

AD-A148 377

A POWER TRANSFORMATION EXPONENTIAL REGRESSION MODEL FOR CENSORED FAILURE. (U) NORTH CAROLINA UNIV AT CHARLOTTE DEPT OF MATHEMATICS D YOUNG ET AL. NOV 84

1/1

UNCLASSIFIED

AFOSR-TR-84-1093 AFOSR-88-0245

F/G 12/1

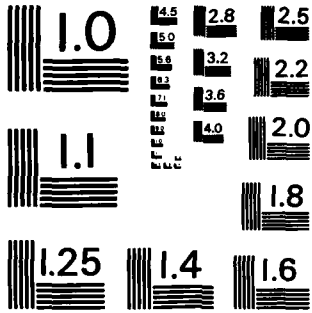
NL



END

FILMED

DTIC




MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED		1b. RESTRICTIVE MARKINGS														
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.														
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE		4. PERFORMING ORGANIZATION REPORT NUMBER(S) TR No. 22														
4. PERFORMING ORGANIZATION REPORT NUMBER(S) TR No. 22		5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR-84-1093														
6a. NAME OF PERFORMING ORGANIZATION University of North Carolina	6b. OFFICE SYMBOL <i>(If applicable)</i>	7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research														
6c. ADDRESS (City, State and ZIP Code) Department of Mathematics Charlotte NC 28223		7b. ADDRESS (City, State and ZIP Code) Directorate of Mathematical & Information Sciences, Bolling AFB DC 20332-6448														
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR	8b. OFFICE SYMBOL <i>(If applicable)</i> NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-80-0245														
8c. ADDRESS (City, State and ZIP Code) Bolling AFB DC 20332-6448		10. SOURCE OF FUNDING NOS. <table border="1" style="width: 100%; border-collapse: collapse; margin-top: 5px;"> <tr> <th style="width: 25%;">PROGRAM ELEMENT NO.</th> <th style="width: 25%;">PROJECT NO.</th> <th style="width: 25%;">TASK NO.</th> <th style="width: 25%;">WORK UNIT NO.</th> </tr> <tr> <td>61102F</td> <td>2304</td> <td>A5</td> <td></td> </tr> </table>			PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.	61102F	2304	A5					
PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.													
61102F	2304	A5														
11. TITLE (Include Security Classification) A POWER TRANSFORMATION EXPONENTIAL REGRESSION MODEL FOR CENSORED FAILURE TIME DATA																
12. PERSONAL AUTHOR(S) David Young* and Mohamed Abdel-Hameed																
13a. TYPE OF REPORT Technical	13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Yr., Mo., Day) NOV 84	15. PAGE COUNT 13													
16. SUPPLEMENTARY NOTATION *Department of Mathematics and Statistics, Brunel University, England.																
17. COSATI CODES <table border="1" style="width: 100%; border-collapse: collapse; margin-top: 5px;"> <thead> <tr> <th style="width: 33%;">FIELD</th> <th style="width: 33%;">GROUP</th> <th style="width: 33%;">SUB. GR.</th> </tr> </thead> <tbody> <tr><td> </td><td> </td><td> </td></tr> <tr><td> </td><td> </td><td> </td></tr> <tr><td> </td><td> </td><td> </td></tr> </tbody> </table>			FIELD	GROUP	SUB. GR.										18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Power transformation model; censored survival data; exponential distribution.	
FIELD	GROUP	SUB. GR.														
19. ABSTRACT (Continue on reverse if necessary and identify by block number) A power transformation regression model is considered for exponentially distributed time to failure data with right censoring, procedures for estimation of parameters by maximum likelihood and assessment of goodness of model fit are described and illustrated.																
<div style="display: flex; justify-content: space-between;"> <div style="writing-mode: vertical-rl; transform: rotate(180deg); font-weight: bold; font-size: 1.5em;">DTIC FILE COPY</div> <div style="text-align: right;">  </div> </div>																
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS <input type="checkbox"/>		21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED														
22a. NAME OF RESPONSIBLE INDIVIDUAL MAJ Brian W. Woodruff		22b. TELEPHONE NUMBER <i>(Include Area Code)</i> (202) 767-5027	22c. OFFICE SYMBOL NM													

AFOSR-TR. 84 - 1093

A142 377

A POWER TRANSFORMATION EXPONENTIAL REGRESSION MODEL
FOR CENSORED FAILURE TIME DATA

by

David Young
Department of Mathematics and Statistics
Brunel University, England

¹
Mohamed Abdel Hameed
Department of Mathematics
University of North Carolina

U.N.C.C. TECHNICAL REPORT NO. 22

November 6, 1984

Approved for public release;
distribution unlimited.

The research of this author is supported by the United States Air Force
Grant Number AFOSR - 80 - 0245.

84 10 93 247

A POWER TRANSFORMATION EXPONENTIAL REGRESSION MODEL
FOR CENSORED FAILURE TIME DATA



by
David Young
Mohamed Abdel Hameed

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

ABSTRACT

A power transformation regression model is considered for exponentially distributed time to failure data with right censoring, procedures for estimation of parameters by maximum likelihood and assessment of goodness of model fit are described and illustrated.

Key Words: Power transformation model, censored survival data, exponential distribution.

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
NOTICE OF TRANSMISSION
This technical report is approved for distribution
MATTHEW J. KENNEDY
Chief, Technical Information Division

1. INTRODUCTION

Suppose that we have n individuals or components which are subject to 'failure' and that the response variable of interest is time to failure. Let T₁*, T₂*, ..., T_n* be independent random variables representing the times to failure of the individuals. We suppose that right censoring may occur because of the need for early termination of the investigation and let T₁, T₂, ..., T_n represent the recorded survival times. Defining censoring indicator variables

$$\omega_i = \begin{cases} 1 & \text{if } T_i^* \text{ is uncensored} \\ 0 & \text{if } T_i^* \text{ is censored} \end{cases} \tag{1.1}$$

we have

$$T_i^* = T_i \text{ if } \omega_i = 1, \quad T_i^* > T_i \text{ if } \omega_i = 0 \tag{1.2}$$

We let

$$n_u = \sum_{i=1}^n \omega_i, \quad n_c = \sum_{i=1}^n (1 - \omega_i) \tag{1.3}$$

denote the numbers of uncensored and censored observations, respectively.

We now suppose that measurements are available on k explanatory variables x₁, x₂, ..., x_k. Setting x' = (x₁, x₂, ..., x_k) the p.d.f. and survival function of T* given x will be denoted by f(t;x) and S(t;x), respectively. If the hazard function f(t;x)/S(t;x) is assumed to be a constant independent of t for any given x, then T* has the exponential distribution with p.d.f.

$$f(t;x) = \begin{cases} \mu_x^{-1} \exp(-t/\mu_x), & t > 0 \\ 0 & \text{otherwise} \end{cases} \tag{1.4}$$

where μ_x denotes the expected value of T* given x.

Various models have been proposed in the literature to represent the dependence of μ_x on x. Fiegl and Zelen (1965) consider the model form

$$\mu_x = \lambda(1 + x'\beta) \tag{1.5}$$

while Greenberg et.al. (1974) use the form

$$\mu_{\underline{x}} = \lambda / (1 + \underline{x}'\underline{\beta}) \quad (1.6)$$

where $\underline{\beta}' = (\beta_1, \beta_2, \dots, \beta_k)$ and λ is a positive constant. Both models require that the condition $\underline{x}'\underline{\beta} < 1$ must be imposed to ensure that $\mu_{\underline{x}} > 0$. An alternative model which does not require a constraint to be imposed on $\underline{x}'\underline{\beta}$ is

$$\mu_{\underline{x}} = \lambda \exp(\underline{x}'\underline{\beta}) \quad (1.7)$$

This model arises for the exponential case from the well-known family of proportional hazard regression models (see e.g. Kay (1977)) in which an assumed underlying hazard function is adjusted by multiplicative exponential factors to allow for the effect of the explanatory variables. Prentice (1973) also discusses the use of censored regression models for the exponential case.

In this paper, we consider the power transformation model given by

$$\mu_{\underline{x}} = \lambda (1 + \delta \underline{x}'\underline{\beta})^{1/\delta} \quad (1.8)$$

We shall refer to δ as the power parameter. It is seen that when $\delta = 1$, the model corresponds to the usual additive regression model for the mean given by (1.5), while if $\delta = -1$ the reciprocal model (1.6) is obtained after appropriate reparameterisation. When $\delta \rightarrow 0$, the model given by (1.7) is obtained. For later work, it is convenient to write the power transformation model in the form

$$\mu_{\underline{x}} = (1 + \delta \underline{x}'_{\star} \underline{\beta}_{\star})^{1/\delta} \quad (1.9)$$

where $\underline{x}'_{\star} = (1, \underline{x}')$, $\underline{\beta}'_{\star} = (\beta_{\star 0}, \beta_{\star 1}, \dots, \beta_{\star k})$ and

$$\beta_{\star 0} = (\lambda^{\delta} - 1) / \delta, \quad \beta_{\star j} = \lambda^{\delta} \beta_j, \quad j=1, \dots, k \quad (1.10)$$

In general, the power parameter δ as well as the coefficient vector $\underline{\beta}_{\star}$ will have to be estimated from the data. In section 2, results are given for maximum likelihood estimation and it is shown how the estimates can be obtained numerically using the statistical package GLIM. The problem of assessment of the goodness of fit of the model is discussed briefly in section 3 and in section 4 a numerical example is used to illustrate the procedure.

2. MAXIMUM LIKELIHOOD ESTIMATION

We shall assume that the stochastic model underlying the censoring mechanism is unknown. The likelihood given the data $(t_i, \omega_i, \underline{x}_i)$, where $\underline{x}_i' = (x_{i1}, x_{i2}, \dots, x_{ik})$, is

$$L = \prod_{i=1}^n \left\{ f(t_i; \underline{x}_i) \right\}^{\omega_i} \left\{ S(t_i; \underline{x}_i) \right\}^{1-\omega_i} \\ = \prod_{i=1}^n \left\{ \mu_{\underline{x}_i}^{-\omega_i} \exp(-t_i/\mu_{\underline{x}_i}) \right\} \quad (2.1)$$

where $\mu_{\underline{x}_i} = (1 + \delta \underline{x}_i' \underline{\beta})^{1/\delta}$ and $\underline{x}_i' = (1, \underline{x}_i)$. If we put

$$\mu_i = t_i \mu_{\underline{x}_i}^{-1}, \quad i=1, 2, \dots, n \quad (2.2)$$

the log likelihood is

$$\log L = \sum_{i=1}^n (\omega_i \log \mu_i - \mu_i) - \sum_{i=1}^n \omega_i \log t_i \quad (2.3)$$

The likelihood equations are obtained by setting

$$\sum_{i=1}^n \frac{(\omega_i - \mu_i)}{\mu_i} \frac{\partial \mu_i}{\partial \delta} = 0, \quad \sum_{i=1}^n \frac{(\omega_i - \mu_i)}{\mu_i} \frac{\partial \mu_i}{\partial \beta_{*j}} = 0, \quad j=0, 1, \dots, k.$$

Since

$$\frac{1}{\mu_i} \frac{\partial \mu_i}{\partial \beta_{*j}} = -x_{ij} \left(\frac{\mu_i}{t_i} \right)^\delta, \quad \frac{1}{\mu_i} \frac{\partial \mu_i}{\partial \delta} = \frac{1}{\delta^2} \left\{ \left(\frac{\mu_i}{t_i} \right)^\delta - 1 - \delta \log \left(\frac{\mu_i}{t_i} \right) \right\} \quad (2.4)$$

the maximum likelihood estimates are found as the solution of the $k+2$ equations

$$\sum_{i=1}^n x_{ij} (\omega_i - \hat{\mu}_i) \left(\frac{\hat{\mu}_i}{t_i} \right)^\delta = 0, \quad j=0, 1, \dots, k \quad (2.5)$$

$$\sum_{i=1}^n (\omega_i - \hat{\mu}_i) \left\{ \left(\frac{\hat{\mu}_i}{t_i} \right)^\delta - 1 - \delta \log \left(\frac{\hat{\mu}_i}{t_i} \right) \right\} = 0 \quad (2.6)$$

where $\hat{\mu}_i = t_i (1 + \hat{\delta} x_{*i}' \hat{\beta}_*)^{-1/\hat{\delta}}$

Routine calculation gives the second order derivatives of the log-likelihood as

$$\frac{\partial^2 \log L}{\partial \beta_{*j} \partial \beta_{*j'}} = - \sum_{i=1}^n x_{ij} x_{ij'} \theta_i^2 \left\{ (\delta+1) \mu_i - \delta \omega_i \right\} = - a_{jj'} \quad (2.7)$$

$$\frac{\partial^2 \log L}{\partial \beta_{*j} \partial \delta} = - \frac{1}{\delta} \sum_{i=1}^n x_{ij} \theta_i \left\{ \omega_i^{-\mu_i} (\theta_i^{-1}) - \mu_i z_i \right\} = - b_j \quad (2.8)$$

$$\frac{\partial^2 \log L}{\partial \delta^2} = - \frac{1}{\delta^2} \sum_{i=1}^n \left[\mu_i z_i^2 + (\omega_i^{-\mu_i}) \left\{ z_i (3 - \theta_i) - \frac{(\theta_i^{-1})}{\delta} \log \theta_i \right\} \right] = -c \quad (2.9)$$

for $j, j' = 0, 1, \dots, k$, where

$$\theta_i = \left(\frac{\mu_i}{t_i} \right)^\delta, \quad z_i = \frac{1}{\delta} (\theta_i^{-1} - \log \theta_i) \quad (2.10)$$

Setting

$$\underline{A} = ((\hat{a}_{jj'})) , \quad \underline{B} = ((\hat{b}_j)) , \quad \underline{C} = ((\hat{c}))$$

where $\hat{a}_{jj'}$ denotes $a_{jj'}$ evaluated at $\beta_* = \hat{\beta}_*$ and $\delta = \hat{\delta}$, etc, the estimated asymptotic covariance matrix of the estimators is

$$\underline{\tilde{V}} = \begin{bmatrix} \underline{\tilde{V}}_{11} & \underline{\tilde{V}}_{12} \\ \underline{\tilde{V}}_{12}' & \underline{\tilde{V}}_{22} \end{bmatrix} = \begin{bmatrix} \underline{A} & \underline{B} \\ \underline{B}' & \underline{C} \end{bmatrix}^{-1}$$

The matrices in the partitioned form are given by

$$\underline{\tilde{V}}_{11} = \underline{A}^{-1} + \underline{A}^{-1} \underline{B} (\underline{C} - \underline{B}' \underline{A}^{-1} \underline{B})^{-1} \underline{B}' \underline{A}^{-1} \quad (2.11)$$

$$\underline{\tilde{V}}_{12} = -\underline{A}^{-1} \underline{B} (\underline{C} - \underline{B}' \underline{A}^{-1} \underline{B})^{-1}, \quad \underline{\tilde{V}}_{22} = (\underline{C} - \underline{B}' \underline{A}^{-1} \underline{B})^{-1}$$

The calculation of the estimates and statistics used for making inferences about δ and β_* is performed straightforwardly using the statistical package GLIM which has been developed for fitting generalised linear models by maximum likelihood for error distributions in the exponential family. The method is similar to that described by Aitken and Clayton (1980) for fitting proportional hazard regression models for censored survival data.

In (2.3), the first term represents the kernel of a log-likelihood treating W_1, W_2, \dots, W_n as independent Poisson variables with $E(W_i) = \mu_i$, and the second term is independent of β_* and δ . From (2.2) we have

$$\{(t_i/\mu_i)^\delta - 1\}/\delta = \mathbf{x}'_{*i} \beta_* \quad (2.12)$$

If δ is known, the value $\hat{\beta}_*(\delta)$ which maximises the log-likelihood, now denoted by $\log L(\hat{\beta}_*(\delta), \delta)$, can be found fitting a Poisson regression model with the user defined link function $g_\delta(\mu_i) = \{(t_i/\mu_i)^\delta - 1\}/\delta$. In the usual case when δ is unknown, the maximum likelihood estimates $\hat{\delta}$ and $\hat{\beta}_*$ can be found using the following iterative procedure which is based on a general method suggested by Pregibon (1980) for selecting the link function in generalised linear models.

Suppose that $\hat{\delta}_1$ denotes an initial approximation to $\hat{\delta}$. Using the first order Taylor series expansion about the value $\delta = \hat{\delta}_1$, gives the approximation

$$g_\delta(\mu_i) \approx \frac{(t_i/\mu_i)^{\hat{\delta}_1} - 1}{\hat{\delta}_1} + (\delta - \hat{\delta}_1) \left\{ \frac{\hat{\delta}_1 (t_i/\mu_i)^{\hat{\delta}_1} \log(t_i/\mu_i) + 1 - (t_i/\mu_i)^{\hat{\delta}_1}}{\hat{\delta}_1^2} \right\} \quad (2.13)$$

An initial fit is made using the model $g_{\hat{\delta}_1}^\wedge(\mu_i) = \mathbf{x}'_{*i} \beta_*$ to yield the estimate $\hat{\beta}_*^{(1)}$ and fitted values $\hat{\mu}_i^{(1)}$ from which we form

$$\hat{z}_i^{(1)} = \left\{ \hat{\delta}_1 \left(t_i / \hat{\mu}_i^{(1)} \right)^{\hat{\delta}_1} \log \left(t_i / \hat{\mu}_i^{(1)} \right) + 1 - \left(t_i / \hat{\mu}_i^{(1)} \right)^{\hat{\delta}_1} \right\} / \hat{\delta}_1^2 \quad (2.14)$$

A fit of the model $g_{\hat{\delta}_1}^\wedge(\mu_i) = \mathbf{x}'_{*i} \beta_* + \hat{z}_i^{(1)} (\delta - \hat{\delta}_1)$ then gives the second approximation $\hat{\delta}_2$, where $\hat{z}_i^{(1)}$, is treated as an additional explanatory variable. The procedure is then repeated until satisfactory convergence is obtained.

If a confidence interval for δ is required, we use the approximate distribution result

$$2 \log L\{\hat{\beta}_* (\hat{\delta}), \hat{\delta}\} - 2 \log L\{\hat{\beta}_* (\delta), \delta\} \stackrel{d}{\sim} \chi_1^2$$

giving

$$P \left[\log L\{\hat{\beta}_* (\delta), \delta\} > \log L\{\hat{\beta}_* (\hat{\delta}), \hat{\delta}\} - \frac{1}{2} \chi_1^2 (1-\alpha) \right] \approx 1-\alpha \quad (2.15)$$

where $\chi_1^2 (1-\alpha)$ denotes the upper 100 α % point of the χ_1^2 distribution. Thus if δ_L and δ_U are the two values of δ for which

$$\log L\{\hat{\beta}_* (\delta), \delta\} = \log L\{\hat{\beta}_* (\hat{\delta}), \hat{\delta}\} - \frac{1}{2} \chi_1^2 (1-\alpha)$$

where $\delta_L < \delta_U$, then the interval (δ_L, δ_U) provides an approximate 100(1- α)% confidence interval for δ . The calculation of this interval will require values of $\log L\{\hat{\beta}_* (\delta), \delta\}$ to be determined for a grid of values of δ .

The covariance matrices for $\hat{\beta}_*$ and $\hat{\delta}$ can also be found straightforwardly using (2.11) since the GLIM output will provide the value of \underline{A}^{-1} , the covariance matrix for $\hat{\beta}_*$ when $\hat{\delta}$ is treated as the true fixed value of δ . Finally, we may wish to test the hypothesis that a specified subset of the explanatory variables have no effect on the expected time to failure. This can be done using a standard likelihood ratio test based on the difference between the maximised log-likelihoods under the full and reduced models, this being the difference between the corresponding deviances in the GLIM output.

3. RESIDUAL PLOTS

A graphical procedure based on suitably defined observed residuals provides a useful preliminary assessment of adequacy of fit of the power transformation model. If the model is correct, the random variable $R = T^*(1 + \delta \underline{x}' \underline{\beta}_*)^{-1/\delta}$ has the standard exponential distribution. Thus if there is no censoring, we define observed residuals by

$$r_i = t_i (1 + \delta \underline{x}'_i \hat{\underline{\beta}}_*)^{-1/\delta}, \quad i=1, 2, \dots, n \quad (3.1)$$

If the model is correct, the $\{r_i\}$ will approximately have the properties expected of a random sample of n observations from the standard exponential distribution. A plot of the i th ordered residual $r_{(i)}$ against the i th smallest exponential score $e_{i,n} = \sum_{j=i}^n j^{-1}$ should then give an approximate straight line relation with unit slope and zero intercept.

To deal with censoring, we note that

$$E\{R | R > t(1 + \delta \tilde{x}'_{*} \tilde{\beta}_{*})^{-1/\delta}\} = 1 + t(1 + \delta \tilde{x}'_{*} \tilde{\beta}_{*})^{-1/\delta} \quad (3.2)$$

Replacing the censored value of R by the maximum likelihood estimate of its expected value, we define modified residuals by

$$r_i^* = \begin{cases} t_i(1 + \delta \hat{x}'_{*i} \hat{\beta}_{*})^{-1/\delta} & \text{if } \omega_i = 1 \\ 1 + t_i(1 + \delta \hat{x}'_{*i} \hat{\beta}_{*})^{-1/\delta} & \text{if } \omega_i = 0 \end{cases} \quad (3.3)$$

A plot of $r_{(i)}^*$ against $e_{i,n}$ should again give an approximate straight line relation if the assumed power transformation model is correct.

4. AN ILLUSTRATIVE EXAMPLE

To illustrate the use of the power transformation model we use the data in table 1 which were taken from Prentice (1973). The data are for groups of advanced lung cancer patients, the groups being defined by two factors, type of chemotherapeutic agent at two levels standard and test, and tumour cell type at four levels, squamous, small, adeno and large. In the original study, four explanatory variables were measured and the analysis based on model (1.7) indicated that only one variable, general medical status on a scale 10,20,...,90 had a real effect. We have only included this variable in our analysis. The survival times are in days and right censored observations are shown with an asterisk.

The GLIM fitting procedure described in section 2 was used to fit the model.

$$\mu_x(j) = [1 + \delta\{\beta_{*0}(j) + x\beta_{*1}(j)\}]^{1/\delta}, \quad j=1,2,\dots,8 \quad (4.1)$$

If the model is correct, the $\{r_i\}$ will approximately have the properties expected of a random sample of n observations from the standard exponential distribution. A plot of the i th ordered residual $r_{(i)}$ against the i th smallest exponential score $e_{i,n} = \sum_{j=i}^n j^{-1}$ should then give an approximate straight line relation with unit slope and zero intercept.

To deal with censoring, we note that

$$E\{R | R > t(1 + \delta \tilde{x}'_i \hat{\beta}_*)^{-1/\delta}\} = 1 + t(1 + \delta \tilde{x}'_i \hat{\beta}_*)^{-1/\delta} \quad (3.2)$$

Replacing the censored value of R by the maximum likelihood estimate of its expected value, we define modified residuals by

$$r_i^* = \begin{cases} t_i(1 + \delta \tilde{x}'_i \hat{\beta}_*)^{-1/\delta} & \text{if } \omega_i = 1 \\ 1 + t_i(1 + \delta \tilde{x}'_i \hat{\beta}_*)^{-1/\delta} & \text{if } \omega_i = 0 \end{cases} \quad (3.3)$$

A plot of $r_{(i)}^*$ against $e_{i,n}$ should again give an approximate straight line relation if the assumed power transformation model is correct.

4. AN ILLUSTRATIVE EXAMPLE

To illustrate the use of the power transformation model we use the data in table 1 which were taken from Prentice (1973). The data are for groups of advanced lung cancer patients, the groups being defined by two factors, type of chemotherapeutic agent at two levels standard and test, and tumour cell type at four levels, squamous, small, adeno and large. In the original study, four explanatory variables were measured and the analysis based on model (1.7) indicated that only one variable, general medical status on a scale 10,20,...,90 had a real effect. We have only included this variable in our analysis. The survival times are in days and right censored observations are shown with an asterisk.

The GLIM fitting procedure described in section 2 was used to fit the model.

$$\mu_x(j) = [1 + \delta\{\beta_{*0}(j) + x\beta_{*1}(j)\}]^{1/\delta}, \quad j=1,2,\dots,8 \quad (4.1)$$

TABLE 1

Data for lung cancer patients: t = days of survival, x = general medical status, G₁ = standard squamous, G₂ = standard small, G₃ = standard adeno, G₄ = standard large, G₅ = test, squamous, G₆ = test, small, G₇ = test, adeno, G₈ = test, large.

G ₁		G ₂		G ₃		G ₄		G ₅		G ₆		G ₇		G ₈	
t	x	t	x	t	x	t	x	t	x	t	x	t	x	t	x
72	60	30	60	8	20	177	50	999	90	25	30	24	40	52	60
411	70	384	60	92	70	162	80	112	80	103*	70	18	40	164	70
228	60	4	40	35	40	216	50	87*	80	21	20	83*	90	19	30
126	60	54	80	117	80	553	70	231*	50	13	30	31	80	53	60
118	70	13	60	132	80	278	60	242	50	87	60	51	60	15	30
10	20	123*	40	12	50	12	40	991	70	2	40	90	60	43	60
82	40	97*	60	162	80	260	80	111	70	20	30	52	60	340	80
110	80	153	60	3	30	200	80	1	20	7	20	73	60	133	70
314	50	59	30	95	80	156	70	587	60	24	60	8	50	111	60
100*	70	117	80			182*	90	389	90	99	70	36	70	231	70
42	60	16	30			143	90	33	30	8	80	48	10	378	80
8	40	151	50			105	80	25	20	99	80	7	40	49	30
144	30	22	60			103	80	357	70	61	70	140	70		
25*	80	56	80			250	70	467	90	25	70	186	90		
11	70	21	40			100	60	201	80	95	70	84	80		
		18	20					1	50	80	50	19	50		
		139	80					30	70	51	30	45	40		
		20	30					44	60	29	40	80	40		
		31	70					283	90						
		52	70					15	50						
		287	60												
		18	30												
		51	60												
		122	80												
		27	60												
		54	70												
		7	50												
		63	50												
		392	40												
		10	40												

where $\mu_x(j)$ denotes the expected survival time for a patient in group j having medical status value x. The use of the initialisation value $\hat{\delta}_1 = 0$ led to the sequence $\hat{\delta}_2 = 0.359$, $\hat{\delta}_3 = 0.432$, $\hat{\delta}_4 = 0.429$ of approximations to $\hat{\delta}$ and $\hat{\delta}$ was therefore taken as 0.43 correct to two decimal places. In table 2, values of $\hat{\mu}_x(j)$ are given for the eight groups and the nine medical status levels used for classifying the patients. The maximum likelihood estimates of $\beta_{*0}(j)$ and $\beta_{*1}(j)$ are also shown together with their estimated standard errors. The results show that at the higher levels of x, the expected survival times are

highest for group G5, and that the differences among the groups G3, G6 and G7 are relatively small. These results support the general finding made by Prentice and it should be noted that for these data, the use of the power transformation model leads to only a moderate improvement in fit compared with the model given by (1.7). The value of $2 \log L\{\hat{\beta}_{*0}(0.43), 0.43\} - 2 \log L\{\hat{\beta}_{*0}(0), 0\}$ is 1.41 which when referred to the χ_1^2 distribution is just significant at the 25% level. It is of course instructive to use the general power transformation model to assess the adequacy of fit of the commonly used model given by (1.7).

TABLE 2

Maximum likelihood estimates of regression coefficients and of expected survival times.

x	G ₁	G ₂	G ₃	G ₄	G ₅	G ₆	G ₇	G ₈
10	27.4	43.0	1.4	49.2	1.9	9.3	23.8	2.5
20	43.6	52.1	5.6	67.4	13.8	15.5	30.1	10.3
30	64.1	62.2	13.0	88.8	39.4	23.6	37.1	24.8
40	89.2	73.4	24.2	113.7	81.2	33.7	45.1	46.8
50	119.2	85.6	39.5	142.1	140.7	46.0	53.9	77.1
60	154.2	98.9	59.2	174.3	219.4	60.4	63.6	116.2
70	194.5	113.3	83.5	210.2	318.6	77.1	74.3	164.7
80	240.1	128.8	112.7	250.0	439.4	96.1	85.9	223.0
90	291.4	145.5	146.9	293.7	582.7	117.6	98.5	291.7
$\hat{\beta}_{*0}(j)$	5.21	8.38	-1.74	8.30	-3.36	2.23	5.81	-1.80
SE $\{\hat{\beta}_{*0}(j)\}$	6.76	4.83	2.61	13.75	3.23	2.68	2.86	4.48
$\hat{\beta}_{*1}(j)$	0.213	0.101	0.214	0.179	0.411	0.150	0.096	0.291
SE $\{\hat{\beta}_{*1}(j)\}$	0.140	0.090	0.072	0.204	0.099	0.067	0.055	0.110

5. EXTENSIONS

The regression model given by (1.8) for the mean or equivalently the reciprocal of the hazard function of the exponential distribution is more flexible than other models that have been previously used. It is therefore likely that except for very large data sets, the model will provide a satisfactory fit if the underlying distribution is exponential, even if the model's link function is not correct.

For distributions other than the exponential the power transformation link function may be used to model the dependence of the hazard function

on explanatory variables. Thus if $h(t; \underline{x})$ denotes the hazard function for time to failure given \underline{x} , we may use the proportional hazards regression model

$$h(t; \underline{x}) = h_0(t; \underline{\theta}) (1 + \delta \underline{x}' \underline{\beta})^{1/\delta} \quad (5.1)$$

where $h_0(t; \underline{\theta})$ represents the hazard function for an individual with $\underline{x} = \underline{0}$ which is assumed to depend on a vector $\underline{\theta}$ of unknown parameters.

A straightforward argument shows that the log-likelihood under the model given by (5.1) is

$$\log L(\underline{\beta}, \underline{\theta}, \delta) = \sum_{i=1}^n (\omega_i \log \mu_i - \mu_i) + \sum_{i=1}^n \omega_i \log \{h_0(t_i; \underline{\theta}) / H_0(t_i; \underline{\theta})\} \quad (5.2)$$

where

$$H_0(t; \underline{\theta}) = \int_0^t h_0(u; \underline{\theta}) du \quad (5.3)$$

is the cumulative hazard function and

$$\mu_i = (1 + \delta \underline{x}_i' \underline{\beta})^{1/\delta} H_0(t_i; \underline{\theta}), \quad i=1, \dots, n \quad (5.4)$$

Suppose now that the underlying distribution depends on two parameters, λ and θ say, and that

$$H_0(t; \lambda, \theta) = \lambda \eta(t; \theta) \quad (5.5)$$

where $\lambda > 0$ and $\eta(t; \theta) \geq 0 \forall t > 0$. Examples of distributions having this property are the Weibull with $H_0(t; \lambda, \theta) = \lambda t^\theta$ and the Gompertz with $H_0(t; \lambda, \theta) = (\lambda/\theta) \exp(\theta t)$, $t > 0$. If (5.5) holds, we may write

$$\mu_i = \eta(t_i; \theta) (1 + \delta \underline{x}_i' \underline{\beta}_{*})^{1/\delta} \quad (5.6)$$

where

$$\beta_{*0} = \frac{\lambda^\delta - 1}{\delta}, \quad \beta_{*j} = \lambda^\delta \beta_j, \quad j = 1, \dots, k \quad (5.7)$$

For θ known, the ML estimate $\underline{\beta}_*(\theta)$ can be found treating the $\{\omega_i\}$ as observations on independent Poisson regression variables, the link function being

$$g(\mu_i) = \left[\left\{ \frac{\mu_i}{\eta(t_i; \theta)} \right\}^\delta - 1 \right] / \delta \quad (5.8)$$

If θ is unknown, a search over a grid of values of θ would be needed to find the global maximum of the log-likelihood.

REFERENCES

Aitken, M and Clayton, D (1980). The fitting of exponential, Weibull and extreme value distributions to complex censored survival data using GLIM. Applied Statistics, 29, No.2, 156-163.

Fiegl, P and Zelen, M (1965). Estimation of exponential survival probabilities with concomitant information. Biometrics, 21, 826-838.

Greenberg, R A, Bayard, S and Byar, D (1974). Selecting concomitant variables using a likelihood ratio step-down procedure and a method of testing goodness of fit in an exponential survival model. Biometrics, 30, 601-608.

Kay, R (1977). Proportional hazard regression models and the analysis of censored survival data. Applied Statistics, 26, No.3, 227-237.

Pregibon, D (1980). Goodness of link tests for generalised linear models. Applied Statistics, 29, No.1, 15-24.

Prentice, R L (1973). Exponential survivals with censoring and explanatory variables. Biometrika, 60, 279-288.

END

FILMED

1-85

DTIC