Hacettepe Journal of Mathematics and Statistics
Volume 45 (5) (2016), 1621 – 1640

A Bayesian approach to Cox-Gompertz model

Nihal Ata Tutkun*† and Haydar Demirhan‡

Abstract

Survival analysis has a wide application area from medicine to marketing and Cox model takes an important part in survival analysis. When the distribution of survival data is known or it is appropriate to assume a survival distribution, use of a parametric form of Cox model is employed. In this article, we take into account Cox-Gompertz model from the Bayesian perspective. Considering the difficulties in parameter estimation in classical setting, we propose a simple Bayesian approach for Cox-Gompertz model. We derive full conditional posterior distributions of all parameters in Cox-Gompertz model to run Gibbs sampling. Over an extensive simulation study, estimation accuracies of the classical Cox model and classical and Bayesian settings of Cox-Gompertz model are compared with each other by generating exponential, Weibull, and Gompertz distributed survival data sets. Consequently, if survival data follows Gompertz distribution, most accurate parameter estimates are obtained by the Bayesian setting of Cox-Gompertz model. We also provide a real data analysis to illustrate our approach. In the data analysis, we observe the importance of use of the most accurate model over the survival probabilities of censored observations.

Keywords: Gompertz, Cox model, Gibbs sampling, Bayesian analysis, full conditional, Newton-Raphson, parametric model.

2000 AMS Classification: 62F15, 62N01, 62N02.

Received : 01.12.2014 *Accepted* : 29.07.2015 *Doi* : 10.15672/HJMS.20158012506

*Hacettepe University Department of Statistics Beytepe Ankara Turkey,
Email: nihalata@hacettepe.edu.tr
†Corresponding Author.
‡Hacettepe University Department of Statistics Beytepe Ankara Turkey,
Email: haydarde@hacettepe.edu.tr

1. Introduction

Survival analysis is a class of statistical methods for studying occurrence and timing of events. An event can be defined as development of a disease, response to a treatment, relapse, or death. Therefore, the time from start of a treatment to response, length of remission, and time to death may be taken as a survival time. The most common approach to model covariate effects on survival times is the Cox's semi-parametric regression model, which takes into account the effect of censored observations [5]. In the Cox model, no particular form of probability distribution is assumed for survival times. However, if it is known, parametric models, such as exponential, Weibull, or Gompertz can be applied.

The Cox model is sensitive to the violations of proportional hazards assumption. The form of baseline hazard rate influences the properties of estimators [2]. Because there is no need to assume a particular form of probability distribution for the survival times, the Cox model is more advantageous than the parametric counterparts if baseline hazard is incompatible with a particular distribution. Hazard function is not restricted to a specific functional form; hence, the model has flexibility and widespread applicability. On the other hand, if the assumption of a particular probability distribution is appropriate for data, inferences based on such an assumption will be more precise. In particular, parameter estimates and estimates of quantities such as relative hazards and median survival times will tend to have smaller standard errors than those obtained without a distributional assumption [4]. Based on asymptotic results, Efron [7] and Oakes [29] showed that parametric models lead to more efficient parameter estimates than the Cox model under certain circumstances [28].

Making special assumptions on the distribution of survival times, such as exponential, Weibull, or Gompertz, leads to parametric regression models. Exponential distribution is widely used in survival studies. It plays a role in lifetime studies analogous to normal distribution in other areas of statistics. It is often referred as purely random failure pattern [26]. Although exponential distribution is characterized by a constant hazard function, its constant hazard rate appears to be restrictive in both health and industrial applications [22]. Weibull distribution is a generalization of exponential distribution. It has a hazard function that is monotone increasing, decreasing, or constant. Therefore, it has broader applications. Although use of exponential or Weibull model may be sufficient for a realistic description of various survival time data, other distributions such as Gompertz are required for more precise results. Gompertz distribution is used to describe mortality curves and later modified by Makeham [27] by addition of a constant hazard function. Only exponential, Weibull, and Gompertz models have the assumption of proportional hazards with the Cox model [2]. Because of the functional form of its hazard rate, Gompertz model is more flexible than Weibull model. Also, it allows to asses the influence of independent variables on both parameters of the distribution [3].

Cox-Gompertz model has a wide application area from automobile industry to medicine. Gompertz distribution is commonly used in actuary, reliability, and life testing as a survival time distribution [1]. Firstly, it is used to fit mortality tables by Gompertz [15]. Spickett and Ark [30] fitted the Gompertz distribution to dose-response data of larval tick populations. Grunkemeier et al. [16] used the Gompertz model for the survival times after a surgery for acquired hearth disease. Classical analysis of Gompertz model for cure rate models was given by Gieser, et al. [13]. Willekens [31] provided connections between the Gompertz, the Weibull and other Type-I extreme value distributions. Fabrizio [8] used Gompertz model for cabinet duration times. Klepper [23] used Gompertz distribution to estimate hazard rate models for the length of time for a particular firm stays in the market. Cantner et al. [3] used the approach of Klepper [23] for German

automobile industry. Jeong and Fine [19] and Jeong [18] used Gompertz distribution to parameterize cumulative incidence function, which is used to estimate the cumulative probability of locoregional recurrences in the presence of other competing events. Gokovali et al. [14] use Gompertz distribution to analyze the determinants of tourists' length of stay at a destination. Launder and Bender [24] developed adjusted risk difference and number needed to treat measures for use in observational studies with survival time outcomes within the framework of the Cox model taking the distribution of confounders into account. The performance of these estimators is assessed by performing Monte Carlo simulations and is also illustrated by means of data of the Dusseldorf Obesity Mortality Study. Ghitanya et al. [12] studied the maximum-likelihood estimates of the parameters by considering a progressively Type-II censored sample from the Gompertz distribution.

Estimation of parameters of Cox-Gompertz model requires use of numerical techniques such as Newton-Raphson (NR). Because NR method requires only first and second partial derivatives of likelihood function, it is very flexible. However, it is highly sensitive to the initial values, it may require a large number of iterations to converge, and it may converge to a local maximum or may not converge in some cases. NR method gives no insight into the distribution of parameters. Moreover, numerical methods such as NR are asymptotic; hence, standard deviations of parameters are obtained only approximately. These are important disadvantages of the classical setting. Another general disadvantage of the classical setting is that ML estimators need not be finite, so it can occur outside the parameter space. Considering these weaknesses, we propose use of a Bayesian approach for estimation of Cox-Gompertz model.

In survival analysis, Bayesian approaches provide a flexible tool via the Gibbs sampling when the full conditional distributions are found in a closed form. Dellaportas and Smith [6] give a Bayesian approach for proportional hazards model with baseline hazard function of exponential and Weibull distributions. Bayesian approaches to the parametric survival models have some advantages over the classical setting. In the Bayesian setting, inference is exact rather than asymptotic. It provides an entire posterior distribution for each element of the model. However, the classical setting yields a point estimate and a precision estimated via an asymptotic method. In addition, the Bayesian approach would give better estimates of variability than the likelihood analysis [9].

Bayesian approaches to some parametric forms of the Cox model are given by Kim and Ibrahim [21]. They consider Cox-Weibull and extreme value regression models, and suggest use of a uniform prior instead of the Jeffrey's. They also derive sufficient conditions for the existence of posterior moment generating functions and those of the posterior distributions to be proper in the case of Cox-Weibull and extreme value regression models. Kim and Ibrahim [21] give Bayesian estimation procedure for an extreme value type I distribution. In their approach, data is a log-completely observed time or log-censoring time. In this study, however, we consider the Gompertz distribution as the distribution of a completely observed or censoring time without any transformations such as log. Then, we propose a Bayesian approach to Cox-Gompertz model. Although the distributional forms of extreme value and Gompertz distributions are similar, their domains are not the same (see for the distributional forms Bender et al. [2] and Kim and Ibrahim [21]. In fact, there are several distributional forms of Gompertz distribution [20, p.25-26, 81-85]. The one used here can be interpreted as a truncated extreme value type-I distribution. Therefore, we give a Bayesian approach for a different parametric model than the one given by Kim and Ibrahim [21]. In addition, we derive full conditional posterior distributions of the model parameters. Because of not using an approximate method to generate random numbers from the full conditionals, our derivations make the application of Bayesian setting more flexible.

In Section 2 the Cox-Gompertz model is illustrated. In Section 3, Bayesian inference for the Cox-Gompertz model is demonstrated and full conditional distributions are given to derive posterior inferences by the Gibbs sampling. In Section 4, a real data analysis is presented to illustrate our approach. We observe that use of classical Cox model can produce notably different estimates of survival probabilities for censored observations. In Section 5, the simulation study on the comparison of estimation accuracies of Cox, Cox-Gompertz models in the classical setting, and Cox-Gompertz model in the Bayesian setting is given. In Section 6, a short discussion is given.

2. The Cox-Gompertz model

A data set, based on a random sample of size n, consists of $(t_j, \delta_j, \boldsymbol{x}_j)$ for $j = 1, \ldots, n$, where t_j is the time on study for the jth individual, δ_j is the event indicator taking 1 if the event has occurred and 0 otherwise, and $\boldsymbol{x}_j$ is the vector of covariates or risk factors for the jth individual. Hazard function for the Cox model is given as follows:

$$h(t;x) = h_0(t)\exp\{\boldsymbol{X}\boldsymbol{\beta}\}, \tag{1}$$

where $\boldsymbol{X}$ is the design matrix including categorical variables or continuous measurements of each individual, $h_0(t)$ is the baseline hazard function obtained for an individual with $x_{ji} = 0$, and $\boldsymbol{\beta}_{[p\times 1]}$ is a vector of unknown parameters. In the absence of tied observations, complete censored-data likelihood is given as follows:

$$\begin{aligned} L(\boldsymbol{\beta}, h_0(t)) &= \prod_{j=1}^{n} h_0(t_j|\boldsymbol{x}_j)^{\delta_j} S(t_j|\boldsymbol{x}_j) \\ &= \prod_{j=1}^{n} h_0(t_j)^{\delta_j} [\exp\{\boldsymbol{\beta}'\boldsymbol{x}_j\}]^{\delta_j} \exp\big\{ - H_0(t_j)\exp\{\boldsymbol{\beta}'\boldsymbol{x}_j\}\big\}, \end{aligned} \tag{2}$$

where $H_0(t)$ is cumulative baseline hazard function and $S(t_j|\boldsymbol{x}_j)$ is survival function [22].

Under Gompertz distribution, the baseline hazard function is defined as follows:

$$h_0(t) = \lambda\exp\{\alpha t\}, \tag{3}$$

where $0 < t \leq \infty$, $\lambda > 0$ is a scale and $-\infty < \alpha < \infty$ is a shape parameter. Cumulative baseline hazard function is as following:

$$H_0(t) = (\lambda/\alpha)[\exp\{\alpha t\} - 1]. \tag{4}$$

Using (3), (4) and the general likelihood function given in (2), likelihood function of the Gompertz model is obtained as following:

$$\begin{aligned} L(h_0(t), \boldsymbol{\beta}|\boldsymbol{t}) \propto \prod_{j=1}^{n} \lambda^{\delta_j} \exp\big\{\delta_j(\alpha t_j + \boldsymbol{\beta}'\boldsymbol{x}_j)\big\} \exp\big\{(\lambda/\alpha)[1 - \exp(\alpha t_j)] \\ \times \exp(\boldsymbol{\beta}'\boldsymbol{x}_j)\big\}. \end{aligned} \tag{5}$$

NR method is a frequently used method to obtain the ML estimates over (5).

3. Bayesian setting for the Cox-Gompertz model

The likelihood function given in (5) is used to obtain a posterior distribution. We consider use of an improper prior distribution to conduct a noninformative Bayesian analysis. Joint prior distribution of $h_0(t)$ and $\boldsymbol{\beta}$ is taken as $p(h_0(t), \boldsymbol{\beta}) \propto \textit{constant}$. Then the joint posterior distribution of $h_0(t)$ and $\boldsymbol{\beta}$ given the data is found from (5) as follows:

$$p(h_0(t), \boldsymbol{\beta}|\boldsymbol{t}) \propto \lambda^{\sum_{j=1}^{n}\delta_j} \exp\left\{\sum_{j=1}^{n}\delta_j(\alpha t_j + \boldsymbol{\beta}'\boldsymbol{x}_j) + (\lambda/\alpha)\sum_{j=1}^{n}[1 - \exp(\alpha t_j)]\exp(\boldsymbol{\beta}'\boldsymbol{x}_j)\right\}. \tag{6}$$

where $p(h_0(t), \boldsymbol{\beta}|\boldsymbol{t}) = p(\alpha, \lambda, \boldsymbol{\beta}|\boldsymbol{t})$.

Gibbs sampling is employed to draw posterior inferences from the posterior given in (6). Full conditional posterior distributions of α, λ, β_i are required to run the Gibbs sampling. The following full conditionals are obtained:

$$\alpha|\lambda, \boldsymbol{\beta} \sim N\left[\frac{3\sum_{j=1}^{n}(\delta_j t_j - \lambda\frac{t_j^2}{2}e^{\boldsymbol{\beta}'\boldsymbol{x}_j})}{\lambda\sum_{j=1}^{n} t_j^3 e^{\boldsymbol{\beta}'\boldsymbol{x}_j}}, \frac{3}{\lambda\sum_{j=1}^{n} t_j^3 e^{\boldsymbol{\beta}'\boldsymbol{x}_j}}\right], \tag{7}$$

$$\lambda|\alpha, \boldsymbol{\beta} \sim Gamma\left[\sum_{j=1}^{n}\delta_j + 1, \frac{\alpha}{\sum_{j=1}^{n}[\exp(\alpha t_j) - 1]\exp(\boldsymbol{\beta}'\boldsymbol{x}_j)}\right], \tag{8}$$

$$\beta_i|\alpha, \lambda, \boldsymbol{\beta}_{-i} \sim N\left[\frac{s_2}{s_1}, \frac{1}{s_1}\right], \tag{9}$$

where $\boldsymbol{\beta}_{-i}$ contains the regression parameters but β_i, $s_1 = (\lambda/\alpha)\sum_{j=1}^{n} c_j x_{ji}^2\ [\exp(\alpha t_j) - 1]$, and $s_2 = \sum_{j=1}^{n}\delta_j x_{ji} - (\lambda/\alpha)\sum_{j=1}^{n} c_j x_{ji}[\exp(\alpha t_j) - 1]$. Derivation of all of these full conditionals are given in the Appendices A1-A3. Implementation of Gibbs sampling using these full conditional distributions is straightforward. Number of iterations is determined such that achievement of convergence is ensured. Convergence check can be made by using the potential scale reduction factor, $\widehat{R}$, given by Gelman [11]. If value of $\widehat{R}$ is close to 1 and less than 1.2 then it is concluded that the convergence is achieved for the relevant parameter [11].

The sufficient conditions for the existence of posterior moment generating function of the model parameters and the propriety of the posterior distribution are mentioned by Kim and Ibrahim [21] for the Weibull and extreme value distribution cases. Kim and Ibrahim [21] assume that one of the parameters of hazard function, corresponding to Weibull distribution, is known; and hence, one of the parameters of hazard function in the extreme value distribution case is also assumed to be known. In addition, they note that if these do not assumed, joint posterior distributions are always improper. On the contrary, all of the parameters of the hazard function of the Gompertz distribution that we are working on are random. Thus, the propriety of our joint posterior distribution is uncertain when looked from the perspective of Kim and Ibrahim [21]. Gelfand and Shau [10] state that if a Gibbs sampler is used on the improper joint posterior, it is possible to use obtained iterates to draw inferences on the lower-dimensional proper posteriors. As a result, if full conditionals are proper, foregoing transition density remains valid. When the full conditionals given in (7)- (9) are investigated, it is seen that they are proper if α and λ are both finite. Therefore, we do not need to ascertain propriety of our joint posterior distribution in another way. Instead, we utilize directly the result given by Gelfand and Shau [10] due to the propriety of the full conditionals.

4. A real data example

A popular data set is taken into account to illustrate and discuss our findings. The data is on lung cancer and given by Lawless [25]. The data set is also used by Gelfand and Mallic [9] and Kim and Ibrahim [21]. Gelfand and Mallic [9] used the data set to illustrate their work on Cox model, for which the baseline hazard, the covariate link, and the covariate coefficients are all unknown. Thus, they investigated four models from the Bayesian perspective. Kim and Ibrahim [21] gave the ML and Bayesian estimates using a uniform prior under the Cox-Weibull model by including an intercept term and assuming one of the parameters of the hazard function is known.

The data set consists three covariates that performance status at diagnosis (measure between 0 and 100), age of patients in years, and months from diagnosis to entry into the study. Three of the 40 observations are censored. There are 3 tied observation pairs. One of them includes one censored and one uncensored observations. The censored one and one of the other two tied pairs were discarded from the data set. These tied observations were not noticed by Gelfand and Mallic [9] and Kim and Ibrahim [21]. In addition they do not mention anything about the tied observations. We fit Cox-Gompertz model under the Bayesian setting. In Gibbs sampling, total number of iterations was taken as 2500, and 10 parallel chains were generated. To filtrate the effect of starting values, burn-in period was taken as the first 500 iterations of each chain. Every 25 iterations were recorded to reduce the autocorrelation in each of the chains. Parameter estimates with their estimated standard deviations for the Cox model in classical setting and the Cox-Gompertz model in both of the classical and Bayesian settings, and potential scale reduction factor, corresponding to each parameter are given by Table 1.

Table 1. Classical and Bayesian parameter estimates (estimated standard deviations) over Cox and Cox-Gompertz models, and values of potential scale reduction factor ($\widehat{R}$) values.

	Classical Estimates		Bayesian Estimates	
	Cox Model	Cox-Gompertz Model	Cox-Gompertz Model	$\widehat{R}$
α		$0.0003\ (7.42 \cdot 10^{-7})$	$-0.0019\ (2.65 \cdot 10^{-7})$	1.003
λ		$0.0196(2.70 \cdot 10^{-7})$	$0.0331\ (5.41 \cdot 10^{-3})$	1.001
β_1	$-0.0130(1.13 \cdot 10^{-4})$	$-0.0504\ (9.12 \cdot 10^{-5})$	$-0.0121\ (2.86 \cdot 10^{-5})$	1.001
β_2	$0.0135(3.16 \cdot 10^{-4})$	$0.0351\ (8.79 \cdot 10^{-5})$	$-0.0076\ (2.97 \cdot 10^{-5})$	1.001
β_3	$-0.0149(1.42 \cdot 10^{-4})$	$0.0219\ (2.65 \cdot 10^{-5})$	$-0.0015\ (4.58 \cdot 10^{-7})$	1.004

$\widehat{R}$ values indicate that the convergence is achieved for all of the parameters. Estimated standard deviations given in Table 1 are obtained by using inverse of the Hessian matrix and the generated Gibbs sequence in the classical and Bayesian settings, respectively. It is seen from the Table 1 that estimated standard deviations of the parameters of Cox-Gompertz model are smaller than that of the Cox model in both of the classical and Bayesian settings. ML and the Bayesian estimates are not far from each other. The Bayesian estimates of the covariate coefficients, which are more precise, are closer to that of the classical Cox model.

To investigate which model is more successful in explaining the censoring, we estimate $P(t_{12} > 231|\boldsymbol{x}_{12})$, $P(t_{15} > 103|\boldsymbol{x}_{15})$ and $P(t_{23} > 25|\boldsymbol{x}_{23})$ over the considered models, where $\boldsymbol{x}_{12}$, $\boldsymbol{x}_{15}$ and $\boldsymbol{x}_{23}$ are the observed values of covariates corresponding to the relevant censored observations. The same approach of Gelfand and Mallic [9] is used to calculate the probabilities in the Bayesian case. ML estimates of the Cox-Weibull model given by Kim and Ibrahim [21] are used. The results and product of these probabilities, referred as overall, are given in Table 2.

Benefit of the parametric approach for this data set is clearly seen in the Table 2 that Cox-Gompertz model is better than the classical Cox model in the estimation of censored survival times. Cox-Weibull model is also unsuccessful. This is an example of the case that the baseline hazard is not compatible with the parametric distribution. The Cox-Gompertz model seems to be more successful in the estimation of survival probabilities in both of settings. When the classical and Bayesian settings of Cox-Gompertz model are compared, the probabilities obtained over the classical estimates for the survival times of 25 and 103 are greater than their Bayesian counterparts. However, the case is just the

Table 2. Survival probabilities for censored observations.

	Classical Estimates			Bayesian Estimates
	Cox	Cox-Gompertz	Cox-Weibull	Cox-Gompertz
$P(t_{12} > 231 \mid \boldsymbol{x}_{12})$	< 0.0001	0.0619	< 0.0001	0.1122
$P(t_{15} > 103 \mid \boldsymbol{x}_{15})$	< 0.0001	0.7083	< 0.0001	0.3795
$P(t_{23} > 25 \mid \boldsymbol{x}_{23})$	< 0.0001	0.9095	< 0.0001	0.7883
Overall	< 0.0001	0.0387	< 0.0001	0.0336

reverse for the survival time of 231. Thus, the Bayesian estimates are more successful for longer survival times for the data set of interest. As for the overall performance, the Bayesian and classical estimates of Cox-Gompertz model are similar in estimating the censored survival times.

Plots of posterior marginal distributions of the parameters are given by Figure 1. Most of the probability mass of all marginal posterior densities of the parameters are less or greater than zero. And all of them are nearly symmetric. We can conclude that all of the parameters have statistically significant effects on the survival times.

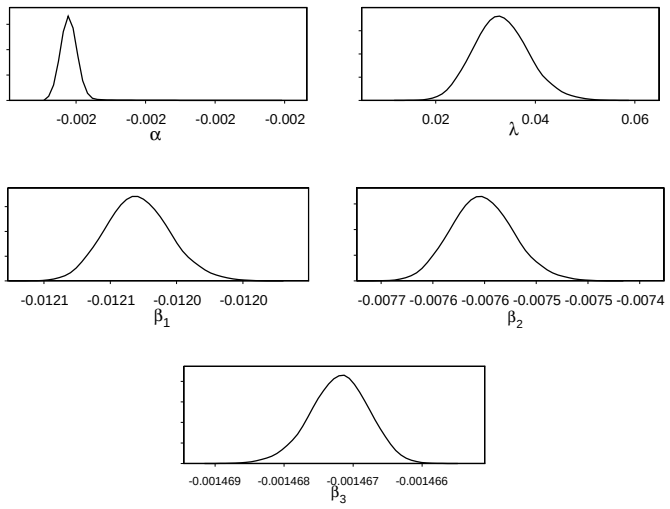

Figure 1. Marginal posterior densities of α, λ and the elements of $\boldsymbol{\beta}$.

5. Simulation study

A simulation study is conducted to investigate the features of our approach and to compare them with classical Cox and Cox-Gompertz models. Two covariates were taken into account. Values of the X_1 is generated from $N(3, 0.1)$ and values of the X_2 is generated from $N(4, 0.5)$. The survival data were generated by using formulas of (10), given by Bender, et al. [2], from the Exponential(λ), Gompertz(α, λ), and Weibull(ν, λ) distributions, respectively.

$$T_j^{\mathrm{E}} = -\frac{\log(U)}{\lambda \exp\{\beta_1 x_{j1} + \beta_2 x_{j2}\}}, T_j^{\mathrm{G}} = (1/\alpha)\log\left[1 - \frac{\alpha \log(U)}{\lambda \exp\{\beta_1 x_{j1} + \beta_2 x_{j2}\}}\right]$$
$$T_j^{\mathrm{W}} = \left[-\frac{\log(U)}{\lambda \exp\{\beta_1 x_{j1} + \beta_2 x_{j2}\}}\right]^{1/\nu}, \tag{10}$$

where $U \sim$Uniform(0,1) and β_i's, $i = 1, 2$, are regression coefficients.

To use a moderate sample size, it is taken as 20. True values of parameters for each survival distribution are given in the third columns of Tables 3-11. Censoring rate is taken as 0 and 0.1, which correspond to cases of no censoring and a moderate rate of censoring, respectively. 1000 independent samples were generated for each of the combinations. Parameter estimates, given by the Tables 3-11 were calculated by averaging the estimates over the generated 1000 samples. Absolute and relative bias, standard deviation and mean square error (MSE) values are reported in Tables 3-11.

It is seen from the Table 3, 4, and 5 that when the survival data are distributed as exponential, parameter estimates and their estimated standard deviations are not affected by the increased censoring for all of three settings. Classical parameter estimates of Cox and Cox-Gompertz models are very different from each other, and estimated standard deviations and MSEs of the parameter estimates of Cox model are smaller than that of Cox-Gompertz model. Absolute and relative biases of parameter estimates of Cox model are smaller than that of Cox-Gompertz model. Thus, Cox model generates better estimates than Cox-Gompertz model in case of exponentially distributed survival data with the classical setting. As for the Bayesian setting, it is interesting that the parameter estimates are similar in all of the cases, in addition their standard deviations and MSEs are close to zero. Absolute biases in the Bayesian setting are somewhat greater than that of Cox model in the classical setting, whereas MSEs are smaller in the Bayesian setting. The cause of this situation is smaller estimated standard deviations of the Bayesian setting. When the classical and Bayesian settings of Cox-Gompertz models are compared, it is seen that absolute and relative biases and the MSEs of the Bayesian setting are smaller than that of the classical setting. As the result, it can be stated the Bayesian approach is neither better nor worse than the classical Cox approach and better than the classical settings of Cox-Gompertz model when the data come from *exponential* distribution. The side effects of the disagreement between the survival distribution and baseline hazard is clearly seen here for the classical settings and obtained smaller variances are neutralized the side effects of the disagreement in the Bayesian setting.

It can be concluded from the Table 6, 7, and 8 that in contrast to the preceding inferences, absolute and relative biases of the parameter estimates obtained over Cox model is greater than that of Cox-Gompertz model for the cases 3 and 4 when the survival data comes from Weibull distribution. In addition, estimated standard deviations and MSEs of the model parameters obtained by Cox model are greater than that of obtained by Cox-Gompertz model for the cases 3 and 4. These situations are just reverse for the case 1. Absolute and relative biases and MSEs of the Bayesian estimates are less than that of Cox and Cox-Gompertz models both. The Bayesian parameter estimates of model parameters are also similar in all of the cases for *Weibull* distributed data. The cause of this can be the conflict between baseline hazard of the Gompertz distribution and the Weibull distributed survival data. When the survival data come from the Weibull distribution, the Bayesian setting is more successful than the classical setting.

When the survival data comes from the Gompertz distribution, see the Tables 9, 10, and 11, the smallest estimated standard deviations are generated by the classical setting of Cox-Gompertz model, whereas the smallest MSEs are given by the Bayesian setting. The classical Cox model produces the worst standard deviations and MSEs among the classical and Bayesian settings of Cox-Gompertz model. This implies that when distribution of the data and underlying baseline hazard agrees, using Cox-Gompertz model is practically reasonable. The smallest absolute biases are seen in the Bayesian setting. Relative biases of the parameter estimates generated by the classical setting of Cox-Gompertz model are greater than that of the Bayesian setting. While the classical Cox model is being affected by the increased censoring, Cox-Gompertz model generates smaller absolute biases for 0 and 0.1 censoring rates in all of the cases. The same inference is valid for the Bayesian approach in the cases 5 and 6. In general, if one has strong information on the distribution of the lifetime data are distributed as *Gompertz*, use of the Bayesian setting for Cox-Gompertz model is a practically reasonable way.

When the overall results are considered, it is concluded that when survival data come from exponential distribution, Cox model in the classical setting gives the best parameter estimates. But if the data come from Weibull distribution, parameter estimates obtained from all of the settings are not sufficient enough. Thus, a Cox-Weibull model can be applied. When the data is distributed as Gompertz, due to the smallest absolute biases and MSEs produced by the Bayesian setting, advantages of the parametric approach over Cox model and advantages of the Bayesian approach over the classical are ascertained.

When the ratio of number of data sets for which NR method were not converged to the total number of the generated data sets is considered, another advantage of the Bayesian approach is clearly seen. Proportion of unconverged iterations for Cox and Cox-Gompertz models are given in Table 12. Cases seen on the first column are the same as the cases defined in the Tables 3, 6, and 9.

It is seen from Table 12 that NR method encounters certain convergence problems for Cox-Gompertz model for exponential and Weibull distributions, because of its dependency to the starting values. Because NR method had not converged in most of the iterations, thus 1000 samples could not be obtained with reasonable number of generations; and hence , some cells of Tables 3 and 6 could not be filled. Convergence of NR method for Cox model under Weibull and Gompertz distributions were less problematic. In general, the Table 12 reflects the problematic dependency of NR method to the starting values for considered models.

6. Discussion

In this article, we consider use of Gibbs sampling to draw posterior inferences for Cox-Gompertz model, when all of the parameters of the hazard function are unknown. We derive required full conditional distributions for all parameters. All of the full conditionals are found to be familiar and proper distributions. Therefore, there is no need to use a random number generation algorithm such as rejection sampling to generate random numbers from full conditionals. This brings in a flexibility to the presented approach.

Main disadvantage of our approach is that if the survival data is not compatible with the Gompertz distribution, it is not as successful as the classical Cox model in the estimation of parameters. This situation is also observed in the simulation study. However, if this is not the case, our approach is more advantageous than Cox model and classical setting of Cox-Gompertz model. It utilizes superiorities of the Bayesian approaches over the classical counterparts, which are mentioned in the Section 1. Because we are treating

Table 3. The ML estimates of parameters of the Cox model over 1000 samples, each was generated from the exponential distribution.

Case	Par.	Values	Cens. Rate	Par.	Estimate	Absolute Bias	Relative Bias	St.Dev.	MSE
			0						
1	λ	0.071							
	β_1	-0.100		β_1	0.124	0.224	224.401	6.690	44.808
	β_2	-0.200		β_2	-0.104	0.096	47.794	1.337	1.797
			0.1						
2	λ	0.071							
	β_1	-0.100		β_1	0.089	0.189	189.043	7.461	55.700
	β_2	-0.200		β_2	-0.097	0.103	51.286	1.507	2.281
			0						
3	λ	0.071							
	β_1	-1.000		β_1	-1.120	0.120	-11.970	7.227	52.242
	β_2	-0.200		β_2	-0.329	0.129	-64.442	1.516	2.315
			0.1						
4	λ	0.071							
	β_1	-1.000		β_1	-1.026	0.026	-2.645	$> 10^4$	$> 10^4$
	β_2	-0.200		β_2	-0.499	0.299	-149.646	$> 10^4$	$> 10^4$
			0						
5	λ	0.071							
	β_1	0.500		β_1	0.248	0.252	50.452	4.071	16.637
	β_2	-1.000		β_2	-0.577	0.423	42.287	0.836	0.878
			0.1						
6	λ	0.071							
	β_1	0.500		β_1	0.103	0.397	79.368	4.793	23.129
	β_2	-1.000		β_2	-0.603	0.397	39.693	0.985	1.128

Par. : Parameter; Cens. : Censoring; St. Dev. : Estimated Standard Deviation; MSE: Mean square error.

all parameters of the hazard function as random, our approach is more precise. The convergence problems of the Gibbs sampling are not as much as NR method, as seen in the simulation study.

Gompertz distribution has many application areas, so does the Bayesian approach to Cox-Gompertz model. Moreover, the Bayesian approach makes the application of the Cox-Gompertz model easier, in all of the mentioned areas, because of the superiorities.

Appendix

A1. Derivation of full conditional distribution of α. Full conditional distribution of α given the other parameters is obtained as

$$p(\alpha|\lambda,\boldsymbol{\beta},\boldsymbol{t}) \propto \exp\left\{\sum_{j=1}^{n}\alpha\delta_j t_j + (\lambda/\alpha)\sum_{j=1}^{n}[1-\exp(\alpha t_j)]\exp(\boldsymbol{\beta}'\boldsymbol{x}_j)\right\}. \tag{11}$$

When we use Taylor expansion of $\exp(\alpha t_j)$ at 0, the following is obtained from eq. (11):

$$p(\alpha|\lambda,\boldsymbol{\beta},\boldsymbol{t}) \propto \exp\left\{\sum_{j=1}^{n}\alpha\delta_j t_j + [\lambda/\alpha]\sum_{j=1}^{n}\left[1-(T_m(\alpha,j)+R_m(\alpha,j))\right]\exp(\boldsymbol{\beta}'\boldsymbol{x}_j)\right\} \tag{12}$$

where m is the order of Taylor expansion, $T_m(\alpha,j) = 1+\alpha t_j+(\alpha^2 t_j^2)/2+\cdots+(\alpha^m t_j^m)/m!$, and $R_m(\alpha,j)$ is the reminder term of the Taylor expansion. Because each term is a function of the rv α, to obtain a tractable full conditional distribution, we need to show that the distribution of $T_m(\alpha,j)+R_m(\alpha,j)$ converges to that of T_m as $m\to\infty$. Let $X_m = \alpha^m t_j^m/m!$ be a sequence of rv's for $m = 1,2,\ldots$ and $Y = \alpha t_j$, where $Y \in \mathbb{R}$.

Table 4. The ML estimates of parameters of the Cox-Gompertz model over 1000 samples, each was generated from the exponential distribution.

Case	Par.	Values	Cens. Rate	Par.	Estimate	Absolute Bias	Relative Bias	St.Dev.	MSE
			0	α	0.022				
1	λ	0.071		λ	0.283	0.211	-295.976	0.061	0.048
	β_1	-0.100		β_1	1.377	1.477	1477.239	4.381	21.371
	β_2	-0.200		β_2	-0.036	0.164	81.815	2.638	6.985
			0.1	α	*	*	*	*	*
2	λ	0.071		λ	*	*	*	*	*
	β_1	-0.100		β_1	*	*	*	*	*
	β_2	-0.200		β_2	*	*	*	*	*
			0	α	0.010				
3	λ	0.071		λ	2.000	1.928	-2700.492	0.178	3.749
	β_1	-1.000		β_1	-6.559	5.559	-555.935	63.643	4081.362
	β_2	-0.200		β_2	-0.001	0.199	99.635	44.223	1955.697
			0.1	α	0.010				
4	λ	0.071		λ	1.997	1.925	-2696.298	0.201	3.746
	β_1	-1.000		β_1	-6.561	5.561	-556.145	64.936	4247.671
	β_2	-0.200		β_2	-0.001	0.199	99.259	45.049	2029.427
			0	α	0.010				
5	λ	0.071		λ	2.111	2.039	-2856.283	0.127	4.175
	β_1	0.500		β_1	-7.004	7.504	1500.855	5628.965	$> 10^4$
	β_2	-1.000		β_2	-0.009	0.991	99.056	4144.209	$> 10^4$
			0.1	α	0.010				
6	λ	0.071		λ	2.084	2.013	-2819.376	0.139	4.072
	β_1	0.500		β_1	-6.923	7.423	1484.535	6196.135	$> 10^4$
	β_2	-1.000		β_2	-0.004	0.996	99.553	4517.634	$> 10^4$

Par. : Parameter; Cens. : Censoring; St. Dev. : Estimated Standard Deviation.
*: 10^6 data sets had been generated, but the convergence could not be achieved for 1000 of them.
MSE: Mean square error.

For a fixed value of k, suppose $|y| < k$. Then, for all $m > k$ the following result is straightforwardly obtained:

$$|y|^{m-k} < k(k+1)(k+2)\cdots(m-1).$$

Thus,

$$0 < \frac{|y|^m}{m!} \leq \frac{|y|^k m - k}{m!} < \frac{k(k+1)(k+2)\cdots(m-1)}{m!} = \frac{|y|^k}{(k-1)!m}.$$

In terms of rv's, we have the following inequality for all values of α:

$$X_m \leq \frac{|Y|^k}{(k-1)!m}. \tag{13}$$

The definition of convergence in probability to zero is as follows:

$$\lim_{m\to\infty} P(|X_m| < \epsilon) = 1. \tag{14}$$

The inequality in (13) implies that if

$$\lim_{m\to\infty} P\left(\frac{|Y|^k}{(k-1)!m} < \epsilon\right) = 1, \tag{15}$$

then eq. (14) is ensured. Because k is a fixed constant, the limit in (15) is straightforwardly equal to one. Thus,

$$X_m \xrightarrow{p.} 0, \text{ as } m \to \infty; \text{ and hence } S_m = \sum_{i=m}^{\infty} X_i \xrightarrow{p.} 0, \text{ as } m \to \infty.$$

Table 5. The Bayesian estimates of parameters of the Cox-Gompertz model over 1000 samples, each was generated from the exponential distribution.

Case	Par.	Values	Cens. Rate	Par.	Estimate	Absolute Bias	Relative Bias	St.Dev.	MSE
			0	α	-0.004			0.002	0.000
1	λ	0.071		λ	0.246	0.175	-244.768	0.058	0.034
	β_1	-0.100		β_1	-0.371	0.271	-270.733	0.009	0.073
	β_2	-0.200		β_2	-0.271	0.071	-35.286	0.005	0.005
			0.1	α	-0.005			0.002	0.000
2	λ	0.071		λ	0.214	0.143	-200.176	0.055	0.023
	β_1	-0.100		β_1	-0.371	0.271	-270.686	0.010	0.073
	β_2	-0.200		β_2	-0.270	0.070	-35.242	0.006	0.005
			0	α	-0.002			0.000	0.000
3	λ	0.071		λ	0.016	0.055	77.072	0.005	0.003
	β_1	-1.000		β_1	-0.371	0.629	62.934	0.014	0.396
	β_2	-0.200		β_2	-0.273	0.073	-36.346	0.008	0.005
			0.1	α	-0.002			0.000	0.000
4	λ	0.071		λ	0.015	0.057	79.510	0.005	0.003
	β_1	-1.000		β_1	-0.371	0.629	62.929	0.015	0.396
	β_2	-0.200		β_2	-0.273	0.073	-36.465	0.009	0.005
			0	α	-0.002			0.001	0.000
5	λ	0.071		λ	0.051	0.021	28.831	0.013	0.001
	β_1	0.500		β_1	-0.371	0.871	174.291	0.010	0.760
	β_2	-1.000		β_2	-0.261	0.739	73.900	0.006	0.546
			0.1	α	-0.002			0.001	0.000
6	λ	0.071		λ	0.043	0.028	39.138	0.012	0.001
	β_1	0.500		β_1	-0.372	0.872	174.331	0.011	0.760
	β_2	-1.000		β_2	-0.261	0.739	73.882	0.006	0.546

Par. : Parameter; Cens. : Censoring; St. Dev. : Estimated Standard Deviation.
MSE: Mean square error.

Let h be a continuous function at zero, if $Y_m \xrightarrow{p.} 0$ as $m \to \infty$ then $h(Y_m) \xrightarrow{p.} h(0)$ as $m \to \infty$ [17, see Theorem 10.2]. Regarding this theorem, if we define $h(S_m)$ as the following:

$$h(S_m) = \exp\Big\{ -[\lambda/\alpha] \sum_{j=1}^{n} \exp(\boldsymbol{\beta}' \boldsymbol{x}_j) S_m \Big\},$$

then $h(S_m) \xrightarrow{p.} 1$, as $m \to \infty$. This result implies that the reminder term in (12) converges to 1 *in probability*; and hence, it converges to 1 *in distribution*.

Right hand-side of (12) is rewritten as follows:

$$\exp\Big\{ \sum_{j=1}^{n} \alpha \delta_j t_j + [\lambda/\alpha] \sum_{j=1}^{n} \exp(\boldsymbol{\beta}' \boldsymbol{x}_j) \big[1 - T_m(\alpha, j)\big] \Big\} \cdot h(S_m). \tag{16}$$

Because the value of n is finite, it concludes from the well-known Slutsky's theorem [17, p. 248] that the expression in (16) converges to the following:

$$\exp\Big\{ \sum_{j=1}^{n} \alpha \delta_j t_j + [\lambda/\alpha] \sum_{j=1}^{n} \exp(\boldsymbol{\beta}' \boldsymbol{x}_j) \big[1 - T(\alpha, j)\big] \Big\} \cdot 1 \tag{17}$$

in distribution as $m \to \infty$. Consequently, the distribution of the remaining expression after the application of Taylor expansion of order m converges to the distribution of original expression in eq. (11). Therefore, it is appropriate to use the Taylor expansion to derive full conditional distribution of α.

Table 6. The ML estimates of parameters of the Cox model over 1000 samples, each was generated from the Weibull distribution.

Case	Par.	Values	Cens. Rate	Par.	Estimate	Absolute Bias	Relative Bias	St.Dev.	MSE
1	ν	1.500	0						
	λ	0.015							
	β_1	-0.100		β_1	0.987	1.087	1087.272	60.900	3710.049
	β_2	-0.200		β_2	-3.607	3.407	-1703.717	12.245	161.545
2	ν	1.500	0.1						
	λ	0.015							
	β_1	-0.100		β_1	-0.070	0.030	30.043	57.020	3251.321
	β_2	-0.200		β_2	-3.170	2.970	-1484.821	11.487	140.772
3	ν	1.500	0						
	λ	0.015							
	β_1	-1.000		β_1	-5.077	4.077	-407.700	14.652	231.293
	β_2	-0.200		β_2	-1.124	0.924	-462.180	2.975	9.703
4	ν	1.500	0.1						
	λ	0.015							
	β_1	-1.000		β_1	-7.710	6.710	-670.990	19.151	411.769
	β_2	-0.200		β_2	-1.964	1.764	-882.138	3.931	18.568
5	ν	1.500	0						
	λ	0.015							
	β_1	0.500		β_1	1.446	0.946	-189.221	12.754	163.557
	β_2	-1.000		β_2	-2.107	1.107	-110.657	2.568	7.820
6	ν	1.500	0.1						
	λ	0.015							
	β_1	0.500		β_1	2.260	1.760	-351.977	15.216	234.637
	β_2	-1.000		β_2	-2.453	1.453	-145.310	3.094	11.682

Par. : Parameter; Cens. : Censoring; St. Dev. : Estimated Standard Deviation.
MSE: Mean square error.

We use the third order Taylor expansion of $\exp(\alpha t_j)$ at 0 to obtain $p(\alpha|\lambda,\boldsymbol{\beta},\boldsymbol{t})$. As the result it is obtained that

$$\begin{aligned} p(\alpha|\lambda,\boldsymbol{\beta},\boldsymbol{t}) &\propto \exp\left\{\sum_{j=1}^{n}\alpha\delta_j t_j + [\lambda/\alpha]\sum_{j=1}^{n}\left[1-(1+\alpha t_j+\frac{\alpha^2 t_j^2}{2}+\frac{\alpha^3 t_j^3}{6})\right]\exp(\boldsymbol{\beta}'\boldsymbol{x}_j)\right\} \\ &\propto \exp\left\{\frac{-1}{2}\left[\frac{\lambda\alpha^2}{3}\sum_{j=1}^{n}t_j^3 e^{\boldsymbol{\beta}'\boldsymbol{x}_j} - 2\alpha\sum_{j=1}^{n}(\delta_j t_j - \lambda\frac{t_j^2}{2}e^{\boldsymbol{\beta}'\boldsymbol{x}_j})\right]\right\} \\ &\propto \exp\left\{\frac{-1}{2}\frac{\lambda\sum_{j=1}^{n}t_j^3 e^{\boldsymbol{\beta}'\boldsymbol{x}_j}}{3}\left[\alpha^2 - 2\alpha\frac{3\sum_{j=1}^{n}(\delta_j t_j - \lambda\frac{t_j^2}{2}e^{\boldsymbol{\beta}'\boldsymbol{x}_j})}{\lambda\sum_{j=1}^{n}t_j^3 e^{\boldsymbol{\beta}'\boldsymbol{x}_j}}\right]\right\} \\ &\propto \exp\left\{\frac{-1}{2\sigma_\alpha^2}(\alpha-\mu_\alpha)^2\right\}. \end{aligned} \tag{18}$$

Then the full conditional distribution of α is obtained normal distribution with mean and variance

$$\mu_\alpha = \frac{3\sum_{j=1}^{n}(\delta_j t_j - \lambda\frac{t_j^2}{2}e^{\boldsymbol{\beta}'\boldsymbol{x}_j})}{\lambda\sum_{j=1}^{n}t_j^3 e^{\boldsymbol{\beta}'\boldsymbol{x}_j}}, \sigma_\alpha^2 = \left[\frac{\lambda\sum_{j=1}^{n}t_j^3 e^{\boldsymbol{\beta}'\boldsymbol{x}_j}}{3}\right]^{-1}, \tag{19}$$

respectively.

To demonstrate appropriateness of the third order Taylor expansion, we consider the mechanism that generates survival times under the Gompertz model. Bender et al. [2] demonstrate that survival times from Gomperts(α,λ) distribution is generated by the transformation of uniformly distributed r.v. U given in eq. (10). We investigate the impact of the value of α on survival times in Gompertz model over eq. (1). Note that in

Table 7. The ML estimates of parameters of the Cox-Gompertz model over 1000 samples, each was generated from the Weibull distribution.

Case	Par.	Values	Cens. Rate	Par.	Estimate	Absolute Bias	Relative Bias	St.Dev.	MSE
1	ν	1.500	0	α	0.051	1.449	96.579	0.022	2.099
	λ	0.015		λ	0.179	0.164	-1090.546	0.044	0.029
	β_1	-0.100		β_1	1.287	1.387	1387.414	0.688	2.398
	β_2	-0.200		β_2	-0.046	0.154	76.980	0.514	0.288
2	ν	1.500	0.1	α	*	*	*	*	*
	λ	0.015		λ	*	*	*	*	*
	β_1	-0.100		β_1	*	*	*	*	*
	β_2	-0.200		β_2	*	*	*	*	*
3	ν	1.500	0	α	0.010	1.490	99.329	18.455	342.805
	λ	0.015		λ	2.222	2.207	-14713.681	0.136	4.890
	β_1	-1.000		β_1	-7.373	6.373	-637.282	5115.662	26170042.877
	β_2	-0.200		β_2	0.000	0.200	99.910	3801.955	14454860.217
4	ν	1.500	0.1	α	0.010	1.490	99.322	12.018	146.645
	λ	0.015		λ	2.178	2.163	-14417.268	0.147	4.699
	β_1	-1.000		β_1	-7.234	6.234	-623.372	3115.142	9704147.295
	β_2	-0.200		β_2	-0.001	0.199	99.713	2252.882	5075475.292
5	ν	1.500	0	α	*	*	*	*	*
	λ	0.015		λ	*	*	*	*	*
	β_1	0.500		β_1	*	*	*	*	*
	β_2	-1.000		β_2	*	*	*	*	*
6	ν	1.500	0.1	α	*	*	*	*	*
	λ	0.015		λ	*	*	*	*	*
	β_1	0.500		β_1	*	*	*	*	*
	β_2	-1.000		β_2	*	*	*	*	*

Par. : Parameter; Cens. : Censoring; St. Dev. : Estimated Standard Deviation.
*: 10^6 data sets had been generated, but the convergence could not be achieved for 1000 of them.
MSE: Mean square error.

eq. (1), $\lambda > 0$, $\exp\{\beta_1 x_{j1} + \beta_2 x_{j2}\} > 0$, and $\log(u) < 0$. Because

$$\lim_{\alpha \to -\infty} T_j^{\mathrm{G}} = \lim_{\alpha \to \infty} T_j^{\mathrm{G}} = 0,$$

survival times goes to zero for greater values of α. For smaller values of $\lambda \exp\{\beta_1 x_{j1} + \beta_2 x_{j2}\}$, the value of U should approach to one to make eq. (1) proper. Only for this case the rate of convergence of T_j^{G} to zero decreases; and hence, we can observe reasonable survival times for greater values of α. Due to the decreased range of reasonable values of U, the probability of having such a situation in practice is small. For greater values of $\lambda \exp\{\beta_1 x_{j1} + \beta_2 x_{j2}\}$, any value of U from $(0, 1)$ interval makes eq. (1) proper. In this case, values of α close to zero give reasonable survival times. Therefore, the rate of convergence will be very fast due to the small values of α; and hence, use of the third order Taylor expansion is appropriate.

A2. Derivation of full conditional distribution of λ. To derive the $p(\lambda|\alpha, \boldsymbol{\beta}, \boldsymbol{t})$, (6) is rewritten by discarding the constants as

$$p(\lambda|\alpha, \boldsymbol{\beta}, \boldsymbol{t}) \propto \lambda^{\sum_{j=1}^{n} \delta_j} \exp\left\{ -(\lambda/\alpha) \sum_{j=1}^{n} [\exp(\alpha t_j) - 1] \exp(\boldsymbol{\beta}' \boldsymbol{x}_j) \right\}. \tag{20}$$

The distribution reached in (14) is gamma with the following shape and scale parameters

$$\sum_{j=1}^{n} \delta_j + 1, \ \alpha \left[\sum_{j=1}^{n} [\exp(\alpha t_j) - 1] \exp(\boldsymbol{\beta}' \boldsymbol{x}_j) \right]^{-1}. \tag{21}$$

Table 8. The Bayesian estimates of parameters of the Cox-Gompertz model over 1000 samples, each was generated from the Weibull distribution.

Case	Par.	Values	Cens. Rate	Par.	Estimate	Absolute Bias	Relative Bias	St.Dev.	MSE
1	ν	1.500	0	α	-0.005	1.505	100.329	0.004	2.265
	λ	0.015		λ	0.342	0.327	-2179.422	0.081	0.113
	β_1	-0.100		β_1	-0.371	0.271	-270.731	0.009	0.073
	β_2	-0.200		β_2	-0.272	0.072	-35.797	0.006	0.005
2	ν	1.500	0.1	α	-0.006	1.506	100.426	0.004	2.269
	λ	0.015		λ	0.301	0.286	-1906.160	0.077	0.088
	β_1	-0.100		β_1	-0.371	0.271	-270.931	0.010	0.073
	β_2	-0.200		β_2	-0.271	0.071	-35.532	0.006	0.005
3	ν	1.500	0	α	-0.002	1.502	100.102	0.001	2.255
	λ	0.015		λ	0.044	0.029	-190.718	0.012	0.001
	β_1	-1.000		β_1	-0.370	0.630	63.018	0.010	0.397
	β_2	-0.200		β_2	-0.271	0.071	-35.593	0.007	0.005
4	ν	1.500	0.1	α	-0.002	1.502	100.113	0.001	2.255
	λ	0.015		λ	0.038	0.023	-153.024	0.011	0.001
	β_1	-1.000		β_1	-0.370	0.630	63.024	0.012	0.397
	β_2	-0.200		β_2	-0.271	0.071	-35.465	0.007	0.005
5	ν	1.500	0	α	-0.002	1.502	100.139	0.001	2.256
	λ	0.015		λ	0.106	0.091	-604.296	0.026	0.009
	β_1	0.500		β_1	-0.372	0.872	174.431	0.010	0.761
	β_2	-1.000		β_2	-0.256	0.744	74.389	0.006	0.553
6	ν	1.500	0.1	α	-0.002	1.502	100.159	0.001	2.257
	λ	0.015		λ	0.094	0.079	-525.466	0.025	0.007
	β_1	0.500		β_1	-0.372	0.872	174.458	0.010	0.761
	β_2	-1.000		β_2	-0.256	0.744	74.359	0.006	0.553

Par. : Parameter; Cens. : Censoring; St. Dev. : Estimated Standard Deviation.
MSE: Mean square error.

A3. Derivation of full conditional distribution of β_i. With the same manner as in Appendix A1, full conditional distribution of a particular regression parameter given the others is obtained by using the Taylor expansion. Then, $p(\beta_i|\beta_{-i},\alpha,\lambda,\boldsymbol{t})$ is obtained by discarding the constants as follows:

$$p(\beta_i|\beta_{-i},\alpha,\lambda,\boldsymbol{t}) \propto \exp\left\{\beta_i \sum_{j=1}^{n} \delta_j x_{ji} + (\lambda/\alpha)\sum_{j=1}^{n}[1-\exp(\alpha t_j)]\exp(x_{ji}\beta_i)c_j\right\}, \tag{22}$$

where $c_j = \exp(\sum_{k=1,k\neq i}^{n} \beta_k x_{jk})$. It is obtained using the second order Taylor expansion of $\exp(x_{ji}\beta_i)$ at 0 that

$$\begin{aligned} p(\beta_i|\beta_{-i},\alpha,\lambda,\boldsymbol{t}) \propto \exp\Big\{&\beta_i\big[\sum_{j=1}^{n}\delta_j x_{ji} - (\lambda/\alpha)\sum_{j=1}^{n} c_j x_{ji}[\exp(\alpha t_j)-1]\big] \\ &-[\lambda/(2\alpha)]\beta_i^2 \sum_{j=1}^{n} c_j x_{ji}^2[\exp(\alpha t_j)-1]\Big\} \end{aligned} \tag{23}$$

by simply arranging (23),

$$p(\beta_i|\beta_{-i},\alpha,\lambda,\boldsymbol{t}) \propto \exp\left\{\frac{-1}{2s_1}\big[\beta_i - s_2/s_1\big]^2\right\}, \tag{24}$$

where $s_1 = (\lambda/\alpha)\sum_{j=1}^{n} c_j x_{ji}^2[\exp(\alpha t_j)-1]$ and $s_2 = \sum_{j=1}^{n}\delta_j x_{ji} - (\lambda/\alpha)\sum_{j=1}^{n} c_j x_{ji} \times [\exp(\alpha t_j)-1]$. Then $p(\beta_i|\beta_{-i},\alpha,\lambda,\boldsymbol{t})$ is approached by the normal distribution with mean s_2/s_1 and variance $1/s_1$.

As for the appropriateness of the second order Taylor expansion, we evaluate the impact of the value of β_i on survival times in Gompertz model as done in Appendix A1. Regarding the second equation in (10), we have the following results for the fixed values

Table 9. The ML estimates of parameters of the Cox model over 1000 samples, each was generated from the Gompertz distribution.

Case	Par.	Values	Cens. Rate	Par.	Estimate	Absolute Bias	Relative Bias	St.Dev.	MSE
1	α	0.000	0						
	λ	0.010							
	β_1	-0.100		β_1	-0.012	0.088	88.380	6.763	45.746
	β_2	-0.200		β_2	-0.362	0.162	-81.164	1.345	1.834
2	α	0.000	0.1						
	λ	0.010							
	β_1	-0.100		β_1	-0.959	0.859	-859.053	11.972	144.077
	β_2	-0.200		β_2	-0.658	0.458	-228.872	2.328	5.631
3	α	0.000	0						
	λ	0.100							
	β_1	-1.000		β_1	-2.257	1.257	-125.700	9.299	88.055
	β_2	-0.200		β_2	-0.627	0.427	-213.332	1.938	3.940
4	α	0.000	0.1						
	λ	0.100							
	β_1	-1.000		β_1	-4.193	3.193	-319.286	14.321	215.294
	β_2	-0.200		β_2	-1.439	1.239	-619.288	3.076	10.999
5	α	-0.001	0						
	λ	1.000							
	β_1	0.500		β_1	0.254	0.246	49.240	5.368	28.872
	β_2	-1.000		β_2	-0.087	0.913	91.284	1.070	1.979
6	α	-0.001	0.1						
	λ	1.000							
	β_1	0.500		β_1	-0.046	0.546	109.288	6.397	41.216
	β_2	-1.000		β_2	-0.168	0.832	83.214	1.289	2.353

Par. : Parameter; Cens. : Censoring; St. Dev. : Estimated Standard Deviation.
MSE: Mean square error.

of α, λ, and β_{-i}:

$$\lim_{\beta_i \to -\infty} T_j^{\mathrm{G}} = \infty \text{ and } \lim_{\beta_i \to \infty} T_j^{\mathrm{G}} = 0.$$

Values close the $-\infty$ are unreasonable and greater values give nearly zero survival times. Positive and larger values of β_i correspond to reasonable survival times for very small values of U; hence, probability of occurrence of this situation is small. Accordingly, small values of β_i will correspond to reasonable survival times in practice. Therefore, the rate of convergence will be very fast; and hence, use of the second order Taylor expansion is appropriate.

References

[1] Ananda, M.M., Dalpatadu, R.J. and Singh, A.K. *Adaptive Bayes estimators for parameters of the Gompertz survival model*, Applied Mathematics and Computation, **75** (2), 167-177, 1996.

[2] Bender, R., Augustin, T., Blettner, M. *Generating survival times to simulate Cox proportional hazards models*, Statistics in Medicine, **24**, 1713-1723, 2005.

[3] Cantner, U., DreBler, K., Kruger, J.J. *Firm Survival in the German Automobile Industry*, Empirica, **33**, 49-60, 2006.

[4] Collett, D. *Modelling Survival Data in Medical Research*, Chapman&Hall, UK, 1994.

[5] Cox, D.R. *Regression models and life tables (with discussion)*, Journal of the Royal Statistical Society : Series B, **34**, 187-220, 1972.

[6] Dellaportas, P. and Smith, A.F.M. *Bayesian inferences of generalized linear and proportional hazards models via Gibbs sampling*, Applied Statistics, **42** (3), 443-459, 1993.

[7] Efron, B. *The efficiency of Cox's likelihood function for censored data*", Journal of the American Statistical Association, **72**, 557-565, 1977.

[8] Fabrizio, C. *New Evidence on the Politics and Economics of Multiparty Cabinets Duration*, Scottish Journal of Political Economy, **49** (3), 249-279, 2002.

Table 10. The ML estimates of parameters of the Cox-Gompertz model over 1000 samples, each was generated from the Gompertz distribution.

Case	Par.	Values	Cens. Rate	Par.	Estimate	Absolute Bias	Relative Bias	St.Dev.	MSE
1	α	0.000	0	α	-0.004	0.004	3854.290	0.000	0.000
	λ	0.010		λ	0.127	0.117	-1170.553	0.000	0.014
	β_1	-0.100		β_1	1.315	1.415	1414.505	0.000	2.001
	β_2	-0.200		β_2	-0.120	0.080	39.925	0.000	0.006
2	α	0.000	0.1	α	-0.004	0.004	4005.207	0.000	0.000
	λ	0.010		λ	0.170	0.160	-1595.591	0.000	0.025
	β_1	-0.100		β_1	1.303	1.403	1403.429	0.000	1.970
	β_2	-0.200		β_2	-0.208	0.008	-3.827	0.000	0.000
3	α	0.000	0	α	-0.004	0.004	395797.522	0.000	0.000
	λ	0.100		λ	0.207	0.107	-106.725	0.000	0.011
	β_1	-1.000		β_1	0.542	1.542	154.200	0.000	2.378
	β_2	-0.200		β_2	-0.132	0.068	33.773	0.000	0.005
4	α	0.000	0.1	α	-0.004	0.004	395723.040	0.000	0.000
	λ	0.100		λ	0.179	0.079	-79.196	0.000	0.006
	β_1	-1.000		β_1	0.709	1.709	170.875	0.000	2.920
	β_2	-0.200		β_2	-0.170	0.030	15.136	0.000	0.001
5	α	-0.001	0	α	-0.003	0.002	-198.367	0.000	0.000
	λ	1.000		λ	0.659	0.341	34.143	0.000	0.117
	β_1	0.500		β_1	2.092	1.592	-318.486	0.000	2.536
	β_2	-1.000		β_2	-1.200	0.200	-19.957	0.000	0.040
6	α	-0.001	0.1	α	-0.007	0.006	-588.075	0.000	0.000
	λ	1.000		λ	0.522	0.478	47.787	0.000	0.228
	β_1	0.500		β_1	2.246	1.746	-349.147	0.000	3.048
	β_2	-1.000		β_2	-1.226	0.226	-22.640	0.000	0.051

Par. : Parameter; Cens. : Censoring; St. Dev. : Estimated Standard Deviation.
MSE: Mean square error.

[9] Gelfand, A.E. and Mallick, B.K. *Bayesian analysis of proportional hazards built from monotone functions*, Biometrics, **51**, 843-852, 1995.

[10] Gelfand,A.E. and Sahu, S.K. *Identifiability, improper priors, and Gibbs sampling for generalized linear models*, Journal of the American Statistical Association, **94**, 247-253, 1999.

[11] Gelman, A. *Inference and monitoring convergence*, in Markov Chain Monte Carlo (eds. W.R. Gilks, S. Richardson, D.J. Spiegelhalter)", Chapman&Hall/CRC, London, 1996.

[12] Ghitanya, M.E., Alqallafa, F., Balakrishnan, N. *On the likelihood estimation of the parameters of Gompertz distribution based on complete and progressively Type-II censored samples*, Journal of Statistical Computation and Simulation, **84** (8), 1803-1812, 2014.

[13] Gieser, P., Chang, M.N., Rao, P.V., Shuster, J.J. and Pullen, J. *Modelling cure rate using the Gompertz model with covariate information*, Statistics in Medicine, **17**, 831-839, 1998.

[14] Gokovali, U., Bahar, O., Kozak, M. *Determinants of length of stay: A practical use of survival analysis*, Tourism Management, **28**, 736-746, 2007.

[15] Gompertz, B. *On the nature of the function expresssive of the law of human mortality, and on a new method of determining the value of life contingencies*, Phil. Trans. R. Soc., **36**, 513-585, 1825.

[16] Grunkemeier, G.L., Jamieson, W.R.E., Miller, D.C. and Starr, A. *Surgery for Acquired Heart Disease*, J Thorac Cardiovasc Surg, **108**, 709-718, 1994.

[17] Gut, A. *Probability: A Graduate Course*, Springer, New York, 2013.

[18] Jeong, J-H., *Parametric regression on cumulative incidence function*, Biostatistics, **8** (2), 184-196, 2007.

[19] Jeong, J-H. and Fine, J. *Direct parametric inference for the cumulative incidence function*, Journal of the Royal Statistical Society : Series C, **55** (2), 187-200, 2006.

[20] Johnson, N.L., Kotz, S. and Balakrishnan, N. *Continuous Univariate Distributions, Vol. 1.*, Wiley, New-York, 1994.

[21] Kim, S.W. and Ibrahim, J.G. "*On bayesian inference for proportional hazards models using noninformative priors*, Lifetime Data Analysis, **6**, 331-341, 2000.

Table 11. The Bayesian estimates of parameters of the Cox-Gompertz model over 1000 samples, each was generated from the Gompertz distribution.

Case	Par.	Values	Cens. Rate	Par.	Estimate	Absolute Bias	Relative Bias	St.Dev.	MSE
1	α	0.000	0	α	-0.003	0.003	2730.118	0.001	0.000
	λ	0.010		λ	0.056	0.046	-458.052	0.011	0.002
	β_1	-0.100		β_1	-0.768	0.668	-668.051	0.017	0.447
	β_2	-0.200		β_2	-0.561	0.361	-180.716	0.010	0.131
2	α	0.000	0.1	α	-0.004	0.004	4442.027	0.001	0.000
	λ	0.010		λ	0.078	0.068	-677.256	0.013	0.005
	β_1	-0.100		β_1	-1.249	1.149	-1149.071	0.023	1.321
	β_2	-0.200		β_2	-0.913	0.713	-356.536	0.014	0.509
3	α	0.000	0	α	-0.004	0.004	355914.065	0.000	0.000
	λ	0.100		λ	0.056	0.044	43.945	0.009	0.002
	β_1	-1.000		β_1	-1.342	0.342	-34.244	0.025	0.118
	β_2	-0.200		β_2	-0.985	0.785	-392.406	0.015	0.616
4	α	0.000	0.1	α	-0.006	0.006	587070.440	0.001	0.000
	λ	0.100		λ	0.080	0.020	19.849	0.011	0.001
	β_1	-1.000		β_1	-2.222	1.222	-122.246	0.035	1.496
	β_2	-0.200		β_2	-1.629	1.429	-714.308	0.021	2.041
5	α	-0.001	0	α	-0.012	0.011	-1070.242	0.006	0.000
	λ	1.000		λ	0.946	0.054	5.370	0.202	0.044
	β_1	0.500		β_1	-0.454	0.954	190.825	0.009	0.910
	β_2	-1.000		β_2	-0.312	0.688	68.771	0.006	0.473
6	α	-0.001	0.1	α	-0.014	0.013	-1309.660	0.006	0.000
	λ	1.000		λ	0.852	0.148	14.843	0.189	0.058
	β_1	0.500		β_1	-0.484	0.984	196.853	0.011	0.969
	β_2	-1.000		β_2	-0.333	0.667	66.712	0.006	0.445

Par. : Parameter; Cens. : Censoring; St. Dev. : Estimated Standard Deviation.
MSE: Mean square error.

Table 12. Table 12: Proportion of unconverged iterations for Cox and Cox-Gompertz models

	Exponential Dist.		Weibull Dist.		Gompertz Dist.	
Case	CM	CGM	CM	CGM	CM	CGM
1	0.190	0.998	0.000	0.998	0.000	0.492
2	0.184	0.999	0.000	0.999	0.002	0.688
3	0.711	0.998	0.000	0.998	0.003	0.708
4	0.769	0.998	0.001	0.999	0.038	0.827
5	0.007	0.998	0.000	0.999	0.000	0.015
6	0.008	0.998	0.000	0.999	0.000	0.075

CM : the Cox-Model; CGM : the Cox-Gompertz Model

[22] Klein, J.P. and Moeschberger M.L. *Survival Analysis : Techniques for Censored and Truncated Data*, Springer, New York, 1997.

[23] Klepper, S. *The Capabilities of New Firms and the Evolution of the U.S. Automobile Industry*, Industrial and Corporate Change, **11**, 645-666, 2002.

[24] Laubender, R.P. and Bender, R. *Estimating adjusted risk difference (RD) and number needed to treat (NNT) measures in the Cox regression model*, Statistics in Medicine, **29**, 851-859, 2010.

[25] Lawless, J.F. *Statistical Models and Methods for Lifetime Data*, Wiley, New York, 1982.

[26] Lee, E.T. and Wang, J.W. *Statistical Methods for Survival Data Analysis*, Wiley, New York, 2003.

[27] Makeham, W.M. *On the law of mortality, and the construction of annuity tables*, Journal of the Institute of Actuaries, **13**, 325-358, 1860.

[28] Nardi, A. and Schemper, M. *Comparing Cox and parametric models in clinical studies*, Statistics in Medicine, **22**, 3597-3610, 2003.

[29] Oakes, D. *The asymptotic information in censored survival data*, Biometrika, **64**, 441-448, 1977.

[30] Spickett, A.M. and van Ark, H. *Fitting the Gompertz function to dose-response data oflarval tick populations*, Onderstepoort J Vet Res., **57** (2), 155-118, 1990.

[31] Willekens, F. *Gompertz in context: the Gompertz and related distributions*, In Forecasting Mortality in Developed Countries - Insights from a Statistical, Demographic and Epidemiological Perspective, edited by E. Tabeau, A. van den Berg Jeths, and C. Heathcote, volume 9 of European Studies of Population, 105-126, Springer, 2002.