

# Parameter Estimation of Shared Frailty Models Based on Particle Swarm Optimization

Oykum Esra Askin<sup>1</sup>, Deniz Inan<sup>2</sup> & Ali Hakan Buyuklu<sup>1</sup>

<sup>1</sup> Department of Statistics, Yildiz Technical University, Istanbul, Turkey

<sup>2</sup> Department of Statistics, Marmara University, Istanbul, Turkey

Correspondence: Oykum Esra Askin, Department of Statistics, Yildiz Technical University, Istanbul, Turkey. E-mail: oykumesra@gmail.com

Received: October 12, 2016 Accepted: December 5, 2016 Online Published: December 13, 2016

doi:10.5539/ijsp.v6n1p48 URL: <https://doi.org/10.5539/ijsp.v6n1p48>

## Abstract

Standard survival techniques such as proportional hazards model are suffering from the unobserved heterogeneity. Frailty models provide an alternative way in order to account for heterogeneity caused by unobservable risk factors. Although vast studies have been done on estimation procedures, Evolutionary Algorithms (EAs) haven't received much attention in frailty studies. In this paper, we investigate the estimation performance of maximum likelihood estimation (MLE) via Particle Swarm Optimization (PSO) in modelling multivariate survival data with shared gamma frailty. Simulation studies and real data application are performed in order to assess the performance of MLE via PSO, quasi-Newton and conjugate gradient method.

**Keywords:** frailty models; unobserved heterogeneity; particle swarm optimization; maximum likelihood estimation

## 1. Introduction

The general concept of survival models is to investigate the impact of risk factors on individual failure times. However, it's impossible to include all important factors into the model. This is due to the lack of information on individual level or the researcher may not be aware of the importance of the factor, or even the existence of it (Hougaard, 1991). Several authors studied the consequences of ignoring the existence of unobserved heterogeneity caused by unobservable risk factors such as Aalen (1988), Lancaster (1979, 1990), Henderson and Oman (1999) and Van den Berg (2001).

In order to account for unobserved heterogeneity, the frailty term was first introduced by (Vaupel, Manton, & Stallard, 1979) for univariate survival data. They indicated that the individual hazard is the product of two terms: an individual level frailty and a baseline hazard function. The multivariate generalization was then introduced by Clayton (1978). The proposed model was a random effect model, which is an extension of a proportional hazards (PH) approach. In a shared frailty model, lifetimes of a group of observations in the same cluster (e.g., individuals in a family) are related to each other (Aalen, Borgan, & Gjessing, 2008). Each cluster shares the same level of frailty. In other words, the observations in a cluster share the same unobservable risks such as genetic structures. The variance of this common frailty is a measure of dependence among lifetimes within a cluster.

There exists a huge amount of literature on estimating parameters according to information of the distributional form of the baseline hazard function. If a parametric form is not assumed for the baseline hazard, different estimation methods are proposed for the semi-parametric frailty models (Clayton & Cuzick, 1985; Gill, 1985; Clayton, 1991; McGilchrist & Aisbett, 1991; Klein, 1992; Nielsen, Gill, Andersen, & Srensen, 1992; McGilchrist, 1993; Rondeau, Commenges, & Joly, 2003; Therneau, Grambsch, & Pankratz, 2003). Also, important works relevant with Bayesian framework were offered by Sinha and Dey (1997), Ibrahim, Chen and Sinha (2001) and Hanagal and Pandey (2015). On the other hand, when the baseline hazard is represented parametrically, model parameters can be estimated with MLE via derivative-based methods such as Newton Raphson, quasi-Newton and conjugate gradient. In spite of having useful properties, these techniques have some drawbacks. The main drawback is that the starting point should be chosen with a value close to a global optimum because they can easily get stuck in a local optimum at multiple solution space. The poor choice of an initial value for any model parameter could lead to finding an estimate that is far away from the optimal solution. When the number of model parameters is too large, choosing suitable initial values becomes harder and researchers are never sure whether the resulting solution is a local or global. Also, biased parameter estimates can be obtained due to the heavy censoring or small sample size.

PSO is a member of stochastic search family inspired by some natural phenomena to solve complicated and high-

dimensional optimization problems. It is a global search strategy that has properties of being robust, fast converge and easy implemented for linear/non-linear functions. In contrast with the derivative-based methods, it is less sensitive to the selection of initial parameter values. Besides, it gives reliable estimates even though the number of parameters being estimated is oversized. Recently, PSO has become an important estimation technique for censored data (see Wang & Huang, 2014). However, to the best of our knowledge, there is no work that uses PSO in frailty models. The main purpose of this study is to present the estimation performance of PSO in modelling multivariate survival data with shared frailty. Two simulation studies with different parameter settings are conducted to evaluate the performances of MLE via PSO, quasi-Newton and conjugate gradient methods. It is shown that PSO is an efficient method to obtain maximum likelihood estimates (MLEs) even though the following data characteristics exist: (i) small number of clusters (ii) large number of observations in a cluster (iii) large number of censored observations (heavy censored). The remainder of paper is organized as follows. Section 2 provides a general concept of shared frailty models. Section 3 provides the basic steps of standard PSO method. Section 4 outlines the MLE via PSO procedure and demonstrates its performance using two simulation studies. Section 5 provides a real data application to confirm the efficiency of procedure.

## 2. Shared Frailty Models

Suppose that there are  $N$  clusters and each cluster  $i$  has  $n_i$  observations ( $i = 1, \dots, N$ ).  $t_{ij}$  is the observed failure time of  $j$ th ( $j = 1, \dots, n_i$ ) observation in  $i$ th cluster. Under a right censoring scheme,  $t_{ij} = \min(c_{ij}, t_{ij}^*)$  where  $t_{ij}^*$  is the failure time and  $c_{ij}$  is the censoring time. Here,  $t_{ij}^*$  and  $c_{ij}$  are independent random variables. The observed censoring indicator  $\delta_{ij}$  is equal to 1 if  $t_{ij}^* < c_{ij}$ , and 0 otherwise. Conditional on frailty  $z_i (> 0)$  and  $X_{ij}$ , the hazard function of  $i$ th cluster has the form

$$h(t_{ij} \setminus X_{ij}, z_i) = z_i h_0(t_{ij}) \exp(\beta' X_{ij}), \tag{1}$$

where  $h_0(\cdot)$  is the baseline hazard function,  $X_{ij}$  is a vector of observed covariates for the  $j$ th observation and  $\beta$  is a vector of regression parameters. It is assumed that survival times (in cluster  $i$ ) are conditionally independent with respect to the frailty. This common frailty is the cause of dependence among lifetimes of within a cluster (Wienke, 2011). The frailties,  $z_i$ , are i.i.d. variables with the common probability density function  $g(z_i)$ .

Let  $(t_{i1}, t_{i2}, \dots, t_{in_i})$  denote  $n_i$  survival times of  $i$ th cluster. The conditional survival function in that cluster can be expressed such as (for a given  $Z_i = z_i; t_{ij} > 0; j = 1, 2, \dots, n_i$ ),

$$S(t_{i1}, \dots, t_{in_i} \setminus z_i, X_i) = S(t_{i1} \setminus z_i, X_{i1}) \dots S(t_{in_i} \setminus z_i, X_{in_i}) = \exp\left(-z_i \sum_{j=1}^{n_i} \Lambda_0(t_{ij}) e^{\beta' X_{ij}}\right), \tag{2}$$

where  $\Lambda_0(\cdot) = \int_0^{\cdot} h_0(s) ds$  is the common cumulative baseline hazard. Here,  $S(\cdot) = \exp(-z_i \Lambda_0(t_{ij}) e^{\beta' X_{ij}})$  denotes the survival function of  $j$ th observation conditional on frailty. The conditional likelihood function of  $i$ th cluster has the form

$$L_i(\psi, \beta \setminus z_i) = \prod_{j=1}^{n_i} h(t_{ij}, \dots, t_{in_i} \setminus z_i, X_i)^{\delta_{ij}} S(t_{i1}, \dots, t_{in_i} \setminus z_i, X_i), \tag{3}$$

where  $\psi$  and  $\beta$  are the vector of baseline hazard parameters and regression parameters, respectively. Unconditional likelihood for observed data can be obtained by integrating Eq.(3) with respect to frailty terms ( $i = 1, 2, \dots, N; j = 1, 2, \dots, n_i$ ),

$$L(\psi, \theta, \beta) = \prod_{i=1}^N \int_0^\infty \prod_{j=1}^{n_i} (z_i h_0(t_{ij}; \psi) e^{\beta' X_{ij}})^{\delta_{ij}} \exp(-z_i \Lambda_0(t_{ij}; \psi) e^{\beta' X_{ij}}) g(z_i) dz_i \tag{4}$$

Various studies have been done on the choice of distribution of frailty random variables. While some authors use continuous distributions such as Gamma (Clayton, 1978; Vaupel et al., 1979), inverse Gaussian (Hougaard, 1984; Whitmore & Lee, 1991; Hanagal & Sharma, 2015), log-normal (McGilchrist & Aisbett, 1991) and positive stable (Hougaard, 1986), other authors use discrete distributions (Caroni, Crowder, & Kimber, 2010; Ata & Ozel, 2012). However, the Gamma distribution is the most common and widely used in literature for determining the frailty effect, which acts multiplicatively on the baseline hazard (Wienke, 2011). Due to its computational convenience, one parameter Gamma distribution (with mean 1 and variance  $\theta$ ) is used as the frailty distribution. The probability density function of one parameter Gamma distribution is as follows

$$g(z) = \frac{z^{1/\theta-1} \exp(-z/\theta)}{\Gamma(1/\theta) \theta^{1/\theta}}. \tag{5}$$

The larger value of  $\theta$  indicates the greater degree of heterogeneity among lifetimes within a cluster. When Gamma distribution is degenerate at value 1, Eq.(1) reduces to the standard PH model. Hence, we can conclude that  $\theta$  takes the value 0 and lifetimes within clusters are independent.

Under the concept of gamma frailty, unconditional survival function for cluster  $i$  is obtained by integrating conditional survival function over the Gamma distribution and can be written as

$$S(t_{i1}, \dots, t_{ini} \setminus X_i) = \int S(t_{i1}, \dots, t_{ini} \setminus z_i, X_i) g(z_i) dz_i = \left[ 1 + \theta \sum_{j=1}^{n_i} \Lambda_0(t_{ij}) e^{\beta' X_{ij}} \right]^{-1/\theta}. \tag{6}$$

The unconditional hazard function of corresponding cluster is,

$$h(t_{i1}, \dots, t_{ini} \setminus X_i) = \frac{\left( \prod_{j=1}^{n_i} h_0(t_{ij}) \right) e^{\sum_{j=1}^{n_i} \beta' X_{ij}}}{\left[ 1 + \theta \sum_{j=1}^{n_i} \Lambda_0(t_{ij}) e^{\beta' X_{ij}} \right]}. \tag{7}$$

Once the parametric form of baseline hazard is specified, the unconditional likelihood function can be easily derived (Wienke, 2011). Taking into account of all clusters and denoting the number of observed events in each cluster as  $d_i = \sum_{j=1}^{n_i} \delta_{ij}$ , one might write the following unconditional likelihood function as (Klein, 1992; Duchateau & Janssen, 2008; Wienke, 2011),

$$L(\psi, \theta, \beta) = \prod_{i=1}^n \frac{\Gamma(D_i + 1/\theta) \prod_{j=1}^{n_i} (h_0(t_{ij}; \psi) e^{\beta' X_{ij}})^{\delta_{ij}}}{\theta^{1/\theta} \Gamma(1/\theta) (1/\theta + \sum_{j=1}^{n_i} \Lambda_0(t_{ij}; \psi) e^{\beta' X_{ij}})^{1/\theta + D_i}}, \tag{8}$$

where  $i = 1, \dots, N$  and  $j = 1, \dots, n_i$ . MLEs of model parameters can be obtained by maximizing Eq.(8).

### 3. Estimation Method: PSO

PSO is a member of Evolutionary Algorithms (EAs) and can be used to overcome some limitations of derivative-based methods. It was first developed by Kennedy and Eberhart (1995) as a population-based method. It is easy to implement for complex and high-dimension, linear/non-linear functions. PSO simultaneously searches a global optimum using more than one candidate solutions in different regions of multiple solution space. Therefore, the problem of being stuck at a local optimum is almost avoided (Aladag, Yolcu, Egrioglu, & Dalar, 2012). Overall, PSO is not affected by the existence of large numbers of unknown parameters as much as derivative-based methods. The algorithm has its own parameters: particle size, inertia weight, acceleration coefficients and velocity vector.

A search in solution space is similar to bird food-finding behavior in a flock. In contrast with the derivative-based methods, the PSO algorithm starts with many candidate solutions (particles). The population of particles is called a swarm, and each particle has a location vector  $P_m$  and a velocity vector  $V_m$ . A particle flies through the solution space in search for a better position and adjusts its own position with reference to the values of personal best (Pbest) and neighbors global best (Gbest). The basic steps standard PSO are given as follows (Egrioglu, Yolcu, Aladag, & Kocak, 2013):

**[Step 1]:** In  $d$ -dimensional search space, stochastically generate a position and a velocity for each particle and then store as the vectors  $\vec{P}_m = \{p_{m,1}, \dots, p_{m,k}, \dots, p_{m,d}\}$  and  $\vec{V}_m = \{v_{m,1}, \dots, v_{m,k}, \dots, v_{m,d}\}$ , respectively. For  $m = 1, \dots, L$  and  $k = 1, \dots, d$ ;  $p_{m,k}$  denotes the  $k$ th position of  $m$ th particle and  $L$  is the total number of particles in a swarm.

**[Step 2]:** Let  $\vec{P}_{best,m}$  represents the vector of best position of  $m$ th particle, and  $\vec{G}_{best}$  is the particle which has the best fitness function value, found so far by all particles. According to the fitness function, determine  $\vec{P}_{best,m}$  and  $\vec{G}_{best}$  that are respectively given by (9) and (10),

$$\vec{P}_{best,m} = \{p_{m,1}^{best}, \dots, p_{m,k}^{best}, \dots, p_{m,d}^{best}\} \tag{9}$$

$$\vec{G}best = \{p_{g,1}^{gbest}, \dots, p_{g,k}^{gbest}, \dots, p_{g,d}^{gbest}\} \tag{10}$$

[Step 3]: Update velocities and obtain particles by using following formulas given below,

$$v_{m,k}^{new} = \omega v_{m,k} + c_1 r_1 (p_{m,k}^{best} - p_{m,k}) + c_2 r_2 (p_{g,k}^{gbest} - p_{m,k}) \tag{11}$$

$$p_{m,k} = p_{m,k} + v_{m,k}^{new} \tag{12}$$

where  $c_1$  and  $c_2$  are the positive acceleration coefficients (learning factors);  $\omega$  is an inertia parameter, and  $r_1$  and  $r_2$  are uniform random variables which are generated within  $[0, 1]$ . Generally, both the values of  $(c_1, c_2)$  are taken fixed at 2 and  $\omega = 0.5 + rand/2$  changes in each iteration (Hu, Eberhart, & Shi, 2004).

[Step 4]: Stop when pre-specified iteration number is met.  $\vec{G}best$  is the vector of optimal solution. Otherwise, return to Step 2.

The Gelman-Rubin (G-R) statistic, modified by Prasad and Souradeep (2012), is used in order to be sure that pre-specified iteration number satisfy the convergence for each position related to unknown model parameter. Let  $T$  represents the total iteration number.  $p_{m,k}$  is the trajectory of  $m$ th particle of  $k$ th position. The variances of within particles ( $W$ ) and between particles ( $B$ ) are calculated for position  $k$  with the following formulas (13) and (14),

$$W_k = \frac{1}{L} \sum_{m=1}^L \sigma_{m,k}^2, \tag{13}$$

where  $\sigma_{m,k}^2 = \frac{1}{T-1} \sum_{j=1}^T (p_{m,k}^{(j)} - \bar{p}_{m,k})$  and  $\bar{p}_{m,k} = \sum_{j=1}^T p_{m,k}^{(j)}$ .

$$B_k = \frac{T}{L-1} \sum_{m=1}^L (\bar{p}_{m,k} - \bar{\bar{p}}_k)^2, \tag{14}$$

where  $\bar{\bar{p}}_k = \frac{1}{L} \sum_{m=1}^L \bar{p}_{m,k}$ .

Using  $W_k$  and  $B_k$ , the variance of the stationary distribution is calculated as,  $\hat{V}_k = [(W_k(T-1))/T] + [B_k/T]$ . An estimated potential scale reduction factor (G-R statistic)  $\hat{R}_k = \sqrt{\hat{V}_k/W_k}$  monitors the convergence ability of performed total iteration. For the value of close to 1 suggests a good convergence achieved for the corresponding parameter  $k$ .

#### 4. Simulation Study

In this section, the estimation performance of MLE via PSO is investigated by two scenarios. In the first simulation study, the performance of PSO is compared with the quasi-Newton method, and then in the second one it is compared with the conjugate gradient method.

For both of simulation studies, the baseline cumulative hazard function is assumed to be Weibull as  $\Lambda_0(.) = \lambda t^p$ . In the first simulation study, Weibull parameters are set to  $(\psi = \lambda, p) = (0.2, 1.5)$  and the frailty variables are taken as Gamma distributed randoms with variance  $\theta = 1.5$ . Three observed covariates  $x^{(1)}$ ,  $x^{(2)}$  and  $x^{(3)}$  are used and randomly generated from Bernoulli distribution with success probability 0.1, 0.5 and 0.9, respectively. Regression parameters are assumed to be  $\beta_1 = \beta_2 = \beta_3 = 1$ . In the second simulation study, Weibull parameters are set to  $(\psi = \lambda, p) = (0.5, 1)$ . The variance of frailty is taken as  $\theta = 1$ . Two observed covariates,  $x^{(1)}$  and  $x^{(2)}$ , are randomly generated from Binomial distribution having the same success probability 0.5 and the number of trials are taken as 2 and 3, respectively. Regression parameters are set to be  $\beta_1 = \beta_2 = 0.5$ .

The model for the failure times  $(t_{i1}, \dots, t_{in_i})$  is considered as,

$$F(t_{ij} \setminus z_i, X_i) = 1 - \exp \left[ -z_i \sum_{j=1}^{n_i} \Lambda_0(t_{ij}) \exp(\beta X_{ij}) \right]$$

where  $\beta X_{ij} = \beta_1 x_{ij}^{(1)} + \beta_2 x_{ij}^{(2)} + \beta_3 x_{ij}^{(3)}$  in the first simulation and  $\beta X_{ij} = \beta_1 x_{ij}^{(1)} + \beta_2 x_{ij}^{(2)}$  in the second simulation. Following the study of Yu (2006), the multivariate survival data are generated as follows:

(1) frailty variables are generated from Gamma distribution with mean 1 and variance  $\theta$ . (2) for the  $j$ th observation in  $i$ th cluster, uniform random variable is generated as  $u_{ij} \sim U(0, 1)$ . Also censoring time  $c_{ij}$  is generated from an exponential distribution  $exp[\kappa]$  with parameter  $\kappa$  to create expected censoring rates.

(3) the failure times are generated by,

$$t_{ij}^* = \left( \frac{-\log(u_{ij})}{z_i \lambda \exp(\beta X_{ij})} \right)^{1/p} \tag{15}$$

(4) the observed failure times are  $t_{ij} = \min(c_{ij}, t_{ij}^*)$ . The observed censoring indicator  $\delta_{ij}$  is equal to 1 if  $t_{ij}^* < c_{ij}$ , and 0 otherwise.

For both of simulations, different size of clusters ( $n_i = 2, 3, 4$ ) and different number of clusters ( $N = 20, 30, 40$ ) are taken into account. Three censoring rates are chosen as 5%, 20% and 40% which represent light, medium and heavy censoring, respectively. Simulations are repeated 100 times. Mean absolute percentage error (MAPE), bias and standard error (SE) vectors are computed for all cases. Let  $\eta$  shows the vector of real parameters and  $\hat{\eta}$  shows the vector of MLEs. In the first simulation study  $\hat{\eta} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\lambda}, \hat{p}, \hat{\theta})$ , and in the second one  $\hat{\eta} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\lambda}, \hat{p}, \hat{\theta})$ . To show the results of 54 different cases less complicated, we report the mean of parameters' MAPEs (mMAPE), biases (mBias) and SEs (mSE) given below,

$$mMAPE(\hat{\eta}) = \frac{\frac{1}{100} \sum_{k=1}^d \sum_{r=1}^{100} \left| \frac{\hat{\eta}_r - \eta}{\eta} \right|^{(k)}}{d}, \tag{16}$$

$$mSE(\hat{\eta}) = \frac{\sum_{k=1}^d \left( \sqrt{\frac{1}{99} \sum_{r=1}^{100} (\hat{\eta}_r - \eta)^2} \right)^{(k)}}{d}, \tag{17}$$

$$mBias(\hat{\eta}) = \frac{\sum_{k=1}^d \left( \sqrt{E(\hat{\eta}) - \eta} \right)^{(k)}}{d}, \tag{18}$$

where  $\hat{\eta}_r$  is the vector of MLEs obtained in  $r$ th repeat and  $E(\hat{\eta}) = \sum_{r=1}^{100} \hat{\eta}_r / 100$ . Besides these measures, the mean of log-likelihood values obtained in each repeat are calculated (mLog.Lik.). The steps of MLE via PSO are given as follows:

**[Step 1]:** PSO parameters are selected as  $(\omega, c_1, c_2, L) = (0.5 + rand/2, 2, 2, 30)$ . T is taken as 1000. In the first simulation study the solution space is 6-dimensional and in the second, it is 5-dimensional. Positions of each particle and their corresponding velocities are generated randomly from uniform distribution in the range of [0,3] and [0,0.5], respectively. Positions of a particle in the swarm for the first and second simulation are illustrated in Figure 1 and 2, respectively.

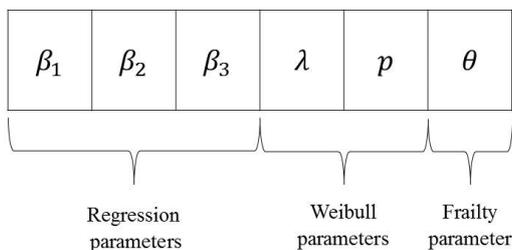


Figure 1. Positions of a particle in Simulation 1

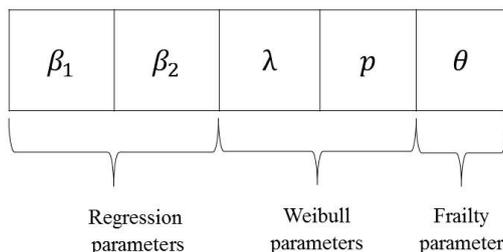


Figure 2. Positions of a particle in Simulation 2

[Step 2]: The unconditional likelihood function given in Eq.(8) is chosen as the fitness function of PSO. For all particles, use the fitness function to calculate the fitness values. Determine  $\vec{P}_{best,m}$  and  $\vec{G}_{best}$ , then update the velocity and position values for each particle using formulas (11) and (12).

[Step 3]: If the termination criteria is met, then stop and go to Step 4. Otherwise return to Step 2.

[Step 4]: If the G-R statistics of all parameter estimations are satisfactory,  $\vec{G}_{best}$  is the vector of MLEs found by using PSO (it is observed that T=1000 is sufficient for converge). To perform the MLE with either the quasi-Newton or conjugate gradient methods, randomly choose a particle from the initial swarm and use it as the initial vector.

Table 1. Simulation Results for Model 1

Cases			Methods								
$\delta$	$n_i$	$N$	PSO				Quasi-Newton				
			mMAPE	mSE	mBias	mLog.Lik	mMAPE	mSE	mBias	mLog.Lik	
5%	2	20	0.751	0.816	0.076	-66.645	1.639	0.871	0.256	-131.932	
		30	0.586	0.506	0.024	-98.473	1.623	0.861	0.250	-209.767	
		40	0.542	0.421	0.028	-134.224	1.577	0.839	0.250	-289.869	
	3	20	0.576	0.499	0.046	-94.536	1.755	0.855	0.280	-166.141	
		30	0.478	0.351	0.013	-143.818	1.703	0.856	0.265	-246.152	
		40	0.454	0.305	0.020	-190.658	1.593	0.847	0.253	-341.636	
	4	20	0.507	0.368	0.019	-123.813	1.788	0.875	0.276	-239.448	
		30	0.460	0.305	0.030	-184.511	1.530	0.866	0.240	-415.138	
		40	0.424	0.257	0.011	-249.729	1.813	0.855	0.281	-506.310	
	20%	2	20	0.844	0.964	0.080	-50.705	1.865	0.881	0.304	-131.768
			30	0.631	0.599	0.054	-76.061	1.631	0.878	0.250	-206.824
			40	0.561	0.485	0.024	-101.141	1.764	0.828	0.275	-206.835
3		20	0.572	0.537	0.024	-73.551	1.681	0.878	0.257	-149.540	
		30	0.530	0.432	0.025	-112.645	1.637	0.860	0.257	-226.845	
		40	0.498	0.354	0.017	-149.404	1.737	0.875	0.276	-284.043	
4		20	0.552	0.461	0.024	-95.459	1.673	0.849	0.268	-167.023	
		30	0.501	0.351	0.036	-140.327	1.538	0.848	0.244	-265.062	
		40	0.444	0.278	0.011	-193.216	1.675	0.873	0.263	-439.114	
40%		2	20	0.898	1.041	0.093	-34.374	1.739	0.853	0.269	-97.330
			30	0.779	0.814	0.078	-51.931	1.602	0.873	0.252	-126.279
			40	0.694	0.669	0.074	-71.023	1.595	0.848	0.257	-168.676
	3	20	0.665	0.732	0.059	-49.059	1.817	0.855	0.282	-106.891	
		30	0.557	0.482	0.023	-76.464	1.747	0.891	0.268	-162.363	
		40	0.506	0.399	0.022	-104.568	1.621	0.878	0.251	-251.969	
	4	20	0.566	0.497	0.029	-64.371	1.665	0.858	0.262	-154.974	
		30	0.516	0.403	0.028	-96.147	1.553	0.868	0.235	-199.620	
		40	0.471	0.336	0.013	-134.302	1.746	0.863	0.276	-262.420	

Table 2. Simulation Results for Model 2

Cases			Methods								
$\delta$	$n_i$	$N$	PSO				Conjugate Gradient				
			mMAPE	mSE	mBias	mLog.Lik	mMAPE	mSE	mBias	mLog.Lik	
5%	2	20	0.442	0.461	0.038	-43.164	1.643	0.848	0.377	-212.481	
		30	0.392	0.303	0.023	-66.949	1.644	0.851	0.379	-214.597	
		40	0.372	0.267	0.021	-90.148	1.609	0.880	0.362	-395.706	
	3	20	0.384	0.269	0.023	-65.337	1.541	0.849	0.351	-210.397	
		30	0.366	0.201	0.011	-98.386	1.642	0.841	0.389	-337.160	
		40	0.347	0.181	0.006	-129.349	1.645	0.842	0.394	-424.842	
	4	20	0.383	0.253	0.018	-83.467	1.498	0.845	0.339	-242.526	
		30	0.354	0.176	0.009	-127.168	1.486	0.861	0.333	-307.569	
		40	0.351	0.148	0.006	-171.539	1.630	0.860	0.374	-518.609	
	20%	2	20	0.466	0.541	0.040	-31.532	1.657	0.837	0.396	-133.936
			30	0.394	0.320	0.038	-46.877	1.714	0.881	0.400	-180.451
			40	0.375	0.251	0.017	-62.962	1.578	0.830	0.362	-234.702
3		20	0.396	0.322	0.024	-42.867	1.735	0.861	0.406	-211.654	
		30	0.364	0.242	0.018	-67.121	1.689	0.891	0.396	-245.108	
		40	0.371	0.215	0.014	-89.857	1.565	0.843	0.357	-275.511	
4		20	0.391	0.285	0.019	-54.476	1.723	0.851	0.414	-173.668	
		30	0.364	0.206	0.018	-85.803	1.721	0.874	0.404	-287.870	
		40	0.357	0.178	0.009	-117.091	1.693	0.858	0.402	-370.715	
40%		2	20	0.546	0.707	0.062	-16.255	1.614	0.871	0.369	-105.767
			30	0.432	0.390	0.045	-26.241	1.631	0.877	0.374	-132.810
			40	0.399	0.300	0.032	-37.778	1.647	0.856	0.380	-162.449
	3	20	0.433	0.366	0.041	-25.335	1.612	0.875	0.367	-111.564	
		30	0.382	0.259	0.021	-37.975	1.569	0.874	0.356	-159.544	
		40	0.367	0.222	0.013	-51.761	1.593	0.828	0.379	-208.918	
	4	20	0.384	0.294	0.027	-32.055	1.557	0.864	0.350	-144.924	
		30	0.366	0.213	0.013	-51.601	1.663	0.866	0.389	-198.873	
		40	0.349	0.174	0.015	-67.963	1.628	0.875	0.373	-256.160	

The results are shown in Table 1 and Table 2. As observed in the tables, MLE via PSO has lower mMAPE and mBias values for all cases. For all cases except two, the values of mSE are lower in the first simulation. In most cases, as the number of clusters increase, a remarkable reduction in mMAPE, mSE and mBias for PSO is seen. However, the quasi-Newton or conjugate gradient methods do not provide the same pattern. For the same level of  $N$  and  $\delta$ , when the size of clusters increase, mMAPE, mSE and mBias decrease in all cases of MLE via PSO except one (in this case, mBias remains constant). It should also be noted that increasing censoring rate has lower impact on the values of mMAPE, mSE and mBias in MLE via PSO procedure.

Also, mLog.Lik values of PSO are quite better than mLog.Lik values of other two methods. To illustrate change of Log.Lik values in each repeat, randomly two cases are chosen among 54 different cases. Figures 3 and 4 show the cases ( $n_i = 4; N = 40; \delta = 20\%$ ) and ( $n_i = 4; N = 40; \delta = 5\%$ ) from first and second simulation study, respectively. As seen in figures, PSO takes the higher and more consistent Log.Lik values compared with derivative-based methods.

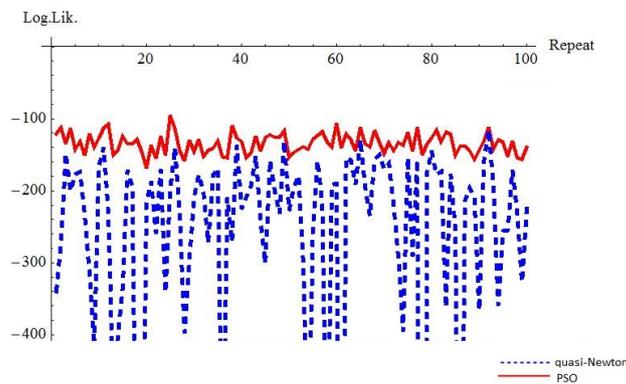


Figure 3. Change of Log.Lik values for case ( $n_i = 4; N = 40; \delta = 20\%$ )

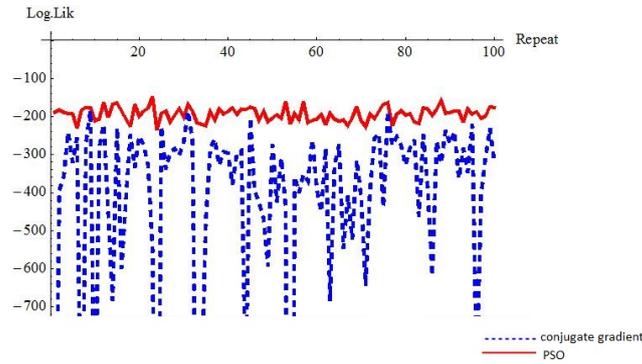


Figure 4. Change of Log.Lik values for case ( $n_i = 4; N = 40; \delta = 5\%$ )

### 5. Real Data Application

The kidney infection data (McGilchrist & Aisbett, 1991) has been widely used in frailty studies. The data set consists of first and second recurrence times of infection for 38 kidney patients using a portable dialysis machine. The event of interest is the occurrence of an infection at the point of insertion of the catheter. The catheter is removed when the infection is observed. Then, the infection is cleared up and the catheter is reinserted for the second time. Each lifetime ( $t_{i1}, t_{i2} > 0$ ) represents the time elapsed (in days) from the catheter insertion to infection. Censoring occurs when the catheter is removed for any reasons except the infection.

We first investigate the goodness of fit of the kidney infection data. Weibull and exponential probability plots (P-P) are shown in Figures 5 and 6. According to P-P plots, Weibull fits the data better than the exponential distribution. Kolmogorov-Smirnov (KS) tests suggest that Weibull is an appropriate distribution on the choice of baseline hazard. Table 3 provides the K-S statistics and associated p-values.

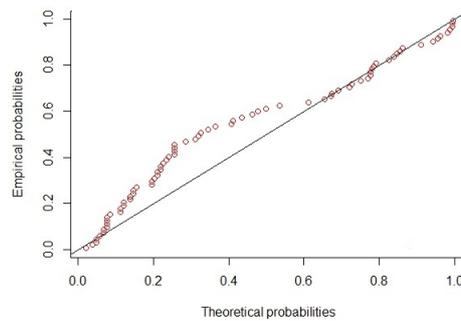


Figure 5. Exponential P-P plot for kidney infection data

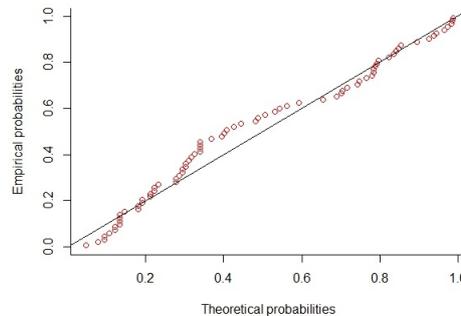


Figure 6. Weibull P-P plot for kidney infection data

Table 3. Kolmogorov-Smirnov statistics and p-values

	MLEs	K-S	p-value
Exponential	$\psi = (0.009)$	0.2239	0.0009
Weibull	$\psi = (0.011, 0.805)$	0.1205	0.2201

Two primary risk factors are included in the model: age and sex of patient (0=male, 1=female). The hazard function of  $i$ th patient conditional on Gamma shared frailty at time  $t_{ij} > 0$  takes the form ( $i = 1, \dots, 38; j = 1, 2$ ),

$$h(t_{ij} | z_i, X_{ij}) = z_i(\lambda p t_{ij}^{p-1}) \exp(\beta_1 x_{ij}^{(age)} + \beta_2 x_{ij}^{(sex)}) \tag{19}$$

The results of MLE with PSO, quasi-Newton and conjugate gradient are reported in Table 4. Standard errors of parameters (Se) are obtained from the inverse of observed information matrix  $I(\beta, \psi, \theta) = -H(\beta, \psi, \theta)$ , where  $H(\beta, \psi, \theta)$  denotes the Hessian matrix. As observed in Table, lower standard errors and higher Log.Lik. values are obtained from PSO. Also, the G-R converge statistics of PSO calculated for 5 parameter estimates are found satisfactory (that is, equal to or less than 1.1).

Table 4. Application results

Parameter	PSO			Q uasi-Newton		C onjugate Gradient	
	Value	Se	G-R statistic	Value	Se	Value	Se
$\beta_1$	0.007	0.025	0.999	-0.088	0.042	-0.088	0.039
$\beta_2$	-2.118	0.577	0.999	0.923	2.066	0.923	1.960
$\lambda$	0.011	0.024	1.019	0.281	0.136	0.281	0.128
$p$	1.293	0.321	1.003	1.146	0.395	1.146	0.375
$\theta$	0.510	0.297	1.030	0.848	0.626	0.846	0.590
Log. Lik.		-332.488			-364.718		-364.776

Note: The initials selected randomly for quasi-Newton and conjugate gradient are (0.257,0.929,0.307,1.180,0.098)

**5. Conclusion**

In this paper, the estimation performance of MLE via PSO is investigated against the estimation performance of MLE via quasi-Newton and conjugate gradient methods for the shared Gamma frailty model. Based on two different simulation studies, PSO procedure outperforms the two derivative-based methods for most cases based on all comparison criteria. Also it is shown that smaller biases are obtained under the situation of heavy censoring. MLE via PSO is also implemented to a real data set and it is shown that this procedure can be preferred for estimating the parameters of parametric shared frailty models.

**Acknowledgements**

Collate acknowledgements in a separate section at the end of the article before the references. List here those individuals who provided help during the research (e.g., providing language help, writing assistance or proof reading the article, etc.).

**References**

Aalen, O. O. (1988). Heterogeneity in survival analysis. *Statistics in Medicine*, 7(11), 1121-1137. <http://dx.doi.org/10.1002/sim.4780071105>

Aalen, O., Borgan, O., & Gjessing, H. (2008). *Survival and event history analysis: A process point of view*. New York: Springer.

Aladag, C., Yolcu, U., Egrioglu, E., & Dalar, A. (2012). A new time invariant fuzzy time series forecasting method based on particle swarm optimization. *Applied Soft Computing*, 12(10), 3291-3299. <http://dx.doi.org/10.1016/j.asoc.2012.05.002>

Ata, N., & Ozel, G. (2012). Survival functions for the frailty models based on the discrete compound Poisson process. *Journal of Statistical Computation and Simulation*, 83(11), 2105-2116. <http://dx.doi.org/10.1080/00949655.2012.679943>

Caroni, C., Crowder, M., & Kimber, A. (2010). Proportional hazards model with discrete frailty. *Lifetime Data Analysis*, 16(3), 374-384. <http://dx.doi.org/10.1007/s10985-010-9151-3>

Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1), 141-151. <http://dx.doi.org/10.1093/biomet/65.1.141>

- Clayton, D. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics*, 47(2), 467-485. <http://dx.doi.org/10.2307/2532139>
- Clayton, D., & Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model. *Journal of the Royal Statistical Society Series A*, 148(2), 82-117. <http://dx.doi.org/10.2307/2981943>
- Duchateau, L., & Janssen, P. (2008). *The frailty model*. New York: Springer.
- Egrioglu, E., Yolcu, U., Aladag, C. H., & Kocak, C. (2012). An ARMA type fuzzy time series forecasting method based on Particle Swarm Optimization. *Mathematical Problems in Engineering*, 2013, 3291-3299. <http://dx.doi.org/10.1155/2013/935815>
- Gill, R. D. (1985). Discussion of multivariate generalizations of the proportional hazards model, by Clayton and Cuzick. *Journal of the Royal Statistical Society Series A*, 148(2), 82-117. <http://dx.doi.org/10.2307/2981943>
- Hanagal, D., & Pandey, A. (2015). Gamma frailty models for bivariate survival data. *Journal of Statistical Computation and Simulation*, 85(15), 3172-3189. <http://dx.doi.org/10.1080/00949655.2014.958086>
- Hanagal, D., & Sharma, R. (2015). Analysis of bivariate survival data using shared Inverse Gaussian frailty model. *Communications in Statistics-Theory and Methods*, 44(7), 1351-1380. <http://dx.doi.org/10.1080/03610926.2013.768663>
- Henderson, R., & Oman, P. (1999). Effect of frailty on marginal regression estimates in survival analysis. *Journal of the Royal Statistical Society Series B*, 61(2), 367-379. <http://dx.doi.org/10.1111/1467-9868.00182>
- Hougaard, P. (1984). Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika*, 71, 75-83. <http://dx.doi.org/10.2307/2336399>
- Hougaard, P. (1986). A class of multivariate failure time distributions. *Biometrika*, 73(3), 671-678. <http://dx.doi.org/10.1093/biomet/73.3.671>
- Hougaard, P. (1991). Modelling heterogeneity in survival data. *Journal of Applied Probability*, 28, 695-701. <http://dx.doi.org/10.2307/3214503>
- Hu, X., Eberhart, R., & Shi, Y. (2004). Recent advances in particle swarm. In Proceedings of the IEEE Congress on Evolutionary Computation (pp. 90-97), Oregon, Portland.
- Ibrahim, J., Chen, M., & Sinha, D. (2001). *Bayesian survival analysis*. New York: Springer.
- Kennedy, J., & Eberhart, R. (1995). Particle Swarm Optimization. In Proceedings of the IEEE International Conference on Neural Networks (pp. 1942-1948), Piscataway, NJ.
- Klein, J. P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics*, 48, 795-806. <http://dx.doi.org/10.2307/2532345>
- Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica*, 47(4), 939-956. <http://dx.doi.org/10.2307/1914140>
- Lancaster, T. (1990). *The econometric analysis of transition data*. Cambridge: Cambridge U. Press.
- McGilchrist, C. A., & Aisbett, C. W. (1991). Regression with frailty in survival analysis. *Biometrics*, 47, 461-466. <http://dx.doi.org/10.2307/2532138>
- McGilchrist, C. A. (1993). REML estimation for survival models with frailty. *Biometrics*, 49, 221-225. <http://dx.doi.org/10.2307/2532615>
- Nielsen, G. G., Gill, R. D., Andersen, P. K., & Sørensen, T.I.A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics*, 19, 25-43. <http://www.jstor.org/stable/4616223>
- Prasad, J., & Souradeep, T. (2012). Cosmological parameter estimation using particle swarm optimization (PSO). *Phys.Rev.D.*, 85(12), 1-13. <http://dx.doi.org/10.1103/PhysRevD.85.123008>
- Rondeau, V., Commenges, D., & Joly, P. (2003). Maximum penalized likelihood estimation in a gamma-frailty model. *Lifetime Data Analysis*, 9(2), 139-153. <http://dx.doi.org/10.1023/A:1022978802021>
- Sinha, D., Dey, & K. (1997). Semiparametric Bayesian analysis of survival data. *Journal of the American Statistical Association*, 92(439), 1195-1212. <http://dx.doi.org/10.2307/2965586>
- Therneau, T. M., Grambsch, P. M., & Pankratz, V. S. (2003). Penalized survival models and frailty. *Journal of Computational and Graphical Statistics*, 12, 156-175. <http://dx.doi.org/10.1198/1061860031365>
- Van den Berg, G. J. (2001). Duration models: Specification, identification and multiple duration. *Handbook of Econo-*

*metrics*, 5, 3381-3460. [http://dx.doi.org/10.1016/S1573-4412\(01\)05008-5](http://dx.doi.org/10.1016/S1573-4412(01)05008-5)

Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16, 439-454. <http://dx.doi.org/10.2307/2061224>

Wang, F., & Huang, P. (2014). Implementing particle swarm optimization algorithm to estimate the mixture of two Weibull parameters with censored data. *Journal of Statistical Computation and Simulation*, 84(9), 1975-1989. <http://dx.doi.org/10.1080/00949655.2013.778992>

Whitmore, G. A., & Lee, M. L. T. (1991). A multivariate survival distribution generated by an inverse Gaussian mixture of exponentials. *Technometrics*, 33, 39-50. <http://dx.doi.org/10.2307/1269006>

Wienke, A. (2011). *Frailty models in survival analysis*. Boca Raton: Chapman & Hall/CRC Biostatistics Series.

Yu, B. (2006). Estimation of shared gamma frailty models by a modified EM algorithm. *Computational Statistics and Data Analysis*, 50(2), 463-474. <http://dx.doi.org/10.1016/j.csda.2004.08.010>

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).