	-6,74E-04	4,27E-03	3,40E-05	-8,05E-04	1,73E-03	6,12E-03	-4,97E-03
	Fit 5	3,54E-03	-8,78E-04	-7,23E-04	4,99E-03	-6,69E-04	-5,95E-04	6,85E-03	5,33E-04	-3,12E-03
	Fit 6	3,67E-03	-9,42E-04	-7,01E-04	5,12E-03	-5,17E-04	-6,35E-04	2,11E-02	-3,96E-03	-1,30E-03

Table 4: Simulation results: average value of MAE over 1000 replications in data generated without and with shared Gamma frailty effect ($\theta = 0; 0.1$ and 0.5); cluster numbers of (a)20 and (b)60 with different censoring levels.

		Without censoring setting			Moderate censoring setting (30%)			Heavy censoring setting (90%)		
	Parameter	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
$\theta = 0$	**Cluster size = 1200, 20 clusters**									
	Fit 1	4,94E-02	3,82E-02	1,95E-02	5,94E-02	4,31E-02	2,25E-02	1,81E-01	9,82E-02	5,24E-02
	Fit 2	5,89E-02	3,83E-02	1,95E-02	6,96E-02	4,32E-02	2,26E-02	2,06E-01	1,02E-01	5,42E-02
	Fit 3	4,96E-02	3,83E-02	1,95E-02	5,98E-02	4,33E-02	2,25E-02	1,82E-01	9,87E-02	5,27E-02
	Fit 4	4,97E-02	3,84E-02	1,95E-02	5,99E-02	4,33E-02	2,25E-02	1,82E-01	9,89E-02	5,27E-02
	Fit 5	4,96E-02	3,83E-02	1,95E-02	5,99E-02	4,33E-02	2,25E-02	1,82E-01	9,89E-02	5,27E-02
	Fit 6	4,96E-02	3,83E-02	1,95E-02	5,98E-02	4,32E-02	2,25E-02	1,82E-01	9,89E-02	5,27E-02
	Cluster size = 1200, 60 clusters									
	Fit 1	5,04E-02	3,77E-02	2,03E-02	6,03E-02	4,32E-02	2,31E-02	1,81E-01	9,87E-02	5,26E-02
	Fit 2	5,92E-02	3,77E-02	2,03E-02	6,89E-02	4,36E-02	2,32E-02	2,02E-01	1,04E-01	5,40E-02
	Fit 3	5,06E-02	3,79E-02	2,03E-02	6,05E-02	4,35E-02	2,31E-02	1,81E-01	9,99E-02	5,28E-02
	Fit 4	5,07E-02	3,81E-02	2,03E-02	6,08E-02	4,37E-02	2,31E-02	1,81E-01	1,00E-01	5,28E-02
	Fit 5	5,07E-02	3,81E-02	2,03E-02	6,07E-02	4,37E-02	2,31E-02	1,81E-01	1,00E-01	5,28E-02
	Fit 6	5,07E-02	3,81E-02	2,03E-02	6,08E-02	4,38E-02	2,31E-02	1,81E-01	1,01E-01	5,30E-02
$\theta = 0,1$	**Cluster size = 1200, 20 clusters**									
	Fit 1	8,53E-02	5,86E-02	2,39E-02	8,82E-02	5,35E-02	2,61E-02	1,89E-01	9,82E-02	5,35E-02
	Fit 2	7,50E-02	5,68E-02	2,37E-02	7,82E-02	5,26E-02	2,60E-02	1,99E-01	1,01E-01	5,47E-02
	Fit 3	5,91E-02	3,81E-02	2,07E-02	6,74E-02	4,24E-02	2,38E-02	1,87E-01	9,96E-02	5,44E-02
	Fit 4	5,84E-02	3,77E-02	2,05E-02	6,69E-02	4,22E-02	2,38E-02	1,86E-01	9,97E-02	5,44E-02
	Fit 5	5,85E-02	3,77E-02	2,05E-02	6,69E-02	4,22E-02	2,38E-02	1,86E-01	9,98E-02	5,44E-02
	Fit 6	5,85E-02	3,77E-02	2,05E-02	6,69E-02	4,21E-02	2,38E-02	1,86E-01	9,96E-02	5,44E-02
	Cluster size = 1200, 60 clusters									
	Fit 1	7,54E-02	6,07E-02	2,49E-02	6,03E-02	4,32E-02	2,31E-02	1,86E-01	1,00E-01	5,46E-02
	Fit 2	7,57E-02	5,99E-02	2,47E-02	6,89E-02	4,36E-02	2,32E-02	2,06E-01	1,03E-01	5,55E-02
	Fit 3	5,64E-02	3,78E-02	2,10E-02	6,05E-02	4,35E-02	2,31E-02	1,87E-01	1,02E-01	5,51E-02
	Fit 4	5,63E-02	3,78E-02	2,10E-02	6,08E-02	4,37E-02	2,31E-02	1,87E-01	1,02E-01	5,50E-02
	Fit 5	5,63E-02	3,78E-02	2,09E-02	6,07E-02	4,37E-02	2,31E-02	1,87E-01	1,02E-01	5,50E-02
	Fit 6	5,63E-02	3,78E-02	2,09E-02	6,08E-02	4,38E-02	2,314-02	1,87E-01	1,02E-01	5,52E-02
$\theta = 0,5$	**Cluster size = 1200, 20 clusters**									
	Fit 1	1,81E-01	1,60E-01	4,88E-02	1,83E-01	1,54E-01	4,75E-02	2,57E-01	1,30E-01	6,06E-02
	Fit 2	1,62E-01	1,56E-01	4,76E-02	1,59E-01	1,50E-01	4,62E-02	2,13E-01	1,27E-01	6,08E-02
	Fit 3	5,94E-02	3,89E-02	1,98E-02	6,56E-02	4,18E-02	2,21E-02	1,94E-01	1,03E-01	5,54E-02
	Fit 4	5,88E-02	3,88E-02	1,98E-02	6,50E-02	4,17E-02	2,21E-02	1,91E-01	1,03E-01	5,55E-02
	Fit 5	5,88E-02	3,88E-02	1,98E-02	6,51E-02	4,17E-02	2,21E-02	1,91E-01	1,03E-01	5,57E-02
	Fit 6	5,87E-02	3,88E-02	1,98E-02	6,51E-02	4,17E-02	2,20E-02	1,92E-01	1,03E-01	5,57E-02
	Cluster size = 1200, 60 clusters									
	Fit 1	1,68E-01	1,65E-01	5,12E-02	1,67E-01	1,58E-01	4,97E-02	2,17E-01	1,24E-01	5,71E-02
	Fit 2	1,68E-01	1,63E-01	5,08E-02	1,65E-01	1,57E-01	4,94E-02	2,16E-01	1,26E-01	5,78E-02
	Fit 3	6,02E-01	3,93E-02	2,14E-02	6,63E-02	4,32E-02	2,35E-02	2,00E-01	1,06E-01	5,26E-02
	Fit 4	5,94E-01	3,93E-02	2,14E-02	6,52E-02	4,33E-02	2,35E-02	1,94E-01	1,05E-01	5,26E-02
	Fit 5	5,94E-01	3,93E-02	2,14E-02	6,52E-02	4,34E-02	2,35E-02	1,94E-01	1,05E-01	5,28E-02
	Fit 6	5,93E-01	3,93E-02	2,14E-02	6,53E-02	4,33E-02	2,35E-02	1,96E-01	1,06E-01	5,31E-02

Table 5: Simulation results: average value of MAE over 1000 replications in data generated without and with shared log-normal frailty effect ($\theta = 0,1; 0.5$ and 1); cluster numbers of (a)20 and (b)60 with different censoring levels.

		Without censoring setting			Moderate censoring setting (30%)			Heavy censoring setting (90%)		
	Parameter	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
$\theta = 0,1$	**Cluster size = 1200, 20 clusters**									
	Fit 1	9,14E-02	6,47E-02	2,45E-02	9,07E-02	5,85E-02	2,52E-02	1,95E-01	1,00E-01	5,41E-02
	Fit 2	7,58E-02	6,25E-02	2,41E-02	7,52E-02	5,73E-02	2,49E-02	2,03E-01	1,01E-01	5,53E-02
	Fit 3	5,89E-02	4,00E-02	2,04E-02	6,57E-02	4,54E-02	2,33E-02	1,92E-01	1,01E-01	5,45E-02
	Fit 4	5,80E-02	3,97E-02	2,03E-02	6,53E-02	4,51E-02	2,32E-02	1,91E-01	1,01E-01	5,44E-02
	Fit 5	5,81E-02	3,97E-02	2,03E-02	6,53E-02	4,52E-02	2,32E-02	1,92E-01	1,01E-01	5,45E-02
	Fit 6	5,80E-02	3,97E-02	2,03E-02	6,53E-02	4,51E-02	2,32E-02	1,92E-01	1,01E-01	5,45E-02
	Cluster size = 1200, 60 clusters									
	Fit 1	7,72E-02	6,50E-02	2,61E-02	7,45E-02	5,75E-02	2,66E-02	1,77E-01	9,96E-02	5,52E-02
	Fit 2	7,56E-02	6,41E-02	2,60E-02	7,31E-02	5,71E-02	2,66E-02	2,00E-01	1,02E-01	5,67E-02
	Fit 3	5,63E-02	3,78E-02	2,14E-02	6,35E-02	4,47E-02	2,43E-02	1,79E-01	1,01E-01	5,52E-02
	Fit 4	5,64E-02	3,78E-02	2,14E-02	6,36E-02	4,46E-02	2,44E-02	1,79E-01	1,01E-01	5,53E-02
	Fit 5	5,63E-02	3,78E-02	2,14E-02	6,36E-02	4,46E-02	2,44E-02	1,79E-01	1,01E-01	5,53E-02
	Fit 6	5,64E-02	3,78E-02	2,14E-02	6,37E-02	4,47E-02	2,44E-02	1,80E-01	1,01E-01	5,54E-02
$\theta = 0,5$	**Cluster size = 1200, 20 clusters**									
	Fit 1	2,19E-01	2,08E-01	6,17E-02	1,97E-01	1,69E-01	5,16E-02	2,30E-01	1,11E-01	5,72E-02
	Fit 2	2,08E-01	2,01E-01	5,97E-02	1,72E-01	1,62E-01	5,03E-02	2,01E-01	1,07E-01	5,65E-02
	Fit 3	5,98E-02	4,02E-02	2,04E-02	6,68E-02	4,54E-02	2,34E-02	1,97E-01	1,02E-01	5,51E-02
	Fit 4	5,93E-02	3,97E-02	2,03E-02	6,64E-02	4,50E-02	2,32E-02	1,97E-01	1,02E-01	5,49E-02
	Fit 5	5,93E-02	3,97E-02	2,03E-02	6,65E-02	4,50E-02	2,33E-02	1,97E-01	1,02E-01	5,51E-02
	Fit 6	5,92E-02	3,97E-02	2,03E-02	6,65E-02	4,50E-02	2,32E-02	1,98E-01	1,02E-01	5,51E-02
	Cluster size = 1200, 60 clusters									
	Fit 1	2,19E-01	2,19E-01	6,59E-02	1,80E-01	1,76E-01	5,53E-02	1,90E-01	1,09E-01	5,89E-02
	Fit 2	2,22E-01	2,16E-01	6,52E-02	1,75E-01	1,74E-01	5,47E-02	1,99E-01	1,11E-01	5,90E-02
	Fit 3	5,95E-02	3,77E-02	2,15E-02	6,80E-02	4,45E-02	2,47E-02	1,90E-01	1,02E-01	5,82E-02
	Fit 4	5,89E-02	3,77E-02	,15E-02	6,76E-02	4,45E-02	2,48E-02	1,89E-01	1,01E-01	5,76E-02
	Fit 5	5,89E-02	3,77E-02	2,15E-02	6,76E-02	4,45E-02	2,48E-02	1,89E-01	1,02E-01	5,77E-02
	Fit 6	5,90E-02	3,77E-02	2,15E-02	6,77E-02	4,44E-02	2,48E-02	1,91E-01	1,02E-01	5,80E-02
$\theta = 1$	**Cluster size = 1200, 20 clusters**									
	Fit 1	3,19E-01	3,16E-01	9,36E-02	2,90E-01	2,70E-01	8,06E-02	2,78E-01	1,33E-01	6,19E-02
	Fit 2	3,14E-01	3,06E-01	9,09E-02	2,72E-01	2,62E-01	7,85E-02	2,06E-01	1,28E-01	6,10E-02
	Fit 3	5,97E-02	3,98E-01	2,02E-02	6,70E-02	4,49E-02	2,36E-02	1,98E-01	1,01E-01	5,52E-02
	Fit 4	5,96E-02	3,97E-01	2,02E-02	6,70E-02	4,48E-02	2,35E-02	1,97E-01	1,01E-01	5,51E-02
	Fit 5	5,97E-02	3,97E-01	2,02E-02	6,70E-02	4,48E-02	2,36E-02	1,97E-01	1,02E-01	5,53E-02
	Fit 6	5,96E-02	3,97E-02	2,02E-02	6,70E-02	4,48E-02	2,35E-02	1,97E-01	1,02E-01	5,53E-02
	Cluster size = 1200, 60 clusters									
	Fit 1	3,31E-01	3,31E-01	9,93E-02	2,81E-01	2,79E-01	8,54E-02	2,11E-01	1,35E-01	6,57E-02
	Fit 2	3,36E-01	3,28E-01	9,84E-02	2,78E-01	2,77E-01	8,47E-02	2,06E-01	1,35E-01	6,55E-02
	Fit 3	5,93E-02	3,79E-02	2,14E-02	6,77E-02	4,35E-02	2,44E-02	1,93E-01	1,04E-01	6,04E-02
	Fit 4	5,93E-02	3,77E-02	2,14E-02	6,77E-02	4,35E-02	2,45E-02	1,91E-01	1,04E-01	5,97E-02
	Fit 5	5,92E-02	3,77E-02	2,14E-02	6,78E-02	4,36E-02	2,45E-02	1,92E-01	1,04E-01	6,00E-02
	Fit 6	5,93E-02	3,77E-02	2,14E-02	6,79E-02	4,36E-02	2,45E-02	1,94E-01	1,05E-01	6,03E-02

References

Abbring, J. & van den Berg, G. (2007),*The unobserved heterogeneity distribution in duration analysis*,Biometrika, 94, pp. 87-99.

Andersen, P. K, Ø . Borgan, R. D. Gill & N. Keiding (1993), *Statistical Models Based on Counting Processes*, Berlin: Springer.

Barbi, E., Caselli G.& Vallin J. (2003), *Hétérogénéité des générations et âge extrême de la vie* ,Institut Nationale d'Études Démographiques / Population, Vol. 58, No. 2006(1).

Chamberlain G. (1985), *Heterogeneity, omitted variable Bias, and duration dependence*, In J. J. Heckman B. Singer(Eds), Longitudinal analysis of labour market data, pp. 3-38, Cambridge University Press.

Clayton, D.G. (1978), *A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence*, Biometrika,65, pp. 141-151.

Cox, D.R. (1972), *Models and Life-Tables. Journal of the Royal Statistical*, Society B, 34, pp. 187-220.

Duchateau, L., Janssen, P., Lindsey, P., Legrand, C., Nguti, R. Sylvester, R. (2002), *The shared frailty model and the power for heterogeneity tests in multicenter trials*, Computational Statistics Data Analysis, 40, pp. 603-630.

Duchateau, L., Janssen, P., Kezic, I. Fortpied, C. (2003), *Evolution of recurrent asthma event rate over time in frailty models*, Journal of the Royal Statistical Society, vol. 52, pp. 355- 363.

Duchateau, L. Janssen, P. (2004), *Penalized partial likelihood for frailties and smoothing splines in time to first insemination models for dairy cows*, Biometrics, vol. 60, pp. 608-614.

Gail, M. H., S. Wieand, & S. Piantadosi. (1984), *Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates*, Biometrika, vol. 71, pp. 431-444.

Hirsch, K. Wienke A. (2012), *Software for semiparametric shared gamma and log-normal frailty models: An overview*, Computer Methods and Programs in Biomedicine, pp. 582-597.

Hougaard, P. (1986b), *Survival models for heterogeneous populations derived from stable distributions*, Biometrika, vol.73(2), pp. 387-396.

Hougaard P. (1995), *Frailty models for survival data*, Lifetime Data Analysis, vol. 1, pp. 255-273.

Hougaard P. (2000), *Analysis of Multivariate Survival Data*, Springer, New York.

Klein, J. (1992), *Semiparametric estimation of random effects using the Cox model based on the EM algorithm*, Biometrics, vol. 48, pp.795-806.

Nielsen, G. G., D., G. R., Andersen, P. K. & Sorensen, T. I. A. (1992), *A counting process approach to maximum likelihood estimation in frailty models*, Scandinavian Journal of Statistics, Theory and Applications, vol. 19(1), pp. 25-43.

Planchet, F. & Thérond P. (2006), *Modèles de Durée - Applications actuarielles*, Economica

Therneau, TM. & Grambsch, PM. (1998), *Penalised Cox models and frailty*, S-plus documentation.

Therneau, T. & Grambsch, P. (2000), *Modelling Survival Data: Extending the Cox Model*, Springer-Verlag, New York.

Trussel, J. & Rodrigues, G. (1990), *Heterogeneity in demographic research*, In: Adams, J., Lam, D. Smouse, P., (Ed.) Convergent Issues in Genetics and Demography. New York NY, Oxford University Press. pp. 111-132.

Vaupel, J.W. and Yashin, A.I. (1985, *Heterogeneity's Ruses: Some Surprising Effects of Selection on Population Dynamics*, American Statistician, vol. 39, pp. 176-185.

Vaupel J. W., Manton K. G. & Stallard E. (1979), *The impact of heterogeneity in individual frailty on the dynamics of mortality*, Demography, Vol. 16, No. 3.

Vaupel J. W. (2002), *Life Expectancy at Current Rates vs. Current Conditions: A Reflexion Stimulated by Bongaarts and Feeney's "How Long Do We Live?"*,Demographic Research, Vol. 7, Art. 8.