FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 7 No. 3, June, 2023, pp 304 - 310
DOI: https://doi.org/10.33003/fjs-2023-0703-1970

EXPLORING THE MODIFIED GAMMA FRAILTY DISTRIBUTION: AN OPTIMAL DESIGN APPROACH USING PYTHON

Abdulazeez, Sikiru Adeyinka

Department of Mathematical Sciences, Kaduna State University, Kaduna..Kaduna State, Nigeria

*Corresponding authors' email: ysabdul94@gmail.com

ABSTRACT

A suction/injection controlled mixed convection flow of an incompressible and viscous fluid in a vertical SurvivalAnalysis is pivotal in understanding the effects of covariates on potentially censored failure times and in the joint modelling of clustered data. It is used in the context of incomplete repeated measures and failure times in longitudinal studies. Survival data are often subject to right censoring and to a subsequent loss of information about the effect of explanatory variables. Frailty models are one common approach to handle such data.Three frailty models are used to analyze bivariate time-to-event data. All approaches accommodate right censored lifetime data and account for heterogeneity in the study population. A Modified Gamma Frailty Model is compared with two existing Frailty Models. The survival-analysis was performed using the Python.The newly derived MGF was analyzed using Python which is more robust when sample size is more than forty.The MGF model performs better than the existing models in the presence of clustering. However the CGF is preferable in the absence of clusters in a given data set.

Keywords: Frailty Models , Censorship, Survival Analysis, Python,Frailty Models

INTRODUCTION

The design of optimal experiments is crucial in the analysis of survival data. When studying time-to-event data or survival data, frailty models, including the Gamma frailty model are essential for capturing unobserved heterogeneity among subjects. Abdulazeez, (2020). Hazard models have become widespread in their use for the analysis of durationtimedata in many scientific disciplines, including biology and medicine (Cox,1972; Kalbfleisch & Prentice, 1980), sociology (Petersen, 1998, Vermunt, 1996), marketing research (Vilcassim & Jain, 1991; Wedel et al., 1995), (Getachew & Bekele 2016) andeconomics (Kiefer, 1988; Lancaster, 1990). These models overcome theproblems of accounting for censored observations of duration and time-varyingexplanatory variables, which arise in applying standard regression type models toduration data. The basic concept in hazard models is the probability of the occurrence of an event during a certain time interval, says t to $t+\Delta t$, given that it has not occurredbefore t, specified as:

$$\lambda(t|N_i(t-),Z_i(t)) = \lim_{\Delta t \to 0} Pr\big(t \le T_{i,N_i(t-)+1} < t+\Delta t | N_i(t-),Z_i(t)\big)/\Delta t$$
$$= \lim_{\Delta t \to 0} Pr(\Delta N_i(t) = 1 | N_i(t-),Z_i(t))/\Delta t \quad (1)$$

The Cox proportional hazards model (Cox, 1972) is commonly used in the analysis of survival time data. An often unstated assumption of the proportional hazards model and of traditional frailty models (with the exception of those that use the compound Poisson distribution (Rakhmawati *et al* (2021)) is that all individuals will experience the event of interest. However, in some situation a fraction of individuals is not expected toexperience the event of interest; that is, these individuals are not at risk. (Anthony *et al* (2019). The terminologyto describe the never-at-risk group varies from field to field, but includes 'long-term survivors' or 'cured' in epidemiology, 'non-susceptibles' in toxicology, 'stayers' in finiteMarkov transition models of occupational mobility,the 'non-fecundable' in fertilitymodels, and 'non-recidivists' among convicted criminals. In epidemiology and medicine,researchers may be interested in analyzing the occurrence of a disease. Many individualsmay never experience that disease; therefore, there exists a fraction in the population thatis protected. Cure models are survival models which allow for a cured fraction in thestudy population.

These models extend the understanding of time-to-event data by allowing for the formulation of more accurate and informative conclusions than previously made. Theseconclusions would otherwise be unobtainable from an analysis that fails to account for acured fraction in the population. If a cured component is not present, the analysis reducesto standard approaches of survival analysis.

In cure models, the population is divided into two sub-populations so that an individualis either cured with probability $1-\phi$, or has a proper survival function $S(t)$, withprobability ϕ. Here, proper survival function means $Lim_{t\to\infty} S(t) = 0$. Individuals regardedas cured will never experience the event of interest and their survival time willbe defined as infinity. Therefore, the hazard and survival functions of cured individuals are set to zero and one, respectively, for all finite values of t.

Longini and Halloran (1996) have proposed frailty cure models that extend standardfrailty models. The frailty random variable in the former has point mass at zero withprobability $1-\phi$ while heterogeneity among those experiencing the event of interest ismodelled via a continuous distribution with probability φ. Price and Manatunga (2001)gave an excellent introduction to this area and applied leukaemia remission data to different cure, frailty and frailty cure models. They found that frailty models are useful inmodelling data with a cured fraction and that the gamma frailty cure model provides abetter fit to their remission data compared to the standard cure model.

In the next section we describe the existing models and a proposed model, then provide an application of the models to an existing data on occupational exposure tagged – IRANIAN data.

MATERIALS AND METHODS

Cox PH models

The notation used for Cox PH models (Cox, 1972), Lee & Song (2001) with one more subscript to capture multiple events is generalized. Let T_{ik}be the total time of the k^{th}event for the i^{th}subject, C_{ik}be the censoring time of the k^{th}event for the i^{th}subject. Let U_{ik}be the observation time, that is,

$$U_{ik} = min(T_{ik}; C_{ik}), \quad (2)$$

$$and$$

$$\delta_{ik} = I(T_{ik} \le C_{ik}) \quad (3)$$

is an indicator of observed k^{th}failure time for subject i. $Z_{ik} = (Z_{1ik}, Z_{2ik}, \ldots . Z_{pik})$is the covariate vector for the i^{th}subject with respect to the k^{th} event, and $Z_i = (Z_{i1}^{'}, Z_{i2}^{'}, \ldots . Z_{ik}^{'})$ denotes the covariate vector for the i^{th} subject, where K is the maximum number of events within a subject. $\beta = (\beta_1, \beta_2, \ldots . \beta_p)$ is a $p \times 1$ vector of unknown parameters. Denote $h_k(t|Z_i(t))$ as the hazard function for the k^{th}event of the i^{th}subject at time t. This is in the context of competing risk.

In general, the hazard function at time t for a subject is defined as the instantaneous probability of failure at time t given the survivorship prior to time t and the covariates:

$$h_{k+1}(t|Z_i(t)) = \lim_{\Delta t \to 0} Pr(t \le T_{i,k+1} < t + \Delta t \,|T_{i,k+1} \ge t, Z_i(t))/\Delta t \quad (4)$$

Note that Cox PH model for the k^{th}event time $T_k$is

$$h_k(t|Z_i(t)) = h_{0,k}(t)\, exp\{\beta^{'} Z_i(t)\} \quad (5)$$

Correlated Gamma Frailty (CGF) Model

This model was introduced by Yashin & Iachine (1995a,b, 1997, 1999a,b) and applied to related lifetimes in many different settings. Examples are found in Rakhmawati et al (2021) ,Pickles et al. (1994), Yashin et al. (1996), Iachine et al. (1998), Iachine (2002), Petersen (1998), Rueten-Budde et al (2019), Wienke et al. (2000, 2001, 2002, 2003a,b, 2005), Zdravkovic et al. (2002, 2004). Zhu &Kosorok(2012)

Let k_0, k_1, k_2 be some real positive values. Set $\lambda_1 = k_0 + k_1 and \lambda_2 = k_0 + k_2$

Let Y_0, Y_1, Y_2 be independently gamma distributed random variables with

$$Y_0 \sim \Gamma(k_0, \lambda_0), Y_1 \sim \Gamma(k_1, \lambda_1), Y_2 \sim \Gamma(k_2, \lambda_2) \quad (6)$$

Consequently,

$$Z_1 = \frac{\lambda_0}{\lambda_1} Y_0 + Y_1 \sim \Gamma(k_0 + k_1, \lambda_1) \quad (7)$$

$$Z_2 = \frac{\lambda_0}{\lambda_2} Y_0 + Y_2 \sim \Gamma(k_0 + k_2, \lambda_2) \quad (8)$$

$$and \quad E(Z_1) = E(Z_2) = 1,$$

$$V(Z_1) = \frac{1}{\lambda_1} := \sigma_1^2,$$

$$V(Z_2) = \frac{1}{\lambda_2} := \sigma_2^2. \quad (9)$$

The following relation holds

$$E(Y_0^2) = V(Y_0) + (E(Y_0))^2$$

$$= \frac{k_0}{\lambda_0^2} + \left(\frac{k_0}{\lambda_0}\right)^2$$

$$= \frac{k_0^2 + k_0}{\lambda_0^2} \quad (10)$$

$$E(Z_1 Z_2) = E\left(\frac{\lambda_0}{\lambda_1} Y_0 + Y_1\right)\left(\frac{\lambda_0}{\lambda_2} Y_0 + Y_2\right)$$

$$= E\left(\frac{\lambda_0^2}{\lambda_1 \lambda_2} Y_0^2 + \frac{\lambda_0}{\lambda_1} Y_0 Y_2 + \frac{\lambda_0}{\lambda_2} Y_0 Y_1 + Y_1 Y_2\right)$$

$$= \frac{\lambda_0^2}{\lambda_1 \lambda_2} \frac{k_0^2 + k_0}{\lambda_0^2} + \frac{\lambda_0 k_0 k_2}{\lambda_1 \lambda_0 \lambda_2} + \frac{\lambda_0 k_0 k_1}{\lambda_2 \lambda_0 \lambda_1} + \frac{k_1 k_2}{\lambda_1 \lambda_2}$$

$$= \frac{k_0 + (k_0 + k_1)(k_0 + k_2)}{\lambda_1 \lambda_2}$$

$$= \frac{k_0}{(k_0 + k_1)(k_0 + k_2)} + 1 \quad (11)$$

$$Cov(Z_1, Z_2) = E(Z_1 Z_2) - E(Z_1) E(Z_2)$$

$$= \frac{k_0}{(k_0 + k_1)(k_0 + k_2)}, \quad (12)$$

This leads to the correlation

$$\rho = \frac{cov(Z_1, Z_2)}{\sqrt{V(Z_1) V(Z_2)}}$$

$$= \frac{k_0}{\sqrt{(k_0 + k_1)(k_0 + k_2)}} \quad (13)$$

Consequently, because of relation

$$k_0 + k_1 = \lambda_1$$

$$= 1/\sigma_i^2 (i = 1,2)$$

It holds that

$$k_0 = \frac{\rho}{\sigma_1 \sigma_2} \quad (14)$$

and

$$k_i = \frac{1}{\sigma_i^2} - k_0$$

$$= \frac{1 - \frac{\sigma_i}{\sigma_j}\rho}{\sigma_i^2} (i, j = 1,2; i \ne j).. \quad (15)$$

To derive the unconditional model, the Laplace transform of gamma distributed random variables is applied. Hence,

$$S(t_1, t_2) = E\{S(t_1, t_2 | Z_1, Z_2)\}$$

$$= E\{S_1(t_1|Z_1) S_2(t_2|Z_2)\}$$

$$= E\left\{e^{-Z_1 \Lambda_1(t_1)} e^{-Z_2 \Lambda_2(t_2)}\right\}$$

$$= E\left\{e^{-\left(\frac{\lambda_0}{\lambda_1} Y_0 + Y_1\right)\Lambda_1(t_1)} e^{-\left(\frac{\lambda_0}{\lambda_2} Y_0 + Y_2\right)\Lambda_2(t_2)}\right\}$$

$$= E\left\{e^{-Y_0\left(\frac{\lambda_0}{\lambda_1}\Lambda_1(t_1) + \frac{\lambda_0}{\lambda_2}\Lambda_2(t_2) - Y_1\Lambda_1(t_1) - Y_2\Lambda_2(t_2)\right)}\right\}$$

$$= \left(1 + \frac{1}{\lambda_0}\left(\frac{\lambda_0}{\lambda_1}\Lambda_1(t_1) + \frac{\lambda_0}{\lambda_2}\Lambda_2(t_2)\right)\right)^{-k0} \left(1 + \frac{1}{\lambda_1}\Lambda_1(t_1)\right)^{-k_1} \left(1 + \frac{1}{\lambda_2}\Lambda_2(t_2)\right)^{-k_2}$$

$$= (1 + \sigma_1^2 \Lambda_1(t_1) + \sigma_2^2 \Lambda_2(t_2))^{\frac{-\rho}{\sigma_1 \sigma_2}} (1 + \sigma_1^2 \Lambda_1(t_1))^{\frac{-1 + \frac{\sigma_1}{\sigma_2}\rho}{\sigma_1^2}} (1 + \sigma_2^2 \Lambda_2(t_2))^{\frac{-1 + \frac{\sigma_2}{\sigma_1}\rho}{\sigma_2^2}} \quad (16)$$

which results in the representation of the Correlated Gamma Frailty model given as

$$S(t_1,.t_2) = \frac{S_1(t_1)^{1-\frac{\sigma_1}{\sigma_2}\rho} S_2(t_2)^{1-\frac{\sigma_2}{\sigma_1}\rho} e^{\beta X_{ij}}}{\left(S_1(t_1)^{\sigma_1^2} + S_2(t_2)^{\sigma_2^2} - 1\right)^{\frac{\rho}{\sigma_1\sigma_2}}},. \quad (17)$$

The Proposed Model – Modified Gamma Frailty (MGF) Model

In order to include heterogeneity in the model, we assume a correlated gamma frailty model. Let Z_j (j = 1*;* 2) be the frailties, andX_j (j = 1*;* 2) vectors of observable covariates of the two individuals of a twin pair. Assume that their individual hazards are represented by the proportional hazards model

$$\lambda(t) = Z_j \lambda_0(t)\, exp\{\beta^T X_j\} (j = 1,2) \quad (18)$$

with a baseline hazard function $\lambda_0(t)$ describing the risk of respiratory infection as a function of age and β denotes the vector of regression parameters. Let the lifetimes of the two twin partners be conditionally independent given their frailties Z_1 and Z_2. Because frailties Z_j(j = 1*;* 2) are usually unobservable, their correlation coefficient used cannot be estimated directly from the empirical data. So a bivariate lifetime model which allows indirect calculation of the parameters is needed. The unconditional bivariate survival function of the correlated gamma frailty model with observed covariates is given by:

$$S(t_1,t_2|X_1,X_2) = S(t_1\,|X_1)^{1-\rho} S(t_2\,|X_2)^{1-\rho} S(t_1\,|X_1)^{-\sigma^2} + \{S(t_2|X_2)^{-\sigma^2} - 1\}^{-\frac{\rho}{\sigma^2}} \quad (19)$$

Where $S(t\ \ |X)$ denotes the marginal univariate survival function, assumed to be equal for both partners in a twin pair. Using a parametric approach we fit a model to the data, such that

$$S(t|X_{ijk}) = \left(1 + [(1+\sigma_1^2\frac{a}{b}(e^{bt}-1))^{\frac{\sigma_2^2}{\sigma_1^2}\rho} - 1]\right)^{-\frac{1}{\sigma_1^2}} e^{\beta_k X_{ijk}} \quad (20)$$

Where a, b,σ_1^2 , σ_2^2, β and ρ are parameters to be estimated. The lifetimes are assumed to be independently censored from the right by independent and identically distributed pairs of non-negative random variables, which are independent of the lifetimes. Thus, observe

$$(T_{i1}, T_{i2}, \Delta_{i1}, \Delta_{i2}, X_{i1}, X_{i2}) \quad (21)$$

with $\Delta_{i1}(i = 1,2,\dots n; j = 1,2)$ as a binary variable with values 1 (event) and 0 (no event). Let the lifetimes follow a distribution (dependent on covariates X_1,X_2) given by the bivariate survival function

$$S(t_1,t_2|X_1,X_2) = P(T_{i1} > t_1, T_{i2} > t_2|X_1,X_2) \quad (22)$$

Starting from this model, we are able to derive the likelihood function given by

$$L(t_1,t_2,\delta_1,\delta_2,X_1,X_2) = \delta_1\delta_2 S_{t_1t_2}(t_1,t_2|X_1,X_2) - \delta_1(1-\delta_2)S_{t_1}(t_1,t_2|X_1,X_2) - (1-\delta_1)\delta_2 S_{t_2}(t_1,t_2|X_1,X_2) + (1-\delta_1)(1-\delta_2)S(t_1,t_2|X_1,X_2) \quad (23)$$

Partial derivatives of the marginal survival functions are given by

$$S_{t_j}(t_1,t_2) = \frac{\partial S(t_1,t_2)}{\partial t_j} \quad (j = 1, 2) \quad (24)$$

and

$$S_{t_1,t_2}(t_1,t_2) = \frac{\partial S(t_1,t_2)}{\partial t_1\, \partial t_2} \quad (25)$$

The model is called the **Modified Gamma Frailty** (MGF) Model.

Numerical Illustration

An application of the models to an existing data on occupational exposure tagged – IRANIAN data is demonstrated here. Relationships between occupational exposures and morbidity, morbidity and job category were analyzed using proportional hazard analysis, allowing for exposure status (never exposed, ever smoked and ever exposed) until the time of carrying out the study. The survival-analysis was performed using the Python programming. Below is Python code framework to estimate parameters in the context of a gamma frailty model using the survival package:

```
pip install lifelines

import numpy as np
import pandas as pd
from lifelines import CoxPHFitter

# Simulate some survival data with gamma frailty (simplified for illustration)
np.random.seed(123)
n = 200
clusters = np.repeat(np.arange(n // 2), 2)
frailty_effect = np.repeat(np.random.gamma(1, 1, n // 2), 2)
baseline_hazard = 0.05
time = -np.log(np.random.uniform(0, 1, n)) / (baseline_hazard * frailty_effect)
censoring_time = np.random.exponential(1/0.03, n)
event = (time <= censoring_time).astype(int)
obs_time = np.minimum(time, censoring_time)
data = pd.DataFrame({'id': clusters, 'obs_time': obs_time, 'event': event})

# Fit a Cox Proportional Hazards model with clusters
cph = CoxPHFitter().fit(data, 'obs_time', 'event', cluster_col='id')
print(cph.print_summary())

pip install statsmodels lifelines

import numpy as np
import pandas as pd
```

```
import statsmodels.api as sm
from lifelines.datasets import load_dd

# Load a sample dataset from lifelines for demonstration
data = load_dd()

# Here, the dataset has the columns 'duration' and 'observed'.
# 'duration' is the observed duration and 'observed' is a binary column indicating if the event was observed or not.

# To fit a PH model with gamma shared frailty:
fml = "duration ~ age + education + + np.log(1 + income) + np.log(1 + income)**2"
phf = sm.PHReg(data["duration"], data[fml], status=data["observed"], ties="efron", strata=data["strata"])
result = phf.fit()
print(result.summary())

# The output will provide parameter estimates, p-values
```

The discrete algorithm was used, since the time-scale (person-years) was discrete. All exposures were first analyzed separately, allowing for age and smoking habits. Two-sided p-values < 0.05 were considered as statistically significant. The relationship between occupational exposures and morbidity was also analyzed simultaneously. Using the stepwise option of Python programming, and allowing for age and smoking habits, specific exposures were included and excluded until the following conditions were met: the significance of the residual Chi-squared was less than 0.25, and the significance of the relative risks was less than 0.10. Using the standard error of the regression coefficient, the 95% confidence intervals were estimated.

The Python programming was also applied in analyzing the Correlated Gamma Frailty Model and the Modified Gamma Frailty Model. Hazard function and survival functions for the exposure data for large and small samples were estimated.

RESULTS AND DISCUSSION

Table 1 shows the results of analysis of the Iranian data and the goodness of fit table. The exponentiated coefficients in the third column of each table of the output shown are interpretable as multiplicative effects on the hazard. In tables 1, for example, holding the other covariates constant, one additional year of age increases the yearly hazard of exposure of worker by a factor of $e^{\beta} = 1.053376$ on average – that is, by 5.3 percent. Similarly, each Forced Ventilatory Function (**FVC**) factor increases the hazard by a factor of 1.059079 or 5.9 percent.

The fifth column is the result of the test of significance of β using the Wald Statistic which is the ratio of the coefficients to the standard error of β. The obtained value is compared with the Z value and a decision is made.

In table 2, holding the other covariates constant, an additional year of age increases the yearly hazard of exposure of worker by a factor of $e^{\beta} = 1.034585$ on average – that is, by 3.5 percent. Similarly, each FVC factor increases the hazard by a factor of 1.001301 or 0.1 percent.

In table 3, holding the other covariates constant, an additional year of age increases the yearly hazard of exposure of worker by a factor of $e^{\beta} = 1.053481$ on average – that is, by 5.3 percent. Similarly, each FVC factor increases the hazard by a factor of 1.062155 or 6.2 percent.

The exposure status (never exposed, exposed and ever smoked), Job category and pack years smoked is considered to be insignificant for the Iranian data using the Cox Model. The CGF captures the exposure status and Job category to be insignificant for the Iranian data while the proposed MGF considers all the variables to be significant for the Iranian data.

Table 1: Regression Coefficients in the Cox Model for the Iranian Study

Covariate	coeff(β)	Exp(coeff(β))	Std error coeff(β)	Z	P	95% C.I for coeff(β)
AGE	0.0520	1.053376	0.0227	2.290749	0.0010	1.0212 - 1.0828
BMI	0.0555	1.057069	0.0119	4.663866	0.0001	1.0321 - 1.0788
EXPOSURE STATUS	- 0.1498	0.86088	0.2122	-0.70594	0.2145*	0.1531 - 1.9214
JOB CATEGORY	0.4337	1.542956	0.3819	1.135638	0.1041*	0.2009 - 1.6029
SYST B P	0.0915	1.095817	0.0286	3.199301	0.0029	1.0899 - 1.1103
PACK YRS SMOKED	- 0.2038	0.815625	0.1914	-1.06479	0.9856*	0.2133 - 1.9339
FVC	0.0574	1.059079	0.0221	2.597285	0.0085	1.0444 - 1.0679
FEV_1	0.0849	1.088608	0.1956	0.434049	0.0007	1.0415 - 1.0940

* Not significant.

Table 2: Regression Coefficients in the Correlated Gamma Frailty Model for the Iranian Study

Covariate	coeff(β)	Exp(coeff(β))	Std error coeff(β)	Z	P	95% C.I for coeff(β)
AGE	0.0340	1.034585	0.0154	2.207792	0.0011	1.0112 - 1.0542
BMI	0.0437	1.044669	0.0231	1.891775	0.0053	1.0221 -1.0658
EXPOSURE STATUS	- 0.0249	0.975407	0.0713	-0.34923	0.6714*	0.0131 -1.0224
JOB CATEGORY	0.0023	1.002303	0.0044	0.522727	0.1304*	0.0019-1.0329
SYST B P	0.0021	1.002102	0.0009	2.333333	0.0039	1.0001 -1.0030
PACK YRS SMOKED	- 0.2138	0.80751	0.0371	-5.7628	0.0035	0.033 - 0.9339

FVC	0.0013	1.001301	0.0006	2.16667	0.0018	1.0003 - 1.0489
FEV_1	0.0362	1.036863	0.0146	2.479452	0.0029	1.0235 -1.0440

* Not significant.

Table3: Regression Coefficients in the Modified Gamma Frailty Model for the Iranian Study

Covariate	coeff(β)	Exp(coeff(β))	Std error coeff(β)	Z	P	95% C.I for coeff(β)
AGE	0.0521	1.053481	0.0167	3.11976	0.0011	1.0215 - 1.0848
BMI	0.0547	1.056224	0.0111	4.927928	0.0001	1.0324 - 1.0798
EXPOSURE STATUS	- 0.1049	0.900415	0.2013	-0.52111	0.0045	0.0312 -0.9814
JOB CATEGORY	0.3987	1.489887	0.4009	0.994512	0.0041	1.0009-1.6029
SYST B P	0.0891	1.09319	0.0271	3.287823	0.0021	1.0021 -1.5003
PACK YRS SMOKED	- 0.2038	0.815625	0.1717	-1.18695	0.0056	0.633 - 0.9939
FVC	0.0603	1.062155	0.0211	2.85782	0.0015	1.0044 - 1.2479
FEV_1	0.0762	1.079178	0.0377	2.02122	0.0029	1.0135 -1.8340

* Not significant.

Table 4: Prognostic Factors of Occupational Exposure using Cox and frailty Models for Iranian study

Prognostic factors	Cox regression	Correlated Gamma Frailty	Modified Gamma Frailty
	HR (CI 95%)	HR (CI 95%)	HR† (CI§ 95%)
Age	1.0530 (1.0212 - 1.0828)	1.0346 (1.0112 - 1.0542)	1.053481(1.0215 - 1.0848)
BMI	1.0571(1.0321 - 1.0788)	1.0447(1.0221 -1.0658)	1.056224(1.0324 - 1.0798)
EXPOSURE STATUS	0.8609(0.1531 - 1.9214)*	0.9754 (0.0131 -1.0224)*	0.900415(0.0312 -0.9814)
JOB CATEGORY	1.5430 (0.2009 - 1.6029)*	1.0023 (0.0019-1.0329)*	1.48989(1.0009-1.6029)
SYST B P	1.0958 (1.0899 - 1.1103)	1.0021 (1.0001 -1.0030)	1.0932(1.0021 -1.5003)
PACK YRS SMOKED	0.8156 (0.2133 - 1.9339)*	0.8075 (0.033 - 0.9339)	0.8156(0.633 - 0.9939)
FVC	1.0591(1.0444 - 1.0679)	1.0013 (1.0003 - 1.0489)	1.062155(1.0044 - 1.2479)
FEV_1	1.0886 (1.0415 - 1.0940)	1.0369(1.0235 -1.0440)	1.079178(1.0135 -1.8340)
AIC#	1,157	751	704

† Hazard Ratio § Confidence interval * Not significant # Akaike Information Criterion

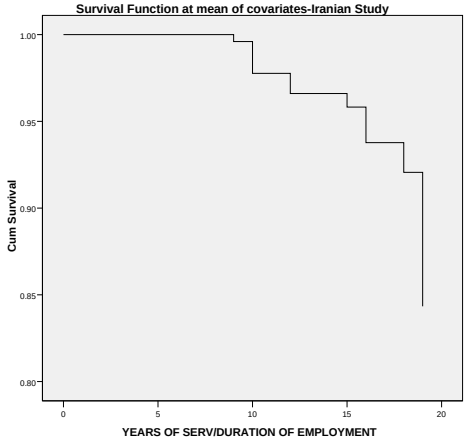

Figure 1: Survival Function at mean of covariates - Iranian Study

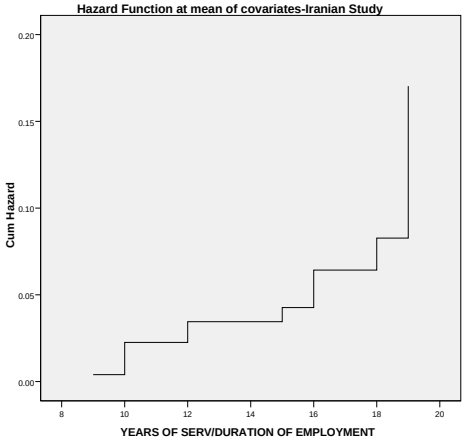

Figure 2: Hazard Function at mean covariates - Iranian Study

CONCLUSION

Interestingly, parameter estimates are quite different depending on distribution of the base-line hazard function. The newly introduced Modified Gamma frailty model offers a very elegantapproach to integrate the concept of clusters into frailty modelling. The survivalfunction is explicitly available and of easy form which allows traditional maximum likelihoodparameter estimation. This is the most important advantage of the suggested modelcompared to the model introduced by Moger & Aalen (2005). Our simulation study revealed insights into the properties of the estimator under the modified gamma frailty model.

The present work contributes to three aspects of Frailty models with censored data. First, It presents several important extensions of the existing models. Secondly, It develops a general asymptotic theory for theFrailty models. Thirdly, It provides simple and efficient numerical method to implement the corresponding inference procedures. It is hoped that this work will facilitate further development and applications of Frailty models.

It has been demonstrated that the MGF is a very generaland powerful approach to the analysis of Frailty models with censored data.This approach can be used to study many other problems. Of great interest would be a non-parametric version

of the correlated compound Poisson frailty model, where the baseline hazard functions are not specified. A part of future research is envisaged in this direction. Another aspect that will be of interest for further research is the problem of identifiability. The identifiability problem is growing with increased censoring, but is reduced by the parametric modelling of the baseline hazard. This study furnishes a structured approach for optimal experiment design using the modified gamma frailty distribution, supported by a demonstrative Python-based simulation.

REFERENCES

Abdulazeez, S.A (2020). An Optimal Design for Inference via Modified Gamma FrailtyDistribution.*KASU Journal of Mathematical Sciences (KJMS)* VOL.1, ISSUE 1.

Anthony I. W.,, Isaac D. E., Victor A. K., (2019): Survival Analysis of Under –five Mortality and Its Associated Determinants in Nigeria: Evidence from a Survey Data *International Journal of Statistics and Applications* 9(2): 59-66

Cox, D.R. (1972). Regression models and life tables, *Journal of the Royal StatisticalSociety, B* **34**, 187-400.

Getachew Y. & Bekele S., (2016). Survival Analysis of Under-Five Mortality of Children and its Associated Risk Factors in Ethiopia. *J Biosens Bioelectron*7: 213.

Iachine, I., Holm, N., Harris, J., Begun, A., Iachina, M., Laitinen, M., Kaprio, J., Yashin, A. (1998). How heritable is individual susceptibility to death? The results of an analysis of survival data on Danish, Swedish and Finnish twins. *Twin Research***1**, 196 – 205

Iachine, I. (2002). The Use of Twin and Family Survival Data in the Population Studies of Aging: Statistical Methods Based on Multivariate Survival Models. Ph.D. Thesis. Monograph 8, Department of Statistics and Demography, University of Southern Denmark.

Kalbfleisch, J. D. and R.L. Prentice, (1980). The statistical analysis of failure timedata(John Wiley & Sons Inc., New York).

Kiefer, N.M., (1988). Economic duration data and hazard functions, *Journal of Economic Literature* **26**, 646-679.

Lancaster, T., (1990). The econometric analysis of transit data (Cambridge UniversityPress, Cambridge, U.K).

Lee Y. & Song J.K. (2001). Hierarchical likelihood approach for frailty models. *Biometrika* 88(1):233–33

Longini, I. M., and Halloran, M. E. (1996). A frailty mixture model for estimating vaccineefficacy. *Applied Statistics*, **45**, 165-173.

Moger, T. A., and Aalen, O. O. (2005). A distribution for multivariate frailty based on the compound poisson distribution with random scale. *Lifetime Data Analysis*, ***11***,41-59.

Petersen J. H (1998). An Additive Frailty Model for Correlated Life Times," *Biometrics* **54** : 646-661.

Pickles, A., Crouchley, R.,Simono, E., Eaves, L., Meyer, J., Rutter, M., Hewitt, J., Silberg,J. (1994). Survival models for developmental genetic data: age of onset of puberty and antisocial behavior in twins. *Genetic Epidemiology***11**, 155 - 170

Price, D. L., and Manatunga, A. K. (2001). Modelling survival data with a cured fraction using frailty models.*Statistics in Medicine, 20*, 1515-1527.

Rakhmawati T.W, Ha I.D, Lee H, Lee Y.(2021). Penalized variable selection for cause-specific hazard frailty models with clustered competing-risks data. *Stat. Med*. 40(29):6541–57

Rueten-Budde A.J, Putter H., Fiocco M. (2019). Investigating hospital heterogeneity with a competing risks frailty model. *Stat. Med*. 38(2):269–88

Vermunt, J.K., (1996). Log-linear event history analysis (Tilburg University Press,Netherlands).

Vilcassim, N. J. and D.C. Jain, (1991). Modeling purchase timing and brand switchingbehavior incorporating explanatory variables and unobserved heterogeneity,*Journal of Marketing Research*, **28**, 29-41.

Wedel, M., W.A. Kamakura, W.S. DeSarbo and F. ter Hofstede, (1995). Implicationsfor asymmetry, non proportionality and heterogeneity in brand switching modelsfrom piece-wise exponential mixture hazard models, *Journal of MarketingResearch*, **32**, 457-463.

Wienke, A., Arbeev, K., Locatelli, I., Yashin, A.I. (2005). A comparison of different correlated frailty models and estimation strategies. *Mathematical Biosciences***198**, 1 – 13

Wienke, A., Christensen, K., Holm, N., Yashin, A. (2000). Heritability of death from respiratory diseases:an analysis of Danish twin survival data using a correlated frailty model. IOS Press, Amsterdam.

Wienke, A., Holm, N., Skytthe, A., Yashin, A.I. (2001). The heritability of mortality due toheart diseases: a correlated frailty model applied to Danish twins. *Twin Research***4**, 266 - 274

Wienke, A., Christensen, K., Skytthe, A., Yashin, A.I. (2002). Genetic analysis of cause of death in a mixture model with bivariate lifetime data. *Statistical Modelling* **2**, 89 - 102

Wienke, A., Lichtenstein, P., Yashin, A.I. (2003a). A bivariate frailty model with a cure fraction for modelling familial correlations in diseases. *Biometrics***59**, 1178 – 1183

Wienke, A., Holm, N., Christensen, K., Skytthe, A., Vaupel, J., Yashin, A.I. (2003b). Theheritability of cause-specific mortality: a correlated gamma-frailty model applied to mortality due to respiratory diseases in Danish twins born 1870 - 1930. *Statistics in Medicine* **22**, 3873 - 3887

Yashin, A.I., Iachine, I.A. (1995a). Genetic analysis of durations: Correlated frailty model applied to survival of Danish twins. *Genetic Epidemiology***12**, 529 – 538

Yashin, A.I., Iachine, I.A. (1995b). Survival of related individuals: an extension of some fundamental results of

heterogeneity analysis. *Mathematical Population Studies***5**, 321-39

Yashin, A.I., Manton, K.G., Iachine, I.A. (1996). Genetic and environmental factors in durationstudies: multivariate frailty models and estimation strategies. *Journal of Epidemiology and Biostatistics***1**, 115 – 120

Yashin, A.I., Iachine, I.A. (1997). How frailty models can be used for evaluating longevity limits:Taking advantage of an interdisciplinary approach. *Demography***34**, 31 - 48

Yashin, A.I.,Iachine, I. (1999a). Dependent hazards in multivariate survival problems. *Journal of Multivariate Analysis***71**, 241 – 261

Yashin, A.I., Iachine, I. (1999b). What difference does the dependence between durations make? Insights for population studies of aging. *Lifetime Data Analysis***5**, 5 – 22

Zdravkovic, S., Wienke, A., Pedersen, N.L., Marenberg, M.E., Yashin, A.I., de Faire, U. (2002). Heritability of death from coronary heart disease: a 36 years follow-up of 20,966 Swedish twins.*Journal of Internal Medicine***252**, 247 – 54

Zdravkovic, S., Wienke, A., Pedersen, N.L., Marenberg, M.E., Yashin, A.I., de Faire, U. (2004). Genetic influences on CHD-death and the impact of known risk factors: Comparison of two frailty models.*Behavior Genetics***34**, 585 – 591

Zhu R., Kosorok M.R. (2012). Recursively imputed survival trees. *J. Am. Stat. Assoc.* 107(497):331–40