Austrian Journal of Statistics
February 2020, Volume 49, 57–71.
http://www.ajs.or.at/
doi:10.17713/ajs.v49i3.1085

Assessing the Goodness of Fit of the Gompertz Model in the Presence of Right and Interval Censored Data with Covariate

Nur Niswah Naslina A.
Universiti Putra Malaysia
&
Universiti Teknologi MARA

Jayanthi A.
Universiti Putra Malaysia

Hani Syahida Z.
Universiti Putra Malaysia

Mohd Bakri A.
Universiti Putra Malaysia

Abstract

This research focuses on assessing the goodness of fit for the Gompertz model in the presence of right and interval censored data with covariate. The performance of the maximum likelihood estimates was evaluated via a simulation study at various censoring proportions and sample sizes. The conclusions were drawn based on the results of bias, standard error and root mean square error at different settings. Following that, another simulation study was carried out to compare the performance of the proposed modifications to the Cox-Snell residuals for both censored and uncensored observations at different combinations of sample sizes and censoring levels. The results show that standard error and root mean square error values of the parameter estimates increase with the increase in censoring proportions and decrease in the number of sample size. This indicates that the estimates perform better when sample sizes are larger and censoring proportions are lower. The performance of the proposed modifications of the Cox-Snell residuals showed that they perform slightly better than existing method.

Keywords: Gompertz model, right censored, covariates, Cox-Snell residuals, proposed modifications.

1. Introduction

Survival analysis consists of statistical procedures used for analysis of data where the outcome variable is time until an event occurs and is often referred to as time to event data. Survival analysis has become a popular tool in observational and experimental studies primarily in the public health, epidemiology, medical and biological sciences (Klein and Moeschberger 2003; Lee and Wang 2003).

Traditional statistical procedures are not equipped to handle the censored observations which is a special type of missing data that occurs in survival analyses when subjects do not experience the event of interest during the follow-up time. Moreover, survival data are not symmetrically distributed or non-normality and typically it will tend to be positively skewed (Collett 2003).

Censoring occurs when we have some information about individual survival time but do not know the survival time exactly. While, for exact or uncensored observations reported when the survival times recorded for the person's that died during the study period which is the times from the start of the experiment until the death. The three most common censoring in survival analysis are right censoring, left censoring and interval censoring. Right censoring occurs when true survival time is equal to or greater than observed survival time or in other way, we can say that the individual is still alive at a given time. Left censoring arises when the individual has experienced the event of interest prior to the start of the study but the exact time of occurrence is unknown. In interval censoring, the true survival time lies within a known time interval instead of being observed exactly (Klein and Moeschberger 2003; Lee and Wang 2003; Kleinbaum and Klein 2012). Interval censored data is very common in medical research where inspection on patients are conducted on different time intervals. So, the lifetime is only known to fall within an interval, $L_i < t_i < R_i$, where L_i and R_i are known as left and right endpoints. In this study, we focus on both right and interval censoring.

Although there are well known methods for estimating unconditional survival distributions, most interesting survival modeling examines the relationship between survival and one or more predictors known as covariates. Residuals are a widely used tool to assess the adequacy of a model. When modeling survival data, it is not as easy to define a residual as for a general linear model. It is common practice to use Cox-Snell residuals to check for overall goodness of fit in survival models (Cox and Snell 1968). Therefore, a set of different residuals has therefore been proposed.

In this paper, we have considered the Gompertz distribution with covariate in the presence of right and interval censored data to study extensively on the performance of this model. A simulation study is carried out to evaluate the maximum likelihood estimation (MLE) procedure for the parameters of the Gompertz model at various censoring proportions and sample sizes by computing the values of bias, standard error (SE) and root mean square error (RMSE). Following that, we had proposed several modifications to the Cox-Snell residuals and analyzed comprehensively their performance at different sample sizes and censoring proportions.

Originally, the Gompertz distribution was developed by a British actuary, (Gompertz 1825), in modeling human mortality and establish actuarial tables. This famous Gompertz theoretical law of mortality states that the death rates increased exponentially with age. Over the past one and a half centuries, many researchers have contributed to the studies of statistical methodology and characterization of this distribution for instance Garg, Rao, and Redmond (1970) studied on the properties of the Gompertz distribution and compare the estimates through the least-squares method and maximum-likelihood estimation. Gordon (1990) considered on the maximum likelihood estimates for the parameters of the mixture of two Gompertz distributions when censoring occurs. Subsequently,Witten and Satzer (1992) addressed the issue of parameter sensitivity of a new method for estimating the model parameters of the Gompertz mortality rate model. Wilson (1994) compared the Gompertz, Weibull and Logistics functions in the analysis of mortality data. Chen (1997) developed an exact confidence interval and an exact joint confidence region for the parameters of the Gompertz distribution. While, Wu, Hung, and Tsai (2004) proposed unweighted and weighted least squares estimates for parameters of the Gompertz distribution under complete set of data and first failure censored data. Lenart (2012) discussed on the comparison method of moments and maximum likelihood estimates from a Gompertz distribution. Kiani, Arasan, and Midi (2012) deliberated on performance of the Gompertz model with time-dependent covariate in the presence of right censored data and applied two confidence interval estimation techniques known as Wald and Jackknife. Kiani and Arasan (2013) was extended the Gompertz model to incorporate time-dependent covariate in the presence of interval-, right-, left-censored and uncensored data. Abu-Zinadah (2014) implemented the maximum likelihood method of estimation for estimating the parameters and performed the goodness-of-fit tests for testing the three-parameters exponentiated Gompertz distribution based on complete and type II censored sampling. Later, El-Din, Abdel-Aty, and Abu-Moussa (2016) estimated the parameters for Gompertz distribution by finding the maximum likelihood method and Bayesian method under three different loss functions. Currently, de Andrade, Chakraborty, Handique, and Gomes-Silva (2019) studied five-parameter model based on a new generalization of the extended Gompertz distribution known as exponentiated generalized extended Gompertz distribution.

Weissfeld and Schneider (1990) conferred on the methods for detecting influential observations for the Weibull model fit to censored data which include the methods of one-step deletion diagnostics, influence functions and curvature diagnostics. Results from Leung, Elashoff, and Afifi (1997) summarised that various methods used to deal with censored data which includes complete data analysis, the imputation techniques, analysis based on dichotomized data and the likelihood-based approach. Farrington (2000) had applied several diagnostic tools such as Cox-Snell, Lagakos (or martingale), deviance, and Schoenfeld residuals for use with proportional hazards models for interval-censored survival data. Sparling, Younes, Lachin, and Bautista (2006) were presented a parametric family of regression models for interval-censored event-time (survival) data that accommodates both fixed and time-dependent covariates. Prinja, Gupta, and Verma (2010) devised that problem of interval censoring arises when time to event may be known only up to a time interval which the situation occurs in a case where the assessment of monitoring is done at a periodical frequency. Kiani and Arasan (2018) discussed on the survival model with doubly interval censored data and time dependent covariate where the life-time is the elapsed time between two related events which means that the first event and the second event are interval censored. Sakurai and Hattori (2018) developed a model-checking procedure based on the cumulative martingale residuals for the interval-censored observations.

2. Methodology

2.1. Gompertz model with right and interval censored data and covariate

Let T be a non-negative continuous random variable which denotes the survival time. The probability density function of the Gompertz is given by,

$$f(t;\gamma;\lambda) = \lambda e^{[\gamma t + \frac{\lambda}{\gamma}(1-e^{\gamma t})]}; t \geq 0, \lambda > 0, \gamma > 0. \quad (1)$$

The corresponding survivor function is given by,

$$S(t;\gamma;\lambda) = e^{[\frac{\lambda}{\gamma}(1-e^{\gamma t})]}. \quad (2)$$

The hazard function is

$$h(t;\gamma;\lambda) = \lambda e^{\gamma t}. \quad (3)$$

The effect of covariates on survival time can be incorporated to the hazard function by letting the parameter λ be a function of the covariates,

$$\lambda = e^{\beta' x}. \quad (4)$$

For data set with a covariate x_i where $i = 1, 2, ..., n$, the hazard function for i^{th} subject can be expressed as,

$$h(t_i;\gamma;\lambda) = \lambda_i e^{\gamma t_i}. \quad (5)$$

where

$$\lambda_i = e^{\beta_0 + \beta_1 x_i}. \quad (6)$$

Therefore, the hazard function is

$$\begin{aligned} h(t_i; x_i; \beta; \gamma) &= \lambda_i e^{\gamma t_i}, \\ &= e^{\beta_0+\beta_1 x_i} . e^{\gamma t_i}, \\ &= e^{\beta_0+\beta_1 x_i+\gamma t_i}. \end{aligned} \tag{7}$$

The probability density function is

$$\begin{aligned} f(t_i; x_i; \beta; \gamma) &= \lambda e^{[\gamma t_i + \frac{\lambda}{\gamma}(1-e^{\gamma t_i})]}, \\ &= e^{\beta_0+\beta_1 x_i} . e^{[\gamma t_i + \frac{e^{\beta_0+\beta_1 x_i}}{\gamma}(1-e^{\gamma t_i})]}, \\ &= e^{\beta_0+\beta_1 x_i+\gamma t_i + \frac{e^{\beta_0+\beta_1 x_i}}{\gamma}(1-e^{\gamma t_i})}. \end{aligned} \tag{8}$$

with the corresponding survivor function given by

$$\begin{aligned} S(t_i; x_i; \beta; \gamma) &= e^{[\frac{\lambda}{\gamma}(1-e^{\gamma t_i})]}, \\ &= e^{[\frac{e^{\beta_0+\beta_1 x_i}}{\gamma}(1-e^{\gamma t_i})]}. \end{aligned} \tag{9}$$

The parameters of this model can be estimated by the method of maximum likelihood (MLE) where $\boldsymbol{\theta = (\beta_0, \beta_1, \gamma)}$ is the vector of parameters.

2.2. Maximum likelihood estimation

To incorporate right and interval censored data to the likelihood function, we need to define the following indicator variables for i^{th} observation,

$$\delta_{E_i} = \begin{cases} 1, & \text{if the } i^{th} \text{ observation is complete} \\ 0, & \text{otherwise.} \end{cases}$$

$$\delta_{R_i} = \begin{cases} 1, & \text{if the } i^{th} \text{ observation is right censored} \\ 0, & \text{otherwise.} \end{cases}$$

$$\delta_{I_i} = \begin{cases} 1, & \text{if the } i^{th} \text{ observation is interval censored} \\ 0, & \text{otherwise.} \end{cases}$$

Then the likelihood function for the full sample consisting of complete, right censored and interval censored data is,

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} [f(t_i)]^{\delta_{E_i}} [S(t_{R_i})]^{\delta_{R_i}} [S(t_{L_i}) - S(t_{R_i})]^{\delta_{I_i}}, \\ &= \prod_{i=1}^{n} \left[e^{\beta_0+\beta_1 x_i+\gamma t_i + \frac{e^{\beta_0+\beta_1 x_i}}{\gamma}(1-e^{\gamma t_i})} \right]^{\delta_{E_i}} \left[e^{\frac{e^{\beta_0+\beta_1 x_i}}{\gamma}(1-e^{\gamma t_{R_i}})} \right]^{\delta_{R_i}} \\ &\times \left[e^{\frac{e^{\beta_0+\beta_1 x_i}}{\gamma}} (e^{\gamma t_{R_i}} - e^{\gamma t_{L_i}}) \right]^{\delta_{I_i}}. \end{aligned} \tag{10}$$

and log-likelihood function is,

$$\begin{aligned} l(\theta) = \sum_{i=1}^{n} \delta_{E_i} &\left[\beta_0 + \beta_1 x_i + \gamma t_i + \frac{e^{\beta_0+\beta_1 x_i}}{\gamma}(1 - e^{\gamma t_i}) \right] \\ &+ \delta_{R_i} \left[\frac{e^{\beta_0+\beta_1 x_i}}{\gamma}(1 - e^{\gamma t_{R_i}}) \right] \\ &+ \delta_{I_i} \left[e^{\frac{e^{\beta_0+\beta_1 x_i}}{\gamma}} (e^{\gamma t_{R_i}} - e^{\gamma t_{L_i}}) \right] \end{aligned} \tag{11}$$

The inverse of the observed information matrix, $[i(\hat{\beta}_0, \hat{\beta}_1, \hat{\gamma})]^{-1}$ can be obtained from the second partial derivative of the log-likelihood function evaluated at $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\gamma}$ which provides the estimates for the variance and covariance.

3. Simulation study and results

3.1. Assessing performance of the parameter estimates

A simulation study using 1000 samples each with n=30,40,50,80 and 100 were conducted for this model for both censored and uncensored observations with fixed covariates, x_i. The covariate values were simulated independently from the standard normal distribution. The values of -5, 0.3 and 0.5 were chosen as the parameters of β_0, β_1 and γ to mimic real life survival data. A sequence of random numbers, u_i, from the standard uniform distribution on the interval $(0, 1)$ was generated to produce lifetimes t_i for $i = 1, 2, ..., n$ subjects. The censoring time, c_i was generated from exponential distribution where the value μ would be adjusted to obtain the desired approximate censoring proportion (cp) for the data with $cp = 0\%, 10\%, 20\%, 30\%, 40\%$ and 50%. The simulated survival time is considered censored if $t_i > c_i$, and will be replaced by the corresponding censoring time. The survival time t_i was generated by,

$$t_i = \frac{1}{\gamma} log\left[1 - \frac{\gamma log(1-\mu_i)}{e^{\beta_0+\beta_1 x_i}}\right]. \tag{12}$$

In order to evaluate the performance of the estimator at different combination of sample sizes and censoring proportions, the bias, standard error (SE) and root mean square error (RMSE) of the parameter were calculated. The bias, SE and RMSE were computed by,

$$bias = E(\hat{\theta}) - \theta, \tag{13}$$

$$SE = \sqrt{E(\hat{\theta} - E(\hat{\theta}))^2}, \tag{14}$$

$$RMSE = \sqrt{SE^2 + bias^2} \tag{15}$$

Table 1: Summary table for bias of the parameters for various n and cp.

Estimates	n	Censoring Proportion					
		0%	10%	20%	30%	40%	50%
$\hat{\beta}_0$	30	-0.2721	-0.2607	-0.2614	-0.2628	-0.2632	-0.2696
	40	-0.1556	-0.1472	-0.1135	-0.1088	-0.1019	-0.0964
	50	-0.1474	-0.1295	-0.1010	-0.0720	-0.0541	-0.0311
	80	-0.1011	-0.0682	-0.0507	-0.0134	0.0454	0.0884
	100	-0.0691	-0.0458	-0.0232	-0.0032	0.0278	0.0727
$\hat{\beta}_1$	30	0.0265	0.0281	0.0314	0.0386	0.0357	0.0315
	40	0.0174	0.0184	0.0122	0.0138	0.0136	0.0134
	50	0.0155	0.0128	0.0118	0.0092	0.0071	0.0134
	80	0.0104	0.0076	0.0065	0.0085	0.0070	0.0059
	100	0.0053	0.0037	0.0033	0.0027	0.0042	0.0035
$\hat{\gamma}$	30	0.0402	0.0410	0.0444	0.0487	0.0510	0.0572
	40	0.0244	0.0253	0.0257	0.0269	0.0282	0.0299
	50	0.0218	0.0213	0.0206	0.0210	0.0217	0.0263
	80	0.0146	0.0132	0.0124	0.0114	0.0088	0.0087
	100	0.0109	0.0099	0.0094	0.0083	0.0078	0.0055

Based on the Table 1, the bias values show an erratic pattern as the sample size and the censoring proportion increase. While, in Table 2, we can see that the standard error values increase as the censoring proportions increase. The Table 3 indicates that the root mean square error values also increase as the censoring proportion increase. This indicates that poorer performance for the parameter estimates at smaller sample sizes and higher censoring proportions, whereas larger sample sizes and lower censoring proportions would have higher accuracy and efficiency of the parameter estimates.

Table 2: Summary table for standard error (SE) of the parameters for various n and cp.

Estimates	n	Censoring Proportion					
		0%	10%	20%	30%	40%	50%
$\hat{\beta}_0$	30	0.7270	0.7596	0.8136	0.8946	0.9419	1.0274
	40	0.5841	0.6003	0.6460	0.6670	0.7003	0.7302
	50	0.5380	0.5547	0.5825	0.6148	0.6455	0.7610
	80	0.4232	0.4391	0.4445	0.4781	0.5168	0.5660
	100	0.3667	0.3760	0.3875	0.4040	0.4317	0.4611
$\hat{\beta}_1$	30	0.2246	0.2375	0.2548	0.2880	0.3069	0.3432
	40	0.1891	0.1971	0.2240	0.2322	0.2444	0.2601
	50	0.1656	0.1723	0.1812	0.1998	0.2148	0.2695
	80	0.1192	0.1295	0.1328	0.1501	0.1707	0.1960
	100	0.1089	0.1153	0.1211	0.1243	0.1342	0.1443
$\hat{\gamma}$	30	0.0919	0.0961	0.1050	0.1172	0.1242	0.1356
	40	0.0742	0.0772	0.0846	0.0883	0.0940	0.0988
	50	0.0680	0.0710	0.0749	0.0801	0.0847	0.1028
	80	0.0534	0.0558	0.0568	0.0619	0.0683	0.0763
	100	0.0462	0.0478	0.0501	0.0522	0.0566	0.0615

Table 3: Summary table for root mean square error (RMSE) of the parameters for various n and cp.

Estimates	n	Censoring Proportion					
		0%	10%	20%	30%	40%	50%
$\hat{\beta}_0$	30	0.7763	0.8031	0.8546	0.9323	0.9780	1.0622
	40	0.6045	0.6181	0.6559	0.6758	0.7077	0.7365
	50	0.5578	0.5696	0.5912	0.6190	0.6478	0.7616
	80	0.4351	0.4444	0.4473	0.4783	0.5188	0.5729
	100	0.3731	0.3788	0.3882	0.4040	0.4326	0.4668
$\hat{\beta}_1$	30	0.2262	0.2392	0.2567	0.2906	0.3090	0.3447
	40	0.1899	0.1980	0.2244	0.2326	0.2448	0.2604
	50	0.1663	0.1728	0.1816	0.2000	0.2149	0.2699
	80	0.1196	0.1297	0.1330	0.1503	0.1708	0.1961
	100	0.1090	0.1153	0.1212	0.1243	0.1343	0.1444
$\hat{\gamma}$	30	0.1003	0.1045	0.1140	0.1269	0.1343	0.1472
	40	0.0781	0.0813	0.0884	0.0923	0.0982	0.1032
	50	0.0714	0.0741	0.0777	0.0828	0.0874	0.1061
	80	0.0553	0.0574	0.0581	0.0630	0.0688	0.0768
	100	0.0475	0.0488	0.0509	0.0528	0.0571	0.0617

3.2. Assessing model fit

Modification of Cox-Snell residuals

Cox-Snell residuals, r_{Ci}, is widely used in the analysis of survival data as discussed by Cox and Snell (1968) to assess a model's goodness-of-fit. A log-cumulative hazard plot of residuals is obtained by plotting the Cox-Snell residual against the cumulative hazard function to assess the model's fit. A well fitting model will exhibit a linear line through the origin with a unit gradient. It should be noted that it will take a particularly ill-fitting model for the Cox-Snell residuals to deviate significantly from this. One criticism of Cox-Snell residuals is that they do not account for censored observations, therefore the adjusted Cox-Snell residuals were devised by Crowley & Hu (1977) in Collett (2003) whereby the standard Cox-Snell residual, r_{Ci} could be used for uncensored observations and $r_{Ci} + \Delta$ whereby $\Delta = log(2) = 0.693$, is used to adjust the residual. The Cox-Snell residuals for the i^{th} individual, $i = 1, 2, ..., n$ is given by,

$$r_{ci} = -log\hat{S}_i(t_i). \tag{16}$$

The modified Cox-Snell residuals by proposed to account for censored data. Crowley & Hu (1977) in Collett (2003) found that the addition of unity to a Cox-Snell residual for a censored observation inflated the residual to too great an extent. Thus, by using the median value of the excess residual, a second version of the modified Cox-Snell residual is,

$$r'_{ci} = \begin{cases} r_{ci}, & \text{for uncensored observations,} \\ r_{ci} + 0.693, & \text{for censored observations.} \end{cases}$$

In this research we propose two modifications to the Cox-Snell residuals as follows

$$r^2_{ci} = \begin{cases} r_{ci}, & \text{for uncensored observations,} \\ r_{ci} + g, & \text{for censored observations.} \end{cases}$$

and

$$r^3_{ci} = \begin{cases} r_{ci}, & \text{for uncensored observations,} \\ r_{ci} + h, & \text{for censored observations.} \end{cases}$$

where g is the geometric mean of data and h is the harmonic mean of data.

Simulation study

A simulation using 1000 samples each with n=30,40,50,80 and 100 using $cp = 0\%, 10\%, 20\%, 30\%,$ 40% and 50% was conducted to compare the residual values. A number of plots based on residuals were used in the graphical assessment of the adequacy of a fitted model. Plot of $ln[-ln(\hat{S}(r_{ci}))]$ vs $ln(r_{ci})$ should be a straight line through the origin with unit slope if the data fits the model. Several modification of the Cox-Snell residuals were used and compare the performance for censored and uncensored data.

- Cox-Snell residuals, r_{ci}
- Modified Cox-Snell, r^1_{ci}
- Replace the median with geometric mean of existing data, r^2_{ci}
- Replace the median with harmonic mean of existing data, r^3_{ci}

The table indicates that as number of sample size increase, the intercept become closer to zero. Similarly as sample size increase, the slope become closer to 1. While for both r and R^2 values closer to 1 indicates a strong relationship. However, when the censoring proportions becomes higher, expected values for intercept and slope go further than zero and 1 respectively. The range for r and R^2 values also become wider as it across higher censoring proportions.

Table 4: Range of intercept for various residuals.

n	cp	Residuals							
		CS		MCS		GMCS		HMCS	
		Min	Max	Min	Max	Min	Max	Min	Max
30	0	-0.4434	0.7074	-0.4418	0.7074	-0.4434	0.7074	-0.4418	0.7074
	10	-0.5032	0.7052	-0.4999	0.7052	-0.5114	0.7052	-0.5049	0.7052
	20	-0.6071	0.6679	-0.6342	0.5596	-0.6837	0.6134	-0.6093	0.6403
	30	-0.6334	0.4586	-0.7819	0.1847	-0.8994	0.3294	-0.6353	0.3862
	40	-0.7406	0.4449	-1.2023	-0.0020	-0.9491	0.2479	-0.7360	0.5308
	50	-0.9274	0.4460	-1.4811	-0.2882	-1.1982	0.2061	-0.9098	0.4677
40	0	-0.5565	0.2631	-0.5572	0.2643	-0.5566	0.2649	-0.5572	0.2643
	10	-0.5791	0.2900	-0.5976	0.2600	-0.6056	0.2680	-0.5847	0.2914
	20	-0.5437	0.3003	-0.5904	0.2216	-0.6064	0.2489	-0.5639	0.2713
	30	-0.5494	0.3688	-0.6542	0.1358	-0.6540	0.2357	-0.5591	0.3518
	40	-0.5860	0.3167	-0.8675	0.0495	-0.7389	0.2180	-0.5888	0.3161
	50	-0.6118	0.3271	-1.1443	-0.0634	-0.9202	0.0743	-0.6406	0.2747
50	0	-0.3182	0.3307	-0.3221	0.3258	-0.3229	0.3258	-0.3221	0.3258
	10	-0.3590	0.3941	-0.3957	0.2789	-0.3849	0.3463	-0.3740	0.3702
	20	-0.3897	0.4280	-0.6187	0.1486	-0.5100	0.3132	-0.3943	0.3981
	30	-0.3899	0.4063	-0.6582	0.0570	-0.5415	0.2698	-0.3931	0.3760
	40	-0.5227	0.3872	-0.7811	-0.0991	-0.6901	0.1930	-0.5253	0.3778
	50	-0.6747	0.5454	-1.2366	-0.2627	-0.8577	0.3193	-0.6355	0.6042
80	0	-0.3636	0.2743	-0.3636	0.2743	-0.3636	0.2743	-0.3636	0.2743
	10	-0.4382	0.2880	-0.4961	0.1998	-0.4959	0.2367	-0.4442	0.2644
	20	-0.4574	0.2349	-0.6484	0.0305	-0.6066	0.1300	-0.4670	0.2032
	30	-0.5841	0.2167	-0.9153	-0.1032	-0.7510	0.0702	-0.5791	0.2005
	40	-0.7151	0.1712	-1.1495	-0.2600	-0.8868	0.0140	-0.6750	0.1465
	50	-0.8315	0.1559	-1.2902	-0.4159	-0.9629	-0.0568	-0.7651	0.1675
100	0	-0.2159	0.2431	-0.2159	0.2431	-0.2159	0.2431	-0.2159	0.2431
	10	-0.2165	0.2427	-0.2506	0.2281	-0.2515	0.2281	-0.2325	0.2423
	20	-0.2763	0.3475	-0.4533	0.1048	-0.3803	0.2437	-0.2793	0.3481
	30	-0.3245	0.3263	-0.5296	0.0500	-0.4251	0.1984	-0.3236	0.3332
	40	-0.4460	0.3178	-0.8253	-0.0609	-0.6152	0.1008	-0.4406	0.3179
	50	-0.5708	0.3374	-1.1290	-0.2312	-0.7591	0.1781	-0.5705	0.3383

Table 5: Range of slope for various residuals.

n	cp	CS Min	CS Max	MCS Min	MCS Max	GMCS Min	GMCS Max	HMCS Min	HMCS Max
30	0	0.4518	1.1746	0.4518	1.1671	0.4518	1.1746	0.4518	1.1671
	10	0.5005	1.1745	0.4008	1.1699	0.4128	1.1727	0.5020	1.1913
	20	0.5203	1.1195	0.4130	1.1246	0.4629	1.1590	0.5299	1.2217
	30	0.5164	1.1159	0.3868	1.0473	0.4819	1.2140	0.5213	1.1200
	40	0.5080	1.1996	0.3893	1.2689	0.5248	1.5320	0.5459	1.2295
	50	0.4581	1.3049	0.3382	1.3125	0.4956	1.5140	0.5038	1.2179
40	0	0.6707	1.2382	0.6699	1.2378	0.6702	1.2382	0.6699	1.2378
	10	0.6635	1.2103	0.6635	1.2187	0.6635	1.2175	0.6652	1.2018
	20	0.6193	1.1785	0.6007	1.1928	0.6086	1.1908	0.6258	1.2304
	30	0.6042	1.2047	0.5728	1.2083	0.5983	1.2083	0.6063	1.2102
	40	0.5809	1.2084	0.5368	1.2056	0.5819	1.2192	0.5781	1.1920
	50	0.5109	1.1698	0.4696	1.2414	0.5695	1.3510	0.5288	1.1929
50	0	0.5105	1.1695	0.5102	1.1759	0.5102	1.1759	0.5102	1.1759
	10	0.5256	1.1914	0.4819	1.1724	0.4891	1.1792	0.5292	1.2249
	20	0.5682	1.1771	0.4746	1.1771	0.5006	1.2103	0.5708	1.2180
	30	0.5665	1.2120	0.4691	1.1382	0.5050	1.1783	0.5697	1.2862
	40	0.5783	1.2510	0.4402	1.2160	0.4958	1.2446	0.5821	1.2826
	50	0.5423	1.2425	0.4126	1.2887	0.4994	1.3424	0.5419	1.2497
80	0	0.6704	1.2657	0.6704	1.2657	0.6704	1.2657	0.6704	1.2657
	10	0.6664	1.3532	0.6569	1.3518	0.6597	1.3509	0.6725	1.3724
	20	0.6414	1.3483	0.6180	1.3304	0.6486	1.3889	0.6650	1.3821
	30	0.5991	1.3061	0.5834	1.3893	0.6416	1.4014	0.6290	1.2800
	40	0.5776	1.4413	0.5310	1.4953	0.6240	1.5625	0.6050	1.4266
	50	0.5535	1.4246	0.5043	1.4914	0.6290	1.5658	0.6104	1.4051
100	0	0.6736	1.1872	0.6736	1.1872	0.6736	1.1872	0.6736	1.1872
	10	0.6826	1.2226	0.6590	1.1969	0.6639	1.1970	0.6885	1.2013
	20	0.6522	1.2321	0.6274	1.1625	0.6510	1.1830	0.6756	1.2387
	30	0.6321	1.3164	0.6234	1.2145	0.6551	1.2303	0.6625	1.2910
	40	0.6201	1.2692	0.5933	1.2423	0.6575	1.3040	0.6534	1.4306
	50	0.6069	1.3254	0.5473	1.2571	0.6798	1.4521	0.6255	1.3462

Table 6: Range of r for various residuals.

		Residuals							
n	cp	CS		MCS		GMCS		HMCS	
		Min	Max	Min	Max	Min	Max	Min	Max
30	0	0.7086	0.9879	0.7086	0.9879	0.7086	0.9879	0.7086	0.9879
	10	0.6320	0.9892	0.5912	0.9863	0.5763	0.9865	0.6300	0.9883
	20	0.6995	0.9857	0.5536	0.9833	0.5253	0.9769	0.6986	0.9846
	30	0.7545	0.9889	0.5864	0.9886	0.5063	0.9920	0.7548	0.9826
	40	0.7288	0.9814	0.6088	0.9851	0.5535	0.9753	0.7342	0.9763
	50	0.7327	0.9764	0.6019	0.9838	0.5075	0.9749	0.7411	0.9750
40	0	0.8800	0.9902	0.8802	0.9904	0.8801	0.9904	0.8802	0.9904
	10	0.8889	0.9922	0.8889	0.9928	0.8889	0.9923	0.8910	0.9914
	20	0.8907	0.9904	0.8624	0.9939	0.8548	0.9915	0.8749	0.9939
	30	0.8965	0.9893	0.8280	0.9892	0.8014	0.9873	0.8674	0.9904
	40	0.8695	0.9891	0.8398	0.9888	0.7950	0.9859	0.8658	0.9893
	50	0.8348	0.9830	0.7724	0.9829	0.7841	0.9851	0.8416	0.9860
50	0	0.7737	0.9914	0.7728	0.9913	0.7731	0.9913	0.7728	0.9913
	10	0.7799	0.9912	0.7673	0.9896	0.7587	0.9897	0.7730	0.9896
	20	0.8230	0.9883	0.7598	0.9873	0.7034	0.9882	0.8182	0.9903
	30	0.8290	0.9922	0.7466	0.9879	0.6880	0.9907	0.8263	0.9905
	40	0.8204	0.9927	0.7230	0.9903	0.6646	0.9876	0.8176	0.9907
	50	0.8159	0.9850	0.5685	0.9862	0.6194	0.9851	0.8110	0.9878
80	0	0.7955	0.9968	0.7955	0.9968	0.7955	0.9968	0.7955	0.9968
	10	0.7979	0.9962	0.7791	0.9961	0.7788	0.9961	0.7955	0.9963
	20	0.7999	0.9952	0.7496	0.9951	0.7453	0.9963	0.8072	0.9957
	30	0.7934	0.9935	0.7168	0.9949	0.7129	0.9953	0.8005	0.9963
	40	0.7662	0.9923	0.6898	0.9864	0.6551	0.9909	0.7839	0.9918
	50	0.7449	0.9924	0.6655	0.9892	0.6352	0.9888	0.7737	0.9902
100	0	0.9079	0.9951	0.9079	0.9951	0.9079	0.9951	0.9079	0.9951
	10	0.9018	0.9944	0.9011	0.9942	0.8991	0.9942	0.9018	0.9945
	20	0.9161	0.9947	0.8795	0.9943	0.8739	0.9940	0.9172	0.9942
	30	0.9147	0.9942	0.8734	0.9954	0.8671	0.9937	0.9189	0.9945
	40	0.8925	0.9934	0.8363	0.9935	0.8048	0.9937	0.9051	0.9945
	50	0.8814	0.9926	0.7691	0.9901	0.7725	0.9916	0.8819	0.9932

Table 7: Range of R square for various residuals.

n	cp	Residuals							
		CS		MCS		GMCS		HMCS	
		Min	Max	Min	Max	Min	Max	Min	Max
30	0	0.697770	0.987440	0.697770	0.987436	0.697770	0.987440	0.697770	0.987436
	10	0.618353	0.988805	0.576048	0.985776	0.560573	0.986021	0.616338	0.987874
	20	0.688421	0.985184	0.537647	0.982726	0.508295	0.976072	0.687429	0.984035
	30	0.745409	0.988324	0.571064	0.988172	0.488007	0.991697	0.745687	0.981917
	40	0.718706	0.980692	0.594314	0.984564	0.537006	0.974398	0.724345	0.975429
	50	0.722784	0.975442	0.587157	0.983185	0.489245	0.973992	0.731506	0.973194
40	0	0.876751	0.989948	0.876980	0.990166	0.876812	0.990151	0.876980	0.990166
	10	0.885874	0.992027	0.885874	0.992614	0.885874	0.992084	0.888045	0.991197
	20	0.887683	0.990177	0.858691	0.993767	0.850850	0.991266	0.871507	0.993702
	30	0.893604	0.988998	0.823379	0.988911	0.795997	0.986989	0.863790	0.990096
	40	0.864034	0.988741	0.835486	0.988464	0.789410	0.985558	0.860477	0.989044
	50	0.830379	0.982441	0.766298	0.982393	0.778252	0.984652	0.837160	0.985614
50	0	0.768924	0.991171	0.767971	0.991125	0.768287	0.991160	0.767971	0.991125
	10	0.775234	0.991007	0.762369	0.989416	0.753523	0.989453	0.768131	0.989416
	20	0.819201	0.988062	0.754662	0.987031	0.697061	0.987945	0.814363	0.990096
	30	0.825339	0.992022	0.741201	0.987684	0.681348	0.990477	0.822573	0.990301
	40	0.816560	0.992531	0.717126	0.990133	0.657451	0.987331	0.813693	0.990440
	50	0.811932	0.984582	0.559555	0.985941	0.611270	0.984772	0.807013	0.987454
80	0	0.792855	0.996725	0.792855	0.996724	0.792855	0.996725	0.792855	0.996724
	10	0.795313	0.996153	0.776185	0.996046	0.775949	0.996098	0.792888	0.996249
	20	0.797255	0.995098	0.746365	0.995031	0.742034	0.996268	0.804742	0.995641
	30	0.790696	0.993427	0.713132	0.994786	0.709131	0.995243	0.797891	0.996212
	40	0.763132	0.992224	0.685788	0.986239	0.650649	0.990746	0.781099	0.991691
	50	0.741628	0.992342	0.661203	0.989110	0.630503	0.988623	0.770766	0.990114
100	0	0.906933	0.995087	0.906933	0.995087	0.906933	0.995087	0.906933	0.995087
	10	0.900811	0.994303	0.900055	0.994094	0.898096	0.994155	0.900780	0.994440
	20	0.915230	0.994669	0.878275	0.994195	0.872588	0.993929	0.916357	0.994125
	30	0.913772	0.994150	0.872123	0.995328	0.865699	0.993610	0.918021	0.994494
	40	0.891353	0.993316	0.834587	0.993446	0.802796	0.993645	0.904082	0.994466
	50	0.880151	0.992479	0.766697	0.989959	0.770193	0.991537	0.880645	0.993152

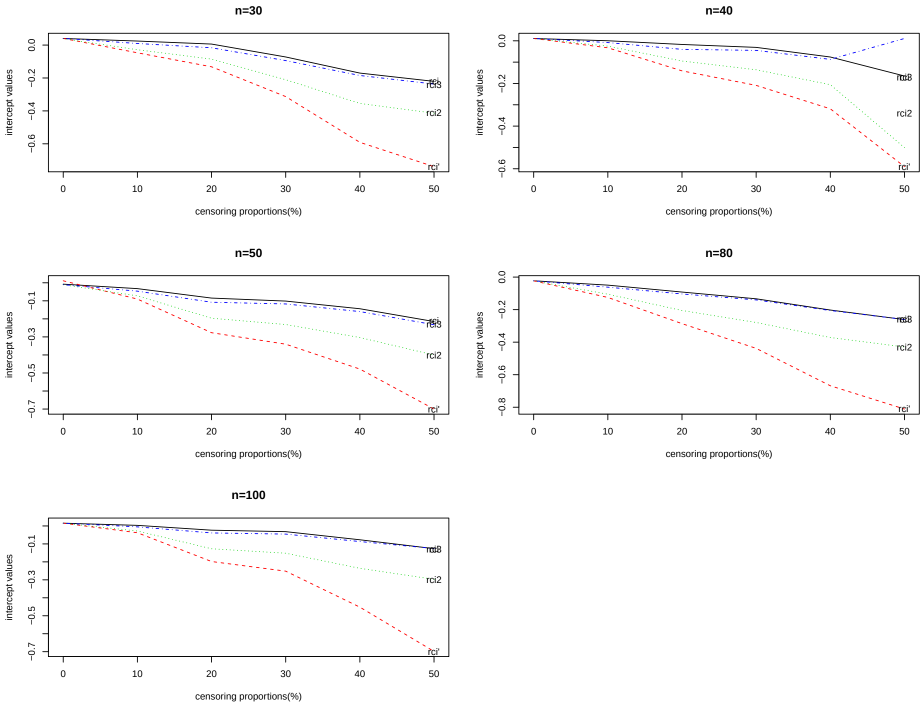

Figure 1: Comparison of residuals for estimated values of intercept.

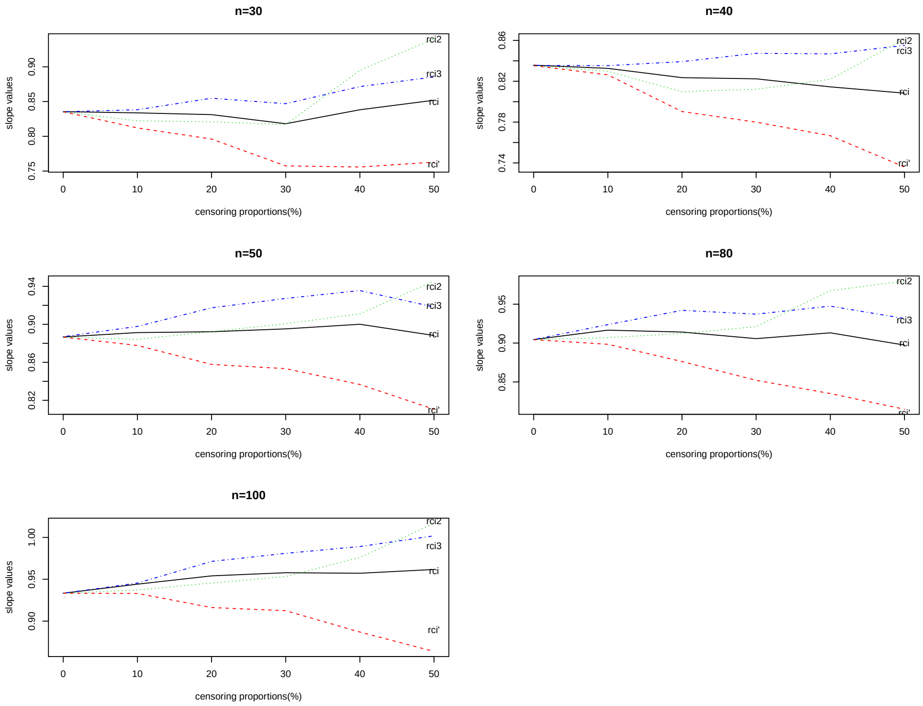

Figure 2: Comparison of residuals for estimated values of slope.

4. Conclusion

Based on the bias, standard error and root mean square error, we can conclude that poorer performance for the parameter estimates at higher censoring proportions and smaller sample sizes. This indicates that the estimates perform better when sample sizes are larger and censoring proportions are lower.

We can conclude that higher number of sample size make intercept closer to zero and slope closer to one. While for the range for r and R2 values becomes wider as the censoring proportion increases. Based on results, we can see that proposed modification of the Cox Snell residual using geometric mean perform slightly better than the existing methods.

Acknowledgment

We would like to thank the Putra Grant, Vot 9595300, Universiti Putra Malaysia for supporting this research project.

References

Abu-Zinadah HH (2014). "Goodness-of-Fit Tests for the Exponentiated Gompertz Distribution." *International Journal of Innovation in Science and Mathematics*, **2**, 372–380. URL https://www.kau.edu.sa/Show_Res.aspx?Site_ID=0009071&LNG=EN&RN=65877.

Chen Z (1997). "Parameter Estimation of the Gompertz Population." *Biometrical Journal*, **39**, 117–124. URL https://doi.org/10.1002/bimj.4710390111.

Collett D (2003). *Modeling Survival Data in Medical Research.* Chapman and Hall/Crc. ISBN 978-1584883258, URL https://www.amazon.com/Modelling-Survival-Medical-Research-Second/dp/1584883251.

Cox DR, Snell EJ (1968). "A General Definition of Residuals." *Journal of the Royal Statistical Society. Series B (Methodological)*, **30**(2), 248–275. URL https://www.jstor.org/stable/2984505?seq=1.

de Andrade TAN, Chakraborty S, Handique L, Gomes-Silva F (2019). "The Exponentiated Generalized Extended Exponential Distribution." *Journal of Data Science*, **17**, 299–330. URL 10.6339/JDS.201904_17(2).0004.

El-Din MMM, Abdel-Aty Y, Abu-Moussa MH (2016). "Statistical Inference for the Gompertz Distribution Based on Type-II Progressively Hybrid Censored Data." *Communications in Statistics - Simulation and Computation*, **46**, 6242–6260. URL https://doi.org/10.1080/03610918.2016.1202270.

Farrington CP (2000). "Residuals for Proportional Hazards Models with Interval-censored Survival Data." *Journal of the International Biometric Society*, **56**, 473–482. URL https://doi.org/10.1111/j.0006-341X.2000.00473.x.

Garg ML, Rao BR, Redmond CK (1970). "Maximum-Likelihood Estimation of the Parameters of the Gompertz Survival Function." *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **19**(2), 152–159. URL https://www.jstor.org/stable/2346545?seq=1.

Gompertz B (1825). "On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies." *Philosophical Transactions of the Royal Society of London*, **115**, 513–583. URL http://links.jstor.org/sici?sici=0261-0523%281825%29115%3C513%3AOTNOTF%3E2.0.CO%3B2-T.

Gordon NH (1990). "Maximum Likelihood Estimation for Mixtures of Two Gompertz Distributions when Censoring Occurs." *Journal Communications in Statistics - Simulation and Computation*, **19**, 733–747. URL https://doi.org/10.1080/03610919008812885.

Kiani K, Arasan J (2013). "Gompertz Model with Time-dependent Covariate in the Presence of Interval-, Right- and Left-censored Data." *Journal of Statistical Computation and Simulation*, **83**, 1472–1490. URL https://doi.org/10.1080/00949655.2012.662979.

Kiani K, Arasan J (2018). "Survival Model with Doubly Interval-censored Data and Time-dependent Covariates." *Proceedings of the World Congress on Engineering*, **I**. URL https://www.semanticscholar.org/paper/Survival-Model-With-Doubly-Interval-Censored-Data-Kiani-Arasan/16509499d65015e3cc2188d1dec25c5b5288f167.

Kiani K, Arasan J, Midi H (2012). "Interval Estimations for Parameters of Gompertz Model with Time-dependent Covariate and Right Censored Data." *Sains Malaysiana*, **115**, 471–480. URL http://www.ukm.my/jsm.

Klein JP, Moeschberger ML (2003). *Survival Analysis: Techniques for Censored and Truncated Data Statistics for Biology and Health*. Springer-Verlag New York. ISBN 978-0-387-95399-1, URL https://www.springer.com/gp/book/9780387953991.

Kleinbaum DG, Klein M (2012). *Survival Analysis: A Self-learning Text*. Springer-Verlag New York. ISBN 978-1-4419-6645-2, URL https://www.springer.com/gp/book/9781441966452.

Lee ET, Wang JW (2003). *Statistical Methods for Survival Data Analysis*. John Wiley and Sons, Inc. ISBN 9780471458548, URL https://onlinelibrary.wiley.com/doi/book/10.1002/0471458546.

Lenart A (2012). "The Moments of the Gompertz Distribution and Maximum Likelihood Estimation of Its Parameters." *Scandinavian Actuarial Journal*, **2014**, 255–277. URL https://doi.org/10.1080/03461238.2012.687697.

Leung KM, Elashoff RM, Afifi AA (1997). "Censoring Issues in Survival Analysis." *Annual Review of Public Health*, **18**, 83–104. URL https://doi.org/10.1146/annurev.publhealth.18.1.83.

Prinja S, Gupta N, Verma R (2010). "Censoring in Clinical Trials: Review of Survival Analysis Techniques." *Indian Journal Community Medicine*, **35**, 217–221. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2940174/.

Sakurai R, Hattori S (2018). "Goodness-of-fit Test for the Parametric Proportional Hazard Regression Model with Interval-censored Data." *Journal of Biostatistics & Epidemiology*, **2**, 115–131. URL https://doi.org/10.1080/24709360.2018.1529347.

Sparling YH, Younes N, Lachin JM, Bautista OM (2006). "Parametric Survival Models for Interval-censored Data with Time-dependent Covariates." *Biostatistics*, **7**, 599–614. URL https://doi.org/10.1093/biostatistics/kxj028.

Weissfeld LA, Schneider H (1990). "Influence Diagnostics for the Weibull Model Fit to Censored Data." *Statistics & Probability Letters*, **9**, 67–73. URL https://doi.org/10.1016/0167-7152(90)90097-Q.

Wilson DL (1994). "The Analysis of Survival (Mortality) Data: Fitting Gompertz, Weibull, and Logistic functions." *Mechanisms of Ageing and Development*, **74**, 15–33. URL https://doi.org/10.1016/0047-6374(94)90095-7.

Witten M, Satzer W (1992). "Gompertz Survival Model Parameters: Estimation and Sensitivity." *Applied Mathematics Letters*, **5**, 7–12. URL https://doi.org/10.1016/0893-9659(92)90125-S.

Wu JW, Hung WL, Tsai CH (2004). "Estimation of Parameters of the Gompertz Distribution Using the Least Squares Method." *Applied Mathematics and Computation*, **158**, 133–147. URL https://doi.org/10.1016/j.amc.2003.08.086.

Affiliation:

Nur Niswah Naslina Azid @ Maarof
[1] Department of Mathematics
Faculty of Sciences
Universiti Putra Malaysia
Selangor, Malaysia
[2] Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
Malaysia
E-mail: niswah@uitm.edu.my

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 49 *Submitted:* 2019-12-19
February 2020 *Accepted:* 2019-12-27