

Weighted Nonlinear Least Squares Technique for Parameters Estimation of the NHPP Gompertz Model

Lutfiah Ismail Al turk*

Statistics Department, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia

Abstract

With the problem of heteroscedasticity an alternative precise estimation method of the nonlinear least squares (NLS) technique is needed. Weighted nonlinear least squares estimation (WNLSE) technique is an alternative that may increase the accuracy of parameters estimation by assigning suitable weights to the time between failures data. In the present study, the traditional maximum likelihood (ML), nonlinear least squares (NLS), and weighted nonlinear least squares (WNLS) techniques are formulated to estimate the three parameters of the NHPP Gompertz model. Empirical weighting method is investigated in NHPP Gompertz model prediction process. Three real software failure data examples are provided to analyze the performance of the three considered methods of estimation. The results of this numerical study indicate the preferences to the WNLSE method with respect to the NHPP Gompertz model's performance, also the value of the weighting factors which give the optimum solution differ according to the nature of software failure data.

Keywords: NHPP Gompertz model; Heteroscedasticity; Maximum likelihood estimation; Nonlinear least square estimation; Weighted nonlinear least squares estimation

Introduction

Modelling the software failure phenomena and estimating the model's parameters are essential matter in the software engineering field. One of the most popular approach to describe the software development process is nonhomogeneous Poisson process (NHPP) software reliability growth models (SRGMs), some of the earlier models belong to NHPP are [1-3], and many others have been proposed afterthought. Two commonly used methods for parameters estimation in NHPP SRGMs are the maximum likelihood estimation (MLE) and the least squares estimation (LSE) [4-6]. The MLE method is very intensively used in the literature because of its statistical properties, using this approach the likelihood function is maximized to determine the parameters that are most likely to produce the observed data. It is usually recommended for large samples, applicable to most models and different types of data, and produces accurate estimates with small estimated variance [7]. The LSE technique has attracted the attention of researchers because of its computational simplicity, least squares estimates are obtained by fitting a parameterized function to the data points via minimizing the sum of the deviations squared, it is simple to be calculated and programmed, but it does not use the information in the entire data set which affects its accuracy, with LSE method the importance of each data sample point is considered the same by assuming that the variance of the error term is constant. More realistic assumptions that suit the reliability data is assigning proper weights to data points through using weighted least squares analysis to try to improve the accuracy of the parameter estimation. Nonlinear least squares estimation (NLSE) and Weighted nonlinear least squares (WNLSE) techniques arise in the cases when the parameterized function is not linear in the parameters [8-11]. NHPP Gompertz model is one of the simplest S-Shaped software reliability models which consider the number of faults per unit of time as independent Poisson random variables, Sakata was the first who employed the Gompertz curve model [12], later the ability of its curve to give a good prediction of the detected cumulative number of faults encourages the researchers to develop NHPP formulation of this model which later has been widely utilized in software engineering field. In the literature, the maximum likelihood (ML), non-linear least squares (NLS), and ordinary least squares (OLS) estimation methods have been considered to estimate the parameters of NHPP Gompertz model [13-15]. The strong non-linearity of the NHPP Gompertz model may cause

inaccurate prediction results when using the MLE method, LSE method require for the variance to be constant across the time range which is not the case with this model. This paper focuses on the weighted nonlinear least squares estimation (WNLSE) method as an alternative approach to overcome the problem of having unequal variance (heteroscedastic) of NHPP Gompertz model, this approach can often be used to increase the effectiveness of parameter estimation. This is accomplished by trying to give each data point its appropriate amount of influence over the parameter estimates based on weighting factors. The rest of this paper is organized as follows: Section 2 provides a brief historical review and some characteristics formulas of NHPP Gompertz model. Section 3 introduces the MLE, NLSE and WNLSE methods for NHPP Gompertz model in case of time data. To show the performance of the WNLSE method three of real data examples are discussed in Section 4. Finally, the last section summarizes the conclusions of the study.

NHPP Gompertz Model

Goel and Okumoto [1] were the first who proposed the NHPP Gompertz model which is later stated to be used by many computer scientists because of the good approximation of this S-Shaped growth model to a cumulative number of software faults observed in testing phase. S-shaped models are used to analyze software reliability where the mean value function versus time plot usually is S-shaped. This model can be expressed either by its mean value function which represents the total cumulated number of errors observed within time is $(0, t_i]$ or by its intensity function which is an appropriate way of specifying how the present failure depends on the past failures, their formulas are respectively as follows:

$$m(t_i; k, b, c) = kc^{b^t} \quad (1)$$

*Corresponding author: Lutfiah Ismail Al turk, Faculty of Sciences, Statistics Department, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia, Tel: 695231526928; E-mail: lturk@kau.edu.sa

Received March 31, 2018; Accepted April 15, 2018; Published April 25, 2018

Citation: Al turk LI (2018) Weighted Nonlinear Least Squares Technique for Parameters Estimation of the NHPP Gompertz Model. J Inform Tech Softw Eng 8: 231. doi: [10.4172/2175-7866.1000231](https://doi.org/10.4172/2175-7866.1000231)

Copyright: © 2018 Al turk LI. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

$$\xi(t_i; k, b, c) = \frac{dm(t_i; k, b, c)}{dt} = kc^{b^t} b^t (inc)(inb) \quad (2)$$

where $k > 0, 0 < b, c < 1, c$ is a constant, b is the shape parameter, and k is the expected initial error content of a software product, the NHPP Gompertz model is a flexible model, demonstrated by the different shapes of its mean value and intensity functions that shown in Figure 1 for different values of parameters, The cumulative failure occurring silhouette in Figure 1a is S-shaped, the growth rate of cumulative failure occurring is small at beginning; then it increases to reach a maximum, and in the end slowed down to give the S-shape form. The shape of the intensity function in Figure 1b is either decreasing or showing an initial raise before start declining (unimodal shape). The rest of this model's characteristics are as follows: The number of remaining errors function based on the mean value function in Equation (1) can be expressed by:

$$N(t_i; k, b, c) = k - m(t_i; k, b, c) = k(1 - c^{b^t}) \quad (3)$$

While, the corresponding error detection rate function is given by using Equations (2 and 3) as follows:

$$D(t_i; k, b, c) = \frac{\xi(t_i; k, b, c)}{\hat{I}(t_i; k, b, c)} = \frac{b^t (inc)(inb)}{-c^{b^t} - 1} \quad (4)$$

Additionally, the conditional reliability and mean time between software failures functions can be respectively written as:

$$D(t_i; k, b, c) = \frac{\xi(t_i; k, b, c)}{\hat{I}(t_i; k, b, c)} = \frac{b^t (inc)(inb)}{-c^{b^t} - 1} \quad (5)$$

$$\mu TBF(t_i; k, b, c) = \frac{1}{\xi(t_i; k, b, c)} = \frac{c^{-b^t} b^{-t}}{k(inc)(inb)} \quad (6)$$

Estimation Procedures for NHPP Gompertz Model

The formulas for determining the estimates of NHPP Gompertz model's parameters for the maximum likelihood (ML), nonlinear least squares (NLS), and weighted nonlinear least squares (WNLS) estimation methods will be presented in this section.

Maximum likelihood estimation method

The maximum likelihood (ML) estimates of the NHPP Gompertz model parameters are provided as follows: Given the mean value and intensity functions in Equations (1 and 2) respectively, the joint density or the likelihood function of t_1, t_2, \dots, t_n of the NHPP Gompertz growth model can be obtained as:

$$L_{NHPP}(k, b, c | \underline{S}) = e^{-m(t_i; k, b, c)} \prod_{i=1}^n \xi(t_i; k, b, c) = \exp(-kc^{b^n}) \left[k^n (inc)^n (inb)^n \prod_{i=1}^n c^{b^{t_i}} b^{t_i} \right] \quad (7)$$

Hence the log-likelihood function is:

$$L_{NHPP}(k, b, c | \underline{S}) = -kc^{b^n} n \ln(k) n \ln(inc) n \ln(inb) \sum_{i=1}^n b^{t_i} \ln(a) \prod_{i=1}^n S_i \ln(b) \quad (8)$$

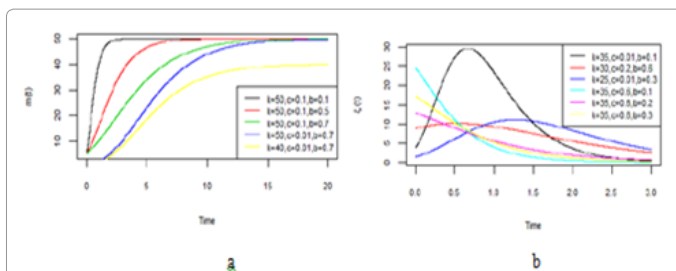


Figure 1: Different mean value and intensity function of the NHPP Gompertz model a) mean value function b) intensity function.

By differentiating Equation (8) with respect to $k, a,$ and b respectively, and equating to zero, yields:

$$k = \frac{n}{c^{b^n n}} \quad (9)$$

$$\frac{n^{b^n n}}{c} = \frac{1}{c} \left[\frac{n}{\ln(c)} + \sum_{i=1}^n b^{t_i} \right] \quad (10)$$

and

$$ns_n b s_n^{-1} \ln(c) = \frac{1}{b} \left[\frac{n}{\ln(b)} + \sum_{i=1}^n S_i + \sum_{i=1}^n S_i b^{t_i} \ln(c) \right] \quad (11)$$

Equations (10) and (11) can be solved numerically to obtain \hat{c}_{MLE} and \hat{k}_{MLE} , then by substituting them in Equation (9), \hat{k}_{MLE} can be found.

Nonlinear least squares and weighted nonlinear least squares estimation methods

LSE is a popular technique and widely used in many fields for function fit and parameter estimation. The LSE method may be simple but very useful in estimating model parameters. It finds values of the parameters such that the sum of the squares of the difference between the fitting function and the experimental data is minimized [16]. Suppose that the following n data points are provided: $(t_1, y_1), (t_2, y_2), \dots, (t_n, y_n)$. The model to be fitted to these data is defined by:

$$y_i = f(x_i, \theta) + \varepsilon_i \quad (12)$$

Where $i = 1, \dots, n, \theta$ is the parameter vector, and ε_i is the error term. In statistics theory ε_i is assumed as independent variables of normal distribution $N(0, \sigma^2)$, where σ^2 is the variance of the normal distribution. Mathematically, the traditional nonlinear least squares estimation method concerns in determining the value of the unknown parameters that minimizes the following quantity:

$$\psi_{NLF} = \sum_{i=1}^n [y_i - f(x_i, \beta)]^2 \quad (13)$$

If $\varepsilon_i \sim N(0, \sigma^2)$, and $\sigma_i^2 (1 \leq i \leq n)$ is not constant, the phenomenon is called heteroscedasticity. In this case the WNLS is recommended as it increases the accuracy of the estimators, the unknown parameters in this estimation method have to be estimated by minimizing the following objective function:

$$\psi_{WNLF} = \sum_{i=1}^n w_i (y_i - f(x_i, \beta))^2 \quad (14)$$

Where $w_i > 0$: is the weighting function, selecting the appropriate weighting function playing an important role in enhancing the obtained prediction results, one of the approach of calculating these factors are the optimal which based on use the inverse of the variance, another way is to specify these factors empirically.

The NLS estimates of the NHPP Gompertz model's parameters $k, c,$ and b are to minimize the following objective function:

$$\psi_{NLF}(k, c, b) = \sum_{i=1}^n \left[t_i - \frac{c^{-b^{t_i}} b^{-t_i}}{k \ln(c) \ln(b)} \right]^2 \quad (15)$$

Differentiating Equation (15) with respect to $k,$ and b and setting the partial derivatives equal to 0, $\left(\frac{\partial \psi_{NLF}(k, c, b)}{\partial k} = 0, \frac{\partial \psi_{NLF}(k, c, b)}{\partial c} = 0, \frac{\partial \psi_{NLF}(k, c, b)}{\partial b} = 0 \right)$

the following three equations will be obtained:

The solution of the Equations (16), (17), and (18) is the NLS estimates \hat{c}_{NLSE} , \hat{c}'_{NLSE} and \hat{b}'_{NLSE} of the NHPP Gompertz model.

Similarly, the WNLS estimates of the NHPP Gompertz model's parameters k , c , and b are to minimize the following objective function:

$$k = \frac{1}{\log c \log b} \frac{\sum_{i=1}^n b^{-2t_{ic}-2b^{t_i}}}{\sum_{i=1}^n t_i b^{-t_{ic}-b^{t_i}}} \quad (16)$$

$$\sum_{i=1}^n \left(t_i - \frac{c - b^{t_i} b^{-t_i} \sum_{j=1}^n t_j b^{-t_{jc}-b^{t_j}}}{\sum_{j=1}^n b^{-2t_{jc}-2b^{t_j}}} \right) \left[b^{-t_{ic}} c^{-b^{t_i}} (b^{t_i} \log c - 1) \right] = 0 \quad (17)$$

$$\sum_{i=1}^n \left(t_i - \frac{c - b^{t_i} b^{-t_i} \sum_{j=1}^n t_j b^{-t_{jc}-b^{t_j}}}{\sum_{j=1}^n b^{-2t_{jc}-2b^{t_j}}} \right) \left\{ c^{-b^{t_i}} \left[t_i \log b (b^{-t_i} + \log c) + b^{-t_i} \right] \right\} = 0 \quad (18)$$

The solution of the Equations (16), (17), and (18) is the NLS estimates (k' NLSE, c' NLSE and b' NLSE) of the NHPP Gompertz model.

Similarly, the WNLS estimates of the NHPP Gompertz model's parameters k , c , and b are to minimize the following objective function:

$$\psi_{WNLS}(k, c, b) = \sum_{i=1}^n w_i \left(t_i - \frac{c^{-b^{t_i}} b^{-t_i}}{k \log c \log b} \right)^2 \quad (19)$$

Differentiating Equation (19) with respect to k , a and b and setting:

$$\left(\frac{\partial \psi k(k, c, b)}{\partial k} = 0, \frac{\partial \psi k(k, c, b)}{\partial c} = 0, \frac{\partial \psi k(k, c, b)}{\partial b} = 0 \right) \quad (20)$$

then solving for k , a and b , yields:

$$k = \frac{1}{\log c \log b} \frac{\sum_{i=1}^n w_i b^{-2t_{ic}-2b^{t_i}}}{\sum_{i=1}^n t_i b^{-t_{ic}-b^{t_i}}} \quad (21)$$

$$\sum_{i=1}^n w_i \left(t_i - \frac{c^{-b^{t_i}} b^{-t_i} \sum_{j=1}^n t_j b^{-t_{jc}-b^{t_j}}}{\sum_{j=1}^n b^{-2t_{jc}-2b^{t_j}}} \right) \left[b^{-t_{ic}} c^{-b^{t_i}-1} (b^{t_i} \log c - 1) \right] = 0 \quad (22)$$

$$\sum_{i=1}^n w_i \left(t_i - \frac{c^{-b^{t_i}} b^{-t_i} \sum_{j=1}^n t_j b^{-t_{jc}-b^{t_j}}}{\sum_{j=1}^n b^{-2t_{jc}-2b^{t_j}}} \right) \left\{ c^{-b^{t_i}-1} \left[t_i \log b (b^{-t_i} + \log c) + b^{-t_i} \right] \right\} = 0 \quad (23)$$

Application of the Estimations Techniques

Description of datasets

To evaluate the predictive capability of the MLE, LSE, and WNLS techniques the following three data sets are used: Data1 is taken from [17], it contains 30 execution times between successive failures in tens of seconds; Data2 is presented by Kim HC [18] consists of 41 time between failures in hours; Data3 were collected from a Philips development centre and consists of 246 inter-failure times in minutes. The three data sets are listed in Tables 1-3. Respectively; read from left to right, and represented graphically in Figure 2.

Selection of weight functions

In WNLS method, suggesting suitable empirical weight functions is critical and play an important role in enhancing the obtained prediction results. In theory, the optimal weight function is $w_i = 1/t_i^2$, where $i = 1, \dots, n$, another type of weighting methods is empirical weight function [19]. Utilising the WNLS method to find the parameter estimates permits the weights to determine the impact of each data item to the final parameter estimates. In fact, the weight for each data item is

| Times between failures in tens of seconds | | | | | | | | | |
|---|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 0.479 | 0.266 | 0.277 | 0.554 | 1.034 | 0.949 | 0.693 | 0.597 | 0.117 | 0.17 |
| 0.117 | 1.274 | 0.469 | 1.174 | 0.693 | 1.908 | 0.135 | 0.277 | 0.596 | 0.757 |
| 0.437 | 2.23 | 0.437 | 0.34 | 0.405 | 0.575 | 0.277 | 0.363 | 0.522 | 0.613 |

Table 1: Data 1 of the times between failures in tens of seconds, number of failures=30.

| Times between failures in hours | | | | | | | | | |
|---------------------------------|-------|-------|--------|--------|--------|--------|--------|--------|-------|
| 5.649 | 3.271 | 11.37 | 9.665 | 4.76 | 41.235 | 2.221 | 0.454 | 4.397 | 6.092 |
| 0.69 | 3.056 | 0.809 | 16.986 | 1.333 | 10.557 | 4.5 | 1.667 | 0.278 | 2.778 |
| 0.061 | 9.883 | 22.5 | 178.64 | 13.294 | 43 | 47.35 | 30.153 | 20.494 | 0.52 |
| 93.5 | 74.48 | 20.8 | 79.5 | 7.033 | 1.967 | 16.617 | 24.417 | 21.167 | 24.8 |
| 336 | | | | | | | | | |

Table 2: Data 2 of the times between failures in hours, 41 single failure data.

provided relative to the weights of the other data item; hence different sets of absolute weights can have same influences. Thus, w_i is taken to be a sequence of positive weights which considers as a random variable. The suggested empirical weight functions in this application are constructed and summarized in Table 4 based on the software testing data and some suggested constant parameters to restrict the parameter estimates in a way that ensure getting enhanced prediction results.

Goodness-of-fit evaluation criteria

The following four criteria will be used to evaluate the performance of the NHPP Gompertz model based on the three different methods of estimation and three real data sets:

- 1 Mean square error (MSE)
- 2 Mean absolute error (MAE)
- 3 Variance
- 4 Root mean square prediction error (RMSPE)

Those measures are useful and widely used in model evaluations to measure the deviation between the predicted and actual values, mathematically are defined as follows:

$$MSE = \sum_{i=1}^n \frac{(\mu TTF(t_i; \theta) - \mu \hat{TTF}(t_i; \theta))^2}{n - p} \quad (24)$$

$$MAE = \sum_{i=1}^n \frac{(\mu TTF(t_i; \theta) - \mu \hat{TTF}(t_i; \theta))^2}{n - p} \quad (25)$$

$$Variance = \sqrt{\sum_{i=1}^n \frac{(\mu TTF(t_i; \theta) - \mu \hat{TTF}(t_i; \theta) - Bias)^2}{n - 1}} \quad (26)$$

$$RMSPE = \sqrt{Variance^2 + Bias^2} \quad (27)$$

Where,

$\mu TTF(t_i; \theta)$: actual mean time between software failures

$\mu \hat{TTF}(t_i; \theta)$: predicted mean time between software failures

n : the number of observations

p : the number of parameters to be estimated.

| Philips failure data | | | | | | | | | | |
|----------------------|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1.054 | 24.8674 | 30.858 | 0.0022 | 10.245 | 0.825 | 1.7927 | 1.0913 | 4.0461 | 1.2304 | 4.8543 |
| 8.1568 | 0.8003 | 0.0737 | 0.8878 | 3.9432 | 3.1141 | 4.0786 | 0.0026 | 1.0979 | 1.6862 | 0.0089 |
| 0.3083 | 0.7192 | 0.0059 | 0.0424 | 0.105 | 0.0114 | 0.1307 | 3.9438 | 2.0688 | 0.0511 | 0.7924 |
| 0.0114 | 0.0043 | 2.8295 | 0.004 | 0.2498 | 0.0023 | 0.0061 | 0.7064 | 0.192 | 0.0545 | 0.0924 |
| 0.6828 | 0.028 | 0.1678 | 0.9687 | 0.0153 | 0.0405 | 0.7636 | 0.1485 | 3.0432 | 0.0174 | 0.8398 |
| 0.1321 | 0.1769 | 0.9607 | 0.0323 | 0.0241 | 1.986 | 2.9412 | 0.6798 | 1.2686 | 0.0272 | 0.0318 |
| 0.012 | 1.0108 | 0.0063 | 0.6236 | 0.0195 | 0.1531 | 0.1219 | 0.0595 | 2.6693 | 0.0352 | 0.0058 |
| 0.0063 | 0.0203 | 0.0833 | 0.0497 | 0.0995 | 1.0135 | 0.6875 | 0.0125 | 0.0106 | 0.2648 | 0.0139 |
| 0.0249 | 0.6875 | 0.1573 | 0.0162 | 0.0138 | 0.0211 | 0.0132 | 0.0152 | 0.0117 | 0.0274 | 0.0089 |
| 3.7143 | 0.0267 | 0.0019 | 0.1991 | 1.955 | 0.0808 | 0.6965 | 0.2034 | 0.0652 | 0.0108 | 0.0368 |
| 0.0012 | 0.0041 | 0.0792 | 4.6968 | 0.1574 | 0.0177 | 0.0856 | 0.6574 | 0.2155 | 0.8006 | 0.0398 |
| 0.0079 | 0.0032 | 0.0994 | 0.0077 | 0.0146 | 0.0234 | 0.0039 | 0.1177 | 0.018 | 0.0122 | 0.005 |
| 0.0035 | 0.0134 | 0.0028 | 0.7002 | 3.895 | 0.2764 | 0.7778 | 0.1614 | 0.8513 | 0.1651 | 0.8563 |
| 0.2975 | 2.9149 | 0.7399 | 1.0617 | 0.0141 | 0.9364 | 0.006 | 0.2109 | 0.0905 | 0.6663 | 0.0438 |
| 3.2155 | 0.007 | 2.0793 | 1.8435 | 0.037 | 9.8811 | 0.0094 | 0.1567 | 1.9834 | 6.7883 | 7.2009 |
| 2.0055 | 0.0676 | 4.7478 | 1.09 | 3.9148 | 0.9403 | 0.1446 | 0.13 | 0.7366 | 0.0058 | 1.0615 |
| 0.0084 | 0.1699 | 0.0011 | 0.0044 | 0.0119 | 0.9526 | 3.9987 | 0.9457 | 0.0301 | 0.0111 | 0.0142 |
| 0.035 | 0.033 | 0.9981 | 7.7357 | 11.188 | 1.762 | 6.11 | 13.196 | 1.675 | 0.0088 | 0.076 |
| 22.137 | 4.8796 | 1.9997 | 3.9976 | 1.9833 | 5.281 | 12.668 | 3.1888 | 11.019 | 2.0366 | 0.8268 |
| 0.9677 | 0.0112 | 0.0967 | 3.1086 | 0.7572 | 0.2747 | 0.0583 | 0.0398 | 0.945 | 4.7504 | 0.0008 |
| 0.003 | 0.2525 | 0.0033 | 2.6702 | 5.3665 | 13.928 | 15.888 | 0.0218 | 15.028 | 13.972 | 12.104 |
| 1.8989 | 26.8023 | 7.9313 | 4.0486 | 0.1535 | 44.866 | 18.207 | 8.0884 | 1.6854 | 7.1911 | 2.8545 |
| 4.0503 | 13.9556 | 13.989 | 0.024 | | | | | | | |

Table 3: Philips failure data 3: execution times between successive failures in minutes, number of failures=246.

And the bias, the average difference between the estimator and the truth, is obtained as follows:

$$Bias = \sum_{i=1}^n \frac{(\mu TTF(t_i; \theta) - \mu \hat{TTF}(t_i; \theta))}{n}$$

The smallest values of the above criteria indicate that the model does fit the data well [20-22].

Data analysis of results

This section evaluates the performance of maximum likelihood, nonlinear least squares, and weighted nonlinear least squares estimation methods for NHPP Gompertz model using MSE, AME, Variance, RMSPE criteria. Three real data sets are used and 10 different empirical weight functions for the WNLSE method, the results are summarized in Table 5, according to the obtained results the following comparisons are done:

Comparison between MLE and NLSE method: For the first and second data sets (Data 1 and Data 2) with n=30 and 41 respectively, the NLSE method gives better performance than the MLE method based on the four selected criteria. While for Data 3 the larger studied data set n=246, the MLE method gives better performance than the NLSE according to the MSE and AME criteria while the other two criteria show that the best prediction results are obtained using the NLSE. In more details the value of the MSE is 21.6463 for MLE method and 22.1516 for the NLSE method, while AME criteria is 2.5233 for the MLE method and 2.8201 for the NLSE method. The values of Variance and RMSPE equal to 4.8800 and 4.9591 respectively for the MLE method but equal to 4.7593 and 4.7441 respectively for the NLSE method.

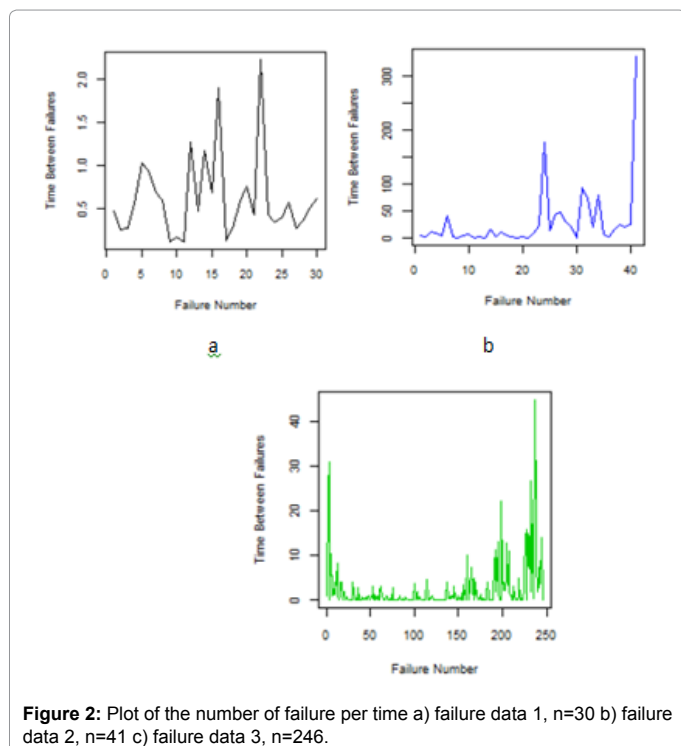


Figure 2: Plot of the number of failure per time a) failure data 1, n=30 b) failure data 2, n=41 c) failure data 3, n=246.

| Empirical weight functions Name | Empirical weight functions Formula (W_i) |
|---------------------------------|--|
| Ewf1 | i^η |
| Ewf2 | $i^{-\eta}$ |
| Ewf3 | $\frac{(\sum_{j=1}^i x_j)^\beta}{i}$ |
| Ewf4 | $\frac{1}{(\sum_{j=1}^i x_j)^\beta}$ |
| Ewf5 | $\frac{1}{(\sum_{j=1}^i x_j)^\beta}$ |
| Ewf6 | $\left(\sum_{j=1}^i x_j\right)^\beta$ |
| Ewf7 | i |
| Ewf8 | $\frac{1}{i}$ |
| Ewf9 | $\frac{1}{i^{-\eta}}$ |
| Ewf10 | $\frac{1}{i^{-\eta}}$ |

Table 4: Empirical weight functions.

where $i = 1, 2, \dots, n$, $\eta = \frac{1}{n}$, and $\beta = 0.5$.

Overall, our results show that the NLSE is superior to the MLE for the data obtained from project with small testing phase, while increasing the time of testing helps to enhance the performance of MLE method.

Comparison between WNLSE and MLE method: The WNLSE estimates based on the 10 suggested empirical weights in Table 4 is compared with the ML estimates considering the four evaluation criteria which results in 40 comparison cases for each data set:

a. For data 1: Only the Ewf5 and Ewf8 give larger Variance and RMSPE for the WNLSE estimates than the MLE estimates since $Variance_{ML} = 0.4728$ and $RMSPE_{ML} = 0.4813$ while

and $Variance_{WNLSE} = 0.5426$ and $RMSPE_{WNLSE} = 0.5719$ for Ewf5 and Ewf8 respectively, therefore 36 of the 40 cases show the superiority of the WNLSE method over the MLE.

b. For data 2: all the 40 cases show that the WNLSE method performs better than the MLE method.

c. For data 3: According to the MSE criteria five of the weight functions show that the WNLSE method gives the best fit model, but with respect to the AME criteria the MLE method gives better prediction results comparing with all the 10 weight functions, while for the Variance and RMSPE criteria all the 10 weight functions show the superiority of the WNLSE method over the MLE method. Overall for this data set, 21 cases show that the best predictive capability comes

when using the WNLSE method and 19 cases demonstrate that the best performance model is obtained when using the MLE method.

Comparison between WNLSE and NLSE method: Also, the majority of the considered cases is for the sake of the WNLSE method against the NLSE method; in more details:

a. For data 1: 36 out of 40 cases exhibit that the WNLSE method has lower evaluation criteria than the NLSE method.

b. For data 2: 25 out of 40 cases illustrate that the performance of the WNLSE method is better than the NLSE method.

c. For data 3: 24 out of 40 cases show that the WNLSE has better prediction performance than the NLSE.

Comparison between the empirical weight functions:

a. Objectively, based on the four selected criteria: Ewf1 gives the best fit model for Data 1.

b. For data 2: the MSE and AME criteria show that Ewf8 gives the best performance model while the Variance and RMSPE demonstrate that Ewf10 gives the best accurate model.

c. For data 3: the MSE and AME criteria show that Ewf5 gives the best performance model while the Variance and RMSPE demonstrate that Ewf4 gives the best accurate model.

In the other hand and according to our subjective judgement, based on the graphical representation in Figures 3 and 4: Ewf7, Ewf2, and Ewf3 give the best fit models for Data 1, 2, and 3 respectively. Figure 3 presents the selected best fit models objectively and subjectively; while Figure 4 shows the empirical weight functions that give the best fit models according to the graphical criteria.

Concluding Remarks

This paper has provided a detailed comparative study between the MLE, NLSE, and WNLSE approaches using three different SRGMs failure data and based on the NHPP Gompertz model. In reliability engineering, not all the failures have the same contribution to the overall system reliability, assigning different weight according to the significance of the failure helps to make the data more representative to the software failure behavior. Our main research objective is to investigate the performance of the WNLSE method in the presence of unequal variance of NHPP Gompertz model. Empirical weight functions based on software testing data is one of approaches that can be used in the WNLSE method. Application results show that the best empirical weight function depends on various factors and differs according to the nature of failure data sets. It also shown that in most cases the WNLSE approach has lower values of the chosen evaluation criteria and consequently better predictive ability comparable with the MLE and NLSE methods.

For the problem of unequal variances of SRGMs associated with reliability data the WNLSE approach is recommended to be checked for other types of growth models. Also obtaining the mathematical formula of the NHPP Gompertz model variance to discover the performance of the optimal weighting approach which may offer better predictive accuracy comparable with the empirical approach is another interesting point to be considered.

Reach to a sound conclusion and choose the best fit model for each project failure data is crucial, to make a reasonable decision subjective and objective methods were used in our examples, it is indicated that different evaluation criteria give different judges about the accuracy of the selected estimation method, also based on the considered objective

| Method of estimation | Data 1 (30) | | Data 2 (41) | | Data 3 (246) | |
|----------------------|-------------|---------------------------|-------------|------------------|--------------|----------------|
| | k | MSE AME Variance RMSPE | k | MSE AME | k | MSE AME |
| | c | | c | Variance RMSPE | c | Variance RMSPE |
| | b | | b | | b | |
| MLE | 30.64931 | 0.2131 | 41.0013 | 2390.637 | 246.0065 | 21.6463 |
| | 0.0405 | 0.4135 | 0.0098 | 40.848 | 0.0386 | 2.5233 |
| | 0.7654 | 0.4728 | 0.9901 | 51.2611 | 0.98 | 4.88 |
| | | 0.4813 | | 52.38 | x | 4.9591 |
| NLSE | 32.481 | 0.2096 | 43.6406 | 1610.832 | 243.2227 | 22.2516 |
| | 0.0311 | 0.392 | 0.0251 | 27.9819 | 0.0201 | 2.8201 |
| | 0.7012 | 0.4618 | 0.9901 | 40.8794 | 0.9766 | 4.7441 |
| | | 0.4681 | | 41.4358 | | 4.7593 |
| WNLSE (Ewf1) | 34.2313 | 0.1702 | 40.7631 | 1593.83 | 231.2387 | 21.5748 |
| | 0.0655 | 0.3579 | 0.0331 | 27.7 | 0.0314 | 2.8293 |
| | 0.7938 | 0.4025 | 0.9913 | 40.2576 | 0.9827 | 4.6426 |
| | | 0.4039 | | 40.6856 | | 4.6482 |
| WNLSE (Ewf2) | 29.3768 | 0.1981 | 39.3751 | 1577.688 | 231.7112 | 21.9059 |
| | 0.0403 | 0.3826 | 0.0342 | 27.5415 | 0.0256 | 2.826 |
| | 0.7125 | 0.445 | 0.9911 | 40.1507 | 0.9798 | 4.6899 |
| | | 0.4499 | | 40.6068 | | 4.6994 |
| WNLSE (Ewf3) | 29.7889 | 0.1874 | 43.2786 | 1592.438 | 243.9065 | 25.2251 |
| | 0.05126 | 0.3695 | 0.0303 | 27.2795 | 0.0297 | 3.9948 |
| | 0.7334 | 0.4418 | 0.9907 | 40.9969 | 0.9909 | 5.8452 |
| | | 0.4493 | | 41.6576 | | 6.0994 |
| WNLSE (Ewf4) | 30.6705 | 0.1824 | 38.9252 | 1577.163 | 243.2387 | 21.3784 |
| | 0.0528 | 0.3701 | 0.0331 | 27.6704 | 0.0346 | 2.9112 |
| | 0.7553 | 0.4179 | 0.9909 | 40.0888 | 0.986 | 4.6048 |
| | | 0.4197 | | 40.5277 | | 4.6048 |
| WNLSE (Ewf5) | 29.755 | 0.2051 | 45.9092 | 1538.828 | 239.0957 | 21.1684 |
| | 0.0414 | 0.3774 | 0.0544 | 24.3231 | 0.04235 | 2.6735 |
| | 0.6922 | 0.4958 | 0.9907 | 44.5169 | 0.984571 | 4.6552 |
| | | 0.5134 | | 46.3769 | | 4.6792 |
| WNLSE (Ewf6) | 35.97407 | 0.1763 | 42.715 | 1846.243 | 242.0957 | 21.7427 |
| | 0.054762 | 0.3683 | 0.015 | 33.1247 | 0.0323 | 3.1125 |
| | 0.7922 | 0.4059 | 0.9901 | 41.903 | 0.9871 | 4.6831 |
| | | 0.4061 | | 41.9105 | | 4.6961 |
| WNLSE (Ewf7) | 26.6588 | 0.1816 | 41.6829 | 1790.294 | 242.4327 | 21.4497 |
| | 0.0576 | 0.3717 | 0.0198 | 31.9693 | 0.05123 | 3.2705 |
| | 0.7422 | 0.4138 | 0.9912 | 41.2448 | 0.991555 | 4.7809 |
| | | 0.4146 | | 41.2462 | | 4.8355 |
| WNLSE (Ewf8) | 28.5632 | 0.2074 | 39.9752 | 1434.6070 | 230.9874 | 26.2192 |
| | 0.0542 | 0.3684 | 0.06 | 23.6119 | 0.045436 | 4.2372 |
| | 0.6901 | 0.5426 | 0.9891 | 44.0092 | 0.993445 | 6.3598 |
| | | 0.5719 | | 46.0809 | | 6.7262 |
| WNLSE (Ewf9) | 31.498 | 0.1847 | 43.9585 | 1567.629 | 243.6426 | 21.6958 |
| | 0.0501 | 0.3708 | 0.0156 | 28.9123 | 0.0279 | 2.8291 |
| | 0.7512 | 0.4248 | 0.9872 | 40.3414 | 0.9822 | 4.6591 |
| | | 0.428 | | 40.8946 | | 4.6658 |
| WNLSE (Ewf10) | 28.2522 | 0.1967 | 40.1092 | 1522.57 | 238.866 | 21.405 |
| | 0.042 | 0.3821 | 0.0148 | 29.5018 | 0.0339 | 2.8293 |
| | 0.7112 | 0.4411 | 0.9865 | 39.1858 | 0.9844 | 4.6199 |
| | | 0.4452 | | 39.5537 | | 4.624 |

Table 5: Goodness-of-fit statistics for different methods of estimation of NHPP Gompertz model.

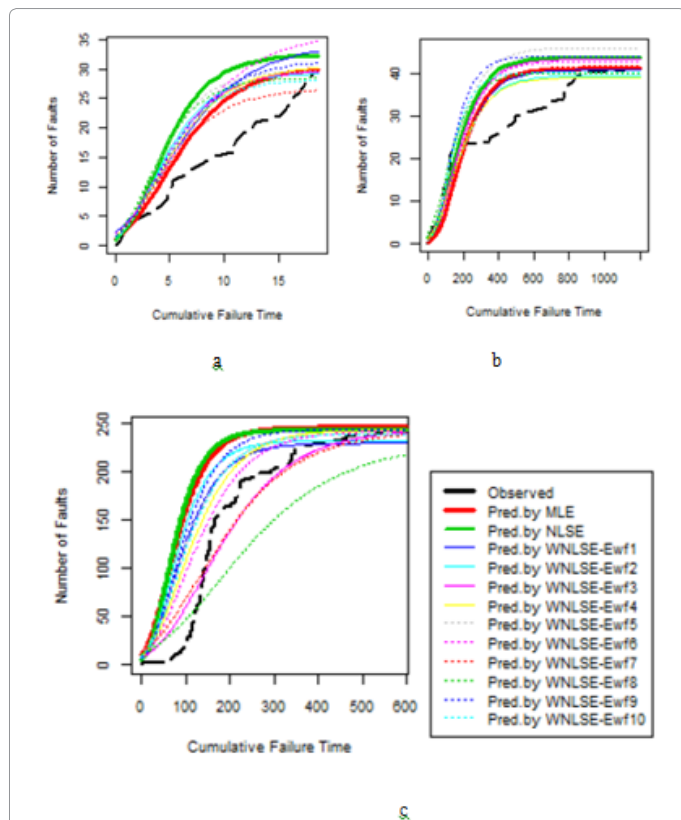


Figure 3: Objective and subjective best fit model a) data 1 b) data 2 c) data 3.

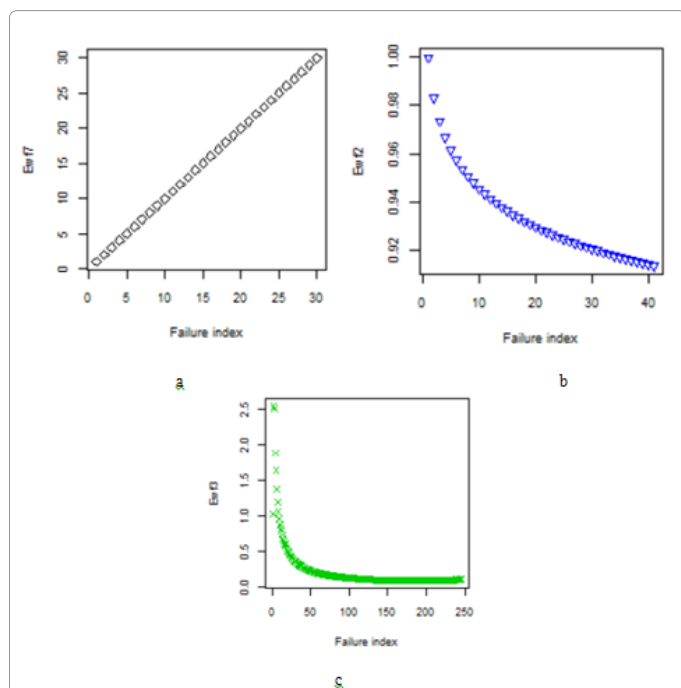


Figure 4: Empirical weight function: Ewf1, Ewf2, and Ewf3 a) data 1 b) data 2 c) data 3.

and subjective methods different best fit models are chosen for each real failure data, In this ambit using the super model approach that may help to get one better accuracy model is worth to be investigated.

Author Information

Lutfiah Ismail Al turk is currently working as Associate Professor of Mathematical Statistics in Statistics Department at Faculty of Sciences, King AbdulAziz University, Jeddah, Kingdom of Saudi Arabia. Lutfiah Ismail Al turk obtained her B.Sc degree in Statistics and Computer Science from Faculty of Sciences, King AbdulAziz University in 1993 and M.Sc (Mathematical statistics) degree from Statistics Department, Faculty of Sciences, King AbdulAziz University in 1999. She received her Ph.D in Mathematical Statistics from university of Surrey, UK in 2007. Her current research interests include Software reliability modeling and Statistical Machine Learning.

References

- Goel AL, OkumotoK (1979) Time-dependent error-detection rate model for software reliability and other performance measures. *IEEE Trans Reliability* 28: 206-211.
- Yamada S, Ohba M, Osaki S (1983) S-Shaped reliability growth modeling for software error detection. *IEEE Trans Reliability* 32: 475-484.
- Goel AL (1985) Software reliability models: Assumptions, limitations and applicability. *IEEE Transactions on Software Engineering* 11: 1411-1423.
- Zhao M, Xie M (1996) On maximum likelihood estimation for a general non-homogeneous poisson process. *Scandinavian Journal of Statistics* 23: 597-607.
- Chang YP (2001) Estimation of parameters for non homogeneous Poisson process: software reliability with change-point model. *Commun Stat Simul Comput* 30: 623-635.
- Rana R, Staron M, Berger C, Hansson J, Nilsson M, et al. (2013) Comparing between Maximum Likelihood Estimator and Non-linear Regression Estimation Procedures for NHPP Software Reliability Growth Modelling. *Joint Conference of the 23rd International Workshop on Software Measurement and the 8th International Conference on Software Process and Product Measurement Ankara*. pp: 213-218.
- Zeepongsekul P, Jayasinghe CL, Fiondella L, Nagaraju V (2016) Maximum-Likelihood Estimation of Parameters of NHPP Software Reliability Models Using Expectation Conditional Maximization Algorithm. *IEEE Transactions on Reliability* 65: 1571-1583.
- Ross GJS (1990) *Nonlinear estimation*.
- Seber GAF, Wild CJ (1989) *Nonlinear Regression*. Wiley, New York.
- Wu S, Shao J (1999) Reliability analysis using the least squares methods in nonlinear mixed-effect degradation models. *Statistica Sinica* 9: 855-877.
- Madsen K, Nielsen HB, Tingleff O (1999) *Methods for Non-Linear Least Squares Problems*. Informatics and Mathematical Modelling, Technical University of Denmark.
- Sakata K (1974) Formulation for predictive methods in software production control: Static prediction and failure rate transition model. *IEICE Trans Inf Syst* J 57: 277-283.
- Satoh D (2000) A discrete gompertz equation and a software reliability growth model. *IEICE Trans Inf & Syst* E83-D: 1508-1513.
- Ohishi K, Okamura H, Dohi T (2009) Gompertz software reliability model: estimation algorithm and empirical validation. *Journal of Systems and Software* 82: 535-543.
- Lutfiah Ismail A (2014) Comparing between maximum likelihood and least square estimators for gompertz software reliability model. *International Journal of Software Engineering & Applications* 5: 51-61.
- Hsu CJ, Huang CY, Chang JR (2011) Enhancing software reliability modelling and prediction through the introduction of time variable fault reduction factor. *Applied Mathematical Modelling* 35: 506-521.
- Littlewood B (1991) *Forecasting software reliability, Bayesian methods in reliability*. Kluwer Academic Publishers, USA.

-
18. Kim HC, Park HK (2010) The comparative study of software optimal release time based on burr distribution. *International Journal of Advancements in Computing Technology* 2: 119-128.
 19. Douglas CM, Elizabeth AP, Geoffrey GV (2012) *Introduction to linear regression analysis*. Wiley, New York. pp: 1-67.
 20. Hwang S, Pham H (2009) Quasi-renewal time-delay fault-removal consideration in software reliability modelling. *IEEE Trans on systems man and cybernetics-Part A: Systems and humans* 39: 200-209.
 21. Huang CY, Kuo SY (2002) Analysis of incorporating logistic testing-effort function into software reliability modeling. *IEEE Transactions on Reliability* 51: 261-270.
 22. Chiu KC, Huang YS, Lee TZ (2008) A study of software reliability growth from the perspective of learning effects. *Reliability Engineering and System Safety* 93: 1410-1421.