Sri Lankan Journal of Applied Statistics, Vol. 8, 2007, p. 19-33

Failure Time Regression to Determine Warranty Policies of Computers

T. R. M. T. DHARMARATHNE AND M. R. SOORIYARACHCHI

Department of Statistics
University of Colombo, Sri Lanka

ABSTRACT

The present era of consumerism has substantially redefined the relationship of buyer and seller. Heightened consumer expectations have placed more responsibility on the manufacturer for the performance of his goods and a warranty or a written affirmation of the quality or performance of the product is a requirement in most cases. However, little attention has been given by the manufacturers for the scientific setting of warranty policies on computers. Thus this paper explores a method to estimate warranty periods based on failure time of computers. The data pool consists of two data sets, namely repair data and sales data from a reputed IT solution company. For each computer, the installation date is taken from sales data and failure data is taken from repair data and the failure time is calculated. A weibull model was found to fit the data well. Cumulative probabilities of failures based on this model were taken in to account in calculating suitable warranty periods. Graphical goodness of fit consisting of probability plots and simultaneous equal precision (EP) confidence bands are used to assess the validity of distributional assumptions. The main problem encountered in this study is setting suitable warranty periods. Life times of the computers were used as the basic modelling variable. The process of solving the problem consists of various statistical methods and techniques. Since there are various engineering methods to calculate warranty periods, these statistical methods show a different angle. To illustrate a valid warranty period certain distributional assumptions were used and those assumptions laid a base for solving the problem.

Key Words: Setting warranty, Graphical goodness of fit, Maximum Likelihood Estimation, Equal Precision (EP) confidence bands, Weibull failure-time regression model

INTRODUCTION

As a general rule, producers, without first making an accurate prior assessment about the life time of their products, seek to increase their warranty periods over and above the industry standard as a means of attracting potential customers. However, there should be prior investigation into, and assessment of, the patterns of failure of products, together with the causes for failure, if one is to fix an accurate and realistic warranty period.

A review of the literature indicates one important statistical approach developed by Kalbfleisch *et al.,* (1991) which has been used to predict warranty claims. Most methods of setting warranty policy were found to be based on engineering techniques. This article illustrates the calculation of warranty periods based on modeling the number of failures within a given time period using a log-linear poisson model. Our method on the other hand models the time to failure using a Weibull parametric model.

The objective of this study was to discover an applicable solution to the warranty - setting policy. Poisson responses used by Kalbfleisch *et. al.* (1991) usually concern the occurrence of the number of events within a specified time. A time to failure analysis seeks greater power by using the time from purchase of the product until failure as the response of interest. Thus we consider a more powerful approach of determining warranty based on the lines of Whitehead (1997). In addition the previous technique is based on several assumptions which may not be valid. Another advantage of our approach is that the model considered can incorporate any number of continuous or categorical explanatory variables whereas the reviewed paper only discusses the possibility of including explanatory variables as an extension to their work.

MATERIALS AND METHODS

Data used

In this paper observations of brand - wise and maintenance - policy wise impact on the failure time of computers were carried out, in order to classify the patterns of life - time. A three year warranty period is the normal industry standard for computers, and an assessment of the merit of this bench mark was carried out to clarify the validity of warranty setting.

The Database has been taken from well reputed IT Solution Company and due to confidentiality it is unable to expose the name of the company and the brands which are used in this study. This paper concentrates on the time to the first failure of computers as this is the basis for determining warranty periods, and therefore it has eliminated repetitive repairs. Validation of data was carried out using a random sample of selected individual entities. Out of 1,752 sales in the year 2004, there were 526 failures which could be categorized as first time failures. Variables measured are brand, warranty, sales date and repair date. A new variable, namely, combination was constructed as follows: brand 1 with Warranty (combination1), brand 1 without Warranty (combination 2), brand 2 with Warranty (combination 3) and brand 2 without Warranty (combination 4). Sales data for 2004 has been taken into consideration and all the data on repairs which occurred from 2004 to the last date of inspection, which was 25/05/2007, has been recorded.

Brief description of methods used

The most widely used measure for reliability of a product is its failure time distribution. Let "T" be used to denote a non negative continuous random variable describing the failure time of a system. Then the distribution of failure time "T" can be characterized by a cumulative distribution function. Often it is possible to model this function using a parametric distribution. The model is fitted using the method of maximum likelihood explained in Meeker and Escobar (1998). Quantile functions can be calculated based on the fitted model. These quantile functions are used to set suitable warranty policies depending on the proportion of failures that the company is willing to tolerate during the warranty period.

Graphical goodness of fit

The topic of graphical goodness of fit has been discussed in detail by Gan *et al.,* (1991) Meeker and Escobar (1998), and O'Connor (2002). For the purpose of checking the distributional assumptions, the plot of failure time (t) *versus* its cumulative distribution function (CDF) denoted by $F(t)$ can be linearized by finding a transformation of $F(t)$ and t such that the relationship between the transformed variables is linear. In this paper three distributions which are commonly found to fit failure time data, namely, the normal, the log-normal and the Weibull are considered. For each distribution the nonparametric estimate of $\hat{F}(t)$ is plotted on the relevant linearized probability scale to asses the departures from a straight line and it measures the graphical goodness of fit. This is even more useful by plotting, in addition, simultaneous confidence bands. Based on the available data; any possible $F(t)$ within these bands is, statistically, consistent with the data.

Modeling failure data using the Weibull distribution

This topic has been discussed by Meeker and Escobar (1998). It is convenient to use a simple alternative parameterization for the Weibull distribution. This alternative parameterization is based on the relationship between the Weibull distribution and the smallest extreme value distribution. Then the Log-Location-Scale Based model for the Weibull distribution is denoted by $\log[t_p(x)] = \mu + \Phi_{sev}^{-1}(p)\sigma$. In the presence of a single factor with I levels this becomes $\log[t_p(x)] = \beta_0 + \beta_i + \Phi_{sev}^{-1}(p)\sigma$ where $i = 1, 2, \ldots\ldots, I$.

This leads to $t_p(x) = \exp[\beta_0 + \beta_i + \Phi_{sev}^{-1}(p)\sigma]$. Here $t_p(x)$ is the pth quantile of the Weibull distribution for explanatory variable values x. The parameter β_0 is the intercept and can be interpreted as the value of μ for the null model where μ is the intercept and σ is the scale parameter of the model. β_i is the effect of the i^{th} level of the factor $i = 1, 2, ..., I$. $\Phi_{sev}^{-1}(p)$ is the p th quantile of the smallest extreme value distribution.

Estimating the parameters of the model

The method of maximum likelihood (ML) is used for estimating the model parameters. In this section, the methods of Ebling (2000) are extended to estimate parameters for covariate models. Ebeling (2000) goes on to show that for an alternatively parameterized Weibull model without covariates given by $\log\left[\mathrm{t_p}(x) \right] = \beta_0 + \phi_{sev}^{-1}(p)\sigma$ where $\mu = \beta_0$ the likelihood function is given by $L(\mu, \sigma) = \prod_{\mathrm{F}} f(t_i) \prod_{\mathrm{C}} S(t_i)$ where F is the set of failures and C is the set of censored indices. Here the density function is given by $f(t_i) = \frac{\theta}{\eta}\left(\frac{t_i}{\eta}\right)^{\theta-1} e^{-\left(t_i/\eta\right)^{\theta}}$ where $\eta = \exp(\mu)$ and $\theta = 1/\sigma$. The survivor function is given by $S(t_i) = e^{-\left(t_i/\eta\right)^{\theta}}$. Taking partial derivatives of $\log L$ with respect to θ and η and setting these equal to zero results in two likelihood equations

$$\sum_{\text{all } i} t_i^{\hat{\theta}} \ln t_i \sum_{\text{all } i} \left(t_i^{\hat{\theta}}\right)^{-1} - \frac{1}{\hat{\theta}} - \sum_{i \in f} \frac{\ln t_i}{r} = 0 \qquad (1)$$

where r is the total number of failures) and

$$\hat{\eta} = \left[\sum_{\text{all } i} \frac{t_i^{\hat{\theta}}}{r} \right]^{1/\hat{\theta}} \qquad (2)$$

Equation (1) has to be solved numerically and $\hat{\theta}$ obtained. Then substituting $\hat{\theta}$ in equation (2) results in $\hat{\eta}$. This leads to $\hat{\mu} = \log \hat{\eta}$ and $\hat{\sigma} = 1/\hat{\theta}$ from the invariance property of maximum likelihood estimations.

The methods explained above are extended by the current authors to a covariate model containing one or more explanatory variables. Maximum likelihood estimates for these parameters requires solving a system of non-linear likelihood equations. Now consider the weibull distribution where the logarithm of the characteristic life is a linear function of one or more covariates and is given by

$\log\left[\mathrm{t_p}(x)\right] = \sum_{m=0}^{K} \beta_m x_{mi} + \phi_{sev}^{-1}(p)\sigma$. Here x_{mi} is the value of the mth explanatory variable X_m for the i^th^ component where i=1,...,n and m=1,….,k. Here x_{0i} =1 for all i. Now the scale parameter for the i^th^ component and shape parameter of the Weibull distribution is given by $\eta_i = \exp\left(\sum_{m=0}^{K} \beta_m x_{mi}\right)$ and $\theta = 1/\sigma$ respectively. Then the likelihood is given by $L(\beta_0,.....\beta_k,\theta)$ =

$$\prod_{i\in F} \frac{\theta}{\exp\left(\sum_{m=0}^{K}\beta_m x_{mi}\right)} \left(\frac{t_i}{\exp\left(\sum_{m=0}^{K}\beta_m x_{mi}\right)}\right)^{\theta-1} \exp\left[-\left(\frac{t_i}{\exp\left(\sum_{m=0}^{K}\beta_m x_{mi}\right)}\right)^{\theta}\right] \prod_{i\in C} \exp\left[-\left(\frac{t_i}{\exp\left(\sum_{m=0}^{K}\beta_m x_{mi}\right)}\right)^{\theta}\right] \quad (3)$$

By taking the log-likelihood and differentiating it with respect to $\beta_m; m = 0,.......,K$ and θ and equating these to zero, the likelihood equations are obtained. These equations have to be solved numerically to obtain the Maximum likelihood estimates of the parameters of obtaining $\hat{\theta}$, $\hat{\sigma} = 1/\hat{\theta}$ can be determined as before. Both continuous and categorical covariates can be incorporated in the model using this same principle.

Comparison of models

Model 1 - Null model is given by $\log[t_p(x)] = \beta_0 + \Phi_{sev}^{-1}(p)\sigma$. Here there is no impact of the factor on the percentile t_p. Model 2- Model with a single factor with I levels each level having different intercepts but same scale parameter is given by $\log[t_p(x)] = \beta_0 + \beta_i + \Phi_{sev}^{-1}(p)\sigma$ where $i = 1, 2, ..., I$. Model 3- I separate models for each level of the factor where each level has a different intercept and a different scale parameter is given by $\log[t_p(x)] = \mu_i + \Phi_{sev}^{-1}(p)\sigma_i$ where $i = 1, 2, ..., I$ .Here I separate lines are fitted to each combination. The likelihood ratio chi-square test as explained in Cox and Hinkley (1978) is used for testing the goodness of fit of the models and selecting the most appropriate model.

The residual analysis is carried out using Cox-Snell residuals. These residuals have been discussed in Meeker and Escobar (1998) and Collett (1993). For testing the Weibull assumption standardized Cox-Snell residuals are plotted on a

Weibull probability plot. If the plot is linear through the origin with a slope=1, the fitted model is adequate.

Setting warranty periods using quantiles of the Weibull model

The Weibull pth quantile can be calculated through the general way of location-scale based distribution using $t_p(x) = \exp[\mu_i + \Phi_{sev}^{-1}(p)\sigma_i]$. This equation gives the warranty period $t_p(x)$ for a proportion p of failures which the company is willing to tolerate for values x of the explanatory variables. Location parameter μ_i and scale parameter σ_i are estimated from smallest extreme value location-scale based distribution and those parameters are used to calculate quantiles. A model with a constant scale parameter and varying location parameter suggest different characteristics with respect to mean value of the model. Compared to other distributions the Weibull distribution is more realistic and addresses the different aspects of failure times. Computer failures can be reflected in numerous ways. Therefore using a more realistic and appropriate model such as Weibull is essential.

Calculation of probabilities of failures associated with the warranty period

To asses the worthiness of a given reliability measure, it is required to calculate the probability of failure, given the failure time of the component. Since the location and scale parameters are known, it is possible to calculate probability for a given failure time, by using $p = \Phi_{sev}\{\frac{\log[t_p(x)] - \mu_i}{\sigma_i}\}$. This equation gives the probability of failure associated with a warranty period of $t_p(x)$ where x gives the values of the explanatory variables.

RESULTS AND DISCUSSION

In this section the initial requirement is to identify the distributional form which the data follows. For this purpose Normal, Log-Normal and Weibull probability plots are plotted with the simultaneous confidence bands. Linearity of plots suggest adequacy of the assumed distributional model and points should be well inside the confidence bands. A straight line is fitted to each plot to ensure the validity of the model. Figure 1 gives the probability plots for the four combinations for the Weibull distribution. Plots (a), (b), (c), and (d) correspond to the four combinations. Similar plots were drawn for the normal and log-normal distributions. However these plots are not presented in this paper.

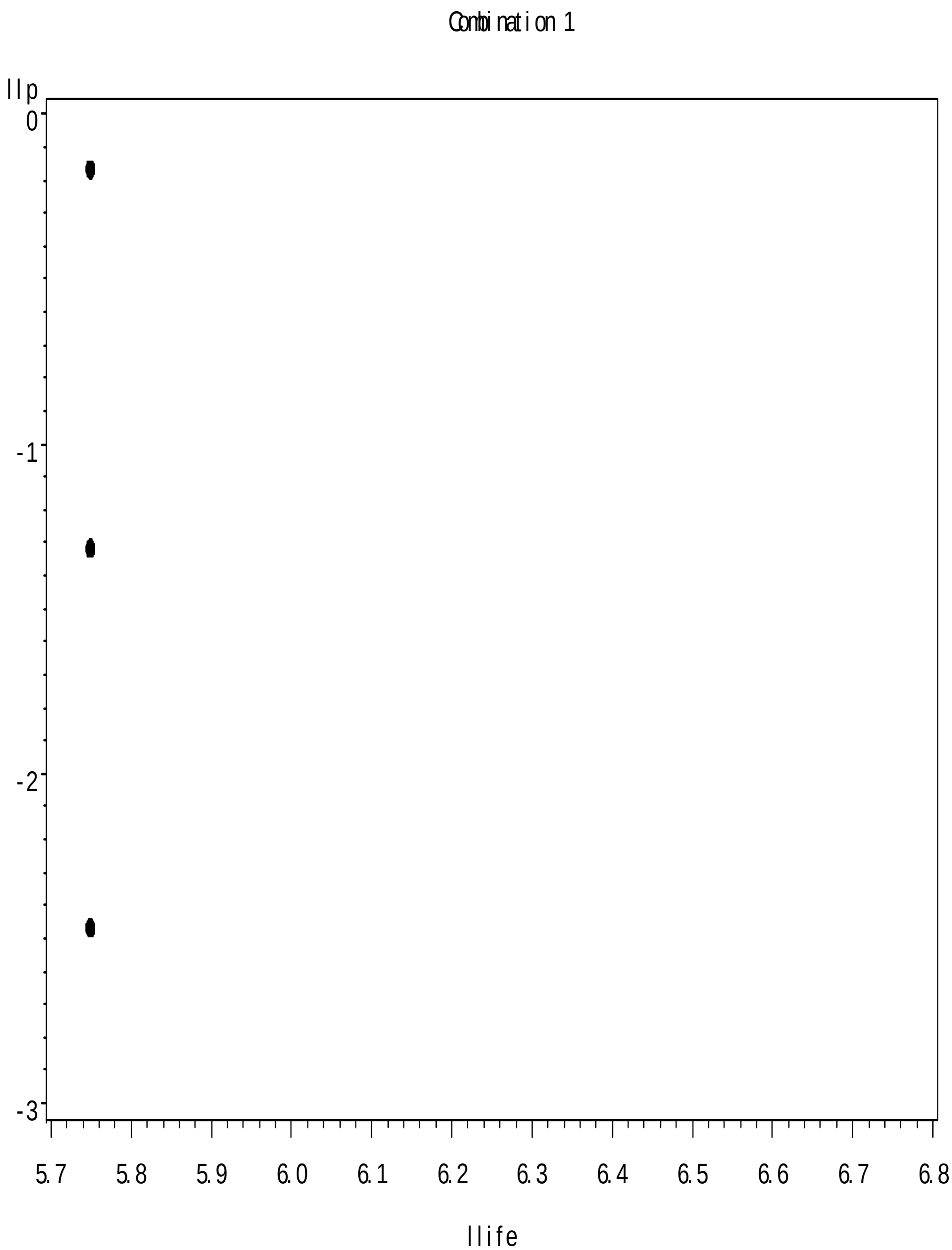

Figure 1(a): Weibull Probability Plots for the four combinations
(Note: llp = *log (-log (1- cumulative probability)) and* llife = *log life*)

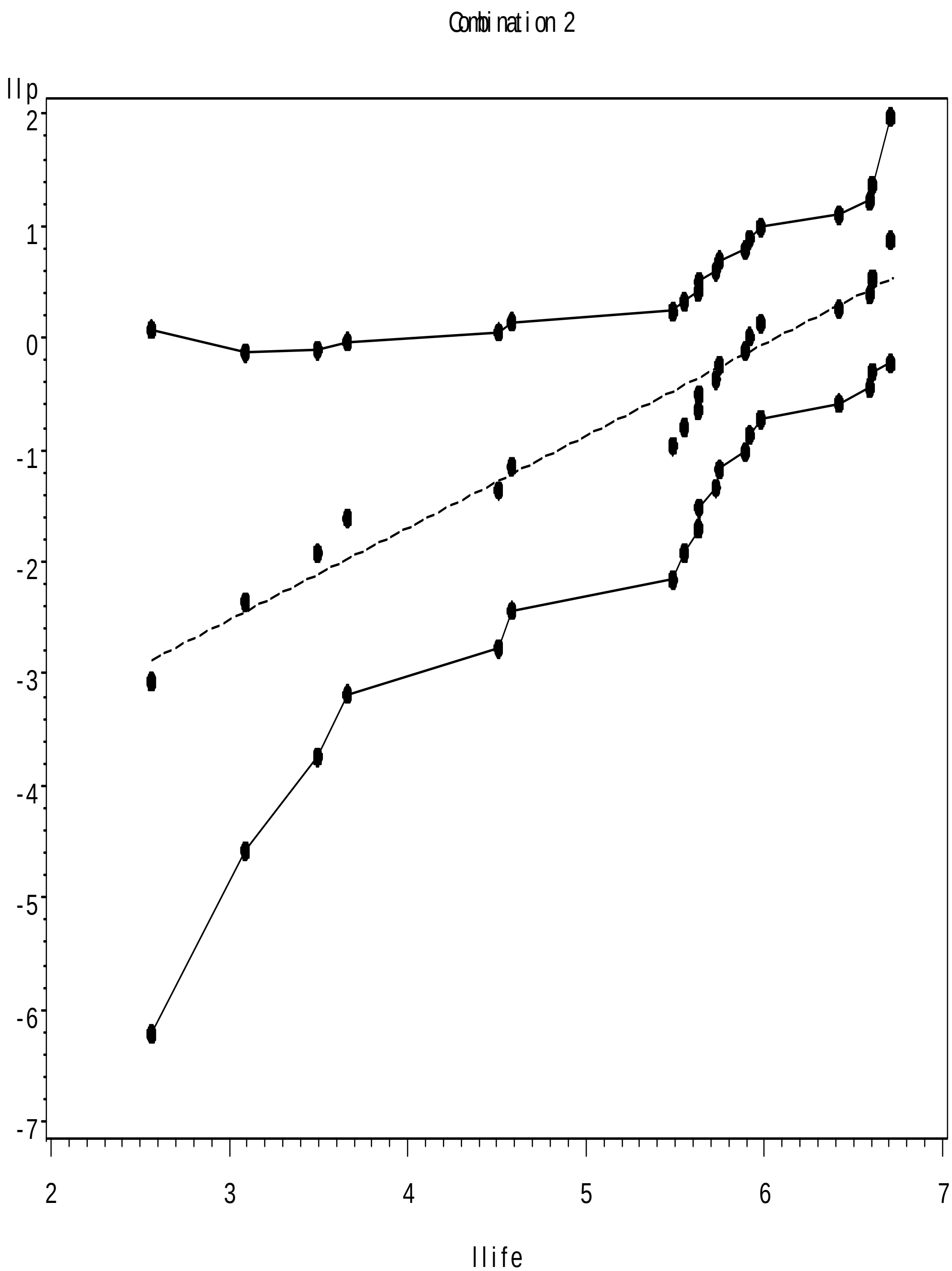

Figure 1(b): Weibull Probability Plots for the four combinations

(Note: llp = *log (-log (1- cumulative probability)) and* llife = *log life*)

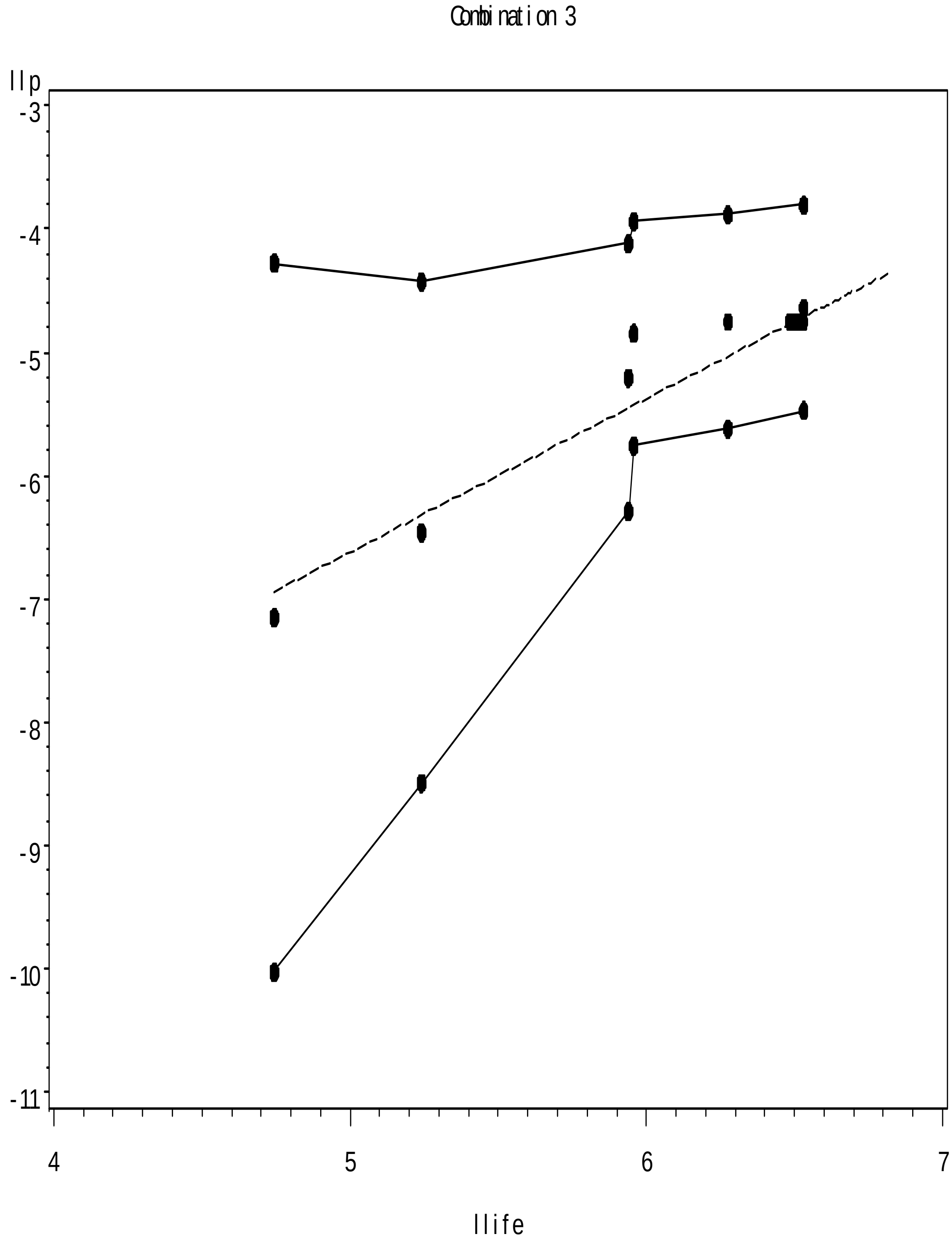

Figure 1(c): Weibull Probability Plots for the four combinations
(Note: llp = *log (-log (1- cumulative probability)) and* llife = *log life*)

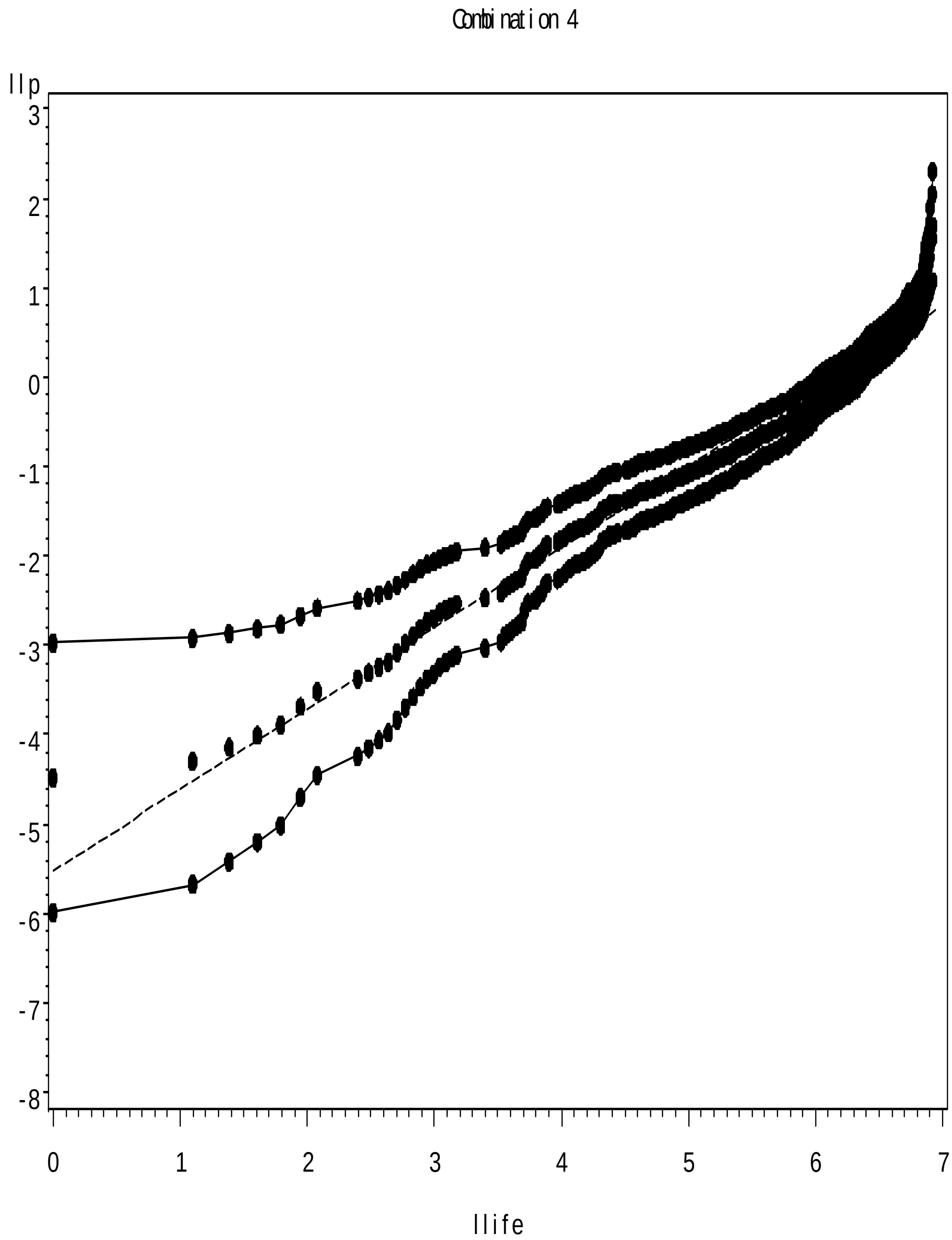

Figure 1(d): Weibull Probability Plots for the four combinations
(Note: llp = *log (-log (1- cumulative probability)) and* llife = *log life*)

For combinations 2, 3 and 4 indicated by plots (a), (b) and (c), inspection showed that the points are closest to a straight line for the Weibull data. Except for a very slight departure at the upper extreme end the fitted line lies well within the simultaneous confidence bands for the Weibull data while for the other distributions the fitted straight lines lie outside the simultaneous confidence bands in most cases. These results indicate that the Weibull model best fits the data for combination 2, 3 and 4. Since there is only one plotting position for combination 1, it is unable to suggest any distribution and assumption of Weibull distribution is fair enough with the overall characteristics of data.

Model comparison is carried out using the likelihood ratio test. The values of -2logL obtained by fitting models 1, 2, 3 are 3076.366, 1628.779 and 1628.255 respectively. The number of parameters corresponding to each of these models are 2,5 and 8 respectively. The value of $-2\log(l)$ for four separate models for the four combinations (model 3) has been calculated as the sum of $-2\log(l)$ for each model. For four separate models (model 3)

$$\sum -2\log(l) = 25.332 + 68.047 + 153.755 + 1381.121 = 1628.255$$

To check whether the combination factor term is significant model 1 is compared with model 2 using the likelihood ratio test. Likelihood ratio (LR) test statistic= $-2\left[\log(l_1) - \log(l_2)\right] = 3076.366 - 1628.779 = 1447.587$ on (5-2) = 3 degree of freedom. As $1447.587 >> \chi^2_{(3),5\%} = 7.82$ the LR test statistic is highly significant suggesting that at 5 % significance level fitting the null model which takes whole data set as one unit is significantly worse than fitting a model with combination as a factor. Thus the combination factor effects failure time. Then to determine whether the scale parameter can be considered as a constant, model 2 is compared with model 3 using the likelihood ratio test as,

$-2\left[\log(l_2) - \log(l_3)\right] = 1628.779 - 1628.054 = 0.524$ on 8-5= 3 degree of freedom.

As $0.524 << \chi^2_{(3),5\%} = 7.82$ this suggest that at 5% significant level fitting one model with combination as a factor is not significantly worse than fitting four separate models. Thus the scale parameter for all four combinations can be considered as a constant. Finally the best fitted model is the model with combination as a factor and common scale parameter given by $\log[t_p(x)] = \beta_0 + \beta_i + \Phi^{-1}_{sev}(p)\sigma \qquad i = 1,2,3$.

For the purpose of diagnostic checking standardized Cox-Snell residuals are plotted in a Weibull probability plot. Here $\log\left[-\log(1-p)\right]$ *versus* $\log(t_p)$ will plot as a straight line if the distribution is Weibull[6]. Here t_p 's are the ordered residuals. Figure 2 gives the Weibull probability plot for the Cox-Snell residuals.

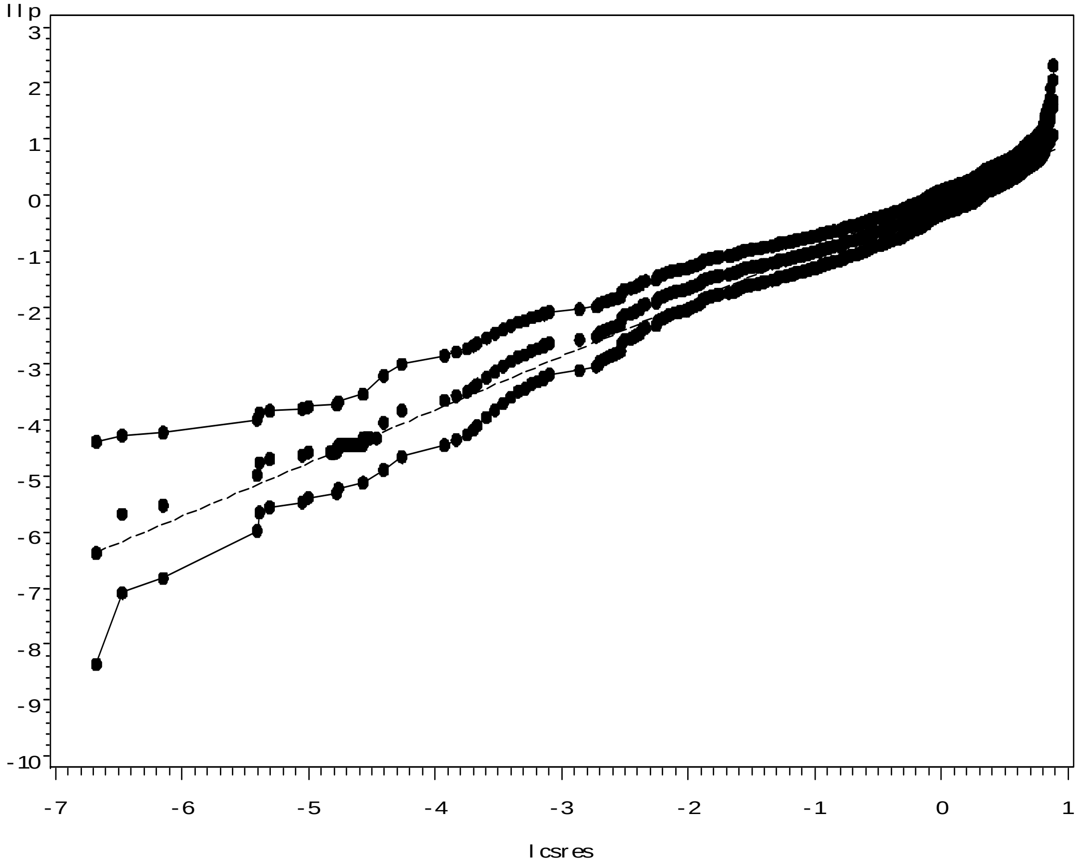

Figure 2: Cox-Snell Weibull probability plot of residuals
(Note: llp = *log (-log (1- cumulative probability))*)

According to Figure 2 it can be clearly observed that the plot is linear and within the simultaneous confidence bands. The figure indicates that the straight line goes through the $(-7,-7)$ coordinates which is the origin and continues to follow a slope=1 straight line. Therefore it suggests the adequacy of the model. These results indicate that the Weibull model with combination as a factor and common scale parameter can model the failure time data well. Therefore this model can be used to estimate quantiles and thereby determine suitable warranty periods for each combination of computers.

For the best model which is the model with combination as a factor, the parametric estimates were β_i ; i=0,1,2,3 are 6.09(0.048), 1.776(0.477), -0.045(0.216) and 4.936(0.334) respectively. The numbers in brackets are the respective standard errors. The constraint used by SAS PROC Lifereg is $\beta_4 = 0$. The parameter estimate (standard error) of σ is 0.941(0.038). Here combination 4 is taken as the base combination and the other combinations are described compared to combination 4. Here it can be seen that the log of percentile of combination 4 are given by $\log[t_p(x)] = 6.09 + 0.941 \times \Phi_{sev}^{-1}(p)$. $\beta_1 = 1.766$ (>0) suggests that the p^{th} quantile of combination 1 is greater than the p^{th} percentile of combination 4 for all $0 < p < 1$. The p^{th} quantile of combination2 (<0) is less than that of combination 4 and most importantly combination 3 gives the highest percentiles compared to other combinations. Also $\beta_3 > \beta_1 > \beta_4 > \beta_2$ indicates a fair guideline of the reliability of the computers. It is difficult to notice a significant difference between combination 2 and 4 with respect to reliability. The parameter estimates indicate that combination 3 is the best and combination 2 is the worst with respect to reliability.

Percentiles of different combinations of computers were calculated using $t_p(x) = \exp\left[\beta_0 + \beta_i + \Phi_{sev}^{-1}(p)\sigma\right]$. These percentiles illustrate the difference between the percentiles of life times of different combinations. Here high reliability of combination 3 was evident and it indicated a clear difference from the other combinations. Combination 1 comes as the second best. Assuming that the sales company will be willing to accommodate a 5% failure rate, the following warranty periods can be suggested for each combination of computers. A 6-month warranty for combination 1, a 1 month warranty for combination 2 and combination 4 and a 10 year warranty for combination 3. For 10,000 computers only 485 failures were recorded for combination 3 and this indicates the high reliability of this combination compared to others. This outcome suggests that combination 3 is very advanced. This may be due to the new technology, research and development process and high level of care in the design phase of this combination. Since the normal industry standard for a computer warranty is a three year period, quantification of the percentage of failures in three years would be very helpful for the top management of the company in the process of decision making. Thus the percentage of failures within three years for each combination was calculated. Of the four combinations, only combination 1 and 3 gives warranty. Therefore it is meaningful only to interpret these two combinations with respect to the industry norm. The calculations clearly showed that the failure probability of combination 3 is very low compared to the combination 1 for a three year warranty period. Results suggest that only 140 failures will occur in combination 3 out of 10000 computers while 3280 failures will occurred in combination 1. This clearly implies the reliability of combination 3 computers. According to the results obtained we can recommend extending the warranty period for combination 3 and reducing the warranty period for combination 1.

CONCLUSION

The results indicate that the computers which have no warranty tend to fail more frequently compared to computers which have warranty. This may be due to the lack of care that is given to the computers which have no warranty periods and lack of continuous monitoring and maintenance that warranty computers experience.

It is seen that the Weibull model best fits the data for combination 2, 3 and 4. Since there is only one plotting position for combination 1 it is unable to suggest any distribution and assumption of Weibull distribution is fair enough with the overall characteristics of data. The brand and warranty combination is an important factor of the defined model. Finally $\log[t_p(x)] = \beta_0 + \beta_i + \Phi_{sev}^{-1}(p)\sigma$ where β_i is the effect of the i^{th} combination β_0 is the intercept of the null model and σ is the common scale parameter was selected as the best model. Although overall results suggest the Weibull model as the best, a slight departure at the upper extreme for combination 4 can be observed. The EP method of simultaneous confidence bands clearly captures that departure. This study focuses on determining methods of calculating warranty periods for different combinations of personal computers and suggests optimized warranty periods that will benefit the business organization. In the decision of purchasing a computer warranty period plays a major role. Therefore great level of care is given to the reliability of computers in the design and manufacturing process. Reliability of a computer is very much dependent on the technology, brand, and ability to adapt to various environments etc. Therefore decision of warranty should be taken with a great level of consideration about the above mentioned factors. This paper concentrates on the time to the first failure of computers, and it has therefore eliminated repetitive repairs. It can be concluded that higher the level of technology and care given higher the reliability of computers.

In the process of deciding on warranty periods, the fitted model and general idea of market concern is essential. Since industrial norm for computer warranty is three years, combination 3 can be further analyzed to check whether the warranty period can be extended. By comparing the economic benefit and the loss with respect to the decision of extending the warranty period it will be possible to determine how best to implement this idea. According to this study we can confidently suggest to increase the warranty period for combination 3.

Overall, it was possible to conclude that the Weibull model with combination as a factor and common scale parameter can provide a substantial solution to the problem of warranty period determination for the different combinations. The method of maximum likelihood (ML) is used for estimating the model parameters. Ebeling (2000) has described methods for estimating model parameters in the absence of covariates. In this paper these methods have been extended by the current authors to a covariate model containing one or more explanatory variables.

REFERENCES

Collet, D. (1993). Modeling Survival data in Medical Research. Chapman and Hall, London.

Cox, D.R. and Hinkley, D.V. (1974). Theoretical Statistics. Chapman and Hall.

Ebling, C.E. (2000). An introduction to reliability and maintainability Engineering. Tata McGraw Hill.

Esterman, M., Gerst, P., Stiebitz, P.H., Ishii, K.(2005). A framework for warranty prediction during product. *Proceedings of IMECE2005.*

Gan, F., Koehler, K., Thompson, J. (1991), Probability Plots and Distribution Curves for Assessing the Fit of Probability Models. *The American Statistician.* 45(1) 14-21

Jung, M., Bai, D.S. (2007) 'Analysis of field data under two-dimensional warranty' *Reliability Engineering & System Safety*. 92(2). 135-143

Kalbfleisch, J.D., Lawless, J.F., Robinson, J.A. (1991) 'Methods for the analysis of warranty claims' *Technometrics* 33.(3) 273-285

Marquez, D. , Neil, M. Fenton, N. (2007) 'A new Bayesian Network approach to Reliability modelling' http://www.agena.co.uk/resources/technology_articles/

Meeker, W.Q. and Escobar, L.A. (1998), Statistical Methods for Reliability Data. John Wiley and Sons.

O'Connor, P.D.T.(2002). Practical Reliability Engineering. John Wiley and Sons.

Whitehead, J. (1997). The Design and Analysis of Sequential Clinical Trials, Revised, 2nd edition. John Wiley and sons.