PROPORTIONAL HAZARDS MODELLING FOR THE ANALYSIS OF RELIABILITY FIELD DATA

BY

ESTELLE WALKER BSc. (Hons.)

Thesis submitted to the CNAA in partial fulfilment of the requirements for the degree of Master of Philosophy.

Sponsoring Establishment

Trent Polytechnic

Collaborating Establishment

GEC - Research

May 1989

ESTELLE WALKER:

Proportional Hazards Modelling for the Analysis of Reliability Field Data.

This thesis reviews the theory of the Proportional Hazards Model for use in analysing reliability field data. Since, much work in reliability is concerned with providing statistical models for the lifelengths or interfailure times of equipment, and hence provide predictions for their performance in the field, reliability field data might be considered most appropriate on which to base ones model.

Proportional Hazards Modelling has great potential in the context of analysing reliability field data, since it assesses significant affects on the lifelengths or interfailure times of equipment due to both internal and external covariates.

Three applications of PHM to the analysis of reliability field data, are presented. The applications illustrate that there is no unique form for reliability data. The form of the data is influenced by the system structure, the system's deployment, repair and maintenance regimes, and the data collection procedures. It is for these reasons that Proportional Hazards Modelling can not be applied in a black box fashion. This thesis demonstrates the questions that require resolving before fitting a model and the difficulties involved in extracting data from poor collection processes.

Four commonly used graphical diagnostic procedures, which assess the fit of the model and indicate whether or not the model's assumptions are violated are highlighted. These procedures are discussed in detail and extended. The additions to the forms applied previously in the literature ease the visual inspection of the plots.

An exploratory data analysis approach to Proportional Hazards Modelling of reliability field data is advised.

OBJECTIVES

The main objectives of the research for this thesis are:

(i) To review the literature on the use of Proportional Hazards Modelling for reliability analysis.

(ii) To investigate the problems involved in applying Proportional Hazards Models to various formats of reliability field data.

(iii) To investigate the implementation of graphical diagnostic procedures for investigating the appropriateness and fit of Proportional Hazards Models.

(iv) To improve the visual inspection of diagnostic plots and hence ease decision making.

(v) To suggest an appropriate route through analysing reliability field data with Proportional Hazards Models.

ADVANCED STUDIES

The following advanced studies were undertaken in connection with the programme of research for this thesis:

(i) Attendance of research seminars at Trent Polytechnic.

(ii) Participation in the following conference:

6th National Reliability Conference, Birmingham 1987.

(iii) Attendance at the following conferences:

'Data Collection and Analysis for Reliability Assessment'. The Institution of Mechanical Engineers, March 1986.

9th Advances in Reliability Technology Symposium, Bradford 1986.

10th Advances in Reliability Technology Symposium, Bradford 1988.

(iv) Presentation and discussion of research findings with industrial collaborator, and data providers.

(v) Attendance of lectures from statistics component of the following degree course at Trent Polytechnic:

BSc.(Hons.) Combined Studies in Sciences, year 3.

(vi) Appropriate reading.

Throughout the period of registration, the author has undertaken some part-time lecturing and tutorials for courses at Trent Polytechnic, viz:

September 1986 - June 1987:	Business Studies I
September 1987 - June 1988:	Business Studies I
1987 - 1988 Various modules:	Post Graduate Diploma in Reliability.

ACKNOWLEDGEMENTS

Work for this thesis has been completed at Trent Polytechnic. The author thanks the institution for the resources provided and their assistance. The author also thanks the staff at the collaborating establishment; GEC - Research.

Thanks go to the companies and individuals who provided the reliability field data analysed in this thesis and the the personnel in these organistions for their help in discussion and feedback: Dr. Tomasz Nowakowski for the data on bus subsystem failures, ICL for the computer hardware failure data, and British Aerospace for the reliability data supplied from a modern weapon system.

Thanks are due to Dr. David Wightman for introducing the author to Proportional Hazards Modelling and the computer programs implemented at Trent Polytechnic, also for his helpful advice throughout the period of study.

The author is grateful to her supervisors, Professor A. Bendell, Dr. M. Baxter, and Dr. M. Phillips for their guidance. Especial thanks to Professor A. Bendell for his efforts in obtaining the field data sets studied by the author.

DECLARATION

During the period of registration for MPhil. the author has not been registered for any other award of the CNAA, nor any award of a University.

CONTENTS

CHAPTER 1

INTRODUCTION TO RELIABILITY

1.1 HISTORY

Reliability is considered to have evolved into a specific field, from the related areas of quality control and machine maintenance, with the appearance of the first papers on reliability in the late 1940's and early 1950's. A review of the development of statistical methods in reliability in the years prior to 1983 is given by Lawless (1983).

Reliability journals first began to appear in the 1950's; IEEE Transactions and Technometrics. These have been followed recently in the 1980's by Reliability Engineering and System Safety (formerly Reliability Engineering), and Quality and Reliability Engineering International.

Much of the statistical work in reliability has investigated the lifetimes of equipment, or their failures over a time period. Parametric families of distributions that could be used to model lifetimes, therefore, began to appear in engineering contexts in the late 1930's and the 1940's. By the 1950's the Weibull and exponential distributions became particularly popular to model lifetimes.

1.2 FAILURE TIME DATA

Functions used extensively in reliability studies are defined below.

Consider an equipment which has not failed by time t. The hazard function, h(t), is the limit of the ratio of the probability of failure in $(t,t+\Delta t)$ to Δt.

$$h(t) = \lim_{\Delta t \to 0^+} \frac{\text{Prob}(\ t<T<t+\Delta t \mid t<T\ )}{\Delta t} \qquad 1.1$$

where T is the time to failure.

The hazard function h(t) gives the probability of instantaneous failure of the equipment having survived to time t, see Cox (1962).

The hazard function can also be considered in terms of other related functions.

For example:

$$h(t) = \frac{f(t)}{R(t)} \qquad 1.2$$

where f(t) is the probability density function of time to failure.

R(t) is the probability that the equipment has survived to time t.

[R(t) is taken to be monotonically non-increasing with $R(0)\equiv1$ and $R(\infty)\equiv0$].

Thus:

$$R(t) = \exp\left[-\int_0^t h(x).dx\right] \qquad 1.3$$

and

$$f(t) = h(t).\exp\left[-\int_0^t h(x).dx\right] \qquad 1.4$$

Another function frequently of interest is the cumulative hazard, H(t);

$$H(t) = \int_0^t h(x).dx \qquad 1.5$$

The exponential distribution has a well developed methodology, and a simple form for the hazard function (constant). This has contributed greatly to its extreme popularity in the reliability field.

In many applications, however, the exponential is not an appropriate model. More complex distributional forms were therefore considered, leading to the popularity of the Weibull distribution being used to model lifetime data. (The exponential is a special case of the Weibull).

Other frequently used distributions are the Gamma, the Generalised Gamma, Log-Normal and Log-Logistic, see

Kalbfleisch and Prentice (1980), Lawless (1982), Cox and Oakes (1984).

1.3 POINT PROCESSES

Point processes are often used to treat repeated failures on the same item of equipment, see Cox and Miller (1965), Cox and Lewis (1966), Feller (1968), Lewis (1972), Thompson (1981), Ascher and Feingold (1984).

Models of this type include homogeneous Poisson processes, non-homogeneous Poisson processes, renewel processes, and superimposed processes.

Cox (1972a) and Gail *et al* (1980) consider models of this type which allow for explanatory variables. They are discussed also by Kalbfleisch and Prentice (1980) and Lawless (1982).

1.4 REGRESSION MODELS

Explanatory variables associated with the response variable (usually the time to failure or the time between failures) are often incorporated in reliability applications.

Regression techniques where maximum likelihood or linear estimaton procedures are usually employed can be used to model the effects of these explanatory factors.

Accelerated Failure Time Models, and Proportional Hazards Models are the regression techniques most commonly applied to failure time data.

In Proportional Hazards models the explanatory variables are assumed to act proportionally on the hazard, whilst the explanatory variables are assumed to act multiplicatively on the time to failure, in Accelerated Lifetime Models.

CHAPTER 2

PROPORTIONAL HAZARDS MODELLING

2.1 HISTORY

The technique of Proportional Hazards Modelling (PHM) has its origin largely in the seminal paper presented to the Royal Statistical Society in March 1972 by Professor Sir D.R. Cox.

PHM combines concepts from biostatistics and reliability theory. It incorporates regression-like arguments for explanatory factors into life-table analysis.

Cox (1972) regarded the application of the technique most likely to be used in 'industrial reliability studies' and in 'medical studies'.

It is indeed true that the medical statistics community readily used the approach of PHM in their studies; the majority of the early papers using the technique are in the medical field. The industrial reliability community, however, took rather longer to apply the technique within their studies.

2.2 BASIC THEORY

PHM is a technique whereby identification of independent effects of variables thought to influence the life length of equipment is possible without the necessity of specifying any particular distributional form for the life lengths of equipment.

The variables associated with the life length of equipment are often termed covariates, explanatory variables, or explanatory factors.

Commonly employed covariates are: operating conditions, material, manufacturer, season, time of day. It is often the case that a covariate such as time since installation for example is incorporated to test for reliability growth or decay.

PHM allows the inclusion of time dependent covariates that can vary with the life experience of equipments.

The method decomposes the observed variation in life length into orthogonal factors, and a common baseline, then identifies which of the factors are significant. Hence, the relative effects of the significant variables can be observed.

The model is structured on the hazard function. Cox(1972) assumes the decomposition of the hazard function into the product of a baseline hazard which is common to all equipments, and an exponential term incorporating the effects of the explanatory variables. This decomposition is written:

$$h(t;z_1,z_2,\ldots.z_k) = h_0(t).\exp(\beta_1 z_1+\beta_2 z_2+\ldots+\beta_k z_k), \qquad 2.1$$

$$t>0, \quad -\infty<\beta_i<\infty, \quad -\infty<z_i<\infty$$

Where t is the survival time for an individual.

The z_i's, $i=1,2,\ldots,k$ are the values of the covariates. They may be naturally measured variables such as temperature, or indicator variables representing for example the presence or absence of a design change.

The β_i's, $i=1,2,\ldots,k$ are the unknown parameters of the model which represent the effect on the hazard of each value of the covariates.

$h_0(t)$ is the common baseline hazard function. The baseline hazard represents the hazard an equipment would experience if the covariates all take the baseline value zero. These may correspond to a natural zero, say temperature, or a nominal zero representing for example a particular design type.

From 2.1 above the effect of the covariates on the hazard is to act multiplicatively on the baseline hazard $h_0(t)$, so for different values of a significant covariate the respective hazard functions are proportional over all time; hence, the name Proportional Hazards Model. Fig 2.1 illustrates graphically the proportionality of the hazard functions.

For all t the hazard experienced by an equipment with covariate value b is twice that experienced by an equipment with covariate value a.

The method estimates the parameters β_i and tests whether these are significantly different from zero, hence whether each covariate has a real effect in explaining the variation in observed life lengths.

Although it has become virtually synonymous with PHM it is not necessary to construct the covariate effects within an exponential term. The formulation is used because there are no restrictions on the coefficients ß, and it ensures that the hazard is always positive.

A fully parametric proportional hazards model can be obtained if one assumes a particular distributional form for $h_0(t)$.

It is often, however, that with complex systems and with the confusing effects of covariates it is difficult or impossible to assume a specified form of the baseline hazard function with any real justification. It is for these reasons that a distribution free approach to modelling the baseline hazard function is the more common procedure. This procedure in which the baseline hazard function is left distribution free was adopted by Cox (1972); the baseline hazard function is an unspecified, non-parametric, non-negative arbitrary function. The form of this function is estimated from the data.

The non-parametric technique requires a different approach from the usual likelihood procedures for the estimation of the coefficients ß.

Cox (1972), utilises, based on heuristic arguments, a 'conditional likelihood' in the estimation of the coefficients ß. This 'conditional likelihood' was subsequently justified within the framework of partial likelihood Cox (1975). It is now generally referred to as Cox's partial likelihood.

2.2.1 PARTIAL LIKELIHOOD CONSTRUCTION

Cox's partial likelihood is based conditionally on the set $\{t_i\}$, $i=1,2,\ldots,k$ of times at which failures occurred. For any time t_i, conditional on the risk set at t_i, R_i, which consists of all items still operational just prior to t_i, the probability that the failure is on the item observed is:

$$\frac{\exp(\beta z_i)}{\sum_{l \in R_i} \exp(\beta z_l)}$$

where β is a row vector of k parameters, and z_i is a column vector of k measured covariate values.

Cox's partial likelihood is then:

$$L(\beta) \quad = \quad \prod_{i=1}^{n} \frac{\exp(\beta z_i)}{\sum_{l \in R_i} \exp(\beta z_l)} \qquad 2.2$$

where n is the observed number of failure points.

2.2.2 ESTIMATION PROCEDURES

We require to find the values of β that maximise the partial likelihood.

The natural logarithm of the partial likelihood is taken, and then the first and second partial differentials with respect to each of the parameters $\beta_1, \beta_2, \ldots, \beta_k$ are obtained. A Newton-Raphson iteration procedure is used in this thesis which employs a Taylor series expansion for each step in the iteration procedure. The procedure starts with initial values of zero for β, and the iteration continues until the convergence criterion is satisfied. In this thesis we use a convergence criterion based on the absolute ratio β_i and the last change $\Delta\beta_i$, $i=1,2,\ldots,k$; convergence is satisfied if $|\Delta\beta_i| \leq |\beta_i| \times 10^{-4}$, $i=1,2,\ldots,k$, for all covariates, see Wightman (1987).

Alternative optimisation procedures are available, for example the expectation maximisation (EM) algorithm, (Cox and Oakes (1984)).

Once the estimates converge, tests of whether each explanatory variable has a significant effect are based upon the asymptotic Normality of the estimates.

Once estimates have been found for $\beta_1, \beta_2, \ldots, \beta_k$, a distribution free estimate of the baseline hazard function may be obtained. From the relationship between the cumulative hazard function and the survivor function we can readily obtain an estimate of the baseline survivor function. There are various approaches to obtaining these estimates suggested in the literature. Cox (1972) considers a baseline hazard function which is taken to be identically zero at points where no failure has occurred.

Cox himself notes that this method is, however, complex. Both Oakes and Breslow in the discussion section of Cox (1972) suggest a non-zero constant value of $h_o(t)$ between the failure times.

Kalbfleisch and Prentice (1973) use a procedure for estimating the baseline hazard, which allows for ties, whereby a likelihood is built in terms of the baseline survivor function. The full likelihood is given by:

$$L(\alpha,\beta) = \prod_{i=1}^{n} \left\{ \prod_{j \in D_i} \left(1 - \alpha_i^{\exp(\beta z_j)} \right) . \prod_{l \in (R_i - D_i)} \alpha_i^{\exp(\beta z_l)} \right\} \qquad 2.2$$

where D_i set of labels associated with individuals failing at t_i

R_i is the set of labels associated with individuals at risk just prior to t_i

$(1-\alpha_i)$ hazard contribution at t_i

n number of failure points

Since the β's have been estimated already from the partial likelihood, 2.2 can be maximised with respect to the α_i's.

The maximum likelihood estimate of the baseline survivor function is:

$$\hat{S}_{o(t)} = \prod_{t_i < t} \hat{\alpha}_i \qquad 2.3$$

The routines employed in this thesis which are extensions of those given in Kalbfleisch and Prentice (1980), use the likelihood formulation 2.2 to obtain an estimate of the baseline survivor function;

$$\hat{S}_o(t) = \prod_{i/t_{(i)}<t} \hat{\alpha}_i$$

which would give an estimate of the cumulative baseline hazard function;

$$\hat{H}_o(t) = - \sum_{i/t_{(i)}<t} \ln \hat{\alpha}_i \qquad 2.4$$

However, the first order approximation to 2.4 is in fact used in the Kalbfleisch and Prentice routines where;

$$\hat{H}_o(t) = \sum_{i/t_{(i)}<t} (1 - \hat{\alpha}_i) \qquad 2.5$$

This is sometimes called the empirical cumulative hazard function (Lawless (1982)).

There are occassions where computational problems arise in a PHM model.

A covariate is monotonic when it is the largest of all the covariate values in the risk set at each failure time; or when it is the smallest of all the covariate values in the risk set at each failure time.

If a covariate z is monotone, then the partial likelihood will be monotone in β, leading to the estimate of $\hat{\beta}=\infty$ or $\hat{\beta}=-\infty$.

Since estimation cannot proceed in such a situation, it is necessary to remove the monotonic covariate from the model.

Multicollinearity in the covariates also inhibits the estimation procedure. This occurs when a covariate is a linear combination of one or more of the other covariates in the model.

It is again necessary to eliminate one of the covariates involved in the multicollinearity to allow estimation of the coefficients to proceed.

2.2.3 TIES

The model by nature is a continuous time model, hence it is assumed that all failed life lengths are distinct. However, it is often the case with reliability data that tied failure times are recorded, this is usually due to the crudity of the time measurement. In the case of tied failure points approximations for the contributions to the partial likelihood at that failure time are available, (Wightman (1987)). The contribution used in this thesis for analysis is that suggested by Breslow (1974):

$$\frac{\exp(\beta s_i)}{\left[\sum_{l \in R_i} \exp(\beta z_l)\right]^{d_i}} \qquad 2.6$$

where s_i is the sum of z over failures at t_i

d_i number of failures observed at t_i

There is no guidance in the literature which would indicate the number of ties for which a discrete version of the model would be more appropriate. The discrete model will be discussed later in section 2.3.6.

In the estimation procedure for the baseline hazard function, whereby 2.2 is maximised with respect to the α_i's, if a single failure occurs at t_i the value $\hat{\alpha}_i$ can be obtained directly. If ties are involved an iterative procedure is required.

2.3 EXTENSIONS TO BASIC MODEL

The basic model as described in section 2.2 can be built upon in a number of ways which will be discussed in this section.

2.3.1 STRATIFICATION

The proportional hazards model requires that an explanatory factor should affect the hazard multiplicatively. Although this may be descriptive of many situations it is unrealistic to expect that all the covariates necessarily fulfill the proportionality assumption. [For example see covariate z_1-event in the example of PHM applied to computer hardware failures, in section 4.5.3].

When a factor does not affect the hazard multiplicatively, stratification could be employed.

Suppose a factor which occurs on q levels violates the proportionality assumption. Individuals can be assigned to q strata based on the level of the factor. The hazard function for an individual in the j'th stratum is given by:

$$h_j(t;z) = h_{oj}(t).\exp(\beta z) \qquad j=1,2,\ldots,q \qquad 2.7$$

Individuals in the same stratum have proportional hazards, but this is not necessarily true for individuals between different strata.

In 2.7 it is assumed that the relative effect of the explanatory variables is the same within all strata, this condition can and may be relaxed with β varying between strata. This is equivalent to applying separate models to each stratum.

A partial likelihood function $L_j(\beta)$ is obtained for each stratum, and the overall partial likelihood of β is approximately the product of these terms. In general:

$$L(\beta) = \prod_{j=1}^{q} L_j(\beta) \qquad 2.8$$

Prentice, Williams and Peterson (1981) introduce models whereby repairable items 'move through strata upon failure'. That is; prior to its first failure an equipment is in stratum 1, after its first and prior to its second failure the equipment is in stratum 2, etc.

2.3.2 COVARIATE FORMULATION

It is usual for simplicity in the application of PHM to reliability problems to include covariate information in the form in which it is collected and measured.

Under the circumstances of the proportionality assumption being violated other formulations for the covariate information may be more appropriate.

Transformations of a covariate z_j may lead to proportional hazards. For example a covariate $x = \ln z_j$, or $x = \sqrt{z_j}$ may be introduced as an alternative. (Davies *et al* (1988)).

It is also possible that a time dependent formulation will be more appropriate. For example $x(z_j, t)$ a function of the measured variable z_j and the basic time metric t may be introduced as an alternative covariate. The choice of transformation is presently arbitrary, (see Wightman, Walker and Bendell (1988), appendix C).

2.3.3 COMPETING RISKS

Competing risk formulations within the PHM methodology are available in the literature. Holt (1978) introduced two models; one which had the same baseline for each cause with cause specific β coefficients, and the other with different baseline and β coefficients for each cause. Kalbfleisch and Prentice (1980) discuss such models and, in the case of common baseline for each cause, find a justification for the procedure of regarding a failure from a particular cause as also a censored event for the other causes.

The above procedure for competing risks is employed in analyses presented in this thesis.

In the analyses presented in this thesis models for competing risks are considered whereby the basic time metric is taken as the time between consecutive failures, irrespective of cause, (this is an alternative to modelling the different failure modes on separate time streams). Given that there are $l+1$ failure modes in the model, l binary dummy variables are introduced into the covariate set to represent the different causes of failure. Censoring events are generated at each failure time for each of the other failure modes.

Interesting results arise from the above model formulation, particularly that of the factorisation of the partial likelihood, (Wightman (1987)). The partial likelihood factorises into two terms; one containing only information from failure mode variables, and the other term being the usual partial likelihood for a model without competing risks, (see also Wightman, Walker and Bendell (1988), appendix C).

Originating from the above result it can be shown that a constant expectated value, conditional on the risk set, of a failure mode indicator covariate is obtained. This will be discussed in further detail in the later section on Schöenfeld's partial residuals, 6.2.6.

2.3.4 DISCRETE PHM

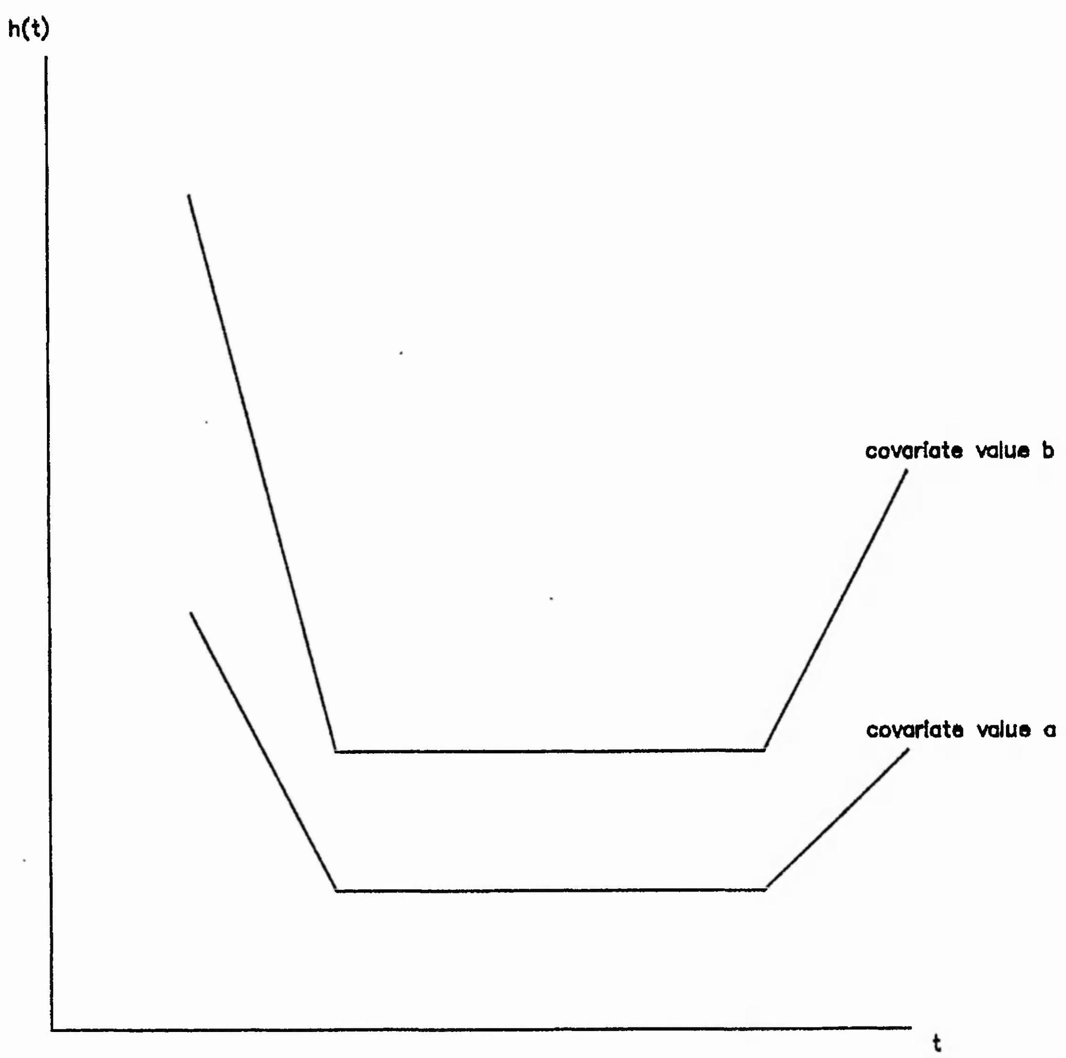
A discrete proportional hazards modelling formulation was introduced by Kalbfleisch and Prentice (1973).

This formulation may be applicable to a continuous time model when there is a large number of ties at failure times, or when failures are grouped into disjoint intervals for example, due to their discovery during routine inspections (Wightman (1987)).

The essence of the discrete proportional hazards model is that the time axis is split into specific time intervals. For each time interval a parameter, h_i, is allocated; the model then contains a finite number of parameters $h_1, h_2, \ldots, h_n$ and β, the vector of covariate coefficients. Likelihood procedures can be implemented for obtaining asymptotic maximum likelihood estimates.

Further discussion and development of discrete models is provided by Prentice and Glöeckler (1978).

Figure 2.1 Hazard functions for different covariate values.



CHAPTER 3

APPLICATION TO VEHICULAR SUBSYSTEM RELIABILITY

This chapter shows an example of an application to the reliability analysis of two vehicular subsystems. The example illustrates the importance of good data recording; and also the necessity of using PHM as an exploratory tool.

3.1 THE INITIAL DATA

PHM was applied to analyse the failure data from two subsystems of a vehicle, supplied by T. Nowakowski (Politechnika Wroclawska).

The data was obtained during field tests of 110 buses operating in 5 different towns/environments.

Each bus was tested over a total of at least 100,000 miles.

Table 3.1 shows the number of buses on test in each of the environments.

TABLE 3.1 Test parameters

Environment	Sample size
1	10
2	50
3	20
4	20
5	10

Figure 3.1 indicates the way in which the failures were recorded.

Table 3.2 shows an example of the form of the data as it was originally provided by the source.

It is stated in an analysis using a modified multiple regression model by the data provider (Nowakowski (1986)) that there are no differences between the effects of the different environments for subsystem A, but subsystem B is strongly influenced by the environment factors.

3.2 ANALYSIS; MODEL 1

The two subsystems were analysed separately.

The basic 'performance' metric, t, in the hazard function (see equation 2.1), was taken to be the miles travelled between events. An 'event' being a recorded failure of the subsystem, or the position at which it leaves the field of observation.

Dummy variables were introduced to model the effects of the environments. Environment 2 was taken as the base since this area has the most buses under observation.

i.e.

	z_1	z_2	z_3	z_4	z_5	
Environment 1	1	–	0	0	0	
Environment 2	0	–	0	0	0	(base)
Environment 3	0	–	1	0	0	
Environment 4	0	–	0	1	0	
Environment 5	0	–	0	0	1	
		(base)				

<u>Table 3.2 Reliability data of Subsystem A</u>

ENVIRONMENT									
1		2		3		4		5	
MTFF	MBF	MTFF	MBF	MTFF	MBF	MTFF	MBF	MTFF	MBF
25502		3501	114366	48469	20199	40462	26898	7464	46485
40335		12946	392	44135	6130	55609	5599	16969	20477
49534		14990	65634	46480	11969	110749	32206	560	90000
90500		18953	62845	41882	23940	60910	63547	16837	45043
		29415	51705	84984		75486	10575	13302	
		64304	3352	14580		79243			
		68455	337	70126		66790			
		71775	34696			82285			
		78703	59782			32657			
		81562	9089			90238			
		85369	3849			103268			
		103126	881			97362			
		108782	4559			112143			
		112311	7403			135083			
		119651	312						
		5788	52407						
		14838	46791						
		16474	90525						
		25793							
		49610							
		68256							
		70838							
		74775							
		81297							
		81321							
		100904							
		106194							
		112248							
		118020							

Since the data was split into two groups; miles to first failure, and miles between failures, an indicator variable was also introduced into the model such that;

$$z_6 = \begin{cases} 1 & \text{first event} \quad \text{(MTFF)} \\ 0 & \text{subsequent event (MBF)} \end{cases}$$

As much censoring information as could be obtained from the data was included in the model. For example, from Table 3.2, looking at environment 1 we are given four miles to first failure (MTFF's). Since we know that there are ten different buses operating in environment 1 (See Table 3.1), there must be six first events censored at 100,000 miles. Also, since there are no miles between failures (MBF's) given there must be another four subsequent events censored from the stage of the first failure up to 100,000 miles.

The results from the proportional hazards analysis for this first model (MODEL 1) can be seen in Tables 3.3 and 3.4, applied to subsystem A and subsystem B respectively.

These show the ß parameter estimates for the significant covariates; their p-value (the probability of obtaining such an extreme effect just due to chance); and the multiplicative effect on the baseline hazard that an indicator variable when equal to 1 will have. The likelihood ratio statistic, L, is also shown for the fitted model, with the tabulated 5% critical value from the Chi-squared distribution, given in brackets.

Table 3.3 Results of PHM analyses. Subsystem A.

MODEL	1			2			3			4		
covariates	β	p	Mult.	β	p	Mult.	β	p	Mult.	β	p	Mult.
z_1 - envir. 1	-	-	-	-	-	-	-	-	-	-	-	-
z_2 - envir. 2	base	-	-	base	-	-	base	-	-	base	-	-
z_3 - envir. 3	-	-	-	-	-	-	-	-	-	-	-	-
z_4 - envir. 4	-	-	-	-	-	-	0.8787	0.0017	2.41	0.9307	0.0011	2.54
z_5 - envir. 5	-	-	-	-	-	-	-	-	-	-	-	-
z_6 - MTFF	-1.2861	0.0000	0.25	-1.1404	0.0000	0.32	-2.8656	0.0000	0.06	*******	*******	*******
z_7 - travelled	*******	*******	*******	*******	*******	*******	-0.0893	0.0000	-	-0.0922	0.0000	-
z_8 - 1st fail.	*******	*******	*******	*******	*******	*******	*******	*******	*******	-2.8185	0.0000	0.06
z_9 - 2nd fail.	*******	*******	*******	*******	*******	*******	*******	*******	*******	-	-	-
z_{10} - 3rd fail.	*******	*******	*******	*******	*******	*******	*******	*******	*******	base	-	-
z_{11} - later fail.	*******	*******	*******	*******	*******	*******	*******	*******	*******	1.3963	0.0039	4.04
	L = 30.900 (3.841)			L = 20.312 (3.841)			L = 237.439 (7.815)			L = 242.961 (9.488)		

Table 3.4 Results of PHM analyses. Subsystem B.

MODEL	1			2			3			4		
covariates	β	p	Mult.	β	p	Mult.	β	p	Mult.	β	p	Mult.
z_1 - envir. 1	-1.7538	0.0014	0.17	-1.7619	0.0014	0.17	-	-	-	-	-	-
z_2 - envir. 2	base	-	-	base	-	-	base	-	-	base	-	-
z_3 - envir. 3	-	-	-	-0.5152	0.0084	0.95	-	-	-	-	-	-
z_4 - envir. 4	-1.8914	0.0000	0.15	-1.7396	0.0000	0.17	-0.8625	0.0170	0.42	-	-	-
z_5 - envir. 5	0.9541	0.0000	2.60	0.8638	0.0000	2.37	0.6384	0.0000	1.89	-	-	-
z_6 - MTFF	-1.0297	0.0000	0.36	-0.7100	0.0000	0.49	-2.0246	0.0000	0.13	*******	*******	*******
z_7 - travelled	*******	*******	*******	*******	*******	*******	-0.0450	0.0000	-	-0.0556	0.0000	-
z_8 - 1st fail.	*******	*******	*******	*******	*******	*******	*******	*******	*******	-2.8438	0.0000	0.06
z_9 - 2nd fail.	*******	*******	*******	*******	*******	*******	*******	*******	*******	-0.9920	0.0000	0.34
z_{10} - 3rd fail.	*******	*******	*******	*******	*******	*******	*******	*******	*******	base	-	-
z_{11} - 4th fail.	*******	*******	*******	*******	*******	*******	*******	*******	*******	-	-	-
z_{12} - 5th fail.	*******	*******	*******	*******	*******	*******	*******	*******	*******	-	-	-
z_{13} - later fail.	*******	*******	*******	*******	*******	*******	*******	*******	*******	2.0908	0.0000	8.09
	L = 161.243 (9.488)			L = 131.374 (11.07)			L = 291.878 (9.488)			L = 353.042 (9.488)		

The likelihood ratio statistic for the models fitted to both subsystems indicate that the model is significant in explaining the observed data.

The covariate, z_6, indicating whether the event was an initial or subsequent event, is highly significant showing that there is a lower hazard for first events (more miles to first failure, than miles between failures).

The results also appear to confirm the hypothesis that there is no difference between the effects of the environments on subsystem A, but that there are significant differences between the effects of the environments on subsystem B.

For subsystem B there is no significant difference between the effects of environment 3 and the base (environment 2); environments 1 and 4 are significantly 'better' (more miles between failures experienced within these environments); and environment 5 is significantly 'worse' (fewer miles between failures experienced in this environment.)

3.3 FURTHER DATA AND MODELS

As indicated is section 3.2 the data provided in the form shown in Table 3.2 masks a great deal of information. We were unable to match the MBF's to the MTFF's for each bus, and hence lost a large amount of censoring information.

The data was immediately sought from the source in a form in which a 'time stream' (time in this instance is really miles travelled) for each bus could be observed. It is then possible to match MBF's to MTFF's and obtain more information for censored events. Table 3.5 shows an example of the form of the data finally obtained.

3.3.1 MODEL 2

A second model (MODEL 2) was applied to this new form of the data with increased censoring information, obtained from the second format of the data, with the same covariates as defined in MODEL 1.

The results obtained, as seen in Tables 3.3 and 3.4, are roughly similar to those for MODEL 1.

For subsystem B environment 3 now also appears 'better' than the base environment 2.

The likelihood ratio statistics, L, again indicate that the models fitted have significant explanatory power.

Despite the increased information in the data, and although having significant explanatory power MODEL 2 does not appear an especially good fitting model, for either subsystem.

Table 3.5 Reliability data for environment 3.

Bus No.	Subsystem A	Subsystem B	Bus No.	Subsystem A	Subsystem B
1	48469	22280	11		18562 20483 90238 94795
2	44135 64334 70464	21205 36596 51467	12		13262 117890
3	46480	36796 41350 53653	13	14580	
4	41882 53851	2753	14		
5			15		
6	84984	63707 84199	16	70126 94066	39142 89041 120184
7		62210	17		33169 66396
8			18		71709 86857
9		82954	19		126800
10		109948	20		40565 67476

3.3.2 MODEL 3

The third model fitted (MODEL 3) includes a covariate for the cumulative number of miles travelled by the bus up to each event. This, if significant, would indicate whether the hazard was increasing or decreasing as the bus travelled more miles.

The inclusion of this additional covariate has quite a dramatic effect; altering the significance of many of the covariates previously fitted.

We can see from the results in Tables 3.3 and 3.4 that for both subsystems the additional covariate, z_7, is highly significant, showing a decreasing hazard as the mileage increases (i.e. reliability growth).

For both subsystems the indicator variable for the event (MTFF/MBF) now shows a greater effect on the hazard.

For subsystem A environment 4 now appears 'worse' than the base and other environments.

For subsystem B the environments 1 and 3 no longer appear to be significantly different from the base (environment 2).

For the models, fitted to both subsystems, there has been a dramatic increase in the likelihood ratio statistics, L. The number of significant covariates have of course also increased which will raise the likelihood ratio statistic.

3.3.3 MODEL 4

Since the indicator variable for the events (MTFF/MBF) shows, in all the models, a highly significant effect, and since there was no real reason to split on MTFF/MBF, other than that being the form in which the data was originally presented, the concept of modelling the failure number was extended, to search for increasing or decreasing hazard as the number of previous failures increases.

For the analysis of subsystem A dummy variables z_8 to z_{11} were introduced to replace z_6, such that:

	z_8	z_9	z_{10}	z_{11}	
1st event	1	0	–	0	
2nd event	0	1	–	0	
3rd event	0	0	–	0	(base)
later event	0	0	–	1	
			(base)		

And in the analysis of subsystem B, dummy variables z_8 to z_{13} were introduced such that:

	z_8	z_9	z_{10}	z_{11}	z_{12}	z_{13}	
1st event	1	0	–	0	0	0	
2nd event	0	1	–	0	0	0	
3rd event	0	0	–	0	0	0	(base)
4th event	0	0	–	1	0	0	
5th event	0	0	–	0	1	0	
later event	0	0	–	0	0	1	
			(base)				

The results from this model (MODEL 4) can also be seen in Tables 3.3 and 3.4.

For subsystem A there appears to be no significant difference between the 2nd and 3rd events (the third events are modelled on the baseline). However, when z_9 (2nd event) was eliminated it had $\hat{\beta}_9 = -0.61988$, and so we can see a gradation within the effects of these dummy variables, which shows the hazard to be increasing as more failures occur.

Again, environment 4 appears 'worse' than the others.

For subsystem B there is also a gradation within effects of the 'failure number' dummy variables. Although z_{11} and z_{12} were eliminated, they were marginal with p-value 0.0265 and 0.0253 respectively. This gradation shows, for this subsystem also, the hazard increases as more failures occur.

None of the environments now have any significant effect, refuting our earlier beief that subsystem B is highly influenced by the environment factor.

The likelihood ratio statistics, L, have again increased.

3.4 GRAPHICAL DIAGNOSTICS FOR MODEL 4

Graphical diagnostics applied to the final fitted model for both subsystems are shown and described in this section.

The diagnostic techniques themselves are extended and explained in greater detail in Chapter 6.

3.4.1 SUBSYSTEM A

The Proportional Hazards Model makes no assumptions about the form of the underlying distribution of the process. The method does, however, obtain an estimate of the baseline hazard $h_0(t)$ at each failure point. It is often possible then to identify the baseline distribution direclty by using the estimate for the hazard in various distributional hazard plots.

Figures 3.2, 3.3. and 3.4 are *Weibull, Log-Normal, and Extreme value hazard plots* respectively. Since the hazard contributions estimated from the fitted model do not yield straight lines on these plots, the underlying baseline distribution does not follow any of these forms. We did not manage to identify a distributional form for the baseline hazard directly from the hazard plots.

Since we have estimates of the baseline hazard function, from the fitted model, we could estimate a distributional form from the relationship with the probability density function; equation 1.2.

Methods based on those in Cox and Snell (1968) are used to obtain residual quantities at each time t_i defined as (see Section 6.3):

$$\hat{e}_i = \hat{H}_0(t_i).\exp(\beta z)$$

where $H_0(t)$ is the cumulative baseline hazard.

The estimated residuals should look roughly like a random sample from the standard negative exponential distribution, if the model is a good fit.

Plotting the natural logarithm of a survivor function $\hat{R}(e)$ estimated from the set of residuals against the residual estimates, produces a graphical goodness-of-fit 'test' for the model, since the plot should result in a straight line with gradient -1. A variance-stabilised form of the plot can be used which employs an angular transformation.

Figure 3.5 is the *variance-stabilised plot for the residuals* estimated from the fitted model. The fit of the model appears reasonable.

Proportionality plots are shown in Figures 3.6, 3.7 and 3.8. The data is stratified on some of the significant binary covariates, and the model run separately for each stratum. Plotting $\ln \hat{H}_0(t)$ v t for each stratum on the same graph should produce plots with constant vertical separation for all t, if the assumption of proportional hazards holds, (Kay (1977)).

We can see that the covariates z_4 and z_{11} appear to violate the proportionality assumption. However, the small number of points in the upper stratum in each of the plots, Figures 3.6 and 3.8, may be disguising the true effect.

Plots based on the influence of individual miles to failure on the $\hat{\beta}_1$ parameter estimates for the significant covariates (Cain and Lange (1984), Reid and Crepeau (1985)) are shown in Figures 3.9 to 3.12

The form presented here shows the estimated normal deviate for the covariate coefficient when each single miles to failure and censoring miles is excluded from the model, one at a time. This is plotted against the order by magnitude of the the miles on the horizontal axis.

We can then examine which miles to failure have miles most influential on the observed significance of the covariates, and which, if any, would if deleted remove the significance of the covariate. For 5% two-tailed tests these points correspond to estimated normal deviates in the range -1.96 to +1.96.

We can see from the plots that there are no such influential points.

3.4.2 SUBSYSTEM B

Plots similar to those described in section 3.4.1 can be seen in Figures 3.13 to 3.23. These refer to the analysis of subsystem B.

Again the underlying distributional form was not directly identified from the hazard plots.

Figure 3.16 shows the overall fit of the model to be somewhat better for subsystem B than it was for subsystem A.

The proportionality plots again show the possibility of non-proportional hazards for those covariates which have very few points in one of the strata.

There are no points that are so influential to any of the covariates that they would alter its significance if deleted.

3.5 SUMMARY

PHM has been used here in an exploratory manner, where we have moved from one model to another adding more information and searching for more explanation.

We have seen that as we moved through the models our previous assumption about the effects of the environments were contradicted; the processes involved are better described by the number of miles travelled and the number of failures that have occurred to each subsystem. Before these covariates were introduced into the model their effects were masked and partially explained by other variables present.

3.6 CONCLUSIONS FROM ANALYSES OF VEHICULAR SUBSYSTEM RELIABILITY

The results of this study indicate that following an exploratory procedure, whereby we step from model to model, we have identified a PHM model that fits the structure of the data reasonably well.

The PHM model has enabled as to identify variables which have a significant effect in explaining the miles between failures for the bus subsystems, as well as the direction and magnitude of these effects.

The main contributions come from the total number of miles travelled, which reduces the hazard as the number of miles increases, and from the number of previous failures which increases the hazard as the number of failures increases..

We have also seen how the presence or absence of some covariates can affect the significance of others which give only a partial explanation of the processes involved.

Figure 3.1 Reliability data scheme.

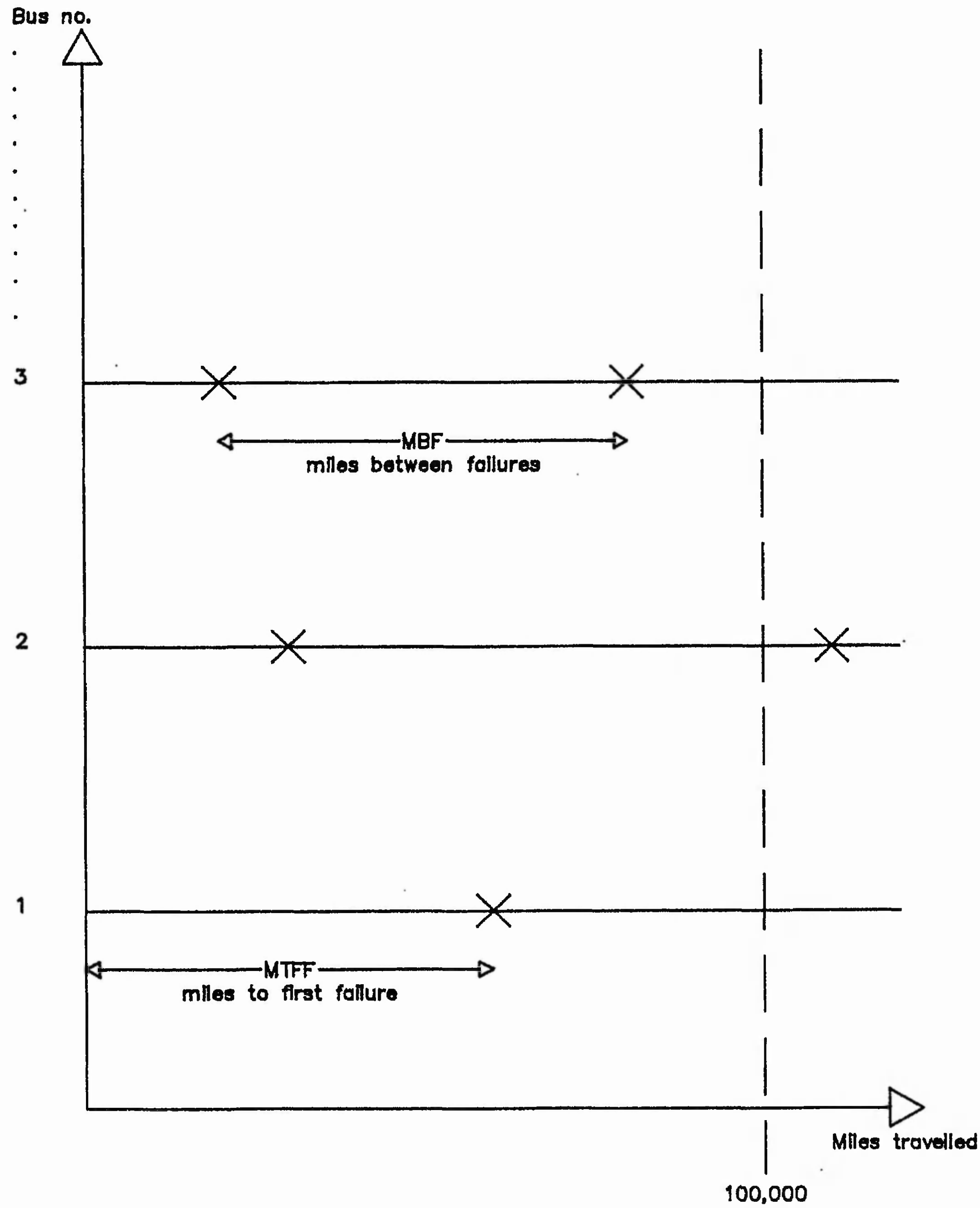


Figure 3.2 Weibull baseline hazard plot. Subsystem A.

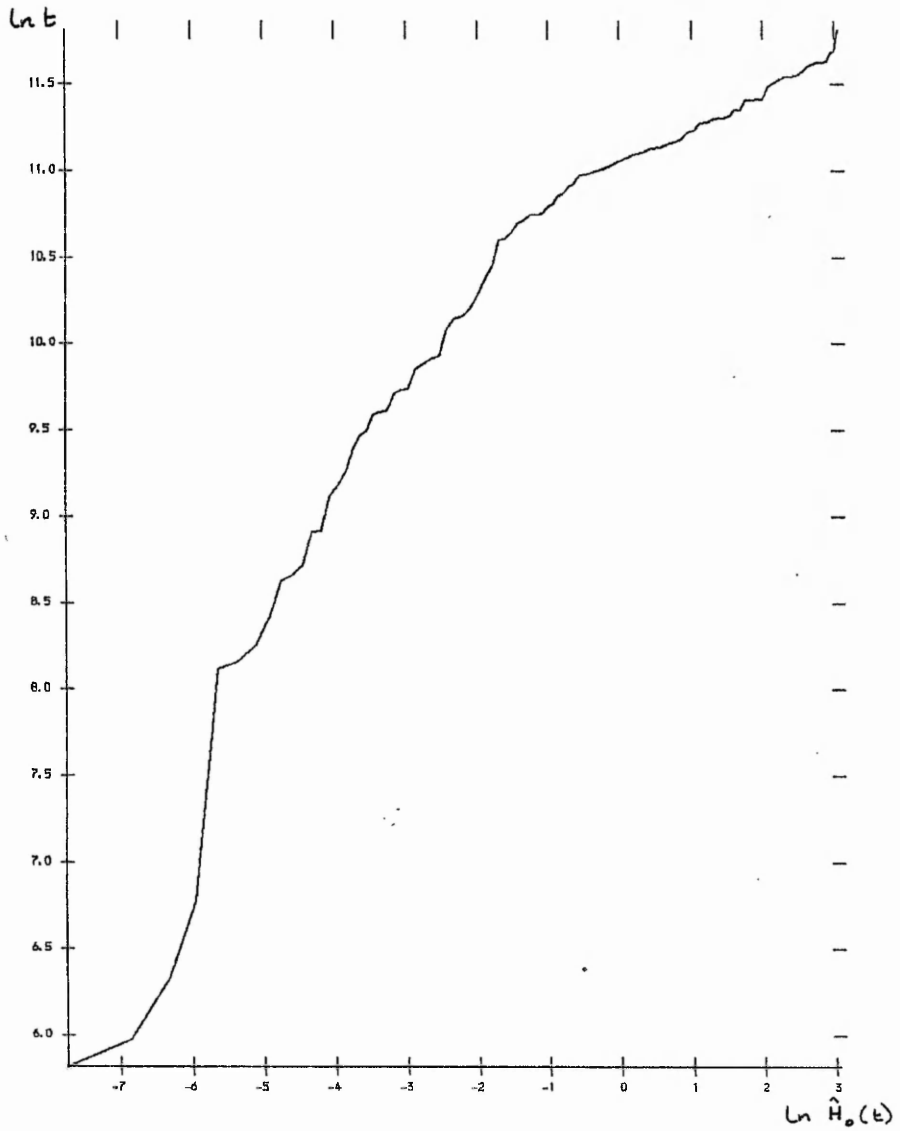
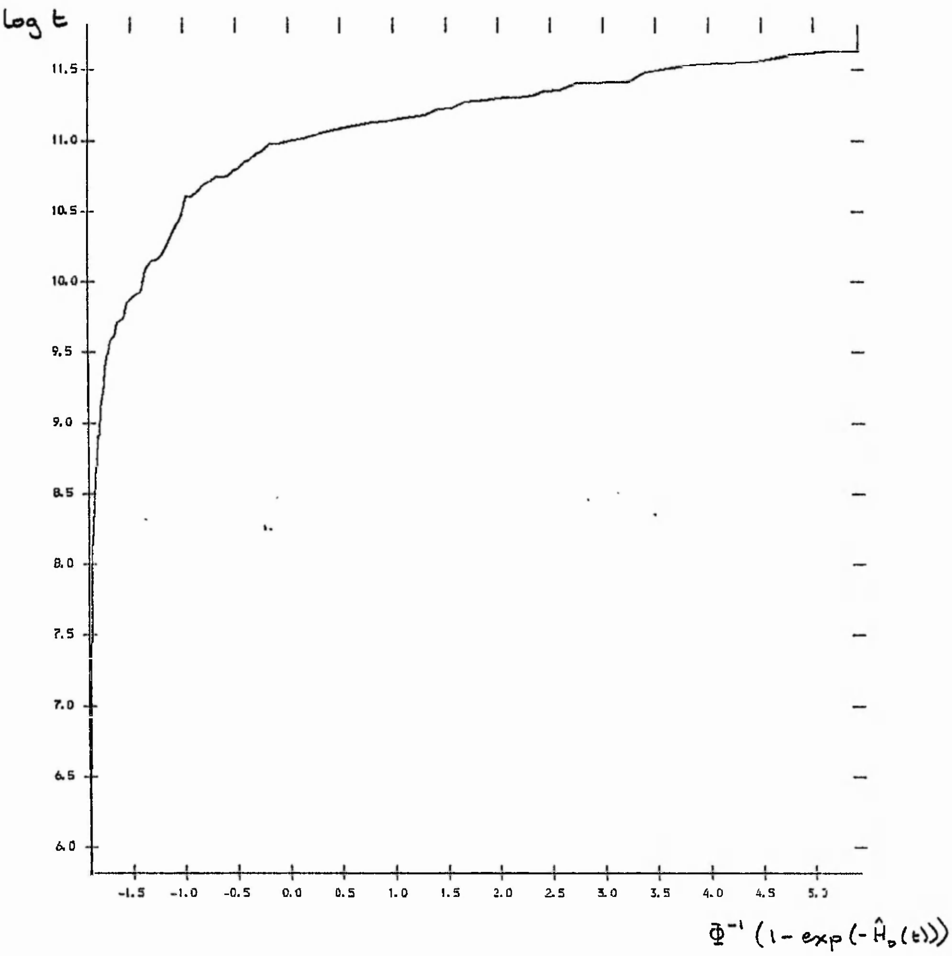


Figure 3.3   Log-Normal baseline hazard plot.   Subsystem A.



<u>Figure 3.4</u> Smallest extreme value hazard plot.
Subsystem A.

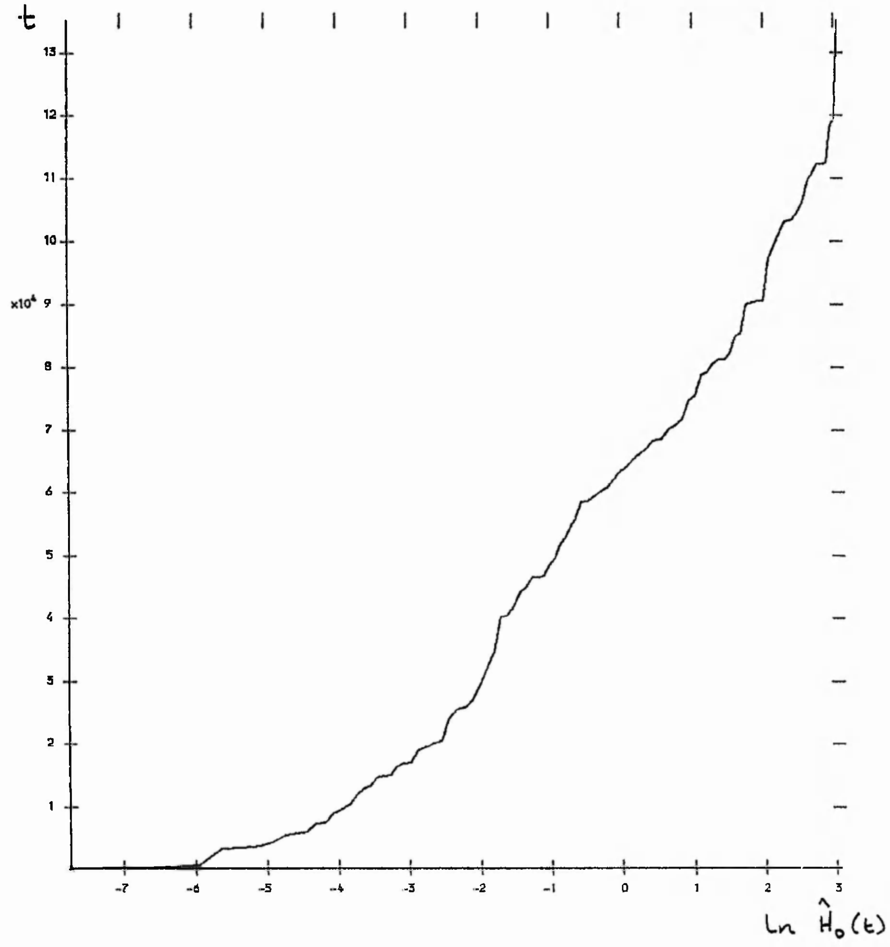


Figure 3.5 Variance-stabilised Cox and Snell type residual plot. Subsystem A.

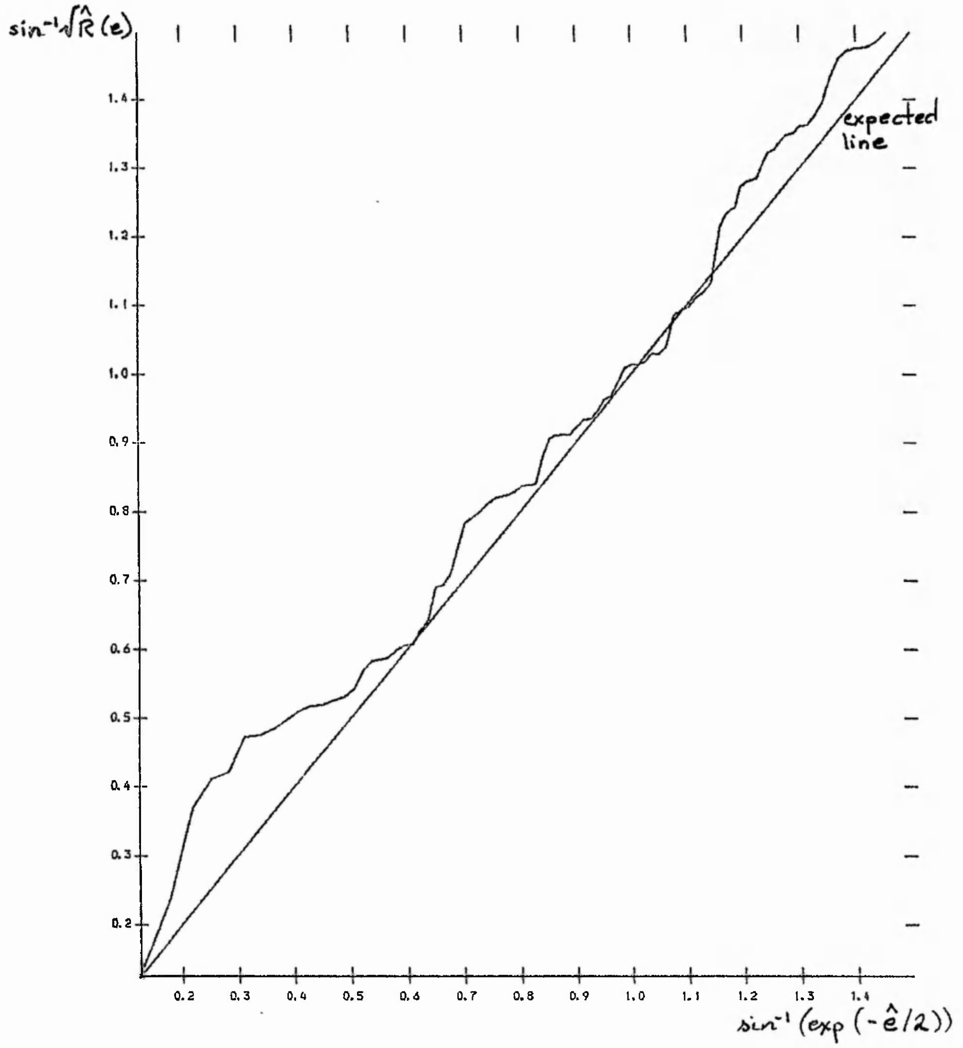


Figure 3.6 Proportionality plot. Subsystem A. Covariate z_4 – 'environment 4.'

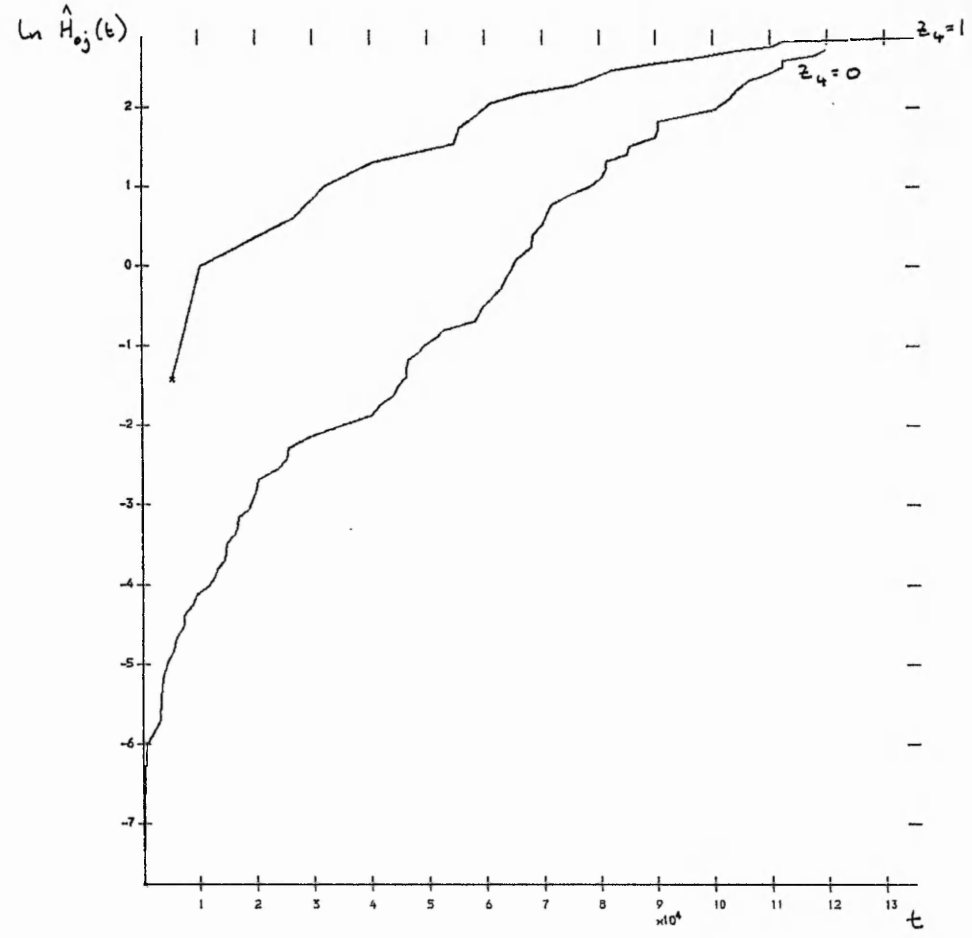
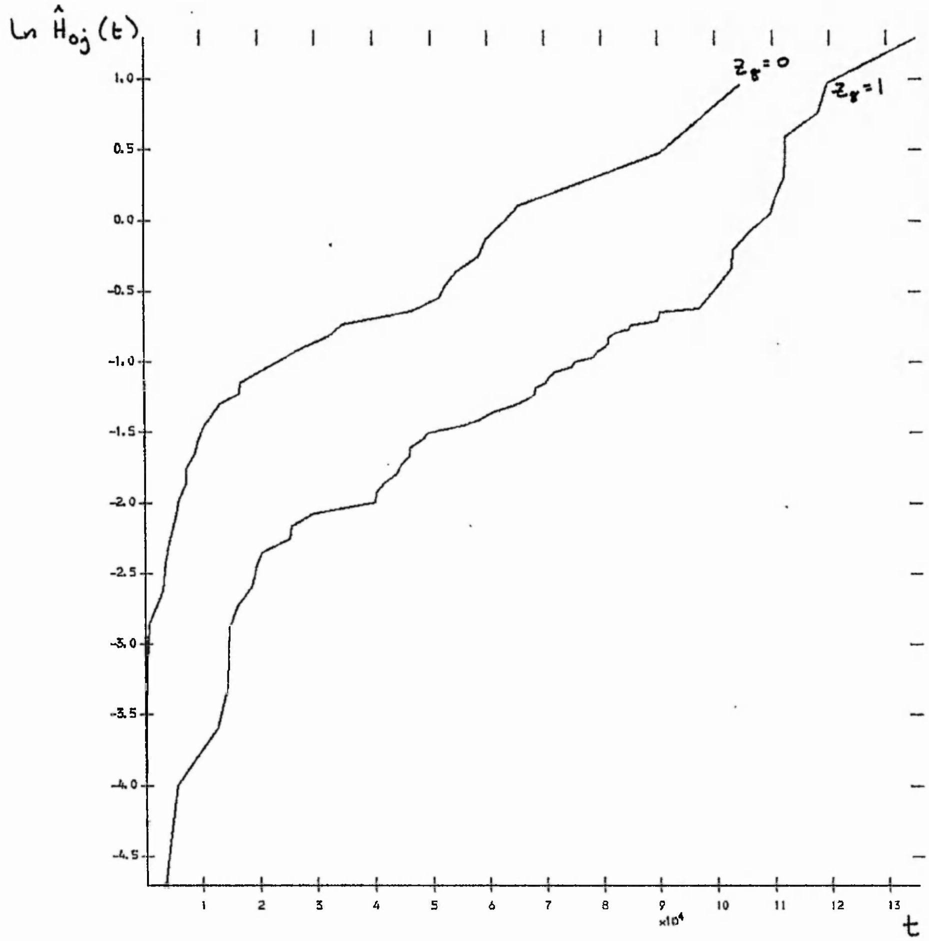


Figure 3.7 Proportionality plot. Subsystem A. Covariate z_8 – 'First failures.'



<u>Figure 3.8</u> Proportionality plot. Subsystem A.
Covariate z_{11} – 'Later failures.'

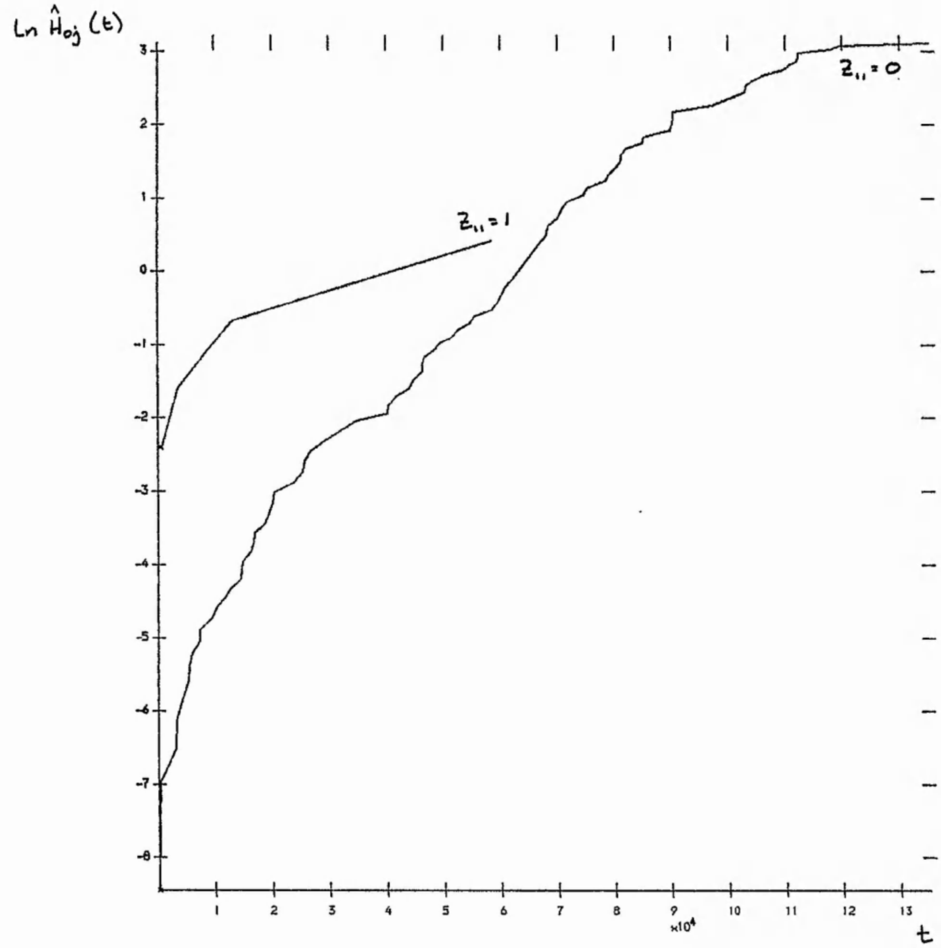


Figure 3.9 Influence plot. Subsystem A.
Covariate z_4 - 'environment 4.'

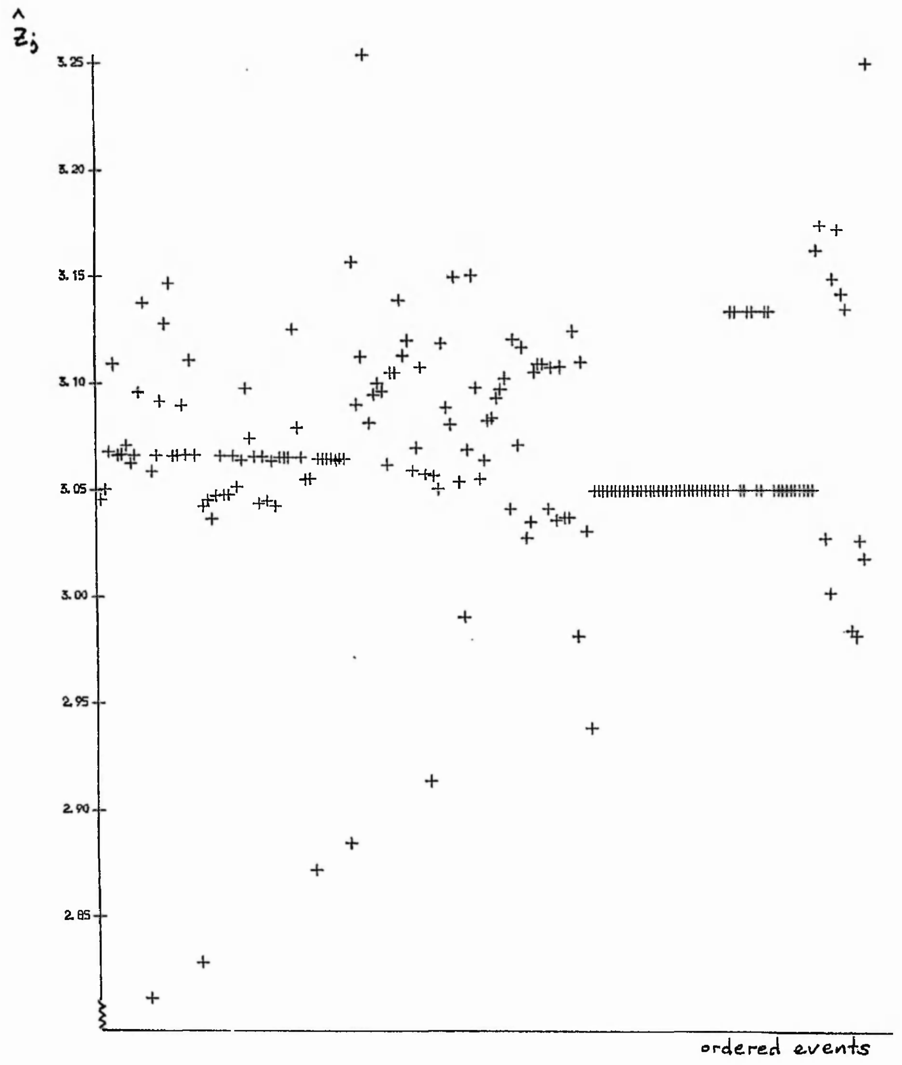
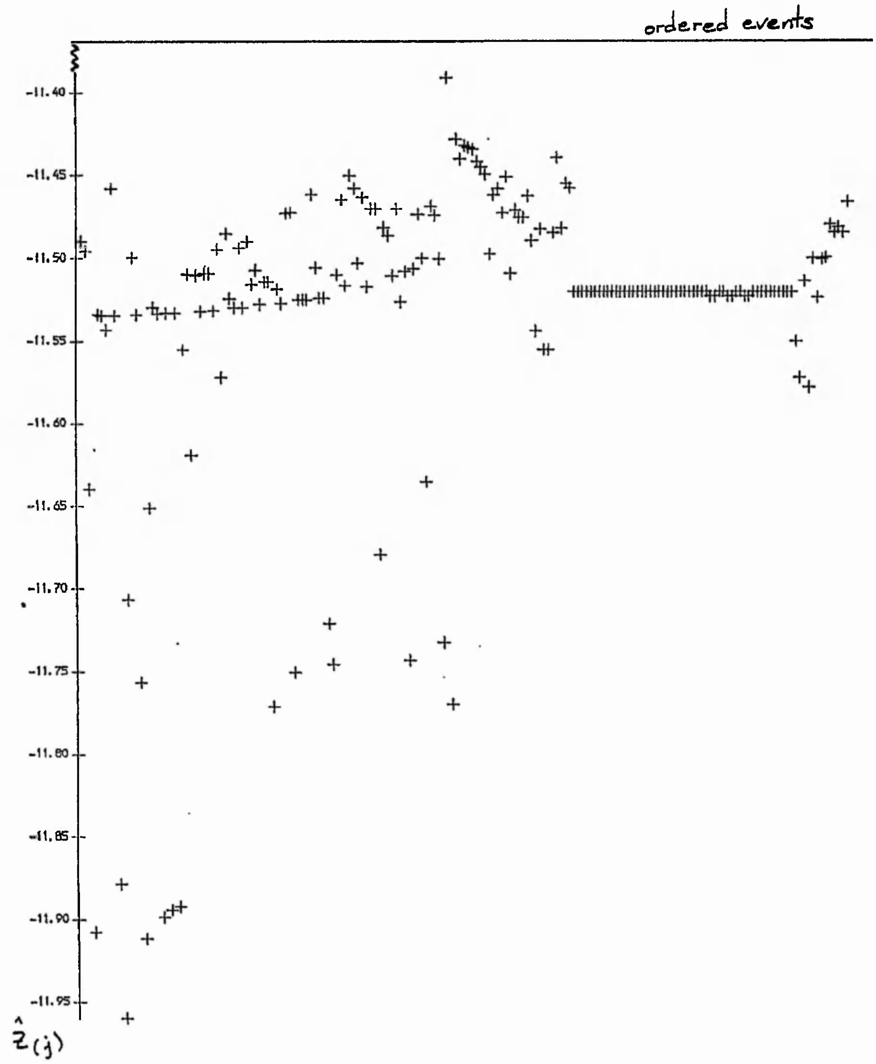


Figure 3.10 Influence plot. Subsystem A.
Covariate z_7 - 'miles travelled.'



<u>Figure 3.11</u> Influence plot. Subsystem A.
Covariate z_8 - 'First failures.'

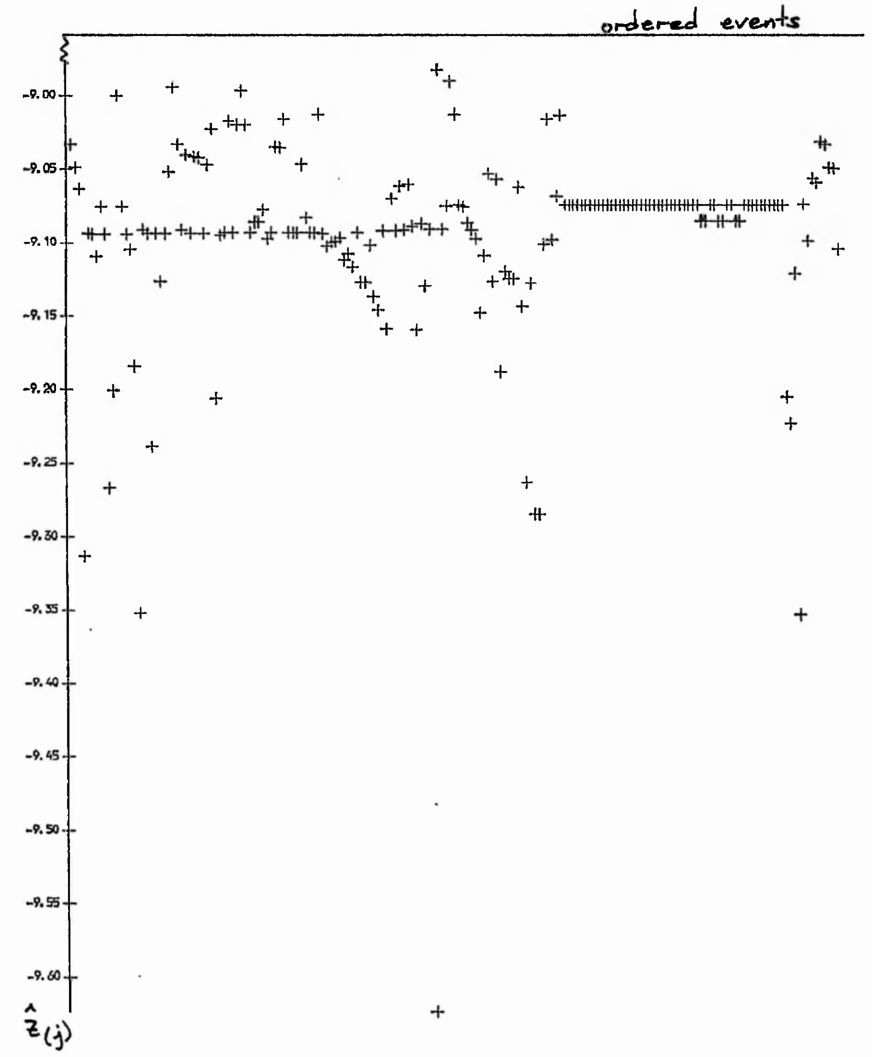


Figure 3.12 Influence plot. Subsystem A.
Covariate z_{11} - 'Later failures.'

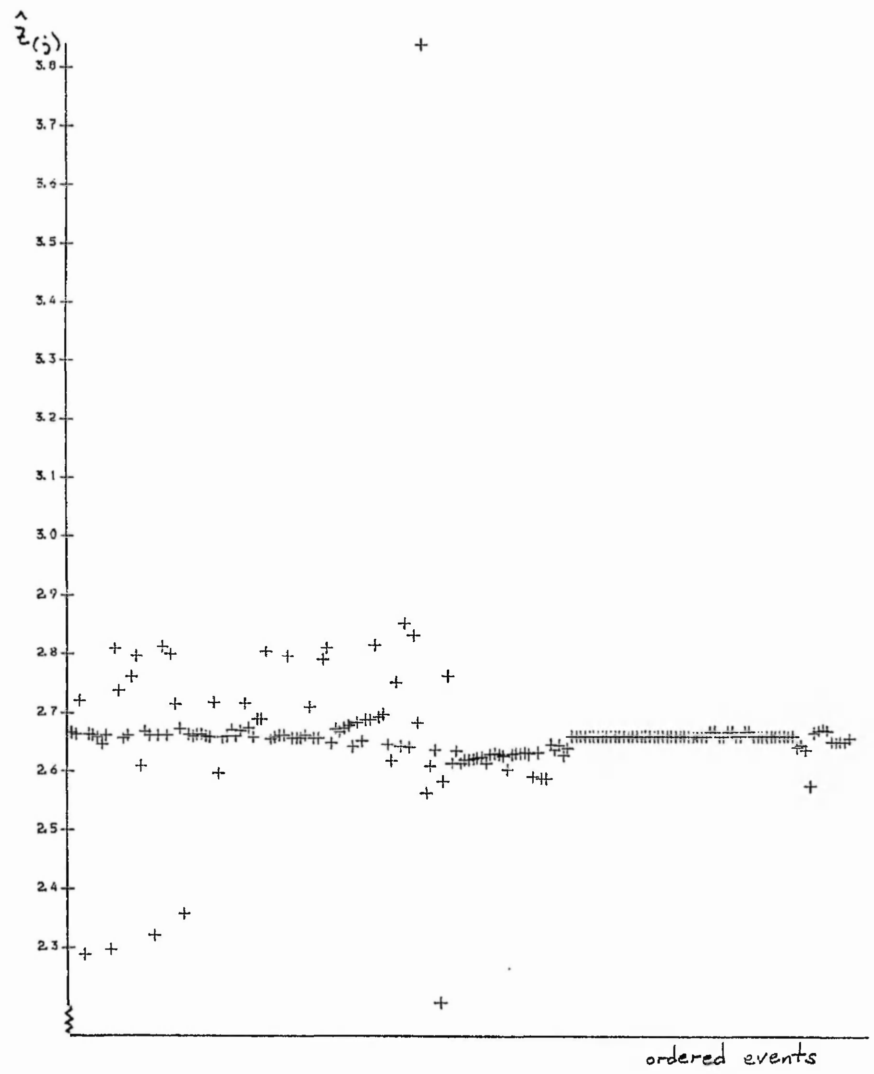


Figure 3.13 Weibull baseline hazard plot. Subsystem B.

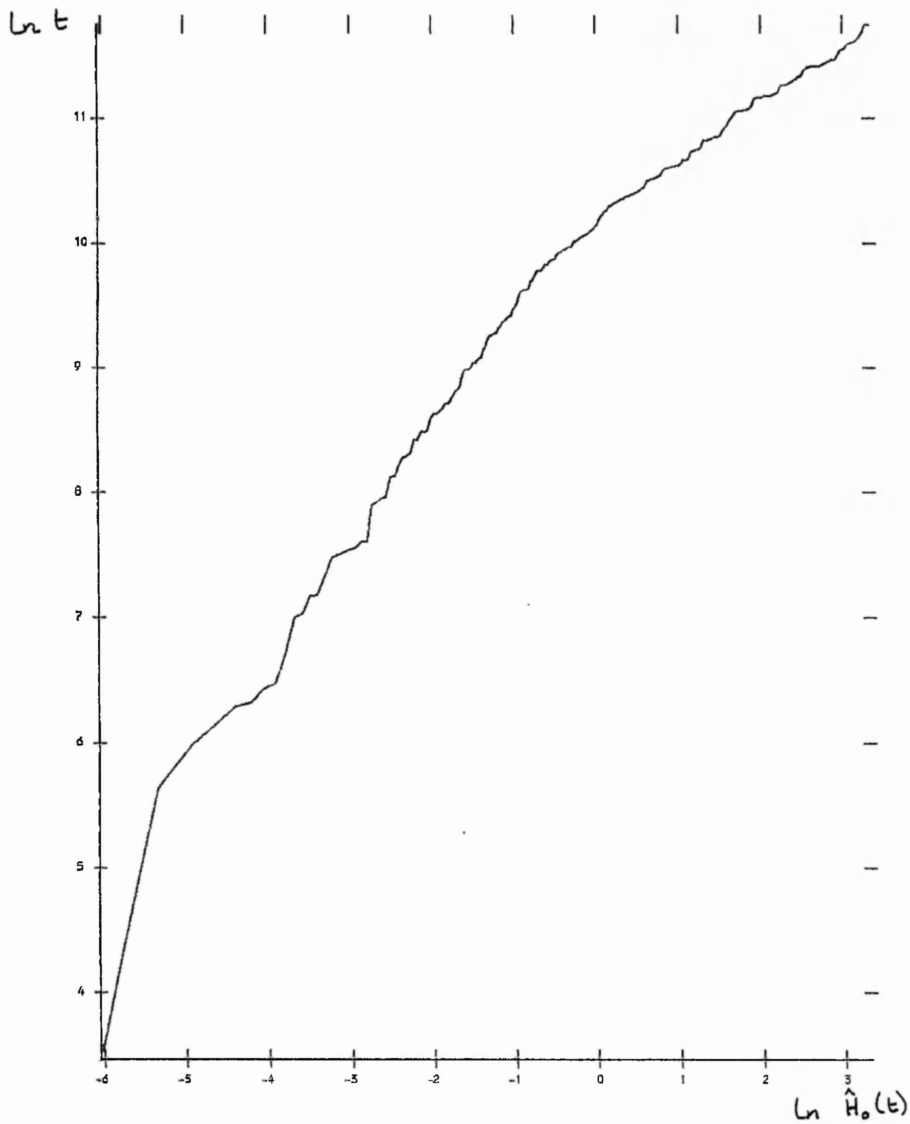


Figure 3.14   Log-Normal baseline hazard plot. Subsystem B.

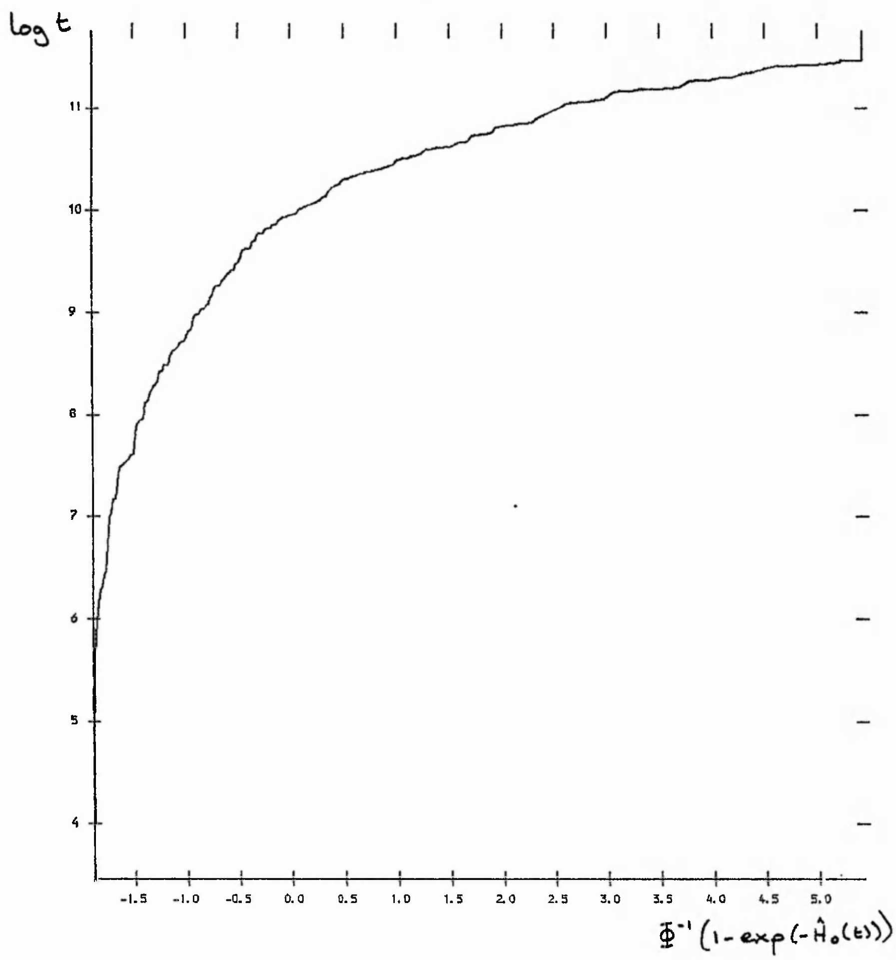


Figure 3.15 Smallest extreme value hazard plot. Subsystem B.

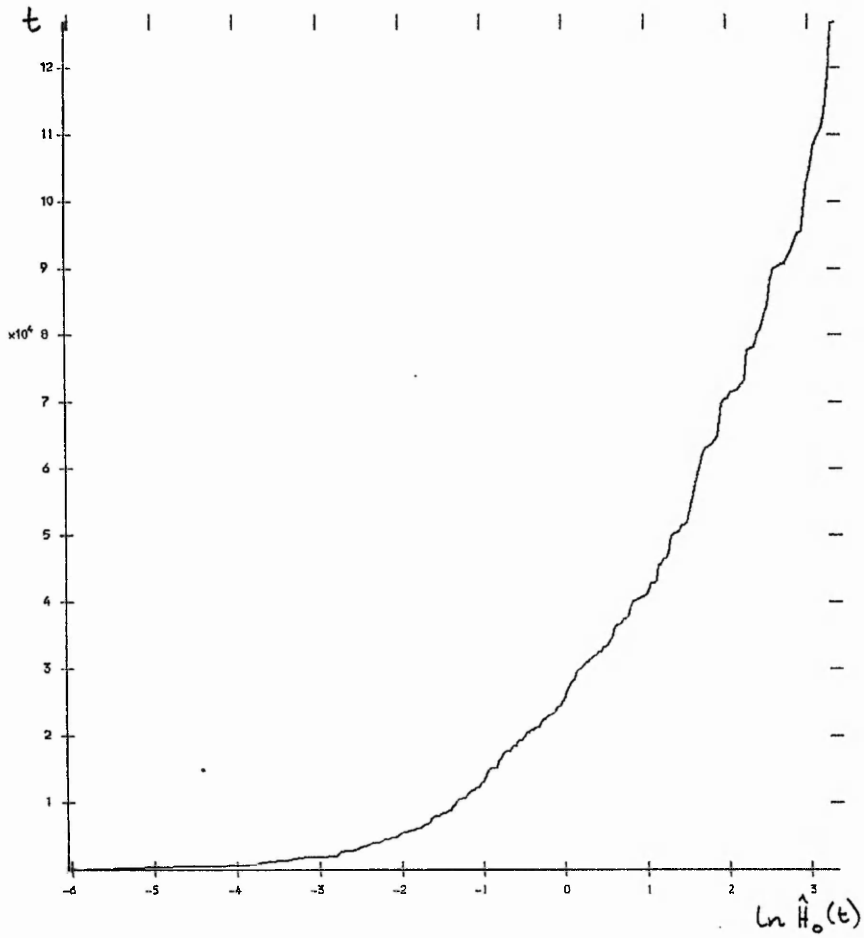


Figure 6.13 Variance-stabilised cox and Snell type residual plot. Subsystem B.

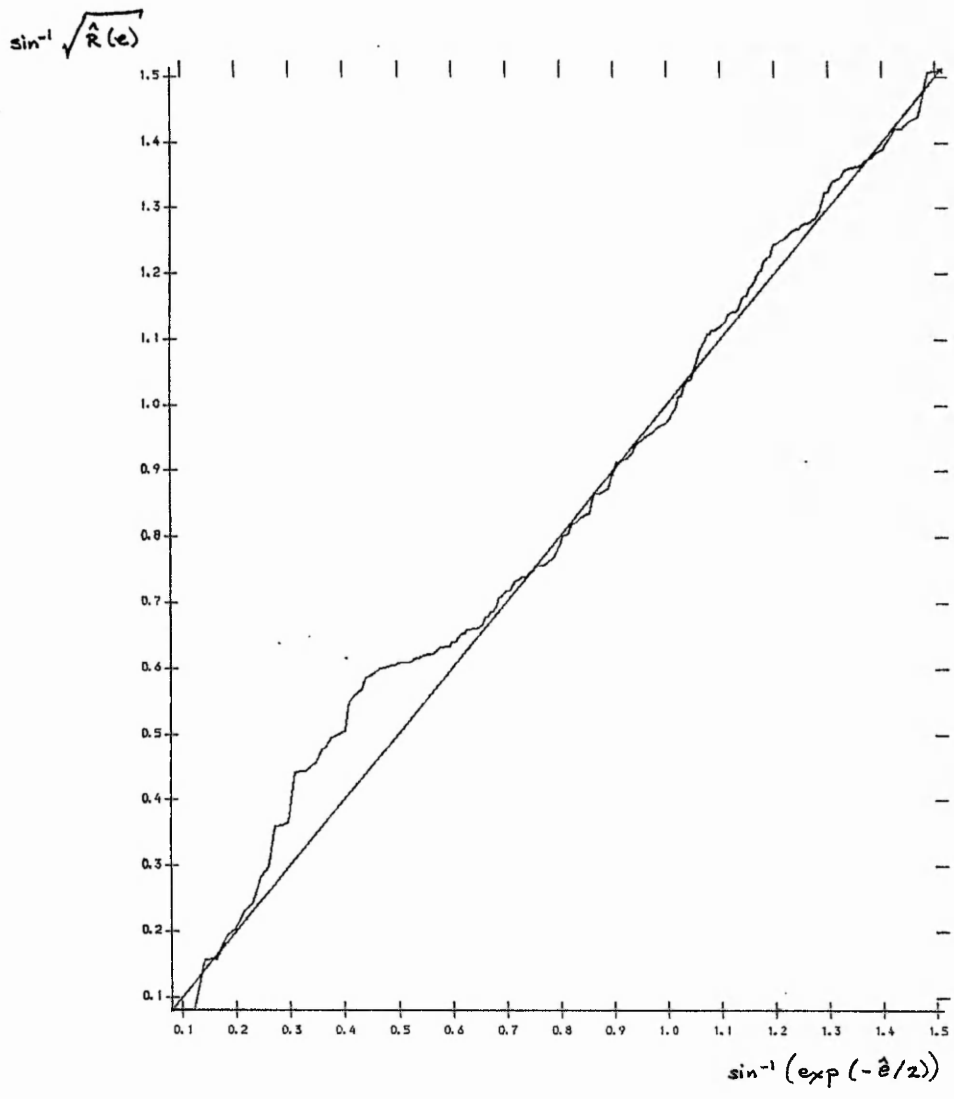
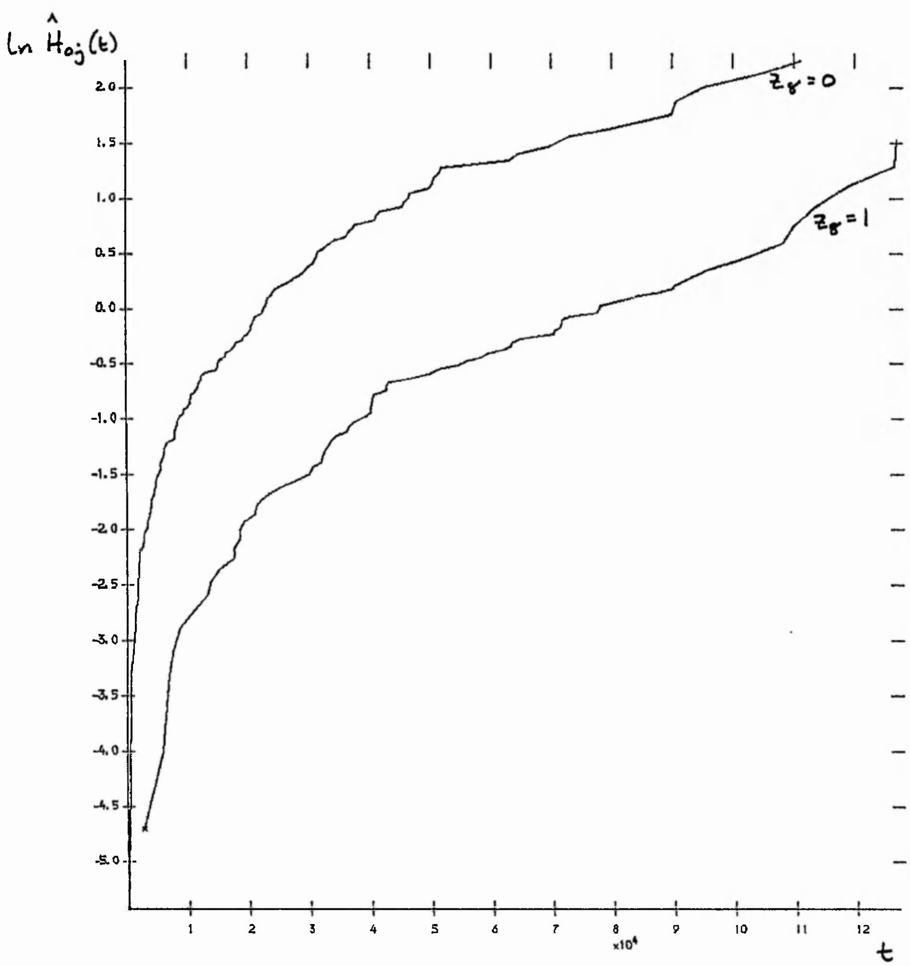


Figure 3.17 Proportionality plot. Subsystem B. Covariate z_8 - 'First failures.'



<u>Figure 3.18</u> Proportionality plot. Subsystem B. Covariate z_9 - 'Second failures.'

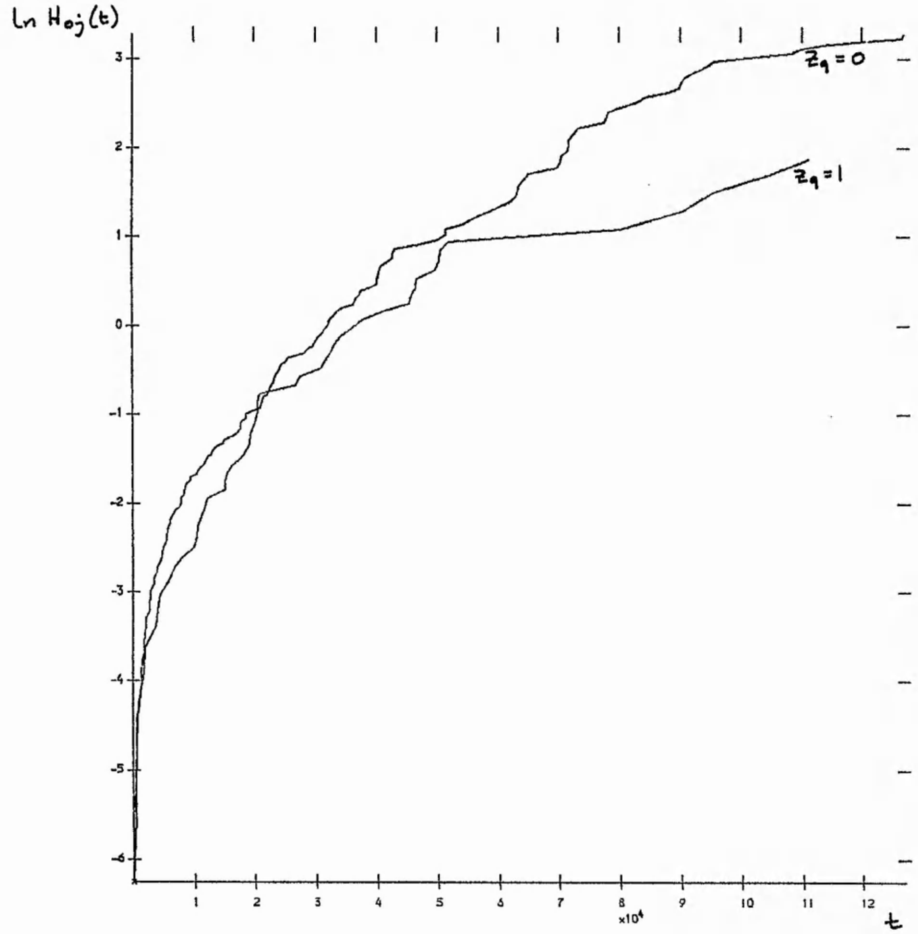


Figure 3.19 Proportionality plot. Subsystem B. Covariate z_{13} - 'Later failures.'

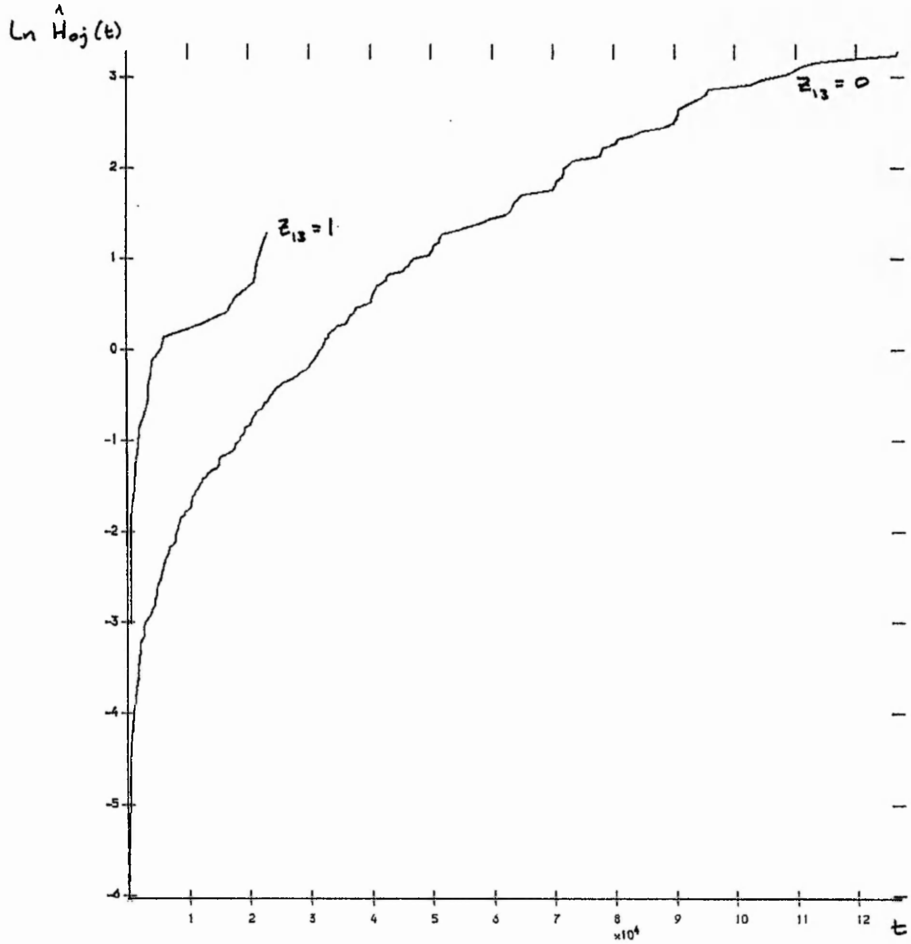


Figure 3.20 Influence plot. Subsystem B. Covariate z_7 – 'miles travelled.'

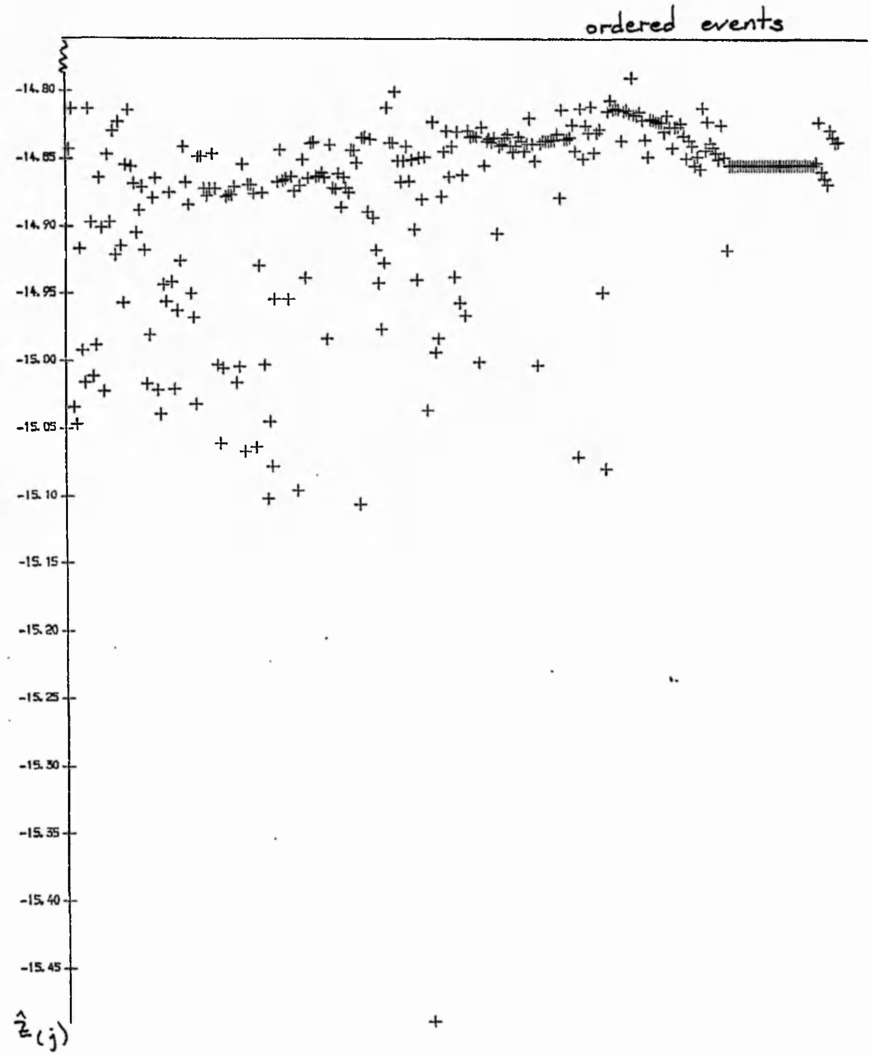


Figure 3.21 Influence plot. Subsystem B. Covariate z_8 – 'First failures.'

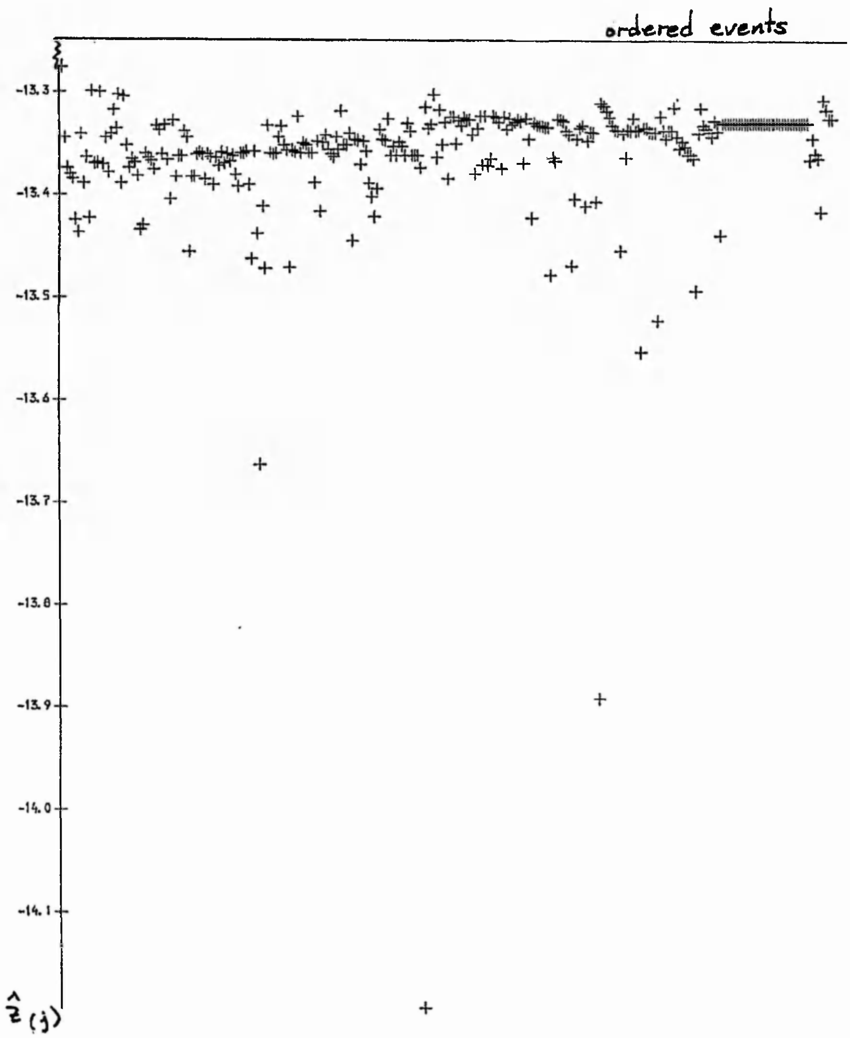
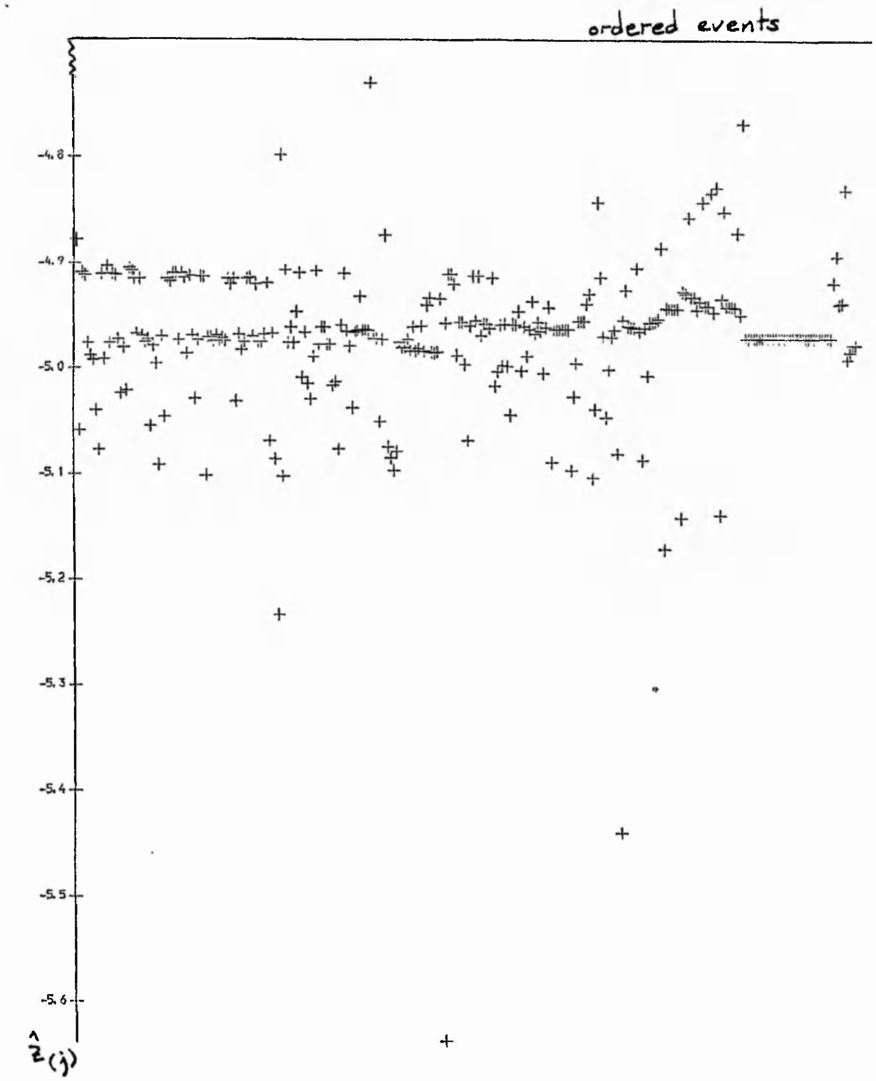
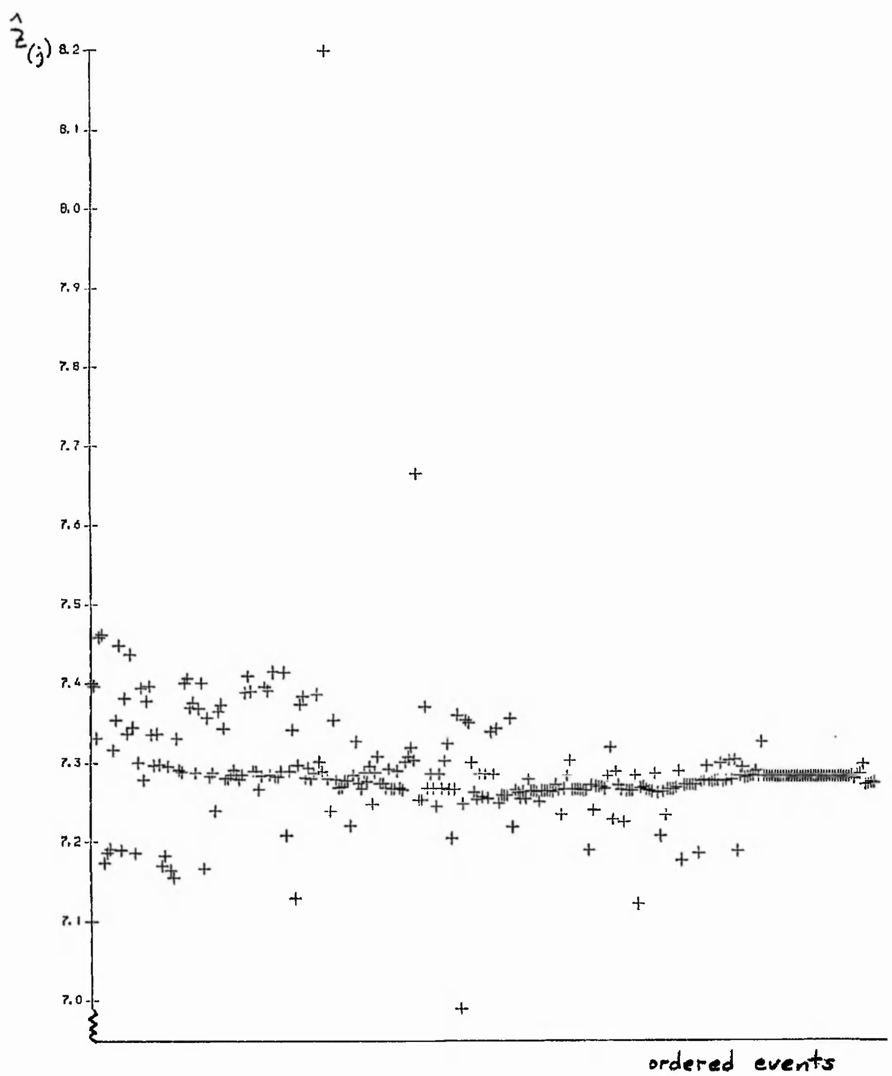


Figure 3.22 Influence plot. Subsystem B.
Covariate z_9 - 'Second failures.'



<u>Figure 3.23</u> Influence plot. Subsystem B.
Covariate z_{13} – 'Later failures.'



CHAPTER 4

APPLICATION TO COMPUTER HARDWARE FAILURES

This chapter shows an example of an application of PHM to computer hardware failure data, (Drury *et al* (1988), Appendix A).

The data came from failure reports of processors within a family of ICL systems.

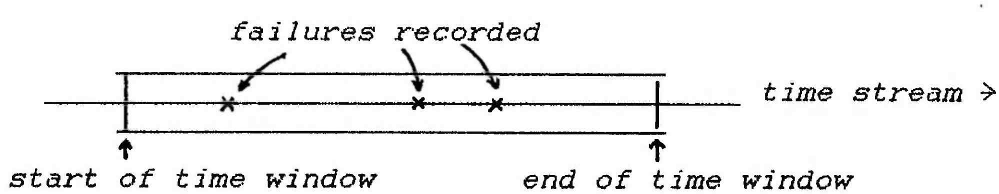
The example shows an approach to a common form of reliability data; 'window structure' where failures have been recorded within a particular time period and the history of failures prior to the start of the 'window' is unknown.

The example also uses the censoring structure for competing risks as discussed in section 2.3.5.

4.1 WINDOW STRUCTURE

This is a structure in reliability data, corresponding to a point process observed within a time window.

A diagram showing the form of a time window over the time stream for an item is shown below.



Data recorded in this manner presents potential problems for analysis since it includes both left and right truncation to observed failure times, at the start and end of the window respectively.

4.2 DATA COLLECTION

Reliability data was collected for processors within a family of computer hardware systems by logging all failures rectified by engineers on a computerised database.

The data was collected within a three month time window; 01/04/84 - 30/06/84, at 364 customer installations.

The records included the information: system identifier, date of failure, date of installation, average usage the system experiences, the number of processors in the system, details of the type of failure the processor experienced, and the quality control system in place at the time of manufacture of the processor(s).

There are four variants of the system, in both single and dual modes; the systems have different numbers of processors depending on the system variant and its operational mode, see Table 4.1.

Table 4.1 Number of processors in each sustem variant.

System type	Number of processors 1	2	4
A	x	x	
B	x		
C	x		
F		x	x

The data provided did not however identify system failures down to the actual failing processor for those systems with multiple processors. It is, therefore, at 'system level' for which processor failures are considered in the modelling of the reliability of the PCB set.

4.3 MODEL STRUCTURE

Since the systems are repairable, a simple starting assumption is that a system is repaired to "as-good-as-new". Therefore, the basic time metric, t, is taken as the time between processor failures occurring on the same system.

We deal with the left truncated events by assuming that such a truncated time to first failure follows the same distribution as subsequent times between failures.

Such a procedure is less wasteful of data than ignoring left truncated times, and would be sound in the case of an exponential baseline.

Treating left truncated events in the above manner is facilitated be measuring the time between the start of the window and the point of first failure, and setting an indicator covariate to show the presence of left truncation.

Since we know from codings in the data that there are a number of different types of failure that a processor can experience, we can use a censoring procedure for competing risks (see section 2.3.5).

Two large groups of events could be identified from the raw data corresponding to failure modes 'No Fault Found' or NFF, and 'component type 272'. Since there are no other failure modes nearly as prevalent all other failure types are grouped together, in this analysis, to form a third failure type 'other'.

We can assume that the three groups of failure mode act as competing risks to the processors.

From the assumed failure mechanism, it follows that a failure identified as any one of the three types, also creates censored events on the system for the two other failure modes.

4.3.1 EXPLANATORY VARIABLES

Covariates introduced into the PHM model were:

Event. An indicator variable, z_1, showing the presence '0' or absence '1' of left truncation.

Age. This is a measured covariate, z_2, giving the age of the system (in months), at the time of failure. Since we are not provided with the exact date of installation the age of the system is calculated from the first day of the month of installation. Any significance of this covariate will show that there is either reliability growth or reliability decay as the system ages.

Average hours per week. The average number of hours use that the system experiences whilst within the time window is taken from actual field returns, and included as the covariate z_3 in the model.

Failure variables. These dummy variables set as in Table 4.2 compare the three different processor failure modes. The failure mode '272' is taken to be the baseline, as this group has the largest number of observations.

Table 4.2 Coding of failure type.

Failure type	Covariate z_4	Covariate z_5
272 (base)	0	0
NFF	0	1
OTHER	1	0

System variables. These dummy variables set as in Table 4.3 indicate the system variant. System variant A is taken to be the baseline since there are more systems of this variant under observation.

Table 4.3 Coding of system type.

System variant	Covariates z_6	z_7	z_8
A (base)	0	0	0
B	0	0	1
C	0	1	0
F	1	0	0

Number of processors. Dummy variables are used here to compare the effect of the number of processors in a system to avoid assuming linear effects, as would be the case if introducing a single covariate taking the value 1, 2 or 4.

The dummy variables are set as in Table 4.4. 1 processor only is taken to be the baseline as most system variants have this configuration.

Table 4.4 Coding for number of processors.

No. of processors	Covariates z_9	z_{10}
1 (base)	0	0
2	0	1
4	1	0

Quality control. Since the start of manufacture a major change in quality control for the processors was introduced. Table 4.5 shows the coding of dummy variables introduced into the model to indicate the quality control to which the processors within the system have been subject.

Table 4.5 Coding of quality control.

Quality control	Covariates z_{11}	z_{12}
Pre-change (base)	0	0
Post-change	0	1
Combination	1	0

Pre-change was taken on the baseline as this was applicable to the majority of the systems.

4.4 SELECTION OF PHM MODEL

As suggested should be the procedure from the example in Chapter 3, the model described in section 4.3 evolved by applying PHM in an exploratory manner.

Within this evolutionary process a number of different structures were considered. Because of technical problems such as multicollinearity and monotonic likelihood (see section 2.2.3 and Bryson and Johnson (1981)) some of these alternative structures had to be neglected or adjusted.

For example, multicollinearity between two covariates; cumulative hours usage to failure, and age of system at failure occurred and was subsequently explained by discussions with the data providers from ICL. The covariate for previous use was then ommitted from the model. Also a covariate identifying the 'power' of the system was found to be a direct linear combination of the dummy variables representing the system variant, also discovered by discussion with ICL, hence the covariate 'power' was eliminated.

In other cases some variables had their specification altered to provide greater information. For example, due to the presence of a few 'young' systems, the redefinition of 'age', from being calculated since the start of the time window to being calculated since installation, was deemed to give more physical explanation.

4.5 RESULTS

The results from the described model, after the usual backwards stepwise elimination procedure, whereby non-significant covariates are excluded one at a time and the model is rerun until all covarites are significant, are given in Table 4.6.

The likelihood ratio statistic, L, is seen to exceed the upper 5% critical value for a χ^2 distribution with nine degrees of freedom, indicating a highly significant model.

Table 4.6 Final significant model after backwards stepwise elimination.

Significant covariates	$\hat{\beta}_i$	p-value	L
z_1 - event	0.6384	0.0001	
z_2 - age	-0.0336	0.0019	
z_3 - av. hrs./wk.	0.0064	0.0030	
z_5 - NFF	0.8345	0.0000	
z_6 - system F	-0.8748	0.0155	99.450 (16.92)
z_8 - system B	-0.7079	0.0083	
z_9 - 4 procs.	1.5974	0.0010	
z_{10} - 2 procs.	0.6417	0.0050	
z_{12} - post-change	-0.6622	0.0015	

4.5.1 SIGNIFICANT COVARIATES

Event. The positive estimate of β_1 suggests that the systems experience a lower hazard during times to left truncated events than they experience during times between subsequent failures. This corresponds to observations of longer left truncated events than time between failures. This is a somewhat counter-intuitive result, since we might have expected left-truncated events to be shorter than full times between failures.

The phenomenon may, however, be explained in three ways:

Firstly, it occurs as a result of an inherent bias in the model caused by the nature of the time window; 71% of the systems under observation had no processor failures within the time window, hence a large number of long right censored left truncated events are compared in the model with the necessarily shorter observed times between failures.

Secondly, as we will see later, the distributional form for the tbf's is found to have a decreasing hazard. Left-truncation, then, may be excpected to increase residual life length.

Finally,as proposed by the data providers bunching of failures, due for example to misdiagnosis, may expain the observed longer times to first failure.

Age. The negative estimate of β_2 indicates a lower hazard experienced by the system at higher ages. Hence, the systems are experiencing reliability growth with increasing times between failures as they get older.

Average hours per week. The sign of the estimate of β_3 indicates that the more use the system experiences per week, the higher the hazard.

The basic time metric in the model is taken from calender time. Although an increase in the usage of the system may result in fewer number of days (or less calender time) between failues, it could nevertheless be possible that more computer time has been used between failures.

Failure variables. The elimination of the covariate z_4 indicates that there is no significant difference between failure modes '272' and 'OTHER'. The significance and sign of $\hat{\beta}_5$ which, after the elimination of the covariate z_4, compares the hazard due to the failure mode 'NFF' with that for the two other modes, indicates shorter times to this mode of failure.

The 'No fault found' failure catagory is most prevalent in the data. This is common to many reliability data sets for which failure mode is recorded.

System variables. The covariate z_7 is eliminated indicating that there is no significant difference between the effects of system variant C and the base system variant A.

Systems of variant F are seen, from the estimate of β_6. to experience a lower hazard than variants A and C. Similarly system variants B experience a smaller hazard.

Number of processors. From the significance of the covariate z_9 it can be seen that systems with four processors experience a higher hazard than those with just one. Similarly, from β_{10}, systems with two processors experience a higher hazard than those on the baseline.

Quality control. The elimination of the covariate z_{11} shows the effect of the combination not to be significantly different to the effect of pre-change processors. However, the systems with processors all included after the change in the quality control procedure are seen to experience a lower hazard than the other systems.

There is some evidence that the new quality control procedures reject bad processors which would previously have been incorporated into a system, since longer times between faiures are associated with processors selected from the new quality control regime.

4.5.2 EFFECT OF SYSTEM VARIANT AND OPERATIONAL MODE COMBINATION

It is of operating importance to know the effects of the combinations of system variant and operational mode. Based on our model the magnitude of the hazard variations can be calculated from the significant β estimates.

The multiplicative effect on the baseline hazard, for each possible combination is shown in Table 4.7.

Table 4.7 Multiples of baseline hazard for system variant/operational mode combinations.

System variant	Number of processors 1	2	4
A	1.00	1.90	–
B	0.49	–	–
C	1.00	–	–
F	–	0.79	2.06

4.5.3 GRAPHICAL DIAGNOSTICS

Some examples of the graphical diagnostic tools are shown in this section. Each technique is described more fully in Chapter 6.

Figure 4.1 is a *Weibull hazard plot* for the baseline hazard estimated from the PHM model. The straightness of the line plotted on these axes indicate that the Weibull is a reasonable distributional form for the underlying process.

The shape and scale parameters for the distribution can be estimated from the plot, and are approximately 0.84 and 1425 days respectively. The shape parameter is less than 1, hence, the baseline shows decreasing hazard.

A variance-stabilised form of the *Cox and Snell residuals* is shown in Figure 4.2. Since the plot lies very close to the expected 45° line the model appears to be a good fit.

Figure 4.3 is the *proportionality plot* obtained for the indicator variable z_1, which shows the presence or absence of left truncation. If the covariate effects the hazard in the assumed manner (proportional hazards) the baseline hazards for the two strata should yeild a constant vertical separation on these axes.
The plot, however, indicates that the assumption of proportional hazards is violated for this covariate. A model stratifying on the covariate may prove to be a worthwhile approach, or since the plots seem to have a reasonable constant separation after about t=30 a time dependent covariate may prove appropriate.

Figures 4.4. and 4.5 show *Schöenfeld partial residuals* plotted against time for the two significant covariates z_2 - age, and z_5 - NFF.

If the assumption of proportional hazards holds the plot of the residuals against t should be scattered about zero for all t. The x's indicate tied points. To ease visual inspection we have applied a moving average based on intervals of 20 tbf's to each graph.

Both plots largely bear out the assumption of proportionlity. and further show no obvious outliers which might have indicated events requiring further investigation.

Figures 4.6 and 4.7 are plots of the influence of individual events on the parameter estimates for the two covariates z_2 - age and z_5 - NFF. They are typical of the plots for naturally measured variables and binary variables respectively. There are no events falling within the range -1.96 and +1.96 which would indicate individual events so influential that their removal from the analysis would alter the significance of the covariate.

4.6 CONCLUSIONS FROM ANALYSES OF COMPUTER HARDWARE FAILURES

This example has shown a PHM model which appears to fit the structure of the data well.

The treatment of left truncated events, by defining a covariate showing its presence or absence, was the least successful aspect of the model.

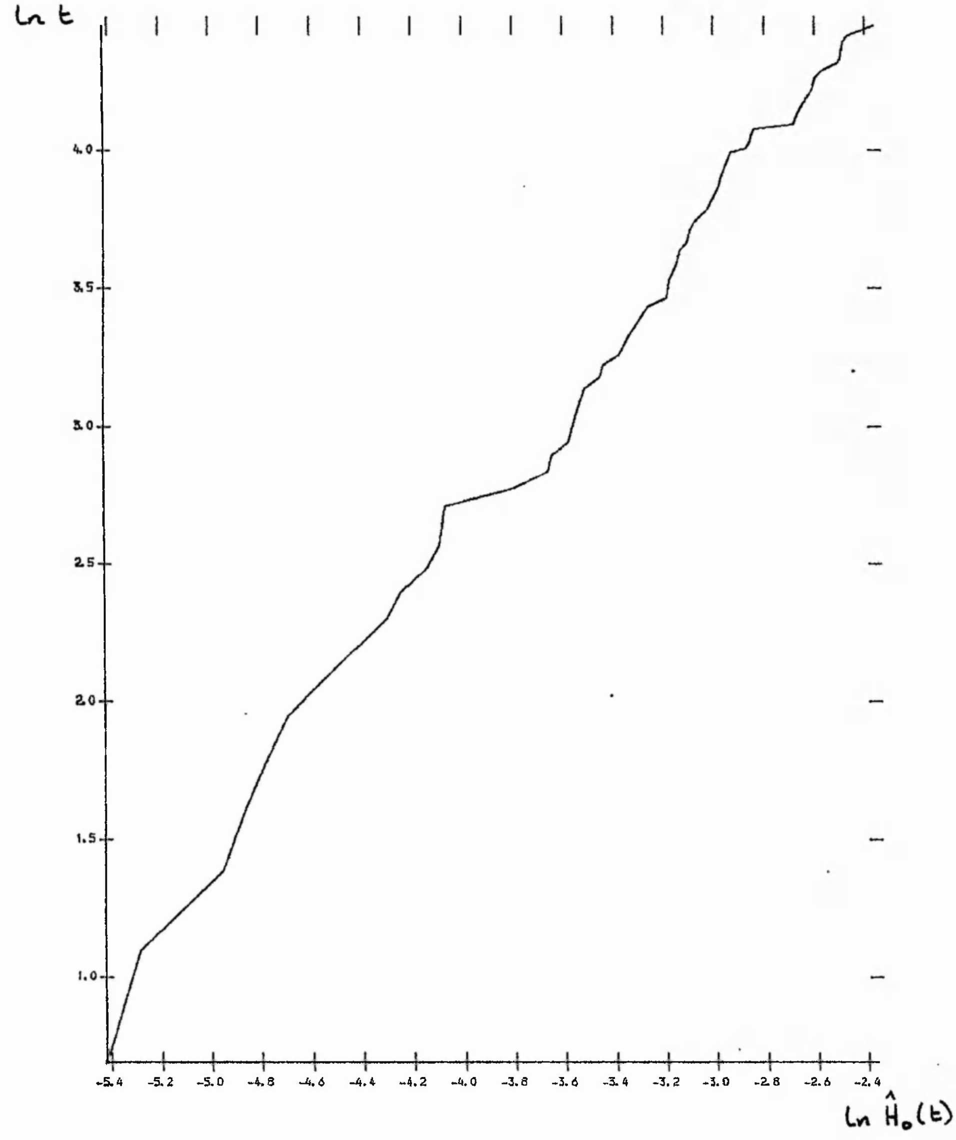
We saw in section 4.5.1 that the significance of the covariate offered a somewhat counter-intuitive result. The bias in the model described in section 4.5.1 may be reduced in examples with a longer duration time window and more observed failures.

There is of course no real reason the believe that the covariate z_1 should effect the hazard multiplicatively, although the treatment would be sound if the tbf's were exponentially distributed.

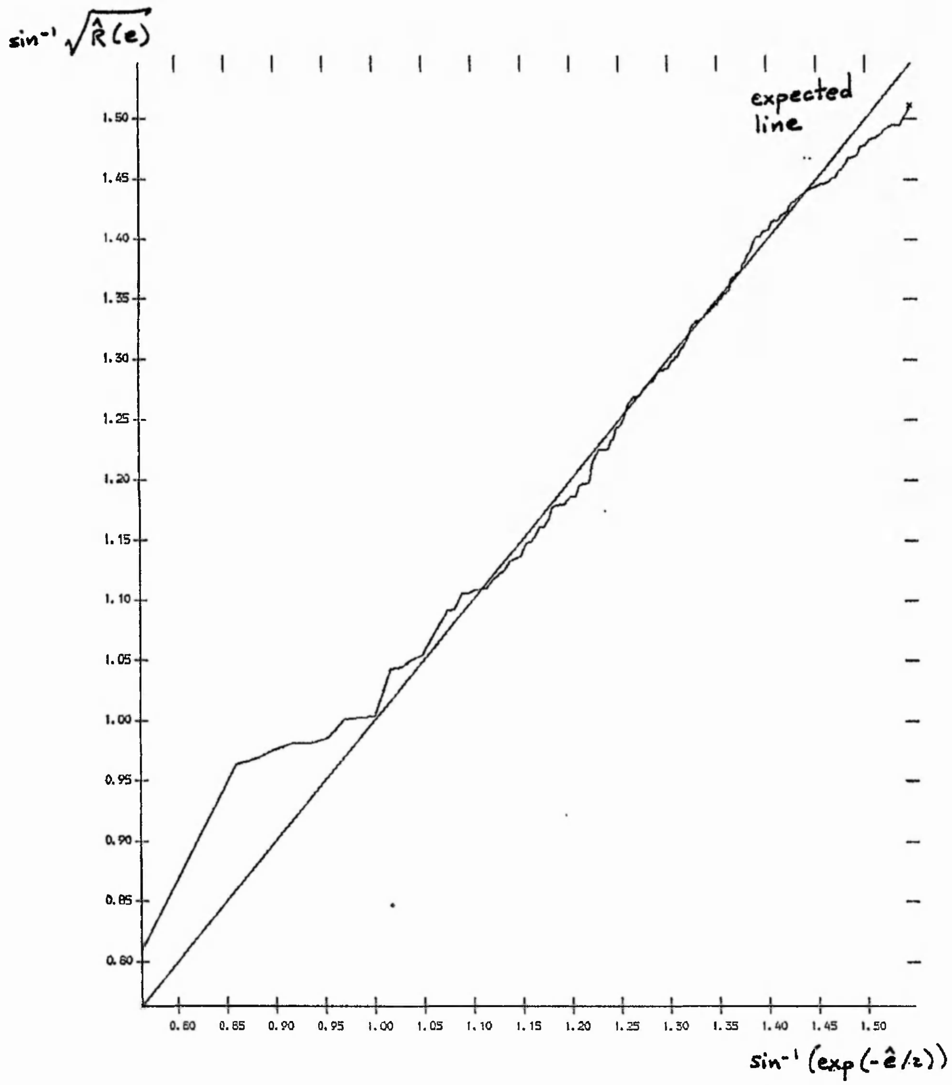
As well as identifying variables having a significant effect on the interfailure times of the systems we have also been able to identify the effects of configuration parameters.

The model has enabled us to identify the underlying structure of the point process, and the distributional form of the baseline.

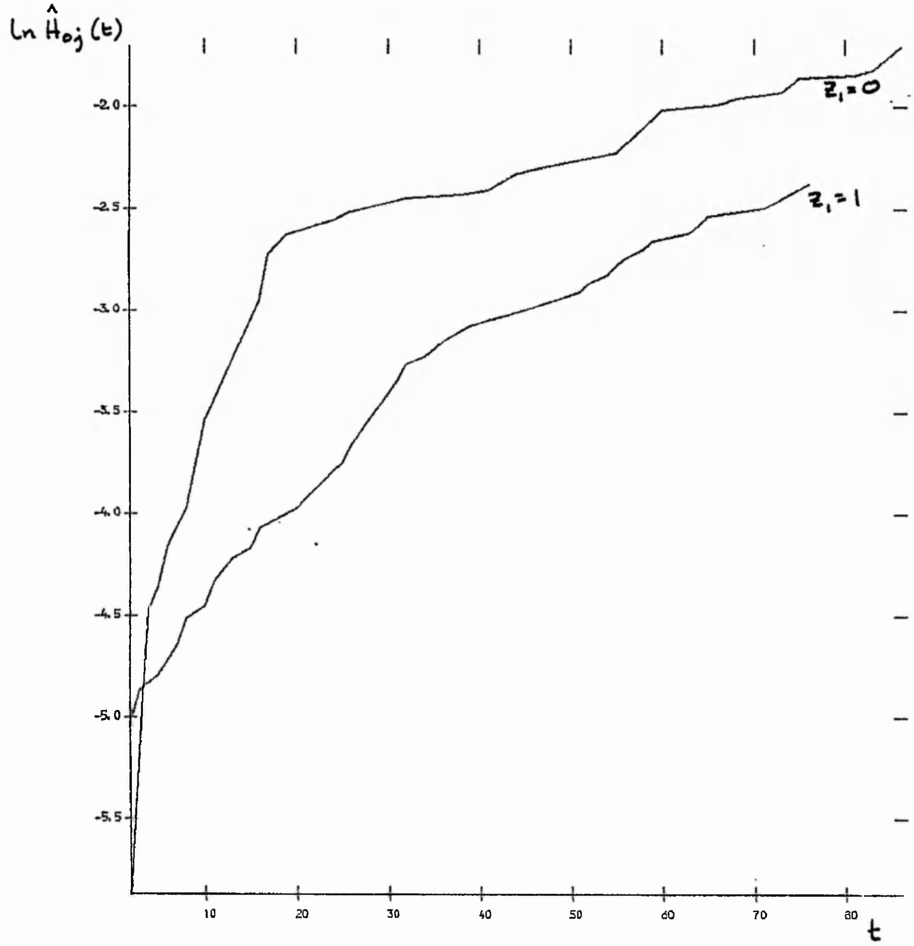
Figure 4.1 Weibull baseline hazard plot.



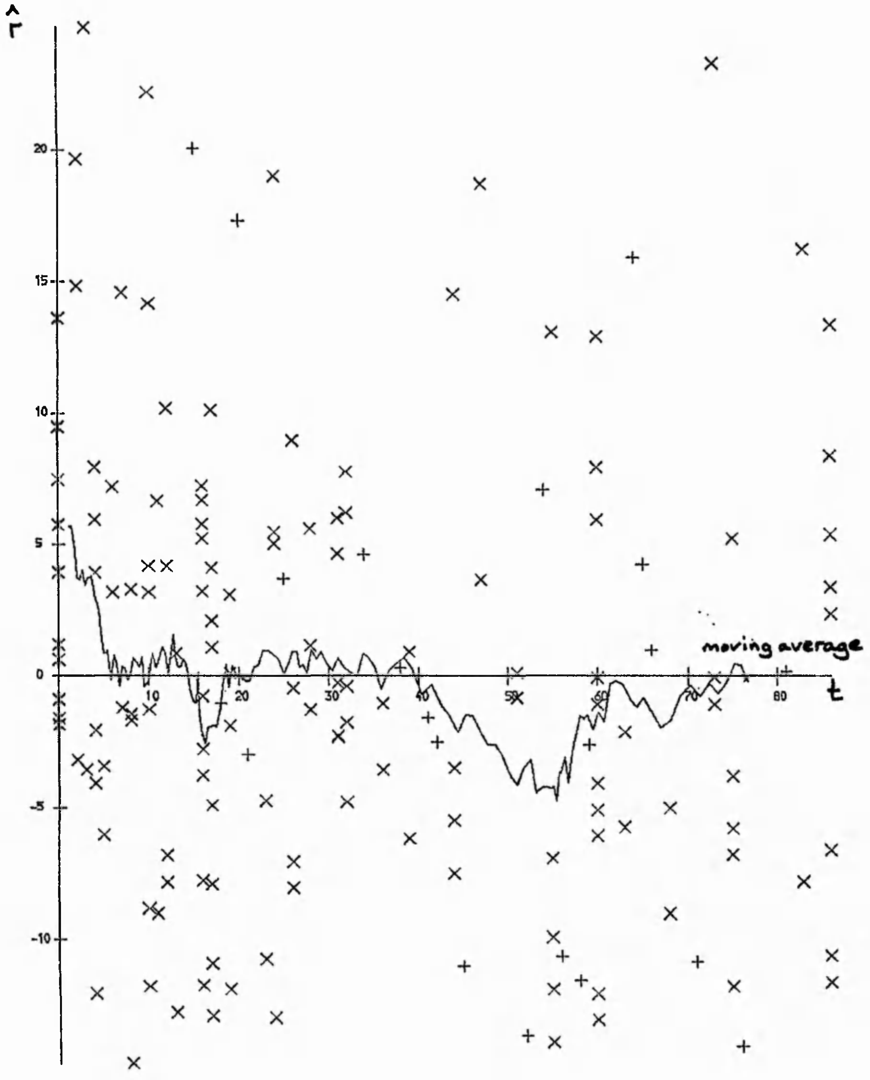
<u>Figure 4.2</u> Variance-stabilised Cox and Snell type residual plot.



<u>Figure 4.3</u> Proportionality plot.
Covariate z_1 - 'Left truncation.'



<u>Figure 4.4</u> Schoenfeld partial residuals. Covariate z_2 – 'age.'



<u>Figure 4.5</u> Schoenfeld partial residuals.
Covariate z_5 – 'No fault found.'

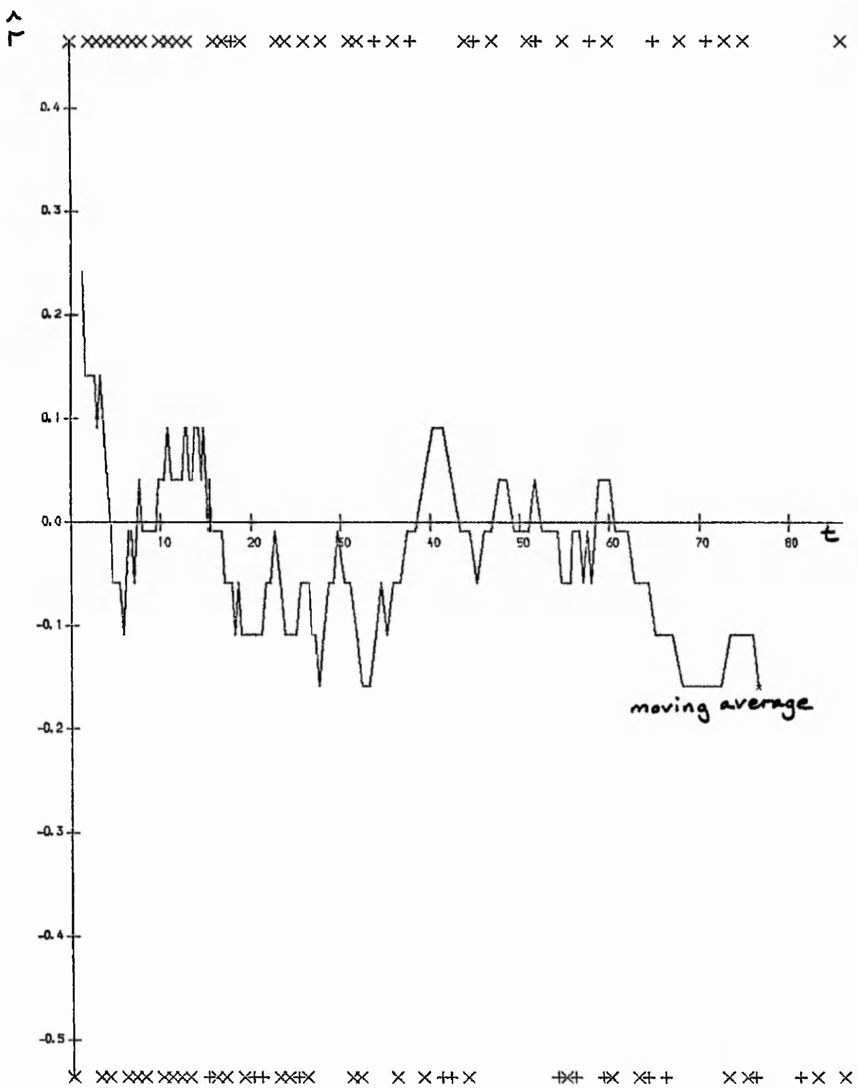


Figure 4.6 Influence plot. Covariate z_2 – 'age.'

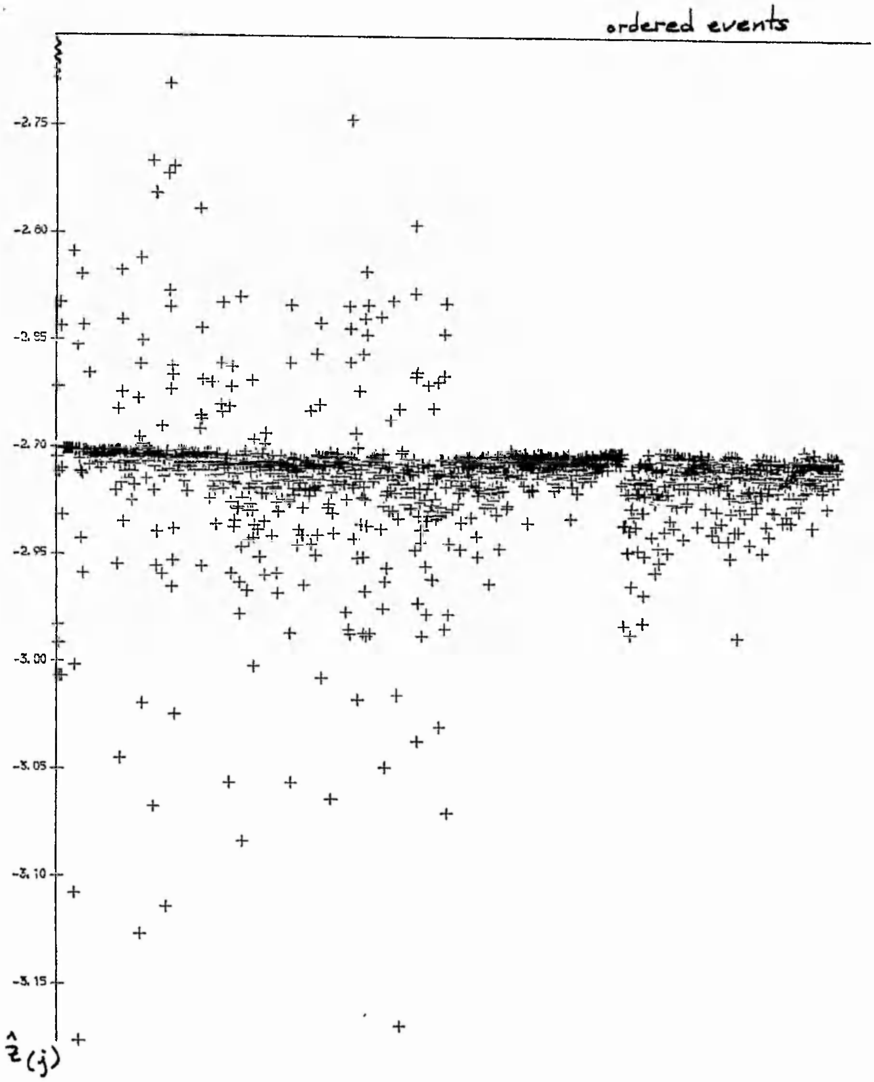
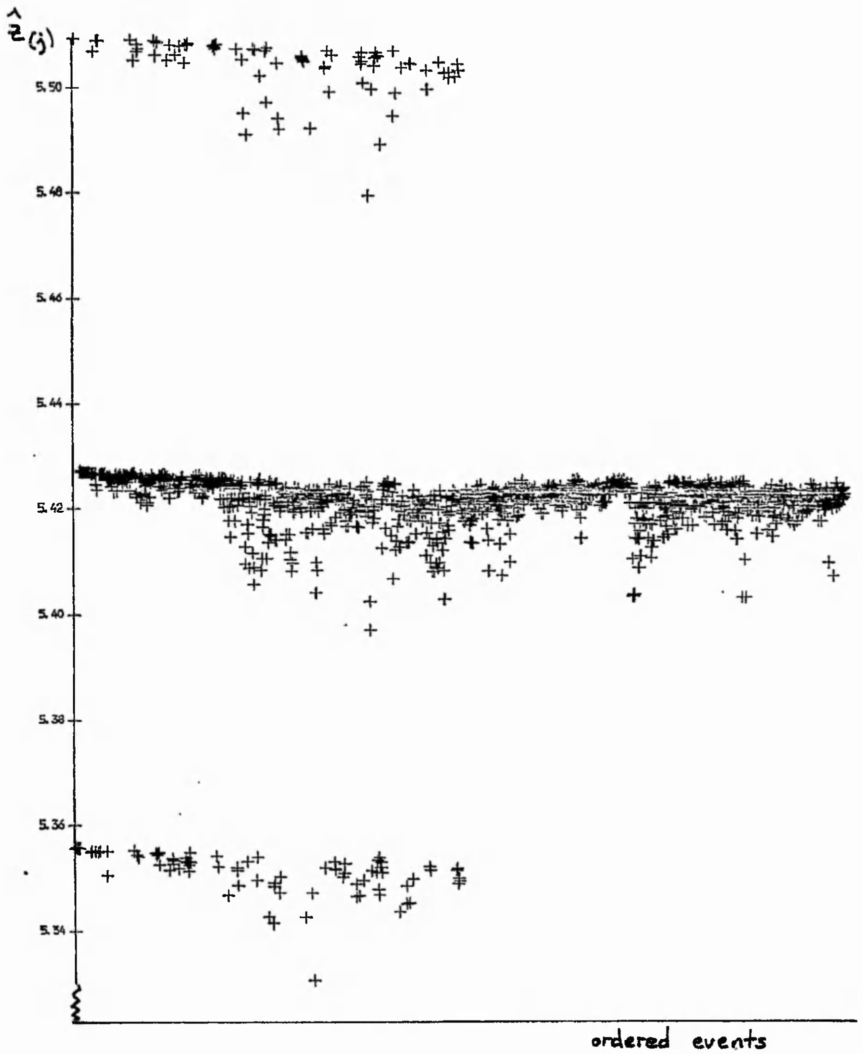


Figure 4.7 Influence plot.
Covariate z_5 - 'No fault found.'



CHAPTER 5

APPLICATION TO A RELIABILITY ANALYSIS OF A CONTEMPORARY WEAPON SYSTEM

Many large complex systems are hierarchical in nature. Failures to such a system may be attributed to any level of the hierarchy; at the lowest level to individual components, or at higher levels to circuit boards, sub-assemblies, or larger modules.

Failure data, therefore, records data at various levels of equipment aggregation, and is itself highly complicated.

The weapon system used as an example in this chapter is such a hierarchical system (Gray *et al* (1988), Appendix B). The analysis of its early field performance is approached through a database whose function was primarily to provide logistic information necessary to the management of equipment spares and resources. It was also hoped it would serve to monitor reliability.

5.1 SYSTEM DESIGN

A complete system comprises subsystems: launcher, optical tracker, generator, and for blind fire systems an additional radar tracker.

The subsystems themselves are of a modular design which is illustrated in Figure 5.1.

Subsystems are interchangeable, so a complete system rarely remains such (with the same individual subsystems) for long: the systems are mobile and are transported frequently.

The susbsytems' modular design is to ease maintenance in a battlefield environment, by combat troops. The principle modules of a subsystem are designed as Line Replacement Units (LRUs). The philosophy of repair by replacement is to ensure maximum operational availability. First line maintenance (on the battlefield) consists of changing a faulty LRU, whilst at second line (in the workshop) the fault will be traced down to the lowest level of the fault, in the hierarchy.

5.1.1 GENERALISED REPRESENTATION OF FAULT STRUCTURE

Using Figure 5.1 which outlines the subsystem structure, and from our knowledge of the maintenance procedures we can construct a generalised representation of the fault structure which is shown in Figure 5.2.

At the lowest level are component faults. Higher in the hierarchy are sub-assembley faults which are the superposition of component faults plus other faults such as interconnection problems. Faults at the next level up are LRU faults which are a further superposition of sub-assembley faults conjugated with other problems that cannot be attributed to sub-assembley or component faults. At the highest level is the subsystem, where we have on this level representation of the total aggregation by superposition of all faults.

5.2 DATA COLLECTION

The database was of a conventional design. The repair technician was required for each failure and repair to complete a descriptive jobcard outlining information such as date, serial number, elapsed time indicator readings (ETIs), and fault classification code.

Information from jobcards was then transcribed into the computer database; building up, in principle, a complete historical record of reliability, maintenance and repair data.

Four files from the database were available detailing: scheduled maintenance, environmental/deployment data, ETI readings, and defect data.

A number of features in the files, however, made data extraction difficult and complicated. For example, inconsistent formatting within the files precluded file merging. Also, free formatting of certain fields in the deployment file made the extraction of information extremely difficult.

Because of the problems described above, this example uses only information readily available in the defect data file, which contains records about faults found, for its analysis.

5.3 STRUCTURE OF DEFECT DATA FILE

Each record of the file refers to an individual fault. The particular subsystem for which data is recorded is identified by its type (i.e. launching unit, optical tracker, etc.) and serial number.

Each record contains fields detailing:

The *date* the fault occurred.

The *holding group* the subsystem was located in at the time of the fault (i.e. a particular squadron or battery). There are six different holding groups.

The type and serial number of the particular *LRU removed* and replaced. If more than one LRU was removed at one time a separate record was generated by each.

The *level in the hierarchy* to which the fault was subsequently traced. A separate record is generated for each identified fault.

ETI reading when the fault occurred. The ETI is a four digit counter which measures the time the unit has spent in normal mode.

5.3.1 ELAPSED TIME INDICATORS

These indicators measure only the elapsed time in normal mode. The subsystems have various levels of operational mode. For example, a radar tracking unit can be operated in modes; high alert,low alert, and normal.

The ETIs themselves are known to be extremely unreliable; often sticking, and sometimes running backwards. The counters are 'throw away' so may be replaced if found to be 'misbehaving'. However, the replacement counter is not necessarily set to the last reading of the discarded counter, or even reset to zero.

The value of the information from the ETIs is, therefore, suspect.

5.4 MODEL FORMULATION

There is a multitude of different PHM model formulations to be chosen from for analysis of any reliabiltiy data. These formulations vary by having different time metrics, censoring structures, and covariates.

The choice of formulations is particularly large when faced with a hierarchical structure such as this, since faults could, theoretically, be followed through at any level.

Initially four formulations of simple, physically plausible PHM models were identified. General covariates that can be included in all the models are; holding group, season the fault occurred, ETI reading, and a time trend.

The four different formulations were identified by considering different possible point processes.

5.4.1 POINT PROCESSES CONSIDERED

Process A. Here the series of events occurring for a particular serial numbered LRU is followed. The basic time metric is taken as the time between the LRU being placed in a subsystem, and its being removed from that same subsystem. The next time between failures (TBF) wil be taken for the same LRU within the next subsystem it enters after repair. Figure 5.3 illustrates this point process.

For this formulation it is necessary to know when the LRU was returned to service, so that its repair time is not included in the TBF calculation.

Covariates could include which particular subsystem the LRU was employed within when the fault occurred.

Process B. In this formulation the position of a certain type of LRU within a particular serial numbered subsystem is followed. The basic time metric here is taken as the time between faults to the same type of LRU in the fixed position of the subsystem. Figure 5.4 illustrates this series of events.

Covariates could include which serial numbered LRU was removed at each fault.

Process C. The types of model employing this point process look at the times to first fault of LRUs ofa given type. This is illustrated in Figure 5.5.

Covariates could include which particular serial numbered subsystem the LRU was in when the fault occurred.

Process D. Models of this formulation look at events at the high level of subsystem. The series of events on a particular serial numbered subsystem is followed. The time metric for this point process is taken as the time between observed faults irrespective of which LRUs are faulty. This series of events is illustrated in Figure 5.6.

Covariates could include which type of LRU is faulty.

5.4.2 NEED FOR CENSORING STRUCTURE

In dealing with the reliability analysis of hierarchical structures, we may like to look at the lifetimes of units at a lower level in the hierarchy than is followed through the basic time metric.

When looking at such events at sub-assembley or lower levels it may be necessary to introduce censoring concepts.

This would be the case in this example since non-faulty sub-assemblies are removed from the field because their parent LRU has been replaced.

Many censoring structures might be identified by interpreting various fault mechanisms in the hierarchical levels. Many of these structures can result in considerable complexity of the model applied.

5.4.2.1 POSSIBLE STRUCTURES

The examples of possible censoring structures described in this section are chosen because they appear reasonable in terms of what is known about the methods of maintenance and fault recording.

In the simplest case consider an LRU with only two levels within its hierarchy. Further, the LRU contains just two sub-assemblies. See Figure 5.7.

There are four failure modes possible for such an LRU, which could be recorded in the data. These are:

Fault recorded to LRU but to neither sub-assembley	(- , -)
Only sub-assembley 1 has fault	(SA1, -)
Only sub-assembley 2 has fault	(- ,SA2)
Both sub-assemblies have faults	(SA1,SA2)

Four censoring structures that might be applicable to such a case are given below, and summarized in Table 5.1.

(i) *Two failure mechanisms.* Only two failure mechanisms [F1] and [F2] are causing the fault in the LRU.

[F1] - only sub-assembley 1 has fault.

[F2] - only sub-assembley 2 has fault.

In such a case (– , –) type faults are ignored as being recording errors. (SA1,SA2) type faults are treated as two independent simultaneous faults; occurring together purely by chance.

[F1] and [F2] are two competing risks, hence a failure of each type also generates a censored event for the other.

(ii) *Three failure mechanisms.* In addition to the two failure mechanisms [F1] and [F2] as above, we could consider a third mechanism [F0] for the case (– , –) where an LRU fails without either sub-assembley being faulty; for example due to a fault in the connectors.

Again, (SA1,SA2) type recorded faults are treated as two independent simultaneous faults.

We now have three competing risks, each mechanism censoring the other two.

(iii) *Four failure mechanisms.* Instead of treating (SA1,SA2) as two independent simultaneous faults, we can define another failure mechanism to explain this mode of failure, [F12], whereby faults of this type are assumed to be simultaneous due to a common external cause; for example a power surge.

There are now four competing risks, accounting for the four possible failure modes in the data, which each censor the other three.

(iv) *Modified four failure mechanisms.* In a particular example of an LRU of the nature assumed in this section, it was found that simultaneous faults to the sub-assembley occurs more frequently than would be expected according to the independent failure model.

We wish then to include a failure mechanism which accounts for simultaneous failure, but unlike [F12] in structure (iii) is not considered a risk competing with [F0], [F1] and [F2]. Hence, a failure mechanism [F12'] can be introduced which itself is never censored, and does not censor the other three mechanisms.

Table 5.1 Effects of various censoring structures.

	Censoring structure			
	(i)		(ii)	
Failure mode	Failure mechanism	Effect on analysis	Failure mechanism	Effect on analysis
(-, -)	-	Ignored	(F0)	censors (F1) and (F2)
(SA1, -)	(F1)	censors (F2)	(F1)	censors (F0) and (F2)
(- ,SA2)	(F2)	censors (F1)	(F2)	censors (F0) and (F1)
(SA1,SA2)	(F1) and (F2)	censors nothing	(F1) and (F2)	censors (F0)

	Censoring structure			
	(iii)		(iv)	
Failure mode	Failure mechanism	Effect on analysis	Failure mechanism	Effect on analysis
(-, -)	(F0)	censors (F1), (F2), (F12)	(F0)	censors (F1) and (F2)
(SA1, -)	(F1)	censors (F0), (F2), (F12)	(F1)	censors (F0) and (F2)
(- ,SA2)	(F2)	censors (F0), (F1), (F12)	(F2)	censors (F0) and (F1)
(SA1,SA2)	(F12)	censors (F0), (F1), (F12)	(F12')	censors nothing

5.5 ANALYSES

In this example we look at just three of the four different tyes of subsystem. These will be referred to as subsystem (1), (2) or (3).

In section 5.4 a number of model formulations were discussed. In this application, however, it was not possible to apply a model using the point process A. This is because information of when an LRU returns to the field after repair is not available.

We consider here the application of some simple models employing the point processes of type D and C.

5.5.1 POINT PROCESS D MODELS

Subsystems (1) and (2) are analysed with such models.

The time metric, t, was taken as the time in days between faults on a particular serial numbered subsystem.

Multiple records on the same day were treated as a single fault event in order to avoid many zero TBFs.

In order to model the TBFs of a number of subsystems together, it is assumed that subsystems of the same type have the same baseline hazard.

Times to first failure, and times since last failure were ignored since the date of entry into service, and the date reports were ceased are unknown.

The models follow a point process at subsystem level.

In this application, because of the large numbers of different types of LRU in the subsystems, covariates were not used to attribute a fault to a lower level. No censorings for competing risks, therefore, were required.

Ten explanatory variabes were introduced to form the covariate set for these models.

Holding group variables. Five binary dummy variables were used to compare the six holding groups. These were set as in Table 5.2. The baseline holding group was selected to be the one with the longest period of reported events.

Table 5.2 Coding of holding group.

Holding group	Covariates z_1	z_2	z_3	z_4	z_5
I (base)	0	0	0	0	0
II	1	0	0	0	0
III	0	1	0	0	0
IV	0	0	1	0	0
V	0	0	0	1	0
VI	0	0	0	0	1

Season dummies. Three dummy variables z_6, z_7 and z_8 were used to compare the seasons spring, summer, and autumn respectively against the baseline season winter.

ETI value. This covariate is set to the actual ETI reading as recorded at the time of the occurrance of the fault. It is used irrespective of any apparent error.

Time trend. This covariate, time in days since an arbitrary start date, allows for the possible discovery of a time trend.

5.5.1.1 SUBSYSTEMS (1)

The results after the usual backwards stepwise elimination process, are given in Table 5.3.

Table 5.3 Significant model for subsystems (1).

covariates	$\hat{\beta}_i$	p-value	Likelihood Ratio Statistic
z_1 - holding group II	0.3732	0.0000	68.695 (9.488)
z_5 - holding group VI	0.5622	0.0000	
z_6 - spring	0.1268	0.0205	
z_{10} - time trend	-0.0006	0.0000	

The likelihood ratio statistic exceeds the tabulated upper 5% critical value for a chi-squared distribution with four degrees of freedom. This indicates that the fitted model provides significantly more explanation, for the observed data, than a model assuming that the covariates have no effect and that the data is homogeneous.

The positive estimates of β_1 and β_5 indicate that subsystems (1) in holding groups II and VI experience a higher hazard than those in the other holding groups.

There may be a number of possible reasons for this that could be investigated; for example these holding groups may transport their equipment over rougher terrain, or their maintenance procedures may not be of such a high standard as the others.

The positive estimate of β_6 indicates that subsystems (1) experience a higher hazard in the spring than in the other seasons. This could possibly be reflecting increased exposure due to exercises.

The negative estimate of β_{10} indicates a decreasing hazard as time passes, by a rate of an approximate 20% reduction per year. The number of subsystems (1) entering the field was known to be rising over the period of observation. If we can assume that the times to first failure are not shortening with calander time, the result indicates that the subsystems (1) are becoming more reliable.

The ETI reading was found to be non-significant. The elapsed time in normal mode was initially considered as important in explaining the fault rates of the units. The finding that the covariate is non-significant in this model is probably mostly due to the unreliability of the ETIs themselves, as discussed earlier in section 5.3.1.

A *Weibull hazard plot* for the estimated cumulative hazard from the model is shown in Figure 5.8. This shows the Weibull to be a reasonable distribution for the baseline hazard. The shape and scale parameters for the distribution are estimated from the plot, and are approximately 0.84 and 11 days respectively. Since the shape parameter is less than 1 a decreasing hazard rate is exhibited.

The plot of the *Cox and Snell type* residuals, Figure 5.9, shows that the model is a good fit. See Chapter 7.

Figures 5.10, 5.11 and 5.12 are *proportionality plots* for the three binary significant covariates. See Chapter 6.

The approximately constant vertical separation in Figure 5.10 shows that the proportionality assumption is not invalid for the covariate z_1 - holding group II.

For holding group VI, the vertical separation in Figure 5.11 changes towards the end of the graphs where there is relatively little data (hence the crossing point). The curves, however, are otherwise reasonably well separated and the proportionality assumption appears plausible for the majority of the data which corresponds to TBFs under 20 days.

The plots in Figure 5.12, which stratifies on the covariate for spring, cross in numerous places, hence do not appear to have a constant vertical separation. The modelling of the saesons in this manner may not be strictly appropriate.

Figure 5.13 shows a *standardised plot of the influence* of individual TBFs on the β_6 parameter estimate, see Chapter 7. The plot has split into two distinct groups; one associated with TBFs in spring, and another associated with TBFs in the other seasons.

Despite the marginality of the significance of the covariate there do not appear to be any individual events that would alter the significance of the covariate if omitted.

5.5.1.2 SUBSYTEMS (2)

Applying the same model structure as was applied to subsystems (1), and commencing with the same initial covariate set, we can analyse the reliability of subsystems (2).

The results after backwards stepwise elimination for this application is shown in Table 5.4.

Table 5.4 Significant mode for subsystems (2).

Covariates	$\hat{\beta}_1$	p-value	Likelihood Ratio Statistic
z_1 - holding group II	0.4009	0.0000	
z_2 - holding group III	0.2566	0.0140	33.71 (9.488)
z_5 - holding group VI	0.5325	0.0000	
z_{10} - time trend	-0.0006	0.0001	

These results, as may be expected, are broadly similar to those obtained earlier for subsystems (1).

Spring, however, is no longer significant and holding group III is now marginally significant.

The results for z_1, z_5 and z_{10} are highly significant as before, and are of the same sign and order of magnitude as in the model for subsystems (1).

The likelihood ratio statistic suggests that this is not such a good-fitting model as that for subsystems (1).

5.5.2 POINT PROCESS C MODEL

For this example a particular type of LRU present in subsystems (3) was selected for analysis. This selection was made because the LRU has just two sub-assemblies and two levels in its hierarchy, so form the simple case discussed in section 5.4.2. The censoring structure (iv) was applied to the model.

Dummy variables were introduced into the covariate set to compare the failure mechanisms. The dummy variables were set as in Table 5.5.

Table 5.5 Coding for failure mechanisms.

Failure mechanism	Covariate z_{11}	z_{12}	z_{13}
[F1] (base)	0	0	0
[F12']	1	0	0
[F0]	0	1	0
[F2]	0	0	1

The covariates z_1 to z_{10} that were used for the previous models were again incorporated into the covariate set.

After backwards stepwise elimination only one covariate; z_{11} remained significant, with $\hat{\beta}_{11} = 1.8713$ and p-value ≈ 0.0000.

The positive estimate of β_{11} suggests that there are shorter times to simultaneous failures than times to failures due to the alternative mechanisms.

5.6 CONCLUSIONS FROM APPLICATION TO A WEAPON SYSTEM

PHM again offers itself as an effective method for the reliability analysis of a system such as this with its ability to incorporate a wealth of auxiliary, or covariate, information.

However, with such a complex hierarchical system and so many possible model formulations which could be applied to the data, it becomes very important to understand the operations of the system under observation, obtain high quality data, and use PHM in an exploratory manner. The complexity of the models requires careful application.

Figure 5.1 Hierarchical nature of subsystems.

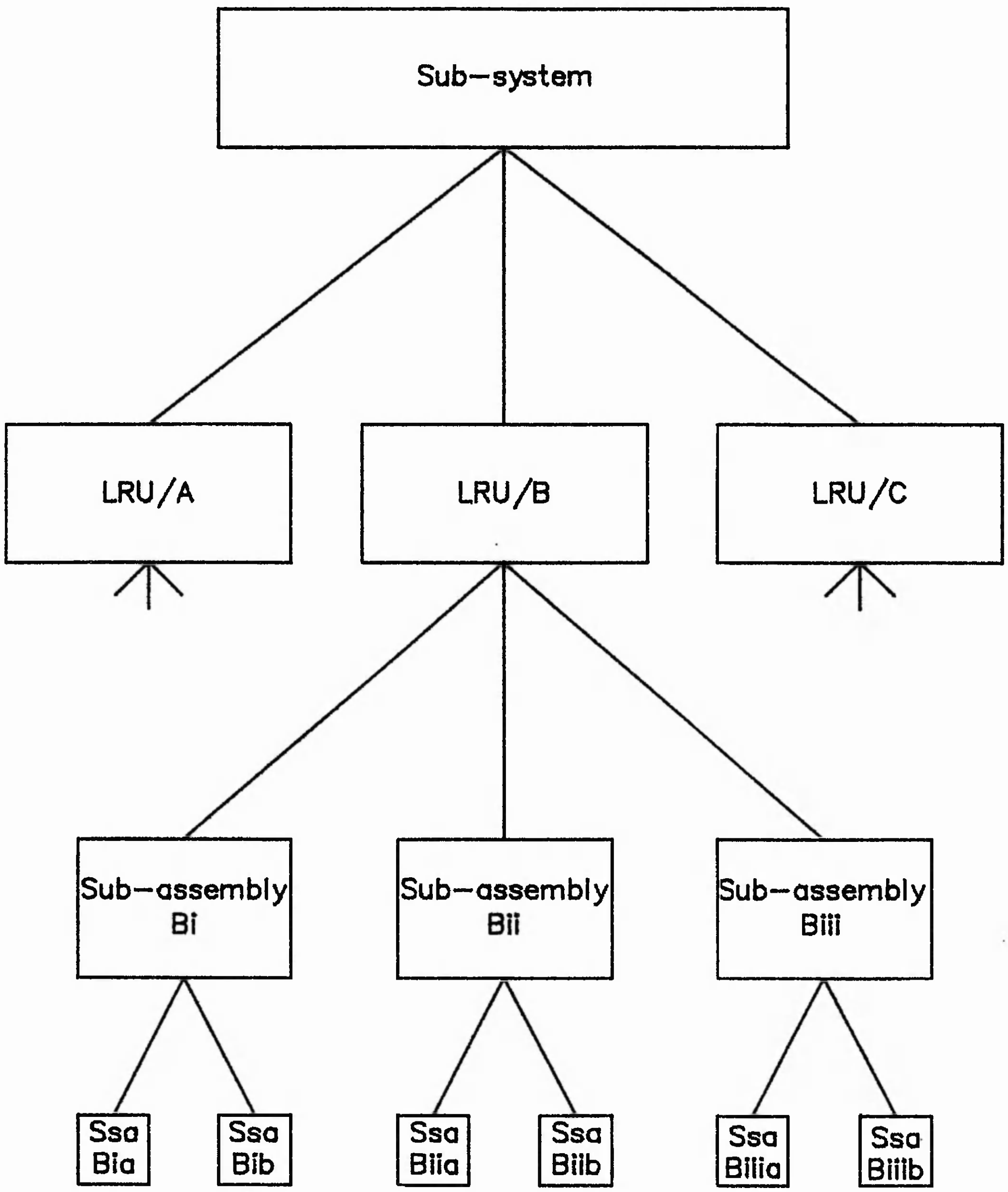


Figure 5.2 Generalised representation of the fault structure.

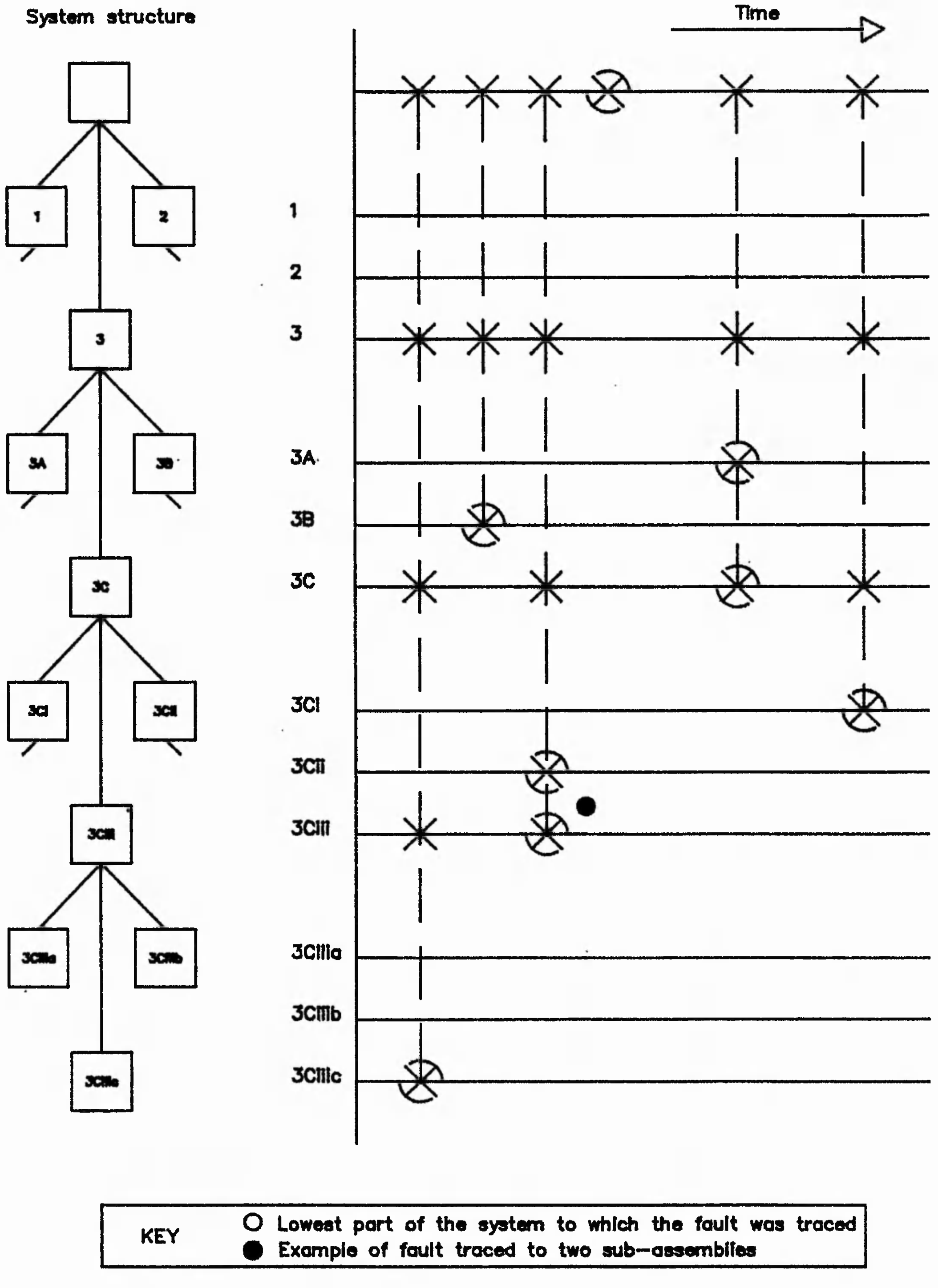


Figure 5.3 Series of events to a particular LRU

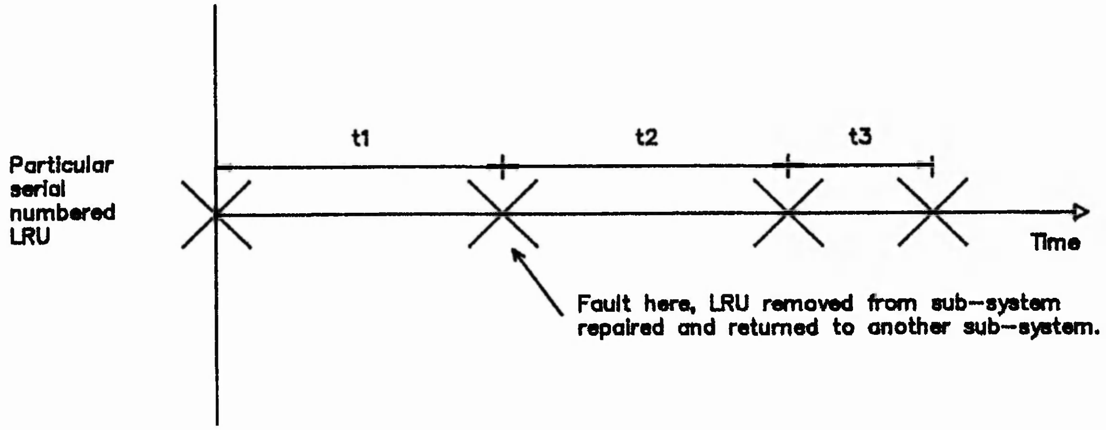


Figure 5.4 Series of events to the position of a certain type of LRU within a particular sub-system.

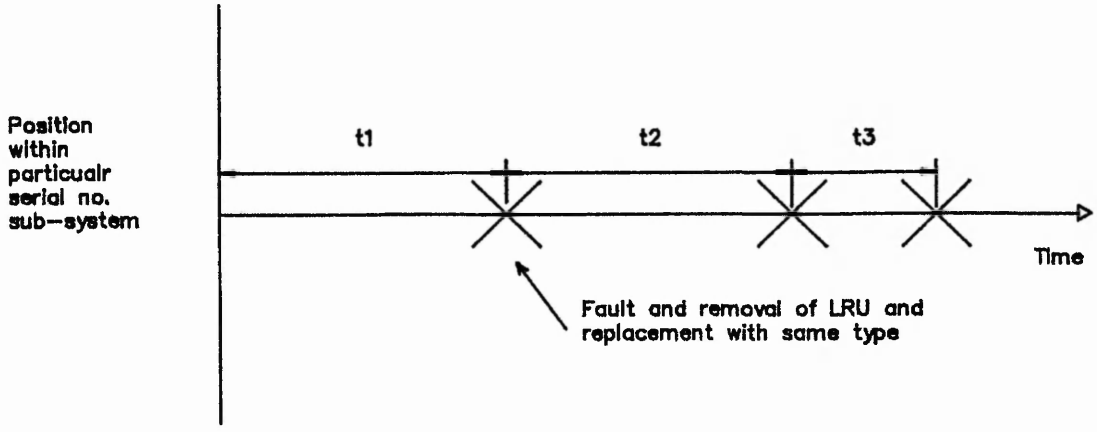


Figure 5.5 Times to first failure of each LRU of a certain type

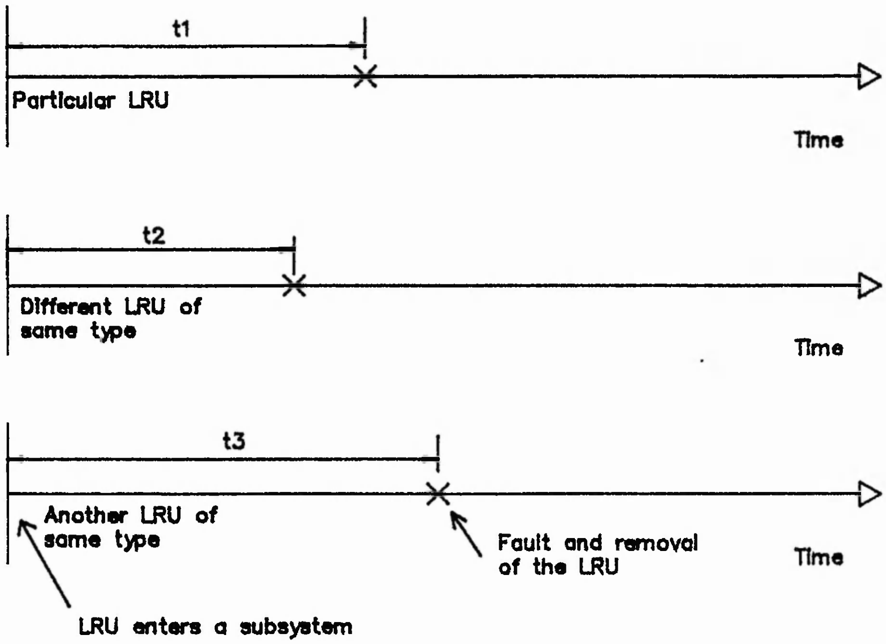


Figure 5.6 Series of events to a particular subsystem

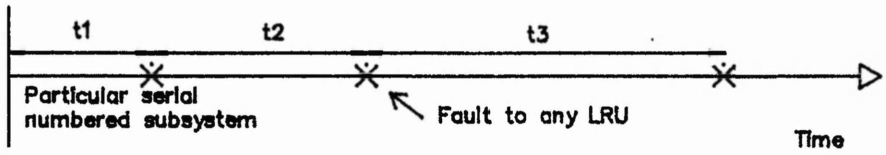


Figure 5.7 LRU with two levels of hierarchy only, and two sub-assemblies

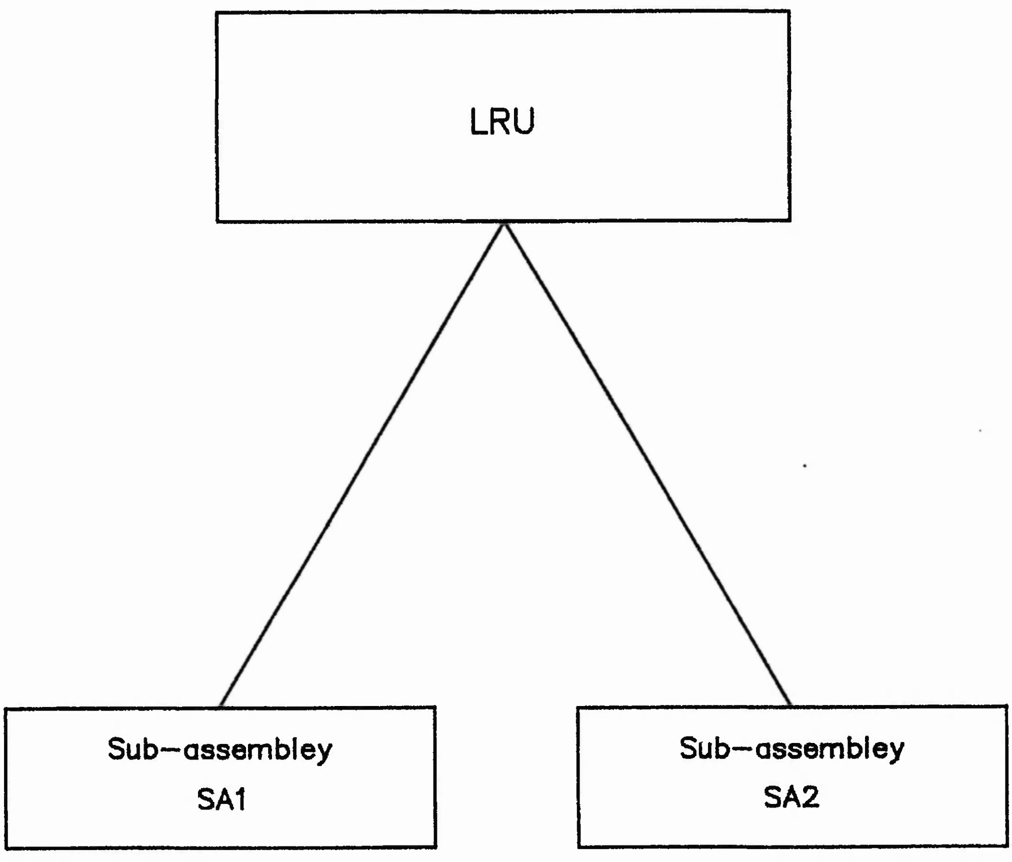


Figure 5.8 Weibull baseline hazard plot.
D-type model. Subsystem (1).

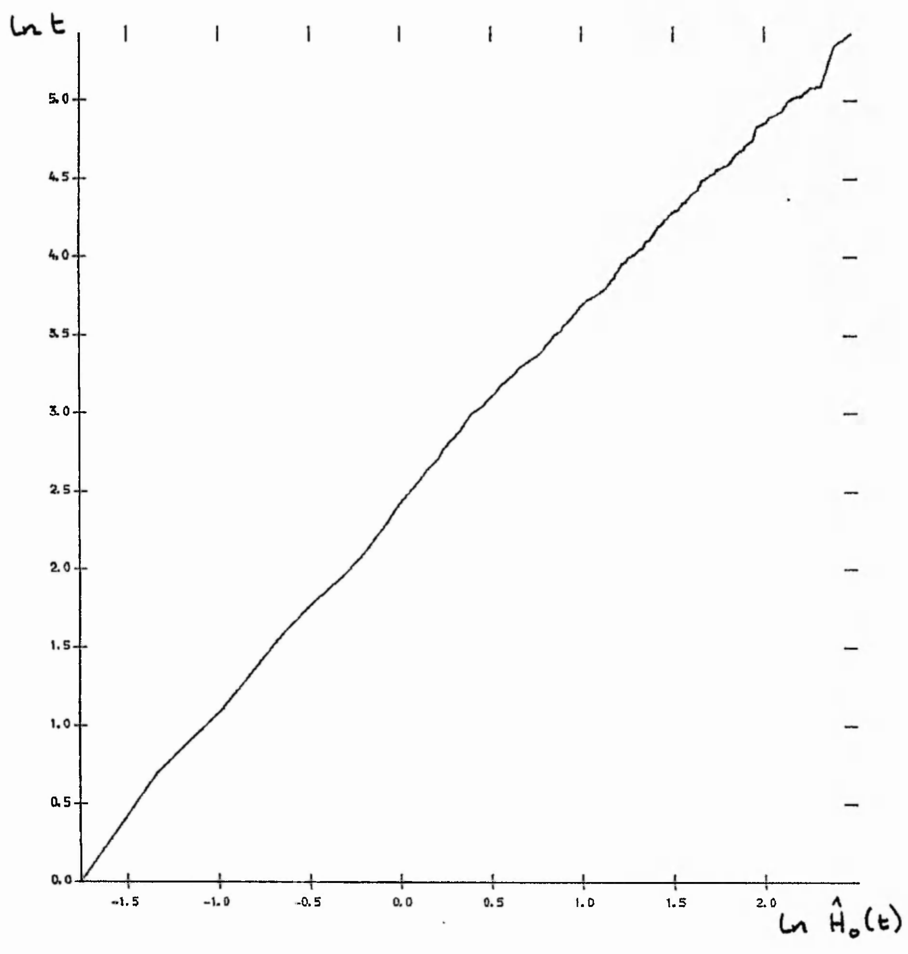


Figure 5.9 Cox and Snell type residual plot.
D-type model. Subsystem (1).

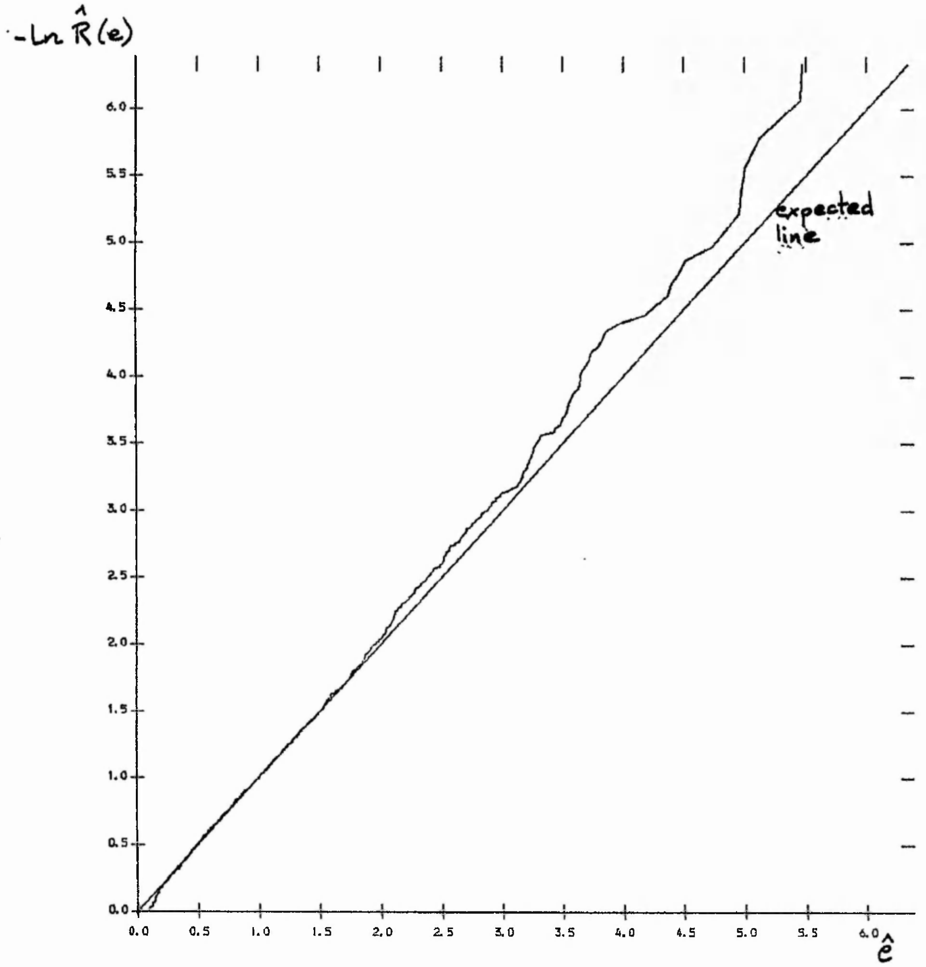


Figure 5.10 Proportionality plot. D-type model. Subsystem (1). Covariate z_1-'Holding group II'

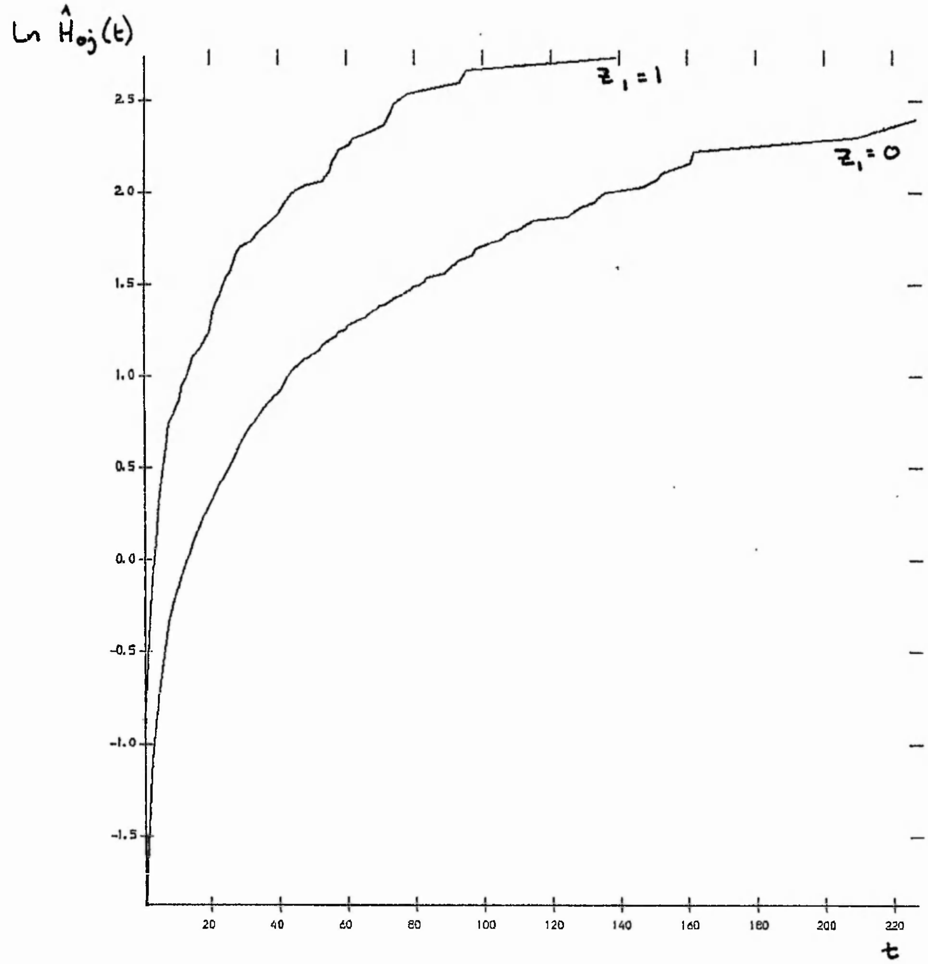


Figure 5.11 Proportionality plot. D-type model. Subsystem (1). Covariate z_5-'Holding group VI'

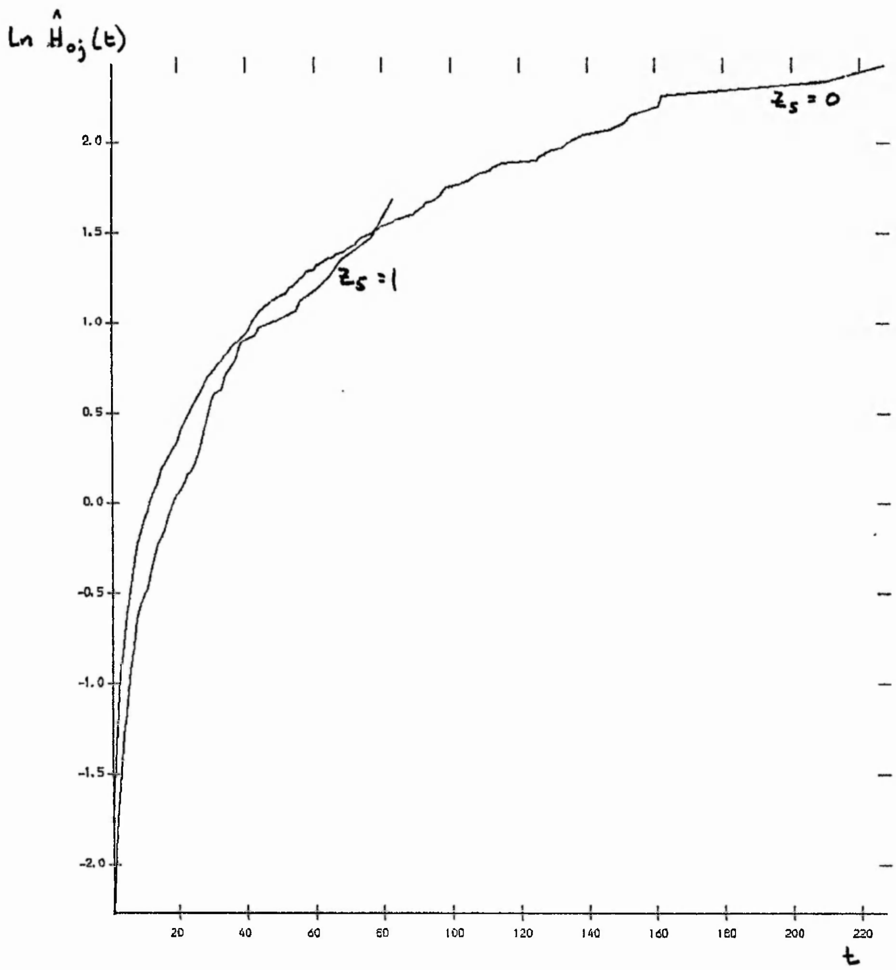


Figure 5.12 Proportionality plot. D=type model. Subsystem (1). Covariate z_6 - 'Spring'

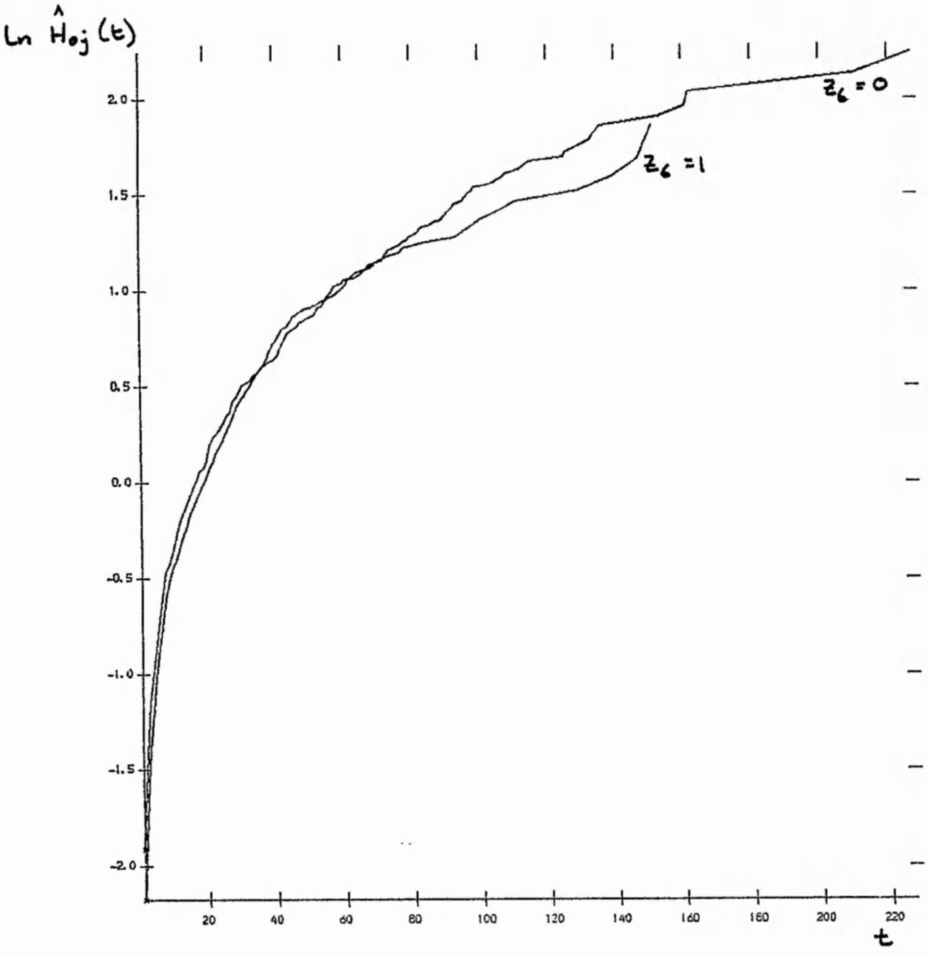
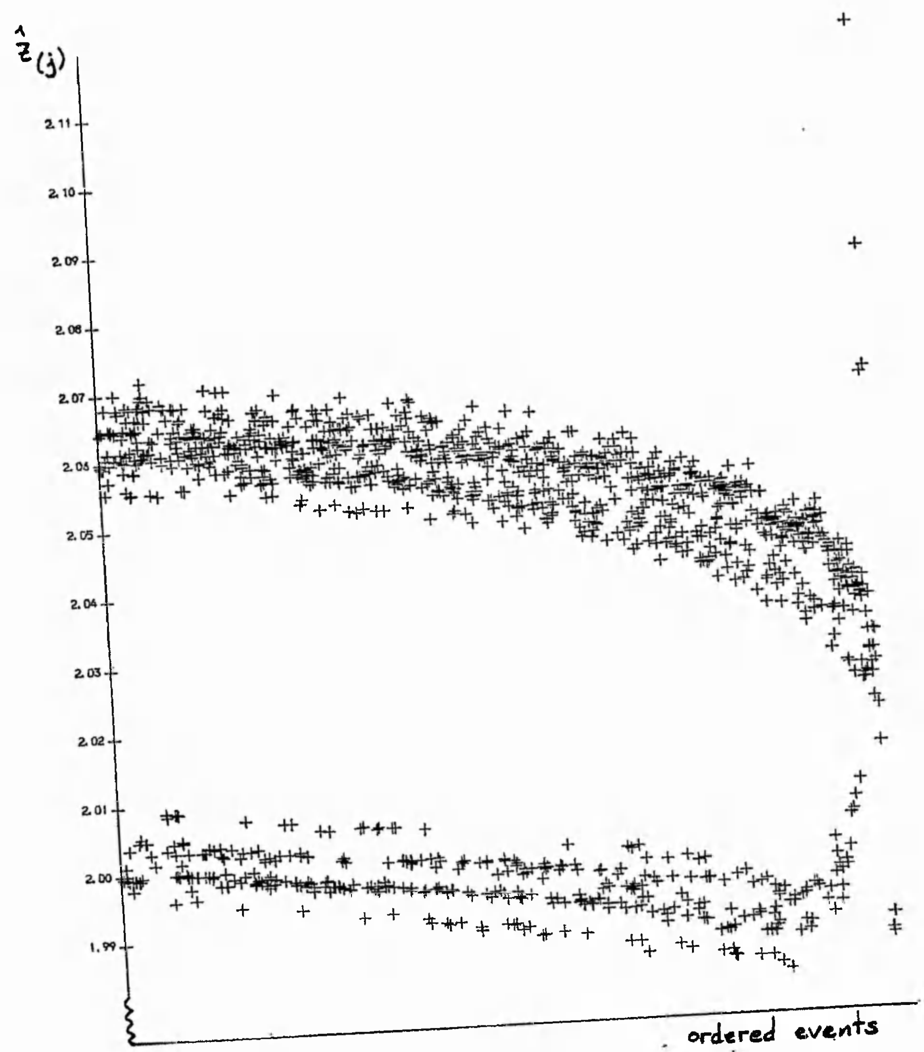


Figure 5.13 Influence plot. D-type model. Subsystem (1). Covariate z_6 – 'Spring'



CHAPTER 6

DIAGNOSTICS FOR PHM

PHM has proved a useful tool for analysing reliability data. It's strength is largely due to its non-parametric approach to analysis.

The method does, however, make the assumption of 'proportional hazards' for the effect of covariates. This assumption should be checked for validity.

In this chapter various diagnostics aids, which have been suggested in the literature, will be considered. These diagnostics can be used to assess the validity of the model's assumption, or to assess the fit of the model.

Many of the diagnostics are based upon graphical techniques.

A graphical method for testing the assumption of proportionality between different levels of a covarite is provided by Kay (1977). This technique is investigated in detail in section 6.1.

Kay (1977) also suggested a technique to assess the fit of a model, through obtaining residual quantities as defined by Cox and Snell (1968). See section 6.3 for detailed investigation into this technique.

Lagakos (1981) defines residual scores similar to the Cox and Snell type residuals defined in Kay (1977). Lagakos adjusts the observed ranks based upon the residual information.

The readjustment should account for the effects of the covariates, if the model is appropriate.

An explanatory variable omitted from the model found to be associated with the adjusted ranking may be correlated with survival time.

An explanatory variable fitted to the model exhibiting an association with the adjusted rank indicates lack of fit of the model.

A Chi-squared goodness-of-fit test is obtained by Schöenfeld (1980), by partitionaing both the covariate space and the time axes. Andersen (1982) proposes a new technique similar to that of Schöenfeld (1980), that invovles partitioning the time axes. A statistic for checking the fit of a model is provided by Moreau *et al* (1985), that in the two-sample problem is the same as that proposed by Schöenfeld (1980).

Schöenfeld (1982) defines residuals which are essentially the difference between the observed value of a covariate and its expected value conditional on the risk set. These are used graphically to examine the proportional hazards assumption, see section 6.2.

Cain and Lange (1984) and Reid and Crépeau (1985) employ essentially the same method to obtain influence functions for the proportional hazards model. These functions are obtained for each covariate and approximate the effect of individual events upon the estimate of the associated β coefficient, see section 6.4.

Storer and Crowley (1985) also discuss a diagnostic for estimating the change in $\hat{\beta}$ due to the deletion of single observations.

Gill and Schumacher (1987) suggest a test of the proportional hazards assumption, for two-level covariates. The test procedures are based on the discrepancy between two-sample tests, e.g. the log rank and a generalised Wilcoxon test. In nonproportional hazards situations the tests might give different answers. Gill and Schumacher (1987) also present a related graphical method for comparing trends in series of events.

Arjas (1987) presents graphical diagnostics which make direct comparisons between observed and expected failure frequencies, as estimated from the model.

The way in which individuals are stratified depends on the aspect of the model being checked. Arjas (1987) investigates a fitted proportional hazards model: a significant covariate omitted from the model: a common baseline assumed for an inappropriate case.

The remaining sections of this chapter concentrate on the four graphical diagnostics for PHM that have been employed in Chapters 3-5.

The techniques are investigated in detail, and improvements made to the presentation of the plots, to ease visual inspection, are also discussed.

6.1 PROPORTIONALITY PLOTTING

The most commonly applied method to test whether a covariate follows the proportionality assumptions is to stratify upon the covariate of interest and, for each stratum (level) of the covariate, plot the logarithm of the cumulative baseline hazard against time, or the logarithm of time, see Kay (1977), Kalbfleisch and Prentice (1980), Aitkin and Clayton (1980), Andersen (1982).

For a binary covariate, z_k say, we have for the hazard function

$$h(t;\underline{z},z_k) = \begin{cases} h_0(t).\exp(\underline{\beta z}).\exp(\beta_k z_k) & , z_k=1 \\ h_0(t).\exp(\underline{\beta z}) & , z_k=0 \end{cases} \qquad 6.1$$

and hence for the logarithm of the cumulative hazards:

$$\log H(t;\underline{z},z_k) = \begin{cases} \log H_0(t) + \underline{\beta z} + \beta_k & , z_k=1 \\ \log H_0(t) + \underline{\beta z} & , z_k=0 \end{cases} \qquad 6.2$$

From equation 6.2 the difference between the log cumulative hazards when $z_k=1$ and $z_k=0$ is a constant;

$$\log H_{k1}(t) - \log H_{k0}(t) = \beta_k \qquad 6.3$$

where $H_{k1}(t)$ is the cumulative hazard when $z_k=1$, and $H_{k0}(t)$ is the cumulative hazard when $z_k=0$.

Now stratifying the data on the level of the covariate z_k and applying the model to the two stratum.

$$\log \hat{H}_{k1}(t) = \log \hat{H}_{01}(t) + \underline{\hat{\beta}z}$$
$$\qquad 6.4$$
$$\log \hat{H}_{k0}(t) = \log \hat{H}_{00}(t) + \underline{\hat{\beta}z}$$

where $H_{01}(t)$ is the baseline cumulative hazard for the stratum for which $z_k=1$, and $H_{00}(t)$ is the baseline cumulative hazard for the stratum $z_k=0$.

From 6.4 and 6.3

$$\log \hat{H}_{01}(t) - \log \hat{H}_{00}(t) = \hat{\beta}_k \qquad 6.5$$

Hence plotting the logarithm of the baseline cumulative hazard against time t for the two strata on the same graph should produce a constant vertical separation equal to $\hat{\beta}_k$.

The stratified models as defined by 6.4 have the same covariate set and covariate coefficients (with the exception of z_k) as are significant in the full model. The plots should still result in a constant vertical separation if the models for each stratum are not constrained in this way. The size of the separation is now dependent on all the covariates. Now for the two strata:

$$\log \hat{H}_{k1}(t) = \log \hat{H}_{01}(t) + \underline{\beta}'\underline{z}'$$
$$\log \hat{H}_{k0}(t) = \log \hat{H}_{00}(t) + \underline{\beta}''\underline{z}'' \qquad 6.6$$

where $\underline{z}'$ is the set of significant covariates in the stratum for which $z_k=1$, and $\underline{z}''$ is the set of significant covariates in the stratum for which $z_k=0$. $\underline{\beta}'$ and $\underline{\beta}''$ are the respective coefficient values.

From 6.6 and 6.3

$$\log \hat{H}_{01}(t) - \log \hat{H}_{00}(t) = \hat{\beta}_k - \underline{\beta}'\underline{z}' + \underline{\beta}''\underline{z}''$$
$$= \text{constant} \qquad 6.7$$

A non-constant vertical separation between the plots will, then, indicate that the covariate z_k does not act proportionaly on the hazard.

6.1.1 CHOICE OF AXES

From equations 6.5 and 6.7 it can be seen that plotting the log baseline cumulative hazards against the failure times t should result in plots having constant vertical separation, hence testing the proportionality assumption. This is the procedure followed by Kay (1977) and Andersen (1982).

Kalbfleisch and Prentice (1980) and Aitkin and Clayton (1980) suggest plotting the log baseline cumulative hazards against log t. Such plots should also have constant vertical separation, and in addition if the lines are straight they will indicate that the baselines are Weibull distributed.

However, since we are looking at the cumulative hazard, more information goes into the plot at larger t's, and this is the area of the plot we should be most interested in. When plotting log t we effectively 'stretch out' the smaller t's and 'bunch up' the larger t's resulting in less of the plot being devoted to the area we are most interested in.

It is for the above reason that we choose to plot against t.

6.1.2 INTERPRETING THE PLOT

The problem with this graphical procedure is that it contains a highly subjective element. It is not easy to decide whether or not there is a constant vertical separation between the log baseline cumulative hazards for each stratum.

Because of the curvature of the plots those with true constant vertical separation appear to close in together as $t \to 0$. See Figure 6.1.

The plots are particularly difficult to interpret when one stratum has only a small number of failures.

6.1.3 PLOTTING SEPARATION

Because of the problem encountered with the visual inspection of the diagnostic as detailed in section 6.1.2, the vertical separation between the two log baseline cumulative hazards is calculated at each failure time and plotted beneath the original graph.

Since the hazard is only estimated at failure times, and the two strata will have different failure times, linear interpolation is employed between the failure points on each stratum. Hence, an estimate of the vertical separation can be calculated at each failure point.

Figure 6.2 is the proportionality plot from the hardware failure example for covariate z_5 - 'No fault found'. The vertical separation at failure points is plotted below.

For constant vertical separation the difference line should be straight and horizontal. Typically, however, the difference line will not be perfectly straight or horizontal, and a means to decide how 'good' the difference line should appear needs to be developed in order to accept the proportionality of the covariate.

A 'runs' and sign' test were chosen to look at the walk of the difference line. At each failure point t_i, the walk is given a '+' sign if the difference line is increasing between t_i and t_{i+1}, or a '-' sign if the difference line is decreasing between t_i and t_{i+1}. If the difference is the same at ti and t_{i+1} no sign is awarded.

The 'runs' test tests the null hypothesis of randomness. When the null hypothesis is true the path of signs should cross (+ to -, or - to +) quite frequently, but when it is not true this happens much less frequently.

The number of runs is defined as the number of blocks separated by a crossover to the other sign.

The 'sign' test is also used, since if the walk is random we could expect about the same number of +'s and -'s.

From the results of the runs and sign test we can see then if there is any significant trend in the vertical separation.

6.1.4 CONFIDENCE BOUNDS FOR THE CUMULATIVE HAZARD

Since the variance is known to alter along the estimates of the log baseline cumulative hazard it would be useful to construct confidence bounds around the estimates.

6.1.4.1 ' LINK'S ' BOUNDS

Initial bounds for the log baseline cumulative hazard were produced, from a transformation of the confidence interval, around the baseline survivor function, which was constructed from an asymptotic estimate of the variance for the survivor function (Link (1984)).

A smoothed version of Breslow's step function estimate for the baseline survivor function; whereby linear interpolation between failure points is employed to estimate a value of the baseline survivor function at times where no failure occurs, is considered.

There are two sources of variation in the estimate of the survivor function; firstly in the estimation of the coefficients β by $\hat{\beta}$, and secondly in the approximation of the integrated hazard function.

The estimates of the variance of the baseline survivor function can be used to construct a confidence interval about the baseline survivor function at various points; based on the asymptotic normality of the survivor function.

A transformation of this interval can be used to construct bounds about the log baseline cumulative hazards. These asymptotic bounds are, however, very wide.

An example of Link's bounds applied to the proportionality plot for the covariate 'No fault found' from the analysis of the ICL hardware failure data can be seen in Figure 6.3. The +'s on the figure are the 95% bounds around the top plot, and the x's those for the lower plot.

Despite the wideness of the bounds, it can be seen that the bounds 'narrow' as t increases indicating less variability in the latter part of the plots.

6.1.4.2 SIMULATED WEIBULL BOUNDS

Simulated bounds were considered in an attempt to narrow the confidence intervals. Assuming a Weibull distributed baseline with parameters estimated from a hazard plot for each of the strata, 90% limits were constructed by simulating twenty groups of fifty failure times from the estimated Weibull distribution.

The parameters of the Weibull distribution are estimated from a Weibull hazard plot.

An example of these simulated bounds as applied again to the proportionality plot for the covariate 'No fault found' from the hardware failure example can be seen in Figure 6.4.

The simulated bounds are narrower than those obtained from the asymptotic variance of the survivor function. We have, however, had to assume a distributional form for the baseline hazard.

The variability is seen to decrease for larger t.

The simulations have also been used to produce 95% bounds around the difference plot. The width of these bounds indicate the sort of variability we might expect to find in the difference walk when the proportionality assumption is valid.

6.2 SCHOENFELD RESIDUALS

Schöenfeld residuals were first introuced as a graphical diagnostic for PHM by David Schöenfeld (1982).

The residuals are known as 'partial residuals' since a set is obtained for each covariate. The partial residuals can be used to test 'locally' the proportional hazards assumption.

6.2.1 DEFINITION OF RESIDUALS

For the k'th covariate at failure time t_i the partial residual r_{ki} is defined:

$$\hat{r}_{ki} = z_{ki} - \hat{E}[z_{ki}|R_i] \qquad 6.8$$

where z_{ki} is the value of the k'th covariate at failure time t_i.

$\hat{E}[z_{ki}|R_i]$ is the conditional expected value of z_{ki} given the risk set R_i at failure time t_i.

when

$$\hat{E}[z_{ki}|R_i] = \frac{\sum_{m \in R_i} z_{km}.\exp(\beta z)}{\sum_{m \in R_i} \exp(\beta z)}$$

The partial residuals are only obtained at failure points.

The partial residuals are not specified for tied failure points. However, since we use Breslow's approximation for the contribution to the partial likelihood in the case of tied failures, we are able to make a slight modification to estimate the partial residuals at each of the tied failure points (Wightman (1987)).

The partial residuals are obtained from elements of the score vector, $U(\beta)$, [$U(\beta) = d \log L(\beta) / d\beta$], vis

$$U(\beta) = \sum_{i=1}^{n} \left[z_{ki} - \frac{\sum_{m \in R_i} z_{km}.\exp(\beta z)}{\sum_{m \in R_i} \exp(\beta z)} \right] = \sum_{i=1}^{n} \hat{r}_{ki}$$

6.2.2 TESTING THE PROPORTIONALITY ASSUMPTION

Since the score vector $U(\beta)=0$ globally the expected value of the residual, $E[\hat{r}_{ki}] \approx 0$. Hence, if proportional hazards holds a plot of $\hat{r}_{ki}$ v t_i will be centred about 0 for all areas of the time scale.

For a binary covariate k, $z_{ki} - \hat{E}[z_{ki} | R_i]$ splits the plot into two bands, above and below the axis, corresponding to the two values of the covariate.

Schöenfeld (1982) gives as an example the residuals for the data of Freirich (Cox, (1972)), where he splits the time axis into three bands; $T>16$, $5<T\leq15$ and $T<5$. Schöenfeld suggests that there is no failure of proportional hazards in the two regions $T>16$ and $5<T\leq15$ since the residuals in these regions approximately sum to zero, but that for the region $T<5$ this condition is not met indicating failure of proportional hazards.

This statement itself cannot be true since the total sum of the residual by definition equals zero.

6.2.3 DIVISION OF THE TIME SCALE

The particular division of the time scale by Schöenfeld (1982) appears to be arbitrary. The gap in the '0' band in the region T<5 may not be too unusual, and if a larger region were investigated the proportionality assumption may not be deemed violated.

Consider a special case of a model with a single binary covariate and an exponential baseline hazard. If from a sample size of n, there are n_1 events for which the covariate value equals 1, then the ratio of the expected number of points in the '1' band to the expected number of points in the '0' band can be estimated. Consider the region $(0,t_x)$; the probability of an individual with covariate value 1 failing in the interval is given by $1-\exp(-e^{\beta}\lambda t_x)$, and the probability of an individual with covariate value 0 failing in the interval is given by $1-\exp(-\lambda t_x)$.

Hence, in the interval $(0,t_x)$;

$$k(t_x) = \frac{[\text{expected no. points in '1' band}]}{[\text{expected no. points in '0' band}]}$$

$$k(t_x) = \frac{\{1 - \exp(-e^{\beta}\lambda t_x)\}}{\{1 - \exp(-\lambda t_x)\}} \times \frac{n_1}{(n - n_1)} \qquad 6.9$$

If t_x is small in relation to the mean failure time then a crude approximation to 6.9 is given by:

$$k(t_x) \approx e^{\beta} \cdot \frac{n_1}{(n - n_1)}$$

For $\beta > 0$ and $n_1 >> (n - n_1)$ a value of $k(t_x) > 10$, say, may not be unusual. Thus at the start of the plots this is quite consistent with gaps in the '0' band in the interval $(0, t_x)$.

6.2.4 APPLICATION OF MOVING AVERAGE

Since there are often in reliability problems a large number of tied failure times the density of partial resiudals at many of the points cannot be seen. Visual inspection is also hindered, in the case of binary covariates, because the two bands are rarely equidistant from the axis.

Because of these reasons and the problem in determining appropriate divisions for the time scale, as discussed in setion 6.2.3, a moving average based on intervals of a large number of failures was applied to the plot to ease visual inspection of the diagnostic. The moving average can then be looked at to assess local fit to the proportionality assumption.

Figure 6.5 is the plot of Schöenfeld residuals against time for the covariate 'average hours use per week' from the hardware failure example. There is a general scatter of residuals for this naturally measured variable. The x's represent tied values hence a density of points greater that 1 in these positions. the moving average line has been applied over groups of twenty observations, this gives a better representation of the scatter about the axis.

Figure 6.6 is the plot of Schöenfeld residuals for the covariate '2 processors' from the hardware failure example. The residuals for this binary covariate have split into two bands above and below the axis. Agian the x's represent tied values. The moving average again based on groups of twenty observations gives a clearer picture with which to assess the local fit to the proportionality assumption.

6.2.5 EFFECT OF CENSORING OBSERVATIONS ON APPEARANCE OF RESIDUAL PLOTS.

For a binary covariate, where a split of residuals into two bands occurs, we can show that an upward or downward trend should be expected in the appearance of the residuals, based on the change in ratio of numbers with covariate equal to 1 and numbers with covariate equal to 0 in the risk set.

When there are no censoring events the appearance of the residuals can easily be seen to depend on the sign of β_j:

Consider the two groups; the n_1 items for which $z_{ji} = 1$, $i = 1, 2, \ldots, n$ hence having covariate combination $\beta_1 z_1 + \beta_2 z_2 + \ldots + \beta_{j-1} z_{j-1} + \beta_j$, and the $(n - n_1)$ items for which $z_{ji} = 0$, $i = 1, 2, \ldots, n$ hence having covariate combination $\beta_1 z_1 + \beta_2 z_2 + \ldots + \beta_{j-1} z_{j-1}$.

Let $a = \exp(\beta_1 z_1 + \beta_2 z_2 + \ldots + \beta_{j-1} z_{j-1})$, and consider the expected number of items surviving past a time t in each group: $n_1.\exp(-a.\exp(\beta_j).H_0(t))$ and $(n-n_1).\exp(-a.H_0(t))$, $z_j=1$ and $z_j=0$ respectively.

Looking at the ratio of the expected number surviving past t in each group:

$$\text{Ratio}_{1.0} = \frac{n_1}{(n-n_1)} \cdot \frac{\exp(-a.\exp(\beta_j).H_0(t))}{\exp(-a.H_0(t))}$$

$$= \frac{n_1}{(n-n_1)} \cdot [\exp(-a.H_0(t))]^{(\exp(\beta_j)-1)}$$

As $t \rightarrow \infty$, $\exp(-a.H_0(t)) \rightarrow 0$

if $\beta_j > 0$ then $\exp(\beta_j)-1 > 0$

hence $[\exp(-a.H_0(t))]^{(\exp(\beta_j)-1)}$ decreases as $t \rightarrow \infty$

if $\beta_j < 0$ then $\exp(\beta_j)-1 < 0$

hence $[\exp(-a.H_0(t))]^{(\exp(\beta_j)-1)}$ increases as $t \rightarrow \infty$

Since $n_1/(n-n_1)$ is constant over all t

$Ratio_{1,0}$ decreases as $t \to \infty$ if $\beta_j > 0$

$Ratio_{1,0}$ increases as $t \to \infty$ if $\beta_j < 0$

Now looking at the contributions from each group to $\hat{E}[z_{ji} | R_i]$.

$$\hat{E}[z_{ji} | R_i] = \frac{\sum_{m \in R_i} z_{jm}.\exp(\beta z)}{\sum_{m \in R_i} \exp(\beta z)} = \frac{\sum_{R_i} \exp(\beta_j)}{\sum_{R_i} [\exp(\beta_j)+1]}$$

Let m_1 = number in risk set at time t_i, for which $z_{jm} = 1$

m_0 = number in risk set at time t_i, for which $z_{jm} = 0$

then
$$\hat{E}[z_{ji} | R_i] = \frac{m_1.\exp(\beta_j)}{m_1.\exp(\beta_j) + m_0}$$

Let

m_1' = number in risk set at time t_{i+1}, for which $z_{jm} = 1$

m_0' = number in risk set at time t_{i+1}, for which $z_{jm} = 0$

then
$$\hat{E}[z_{ji+1} | R_{i+1}] = \frac{m_1'.\exp(\beta_j)}{m_1'.\exp(\beta_j) + m_0'}$$

Investigating the relationship (R) between $\hat{E}[z_{ji} | R_i]$ and $\hat{E}[z_{ji+1} | R_{i+1}]$:

$$\frac{m_1.\exp(\beta_j)}{m_1.\exp(\beta_j) + m_o} \quad (R) \quad \frac{m_1'.\exp(\beta_j)}{m_1'.\exp(\beta_j) + m_o'}$$

$$m_1(m_1'.\exp(\beta_j) + m_o') \quad (R) \quad m_1'(m_1.\exp(\beta_j) + m_o)$$

$$\frac{m_1}{m_o} \quad (R) \quad \frac{m_1'}{m_o'}$$

We already know the relationship (R) since $m_1/m_o = Ratio_{1,o}$ at time t_i, and m_1'/m_o' = Ratio at time t_{i+1}.

If $\beta_j > 0$, $m_1/m_o > m_1'/m_o'$, hence $\hat{E}[z_{ji}|R_i] > \hat{E}[z_{ji+1}|R_{i+1}]$
If $\beta_j < 0$, $m_1/m_o < m_1'/m_o'$, hence $\hat{E}[z_{ji}|R_i] < \hat{E}[z_{ji+1}|R_{i+1}]$

The appearance of the residuals on both bands is thus determined; increasing if $\beta_j > 0$, and decreasing if $\beta_j < 0$.

Figure 6.7 shows the upward trend in each band of the residuals as we would expect for a binary covariate with a positive estimate of ß. The figure shows the plot for such a covariate; 'left truncation' from the hardware failure example.

Figure 6.8 shows the downward trend we would expect for a binary covariate with a negtive estimate of ß. The figure shows the plot for such a covariate 'system B' from the hardware failure example.

However, with the inclusion of censoring events it is no longer possible to determine whether the $Ratio_{1,0}$ is increasing or decreasing because there is no longer a constant term $n_1/(n-n_1)$.

Experience has shown that the appearence of these residual plots is highly affected by the pattern of censoring observations: the effect has even altered the trend we might expect from the sign of β_j, given no censoring.

Figure 6.9 shows the Schöenfeld residuals for a model with censored times. This plot has increasing trend in the two groups of residuals despite the estimate of the coefficient β being negative. The plot shows the residuals for the covariate 'system F' from the hardware failure example.

Wightman (1987) observing the apparent dependence on the pattern of censoring events, plots the censoring observations at their covariate values on the same figure.

6.2.6 RESIDUALS FOR COVARIATES FROM COMPETING RISKS

The Schöenfeld partial residuals are extracted from elements of the score vector $U(\beta)$.

$$U(\beta) = \frac{d \log L}{d \beta_j} = \sum_{i=1}^{n} \left[z_{ji} - \frac{\sum_{m \in R_i} z_j . \exp(\beta z)}{\sum_{m \in R_i} \exp(\beta z)} \right]$$

Now, in the case of a model involving competing risks the partial likelihood factorises into two terms; one which is the usual partial likelihood as obtained from a simple model without competing risks, and the other containing only information for the failure mode covariates, see Wightman (1987) and section 2.3.5.

For a failure mode covariate the partial residual is obtained via the factor of the partial likelihood involving the failure mode terms, L_f.

$$L_f = \prod_{i=1}^{n} \left[\frac{\exp(\alpha_1 x_{1i} + \alpha_2 x_{2i} + \ldots + \alpha_l x_{li})}{[\,1 + \exp(\alpha_1) + \exp(\alpha_2) + \ldots + \exp(\alpha_l)\,]} \right]$$

where l = number of covariates representing failure modes.

x_{ji} , $j=1,2,\ldots,l$ = value of j'th failure mode covariate.

α_j , $j=1,2,\ldots,l$ = associated parameter for covariate x_j.

$$\log L_f = \sum_{i=1}^{n} [(\alpha_1 x_{1i}+\ldots+\alpha_l x_{li}) - \ln\{1+\exp(\alpha_1)+\ldots+\exp(\alpha_l)\}]$$

$$U(\alpha_j) = \frac{d \log L_f}{d\,\alpha_j} = \sum_{i=1}^{n} \left[x_{ji} - \frac{\exp(\alpha_j)}{[1+\exp(\alpha_1)+\ldots+\exp(\alpha_l)]} \right] \qquad 6.10$$

The element corresponding to the expected value of x_{ji} conditional on the risk set R_i, is constant over all t_i:

$$\hat{E}[x_{ji} | R_i] = \frac{\exp(\alpha_j)}{[1+\exp(\alpha_1)+...+\exp(\alpha_l)]}$$

The Schöenfeld partial residuals for a failure mode covariate involved in competing risks therefore split into two bands above and below the axis, and remain parallel to the axis over all t_i.

Figure 6.10 shows the Schöenfeld residual for the competing risks covariate 'No fault found' from the hardware failure example. The residuals in each band have a constant value.

6.2.7 INVESTIGATION OF FORM OF TIME DEPENDENCE

Pettit and Bin Daud (1987) use Schöenfeld's residuals to investigate the nature of time dependence that may be applicable for a covariate which violates the proportional hazards assumption.

A proportional hazards model with a time dependent covariate x becomes:

$$h(t;\underline{z},x) = h_0(t).\exp(\underline{\beta z} + \alpha g(t)x) \qquad 6.11$$

where α is the associated unknown parameter of the covariate x

$g(t)$ is some function of the time metric t.

By fitting the model without the time dependence for the covariate x, $\alpha g(t)$ is approximated by the coefficient $\hat{\beta}(x)$.

For g(t) varying about zero slowly then:

$$\hat{E}[r_i(x)] \approx \alpha g(t_i).A_i(x)$$

where $r_i(x)$ is the residual at failure time t_i for the covariate x

$A_i(x)$ is the variance for the coefficient of the variable x

Hence, plotting $r_i(x)/A_i(x)$ v t_i may give some indication of the form of g(t).

Pettit and Bin Daud use various smoothing techniques for these plots to build an overall picture of the form of g(t).

However, from section 6.2.5 the residuals are influenced by the censoring observations, which will therefore effect the form of the plots. The form of g(t) can then only be reliably pictured in the plots when there are no censoring observations.

6.3 COX AND SNELL TYPE RESIDUALS

A graphical procedure for assessing the goodness-of-fit of the Proportional Hazards Model is provided by a plot of Cox and Snell type residuals, (Cox and Snell (1968)).

Methods based on those in Cox and Snell (1968) are used to obtain residual quantities which should, if the model fitted is appropriate, be consistent with a sample from the standard negative exponential distribution.

6.3.1 DEFINITION OF RESIDUALS

The residual quantities are defined as (Kay (1977)):

$$\hat{e}_i = \hat{H}(t_i;z_i)$$

$$= \hat{H}_0(t_i).\exp(\hat{\beta}z) \qquad 6.12$$

where $\hat{e}_i$ is the estimated residual at the failure time t_i.

$H_0(t_i)$ is the cumulative baseline hazard obtained from fitting the model.

In forming residuals based on the cumulative hazard there is a problem of how to incorporate tied failure points, and censored events.

An estimate of the hazard based on the failure points, which allows for tied failures, is used (see section 2.3.1).

Linear interpolation is employed between failure points to obtain an estimate of the cumulative hazard at censored times. The value of the censored residual is obtained from the resulting estimate of the cumulative hazard. Censored points occurring after the last failure are allocated a 'cumulative' hazard value equal to that of the last failure point.

6.3.2 DISTRIBUTION OF THE RESIDUALS

The estimated residuals, if there is no censoring, should look roughly like a random sample from the standard negative exponential distribution.

Consider 6.12

$$\hat{e}_i = \hat{H}(t_i) = \int_0^{t_i} \hat{h}(x)\, dx \qquad 6.13$$

hence,

$$\frac{d\,\hat{e}_i}{dt} = \hat{h}(t) \qquad 6.14$$

A probability density function f(t) is given by:

$$f(t) = \frac{-\,d\,S(t)}{dt} \qquad 6.15$$

where S(t) is the probability of an item surviving past time t, given that it has not failed prior to t. (The survivor function).

From the proportional hazards model:

$$S(t) = \exp\left[- \int_0^t \hat{h}(x)\ dx \right]$$

hence

$$\frac{d\ S(t)}{dt} = \exp\left[- \int_0^t \hat{h}(x)\ dx \right] . [- \hat{h}(t)]$$

then for the probability density function f(t) in 6.15

$$f(t) = \hat{h}(t).\exp\left[- \int_0^t \hat{h}(x)\ dx \right]$$

substituting from 6.13 and 6.14

$$f(t) = \frac{d\ e_i}{dt} . \exp(-e_i)$$

hence

$$\int f(t)\ dt = \int \exp(-e_i)\ d\hat{e}_i$$

The $\hat{e}_i$, therefore, are seen to have negative exponential distributional form with parameter 1.

6.3.3 USING THE RESIDUALS TO ASSESS THE FIT OF THE MODEL

We have seen in section 6.3.2 the distributional form the residuals can be expected to exhibit if the model fitted is appropriate. It is this property that forms the basis of this graphical assessment for the fit of the model.

A residual e_i for a censored observation at t_i is treated as a censored observation.

The residuals are reordered by size, and a product limit survivor function R(e) is obtained from the set of censored and uncensored residuals.

If the model is appropriate, then a plot of $-\ln \hat{R}(e_i)$ v $\hat{e}_i$ should be roughly linear with slope 1.

Consider a function $g(x) = \exp(-x)$:- c.f. form of the residuals.

A survivor function at x' is given by:

$$S(x') = \int_{x'}^{\infty} \exp(-x)\ dx = \exp(-x')$$

hence $$\ln S(x') = -x'$$

Our product limit survivor function $\hat{R}(e)$ should then display the property:

$$-\ln \hat{R}(e_i) = \hat{e}_i$$

Hence, if the plot is a straight line with gradient 1, then the residuals appear to be from a unit negative exponential indicating that the model is a good fit.

6.3.4 DEVIATION FROM THE EXPECTED LINE

Because of the treatment of the censored events after the last failure, over-estimation of ln R(e) can result in the latter part of the plot; moving it away from the expected straight line.

Figure 6.11 shows a Cox and Snell type residual plot for the hardware failure example. The deviation from the expected 45° line is increasing at the latter end of the plot.

It is not apparent, in the literature, to what extent agreement with the anticipated line should be expected.

Since the expected form of the residuals is known to be unit negative exponential, bounds can be simulated. Figure 6.12 shows 95% simulated bounds around the Cox and Snell type residual plot for the hardware failure example. The bounds are seen to widen as e increase.

The bounds in the plot have been estimated by simulating sets of residuals from the unit negative exponential distribution. For the 95% bounds shown in Figure 6.12 200 sets of residuals were simulated. For each residual in each set the corresponding ln R(e) was estimated. The bounds were formed by using the 6th and 195th largest estimated ln R(e) at any point on the residual axis.

This simulation technique could be used to form templates for Cox and Snell residual plots, showing bounds at various confidence levels. Since the number of residuals simulated within each set, equal to the number of events in the Cox and snell plot, will effect the width of the bounds separate templates will need to be prepared for different sample sizes.

Figure 6.13 shows such a template. This has been prepared for a sample size of 200. 99.9%, 99% and 95% bounds, indicated by +'s, x's and o's respectively, have been formed by simulating 2000 sets of residuals.

Computer time may, however, prove too costly to prepare such templates as a matter of course.

6.3.5 VARIANCE STABILISATION

Since the survivor function R(e) is a binomial probability; with the probability of any item surviving past a time t equal to p. The variance of the proportion of a sample size n surviving is given by p(1-p)/n.

In this instance the sample size n is not constant (reduction by 1 as each item 'dies'), therefore the variance in R(e) is not constant. Hence the variability in the plot $-\ln \hat{R}(e)$ v e increases as e gets larger.

An angular transformation; $\sin^{-1}\sqrt{x}$ has the effect of stabilising the variance in an estimated proportion x with binomial variace p(1-p)/n (Bartlett (1947)).

A variance stabilised form of the Cox and Snell residuals can then be presented by plotting $\sin^{-1}\sqrt{R(e_i)}$ v $\sin^{-1}(\exp(-e_i/2))$, (Aitkin and Clayton (1980)). The variance of the new varianced stabilised function is given approximately by 1/4N, when $\sin^{-1}$ is measured in radians, and N is the size of the original sample, (Bartlett (1947)).

Figure 6.14 is the variance-stabilised plot of the Cox and Snell type residuals for the hardware failure example. Because of the variance stabilisation the plot is seen to adhere more closely to the expected 45° line.

6.4 INFLUENCE FUNCTIONS

Empirical influence functions can be used in an informal manner to identify observations which may greatly effect statistical inferences regarding the covariates.

The technique approximates the influence of individual observations on each of the ß coefficients.

6.4.1 CALCULATION OF INFLUENCE

The influence can be obtained in an exact manner by dropping each observation in turn and refitting the model.

In most practical applications, however, this leads to a prohibitive requirement of computer time.

Cain and Lange (1984) employ a first order approximations, based on a Taylor series expansion, to $\hat{\beta} - \hat{\beta}_{(j)}$.

Where $\hat{\beta}$ estimate of β with all observations.

$\hat{\beta}_{(j)}$ estimate of β with j'th observation removed.

Hence, $\hat{\beta} - \hat{\beta}_{(j)}$ is the influence of the j'th observation on the estimate of β.

This representation of the influence is found to comprise Schöenfeld's partial residual (see section 6.2.1) and a further component which is the effect an item has on the β coefficient via all the risk sets that contain the item. (Wightman (1987)).

6.4.2 GRAPHICAL REPRESENTATION OF INFLUENCE

The influence function can be represented graphically by plotting the standardised influence against the rank of survival time (Cain and Lange (1984)), or by plotting influence against covariate value (Reid and Crepeau (1985)).

However, in order to identify single observations which may be so influential as to alter the significance of the covariate if omitted, we plot the estimate of the changed z-score against the rank of the observation.

Estimating the change in the z-score can be achieved if we can assume that the variance-covariance matrix is not fundamentally altered when one observation is omitted. This assumption is already made in the calculation of the estimated influence function. (Wightman (1987)).

$$z - \hat{z}_{(j)} = \frac{\hat{I}_{(j)}}{St(\hat{\beta})}$$

where z is the z-score obtained when all the observations are included

$\hat{z}_{(j)}$ is the estimated z-score when the j'th observation is omitted

$\hat{I}_{(j)}$ is the calculated influence of the j'th observation

$St(\hat{\beta})$ is the standard deviation of $\hat{\beta}$ obtained from the variance-covariance matrix

Figure 6.15 shows the influence plot for the covariate 'average hours use per week' from the hardware failure example, z_j v rank. The majority of the points occur around the value of the original z-score. These are observations with little influence, and are very often censored events. There are no observations in this example so influential as to alter the significance of the covariate if omitted (values in the region <1.96 since +'ve β).

Figure 6.16 shows the influence plot for the covariate '4 processors' from the hardware failure example. For this binary covariate there are three main groups of influence points. The central group centred about the original z-score largely comprises censored events. The other two groups are largely due to the two levels of the covariates on the failure times.

Figure 6.1 Vertical separation in proportionality plotting

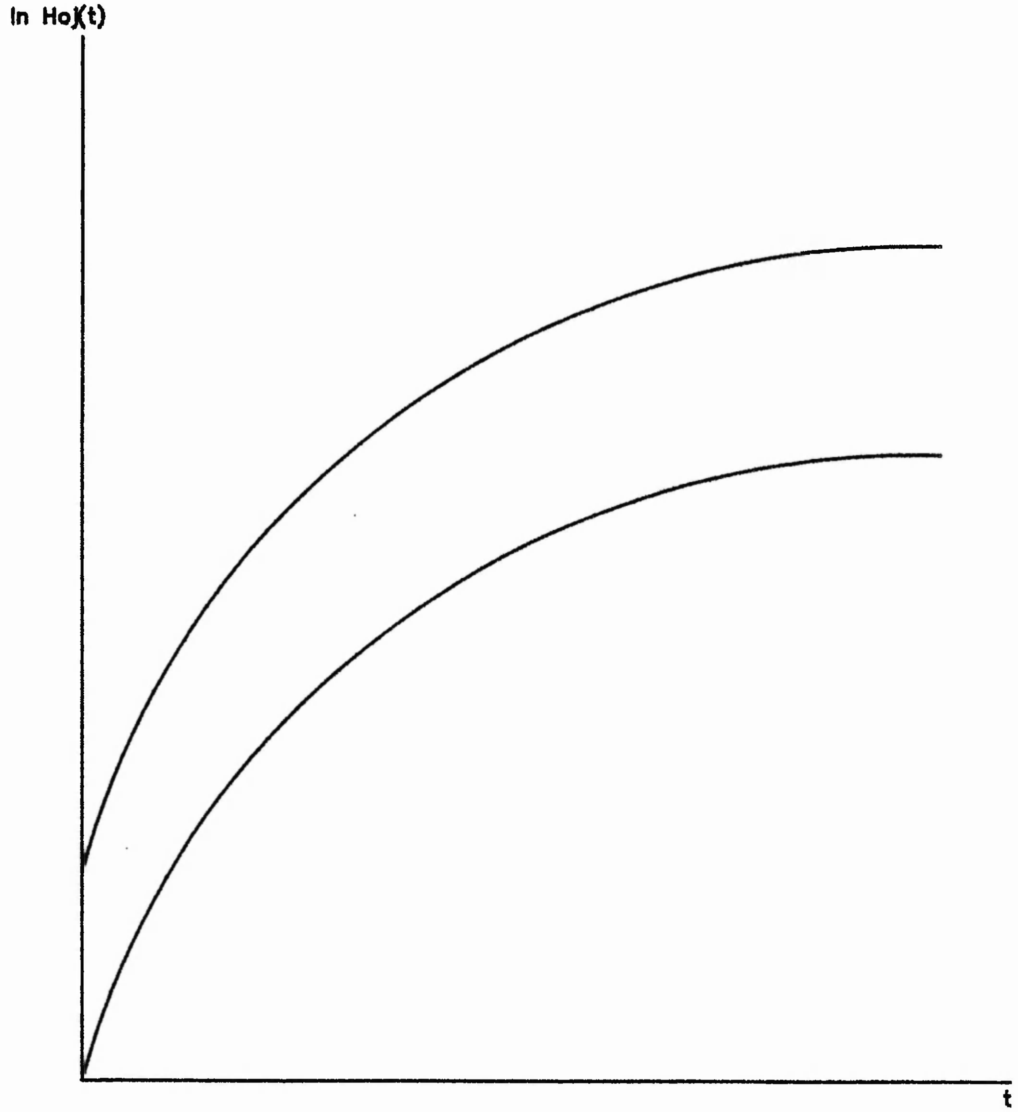


Figure 6.2 Vertical separation plotted below proportionality plot, from the hardware failure example. Covariate 'No fault found.'

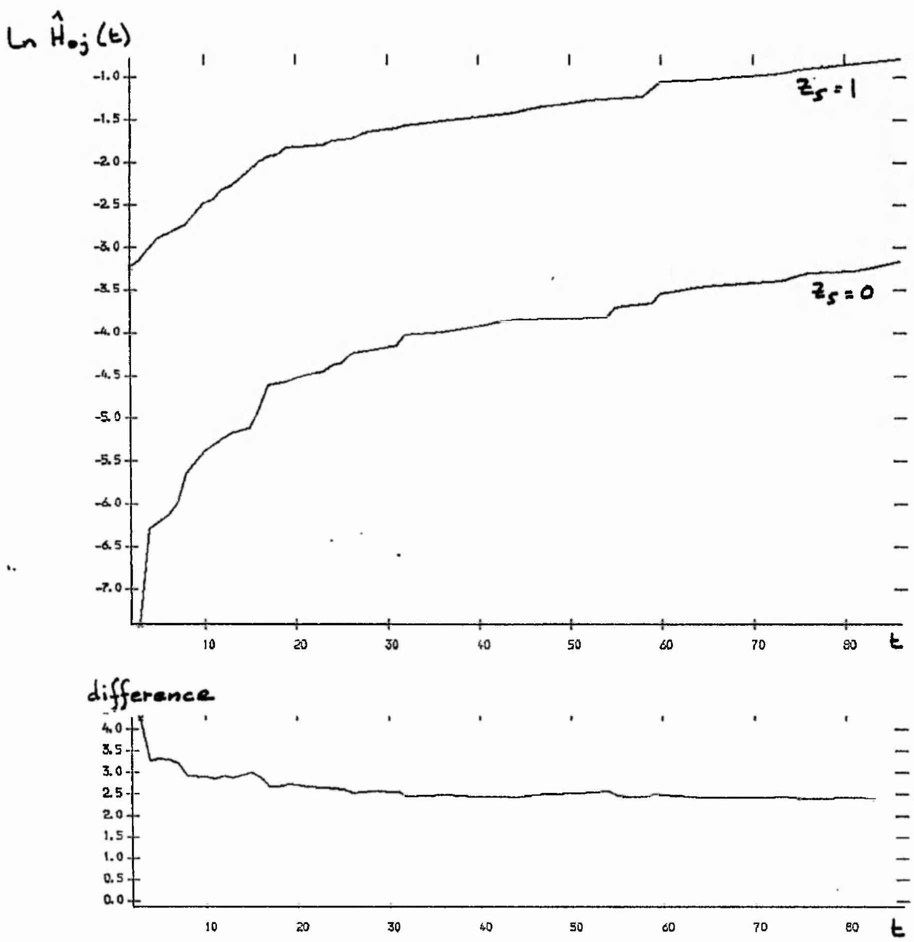


Figure 6.3 95% Link's bounds around proportionality plot for the hardware failure example. Covariate 'No fault found'.

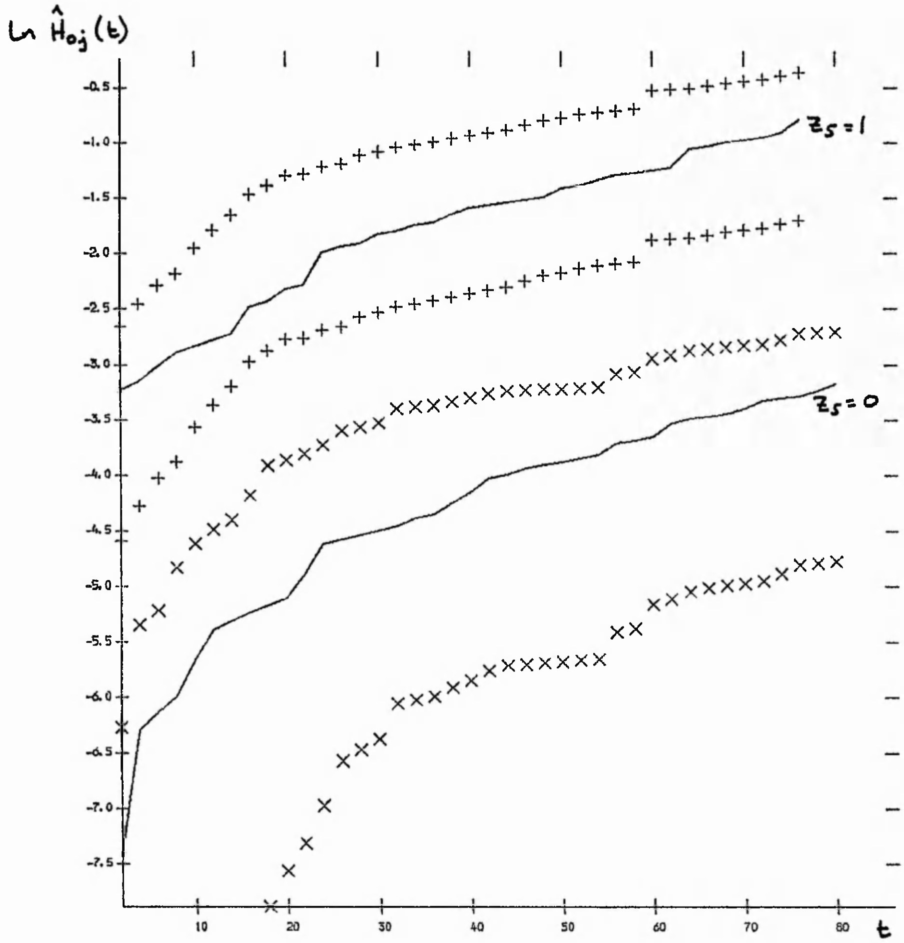


Figure 6.4 95% Simulated Weibull bounds around proportionality plot for the hardware failure example. Covariate 'No fault found.'

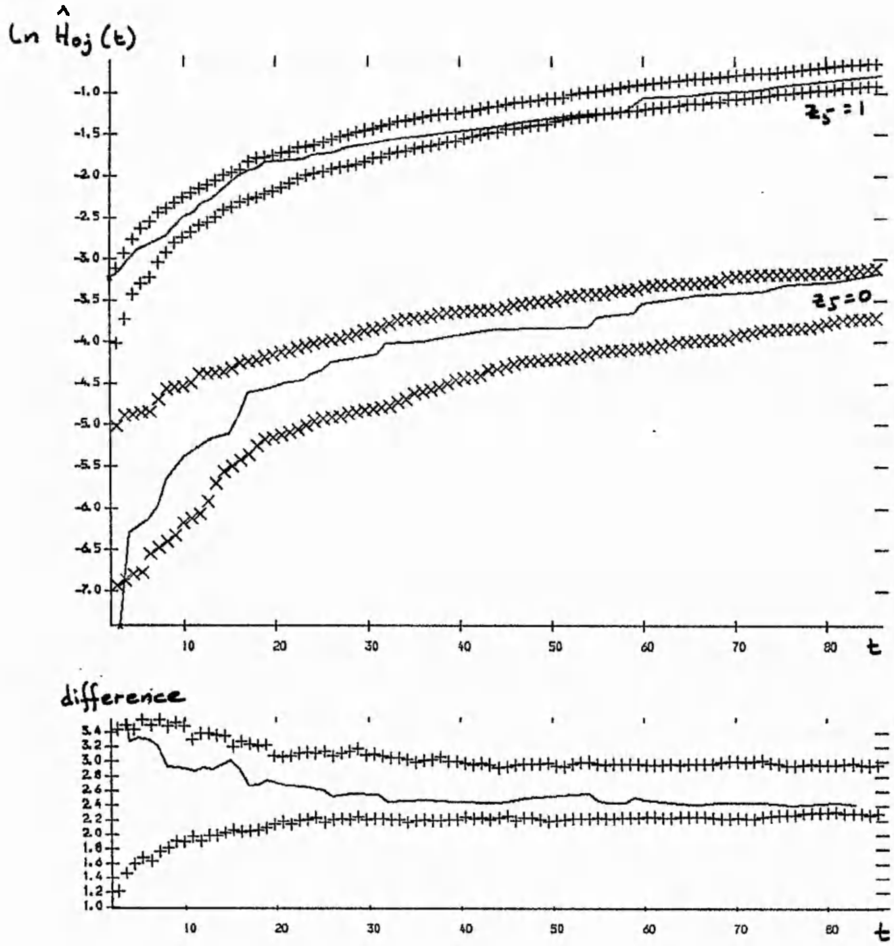


Figure 6.5 Schoenfeld partial residuals for covariate 'average hours use per week' from the hardware failure example. Moving average over groups of twenty observations.

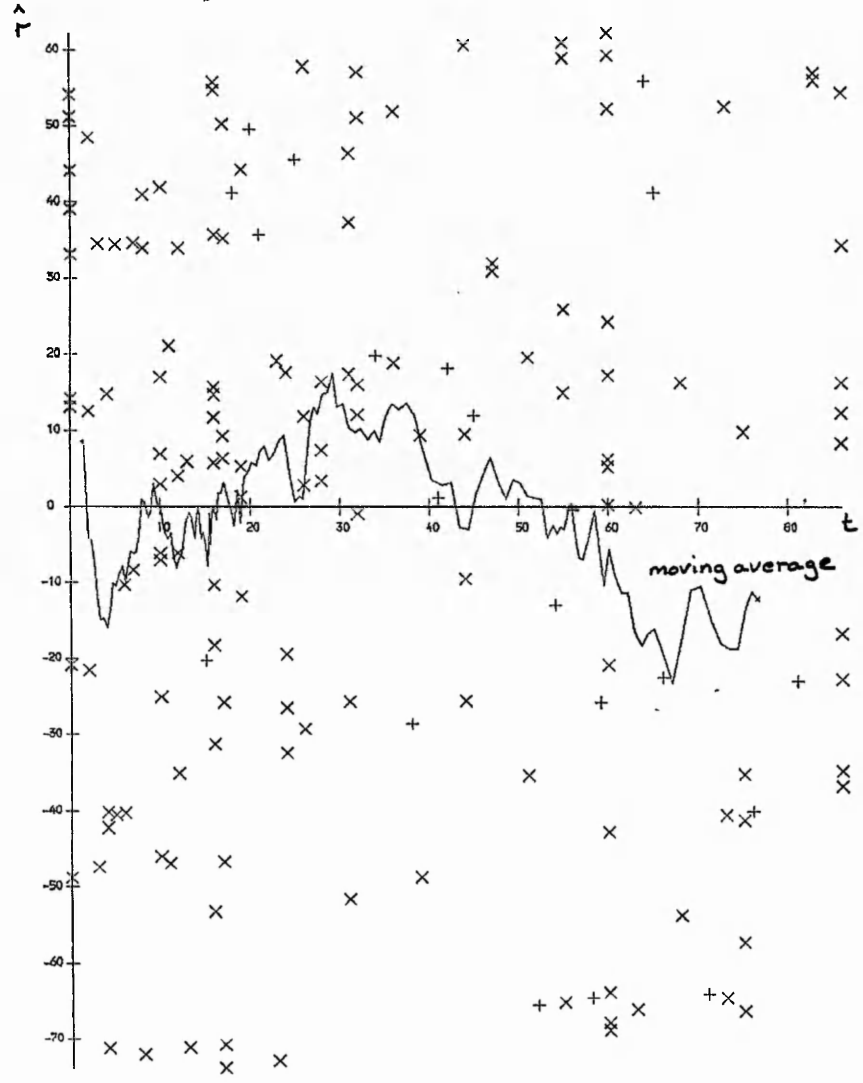


Figure 6.6 Schoenfeld partial residuals for covariate '2 processors' from the hardware failure example. Moving average over groups of twenty observations.

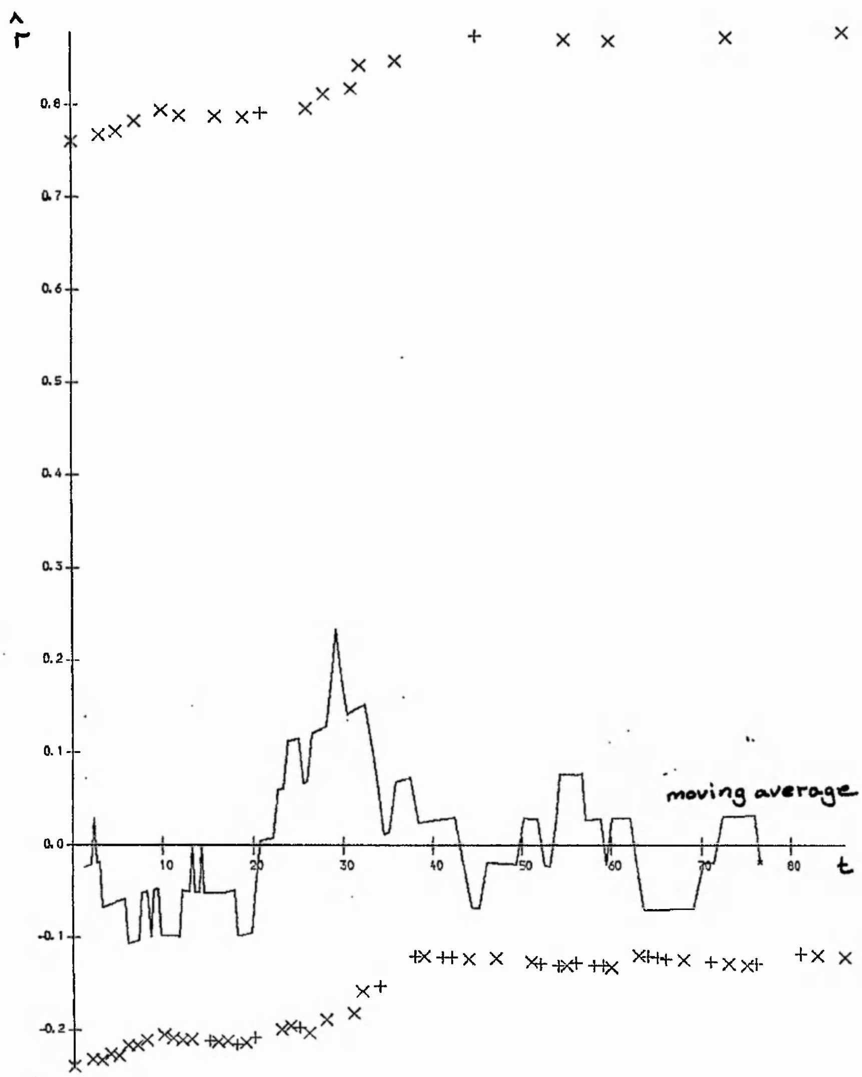


Figure 6.7 Schoenfeld partial residuals for covariate with positive estimate of β. 'Left truncation' from the hardware failure example.

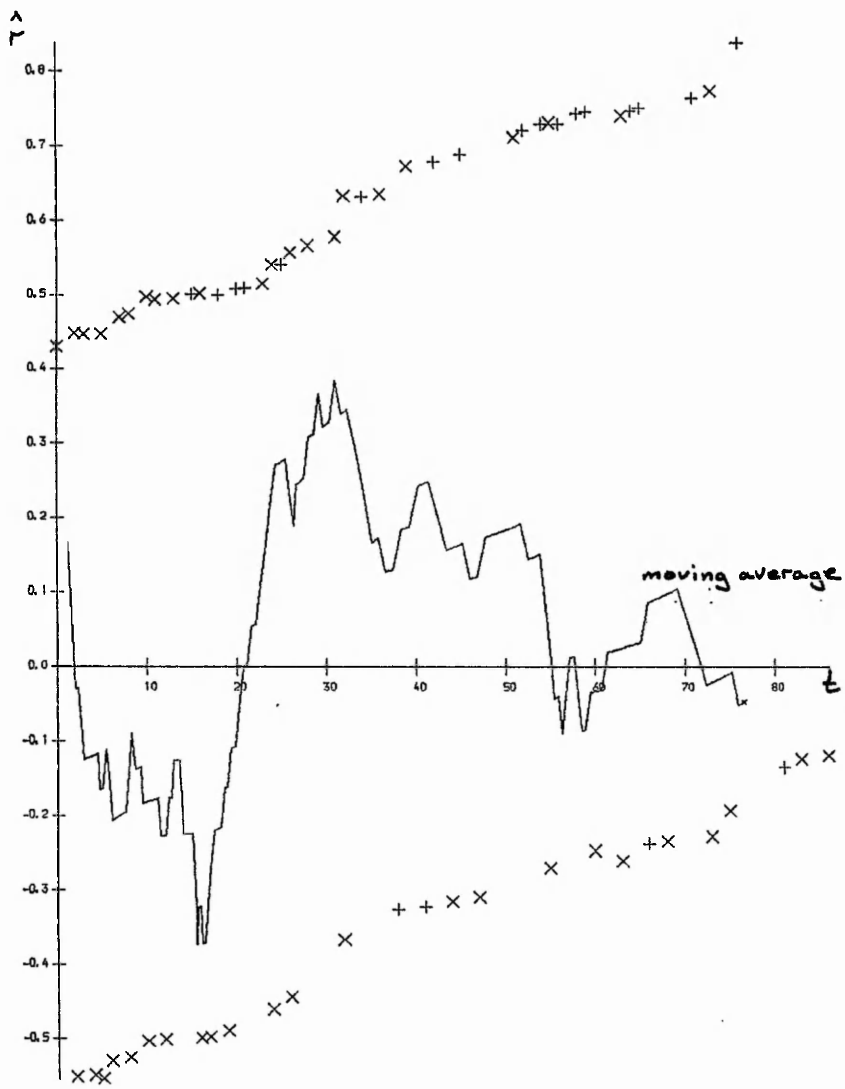


Figure 6.8 Schoenfeld partial residuals for covariate with negative estimate of β. 'System B' from the hardware failure example.

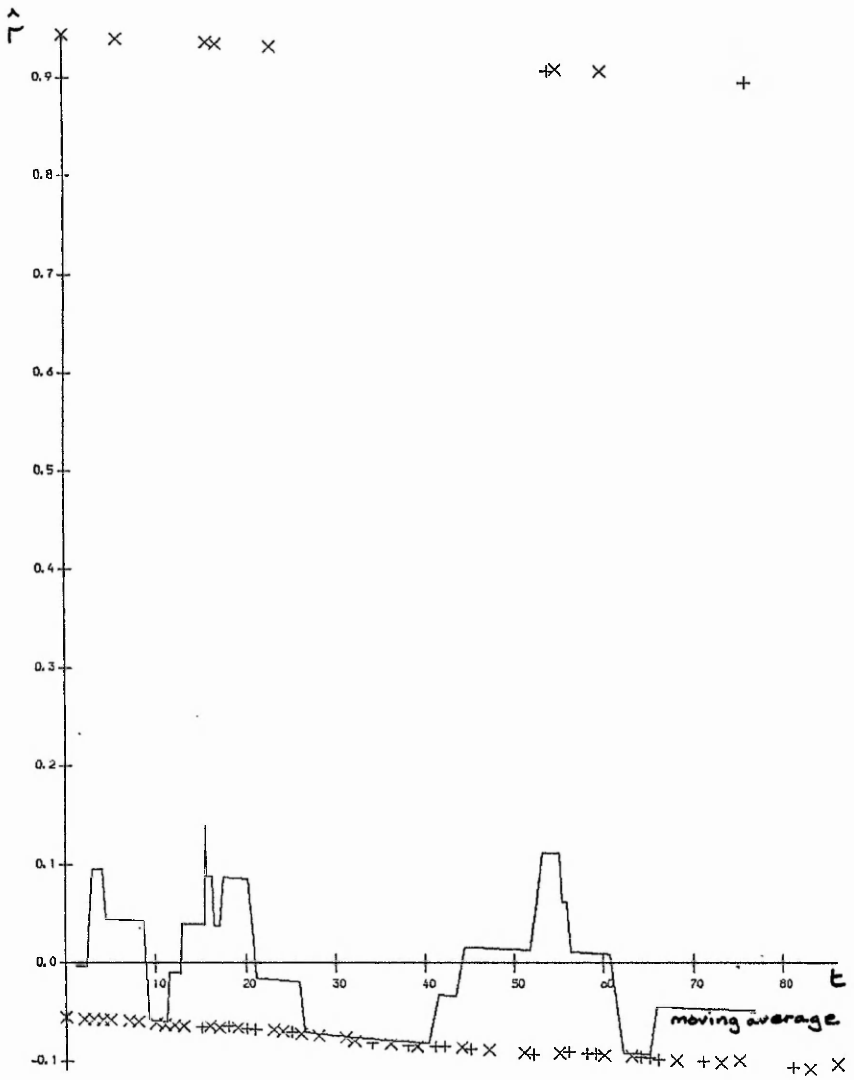


Figure 6.9 Schoenfeld partial residuals for covariate with negative estimate of β. 'System F' from the hardware failure example.

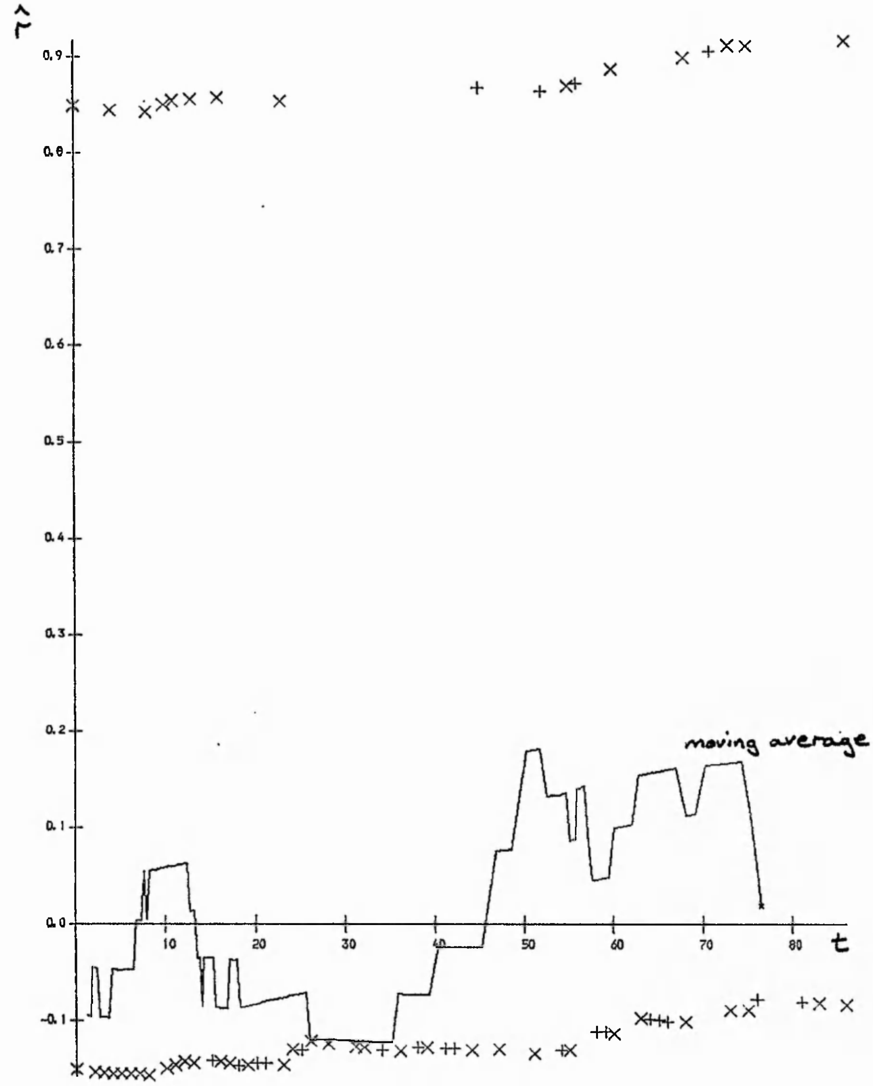


Figure 6.10 Schoenfeld partial residuals for competing risks covariate 'No fault found', from the hardware failure example.

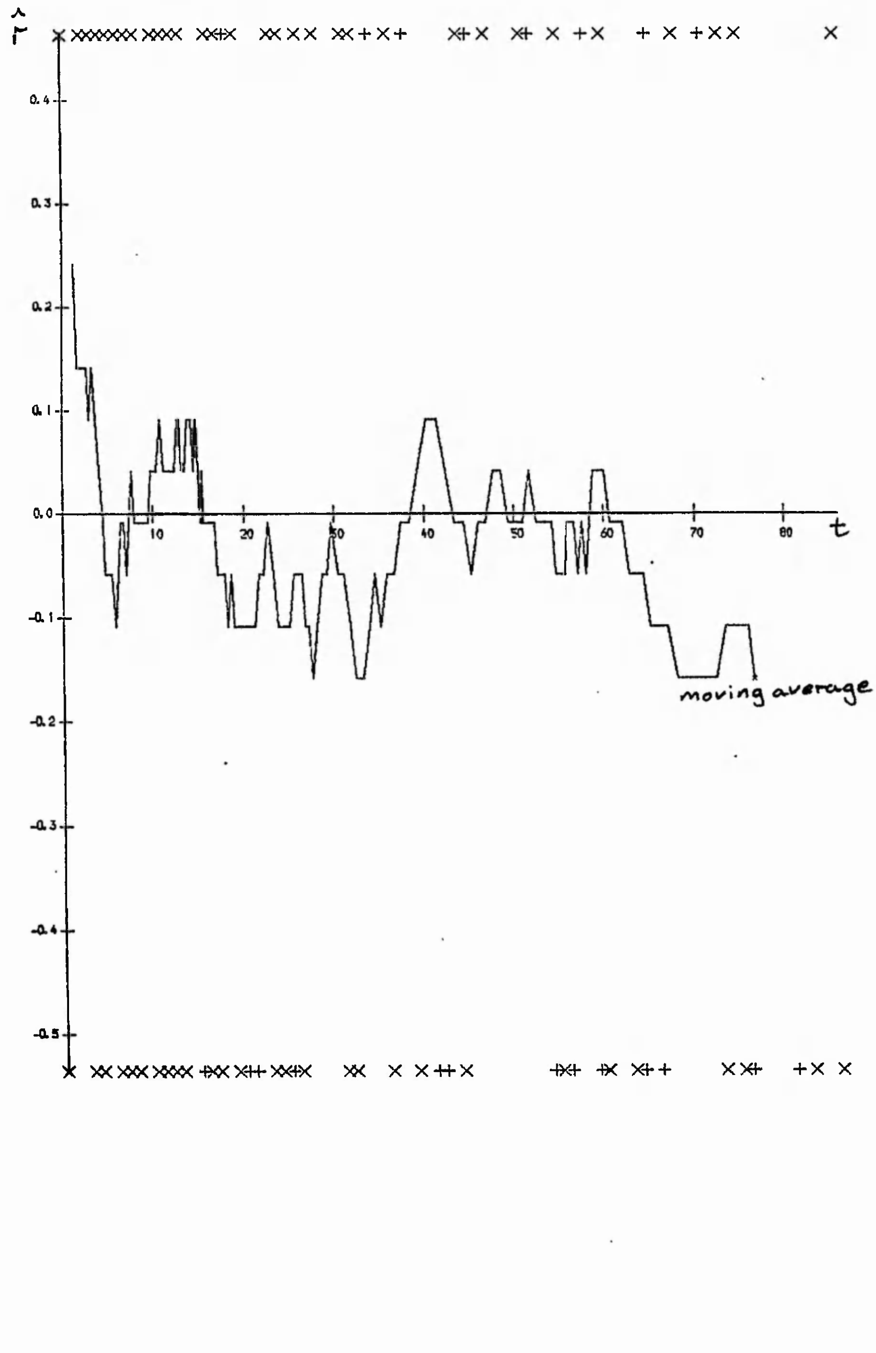


Figure 6.11 Cox and Snell type residual plot for the hardware failure example.

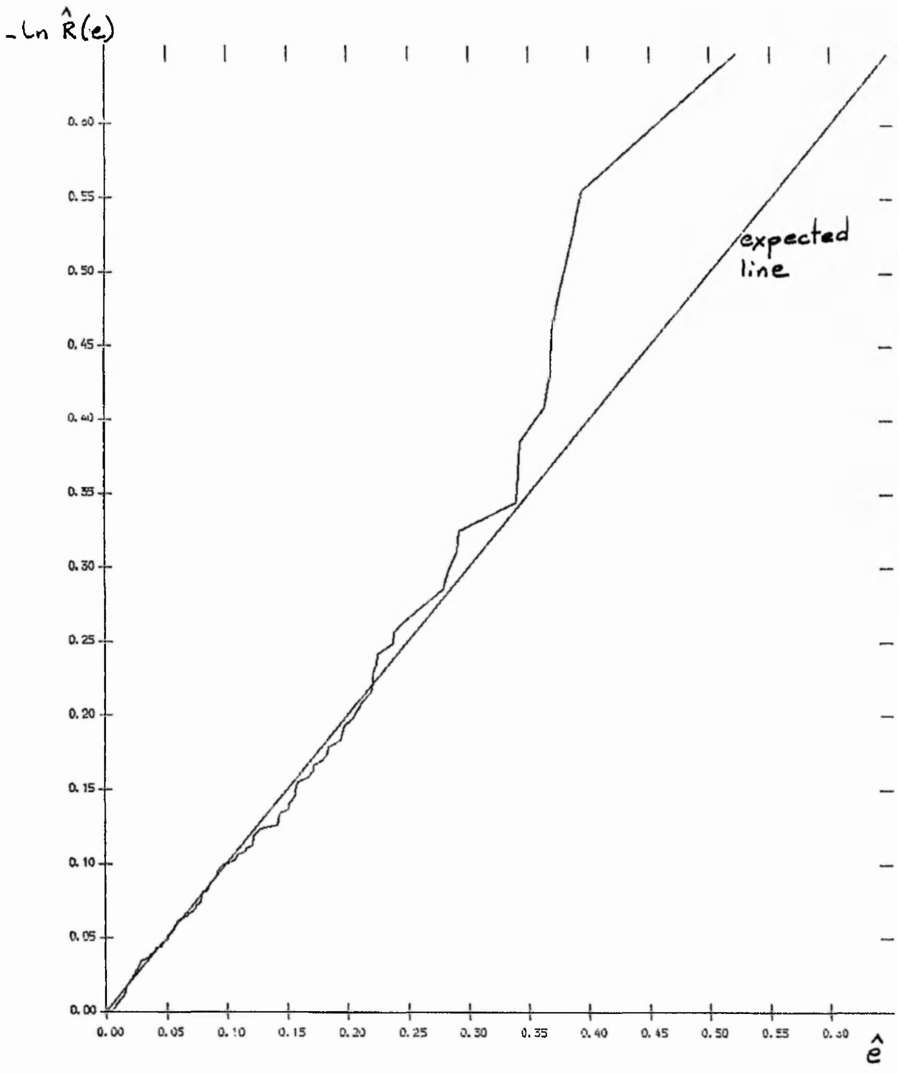


Figure 6.12   95% Simulated bounds around the Cox and Snell type residual plot for the hardware failure example.

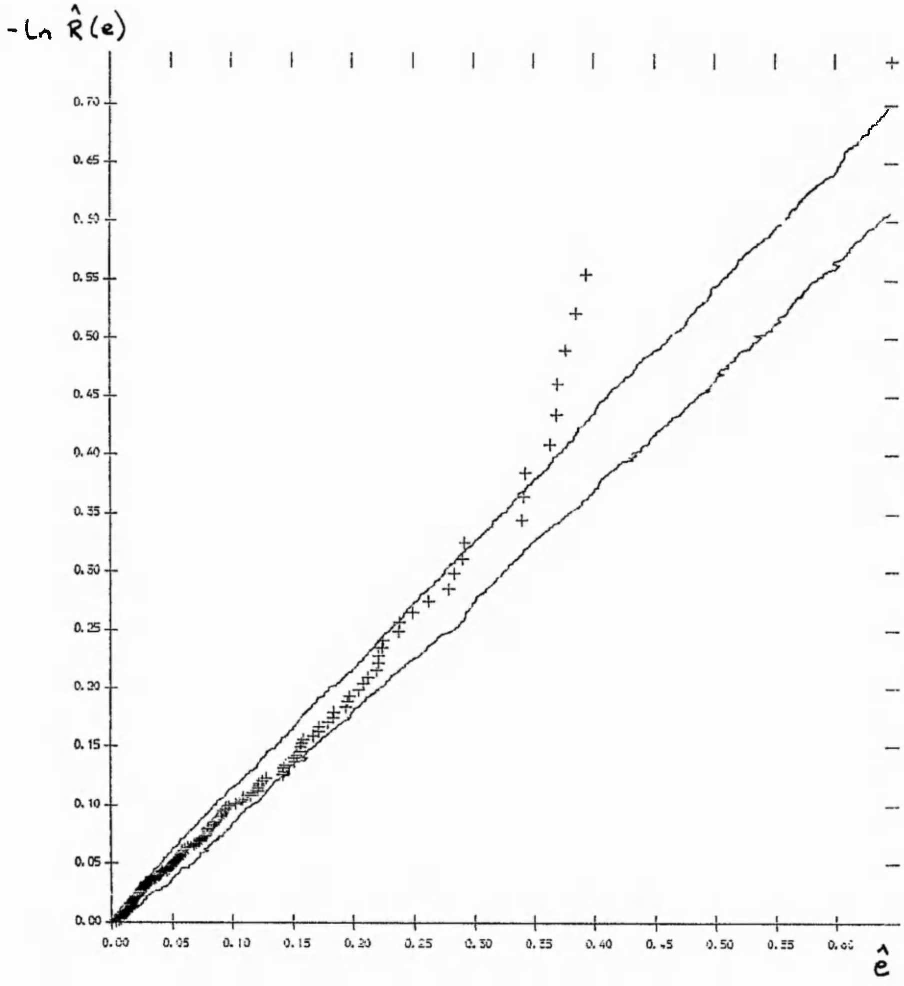


Figure 6.13 Simulated confidence bound template for Cox and Snell residual plot.

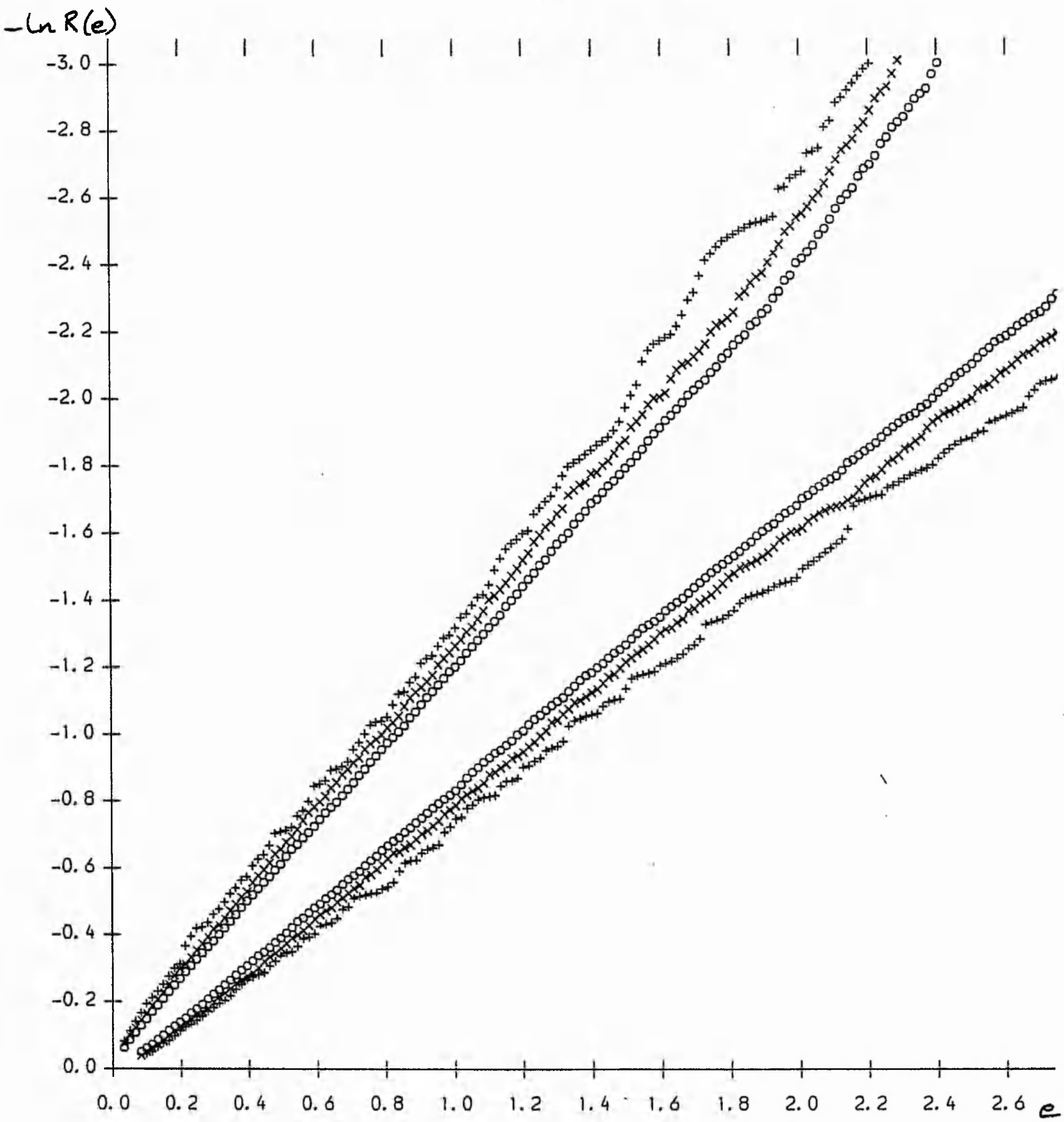


Figure 6.14 Variance-stabilised plot of Cox and Snell type residuals for the hardware failure example.

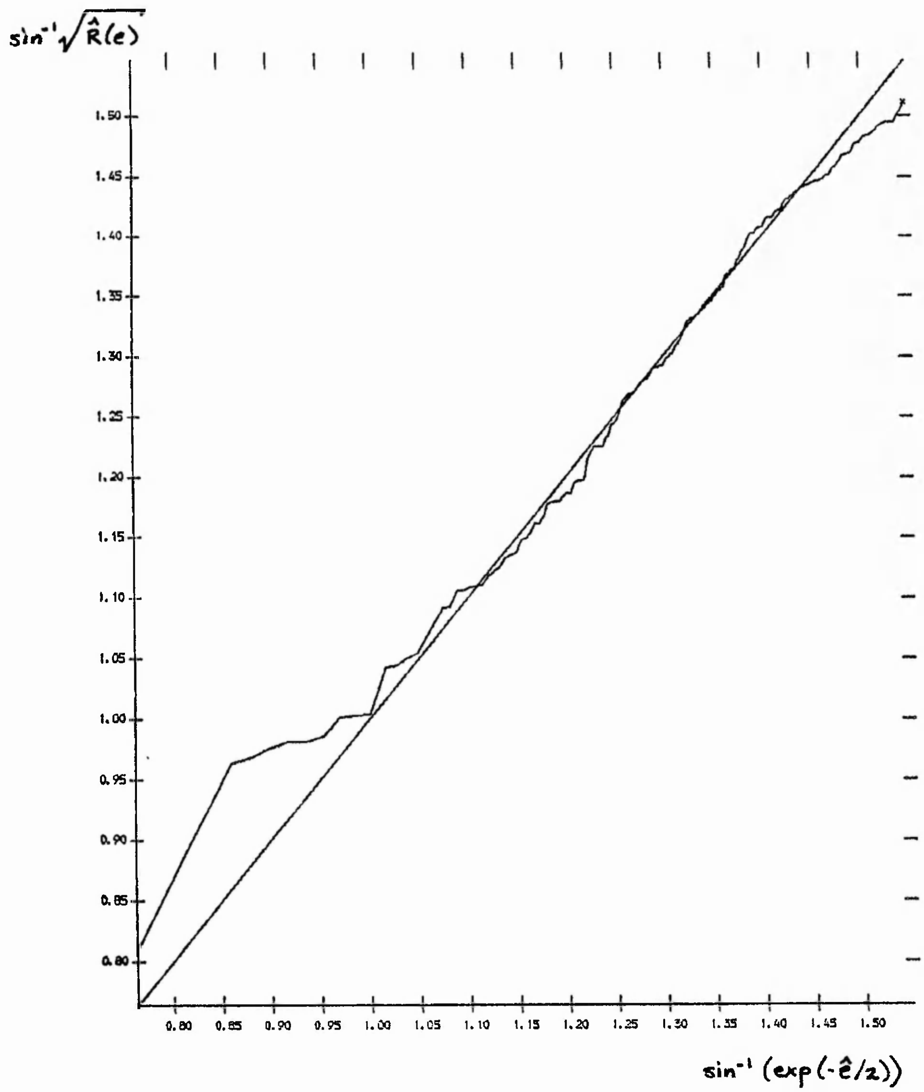


Figure 6.15 Influence plot for the covariate 'average use per week' from the hardware failure example.

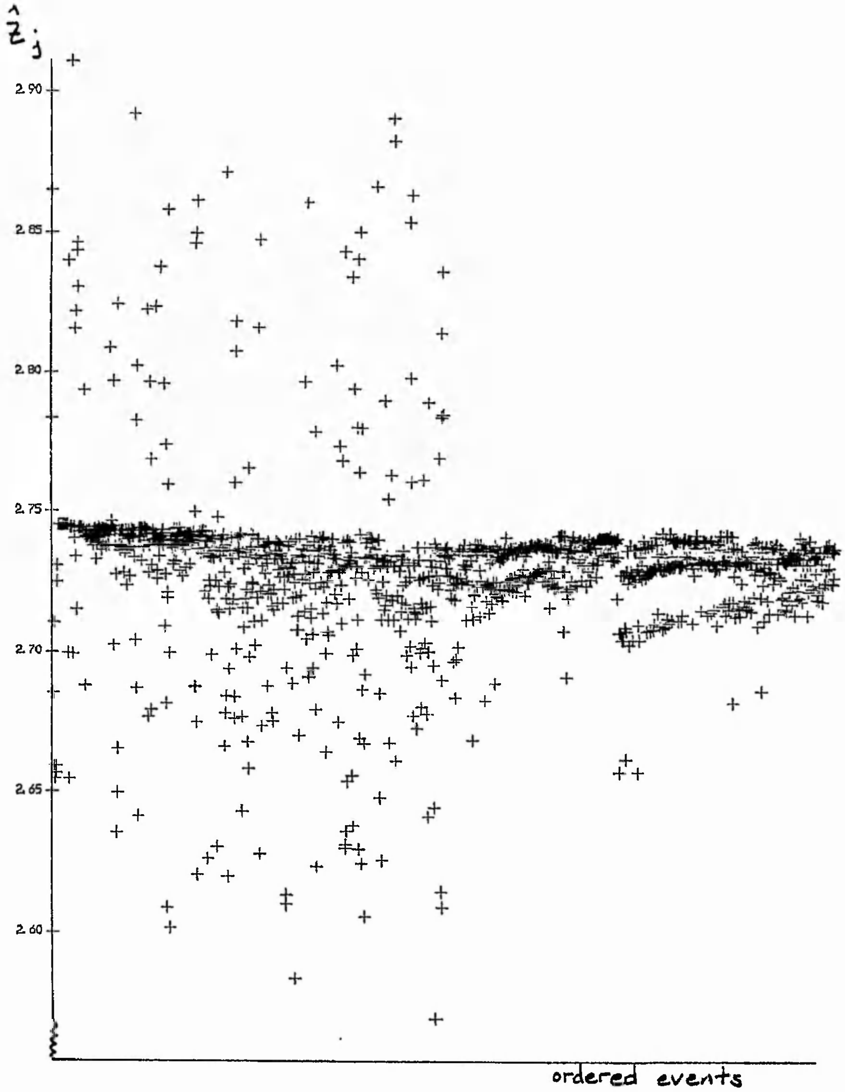
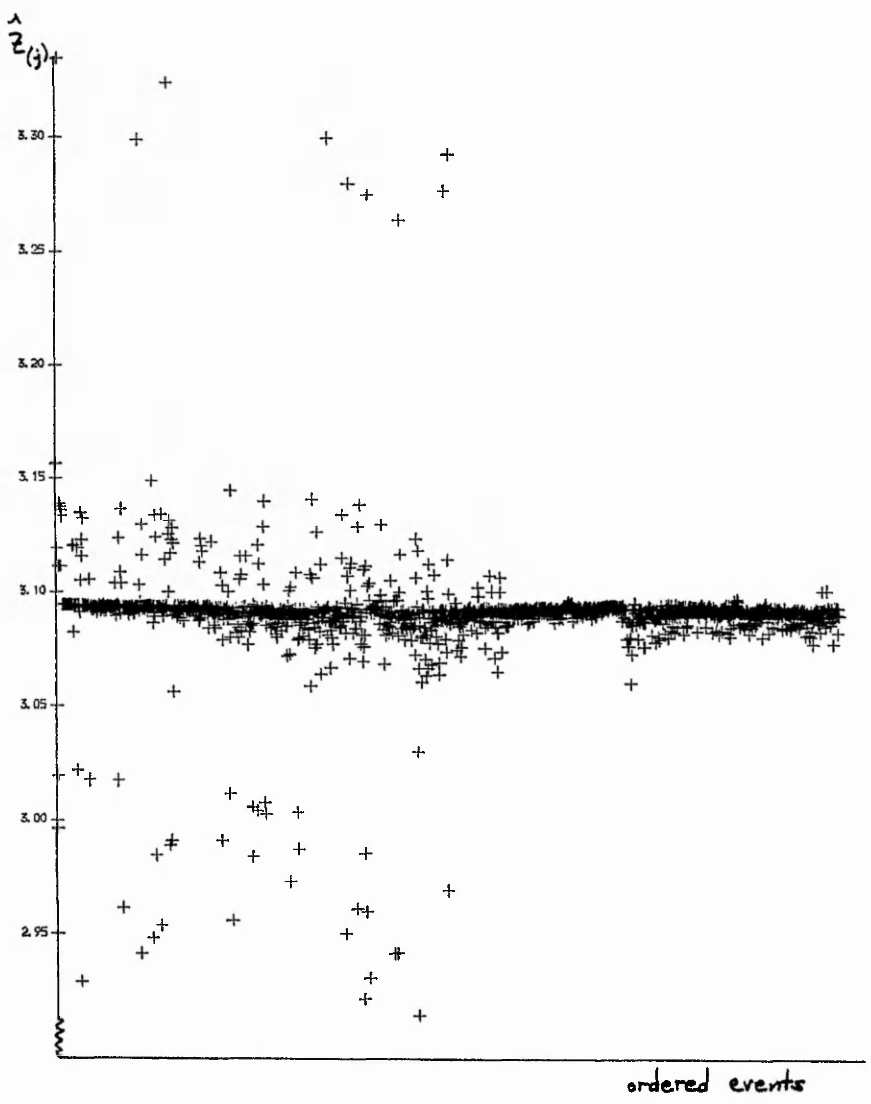


Figure 6.16 Influence plot for the covariate '4 processors' from the hardware failure example.



CHAPTER 7

CONCLUSIONS AND CONTRIBUTIONS TO KNOWLEDGE ACHIEVED

7.1 SUMMARY OF APPROACHES TO VARIOUS FIELD DATA STRUCTURES

Chapters 3-5 have given examples of the application of PHM to reliability field data.

Each data set has had its own nuances, and provided different problems for their analyses.

The data structures reflect the data collection processes and procedures as well as field deployment and failure phenomena.

The three examples, therefore, have led to illustrate that there is no universal form of reliability data, and that any data set has its own unique features.

7.1.1 IDENTIFICATION OF POINT PROCESSES

Appropriate structure for modelling and analysis centres about the identification of appropriate point processes to describe failures of repairable systems.

A careful choice of the basic time metric, t, must be made. This may be reflective of the maintenance and repair procedures.

It is often assumed, for simplicity, that an item is repaired to as-good-as-new.

It is of course also necessary to ensure that the choice of the time metric is readily obtainable from the data.

In a hierarchical system the choice of different point processes, on which to base PHM, is greatly increased. This is because the series of events may now be applied to, and observed on, any level of the hierarchy.

7.1.2 CENSORING STRUCTURES

Having adopted the use of an appropriate point process there is now a need to adopt a sensible censoring structure.

Censoring is used when an item leaves the field of observation without having failed at this point.

A great deal of information can be lost by not using a censoring structure.

The simplest form of censoring is used in introducing a censored observation for right truncated failure times. For example, if we know when failure recording ceased; when using a metric, t, of times between failures, a censored time is used from the last failure to the termination of recording.

For systems with competing risks censoring can be used to a greater extent.

Partial or complete censoring could be used in such situations. This is because although an item has failed a distinction can be made between modes of failure.

Again, this is more complex in hierarchical systems since numerous failure mechanisms may be assumed at various hierarchical levels.

7.1.3 EXPLORATORY USE OF PHM

Since there is a multitude of PHM models which might be applied to any reliability field data, PHM can be used as a powerful exploratory tool.

In practice we use PHM, moving from one model to another searching for a better fit, and more explanatory power.

Although the method identifies relationships between the life length of equipment and covariates fitted, these need not be causal relationships. Such relationships identified may be masking underlying patterns which serve for more explanation to the observed data.

The example in Chapter 3 illustrates this phenomenon clearly.

7.1.4 TRANSFORMATIONS

It is usual practice in reliability analysis to include covariate infromation in the form it was recorded.

However, it may be the case that alternative formulations for the covariate are more appropriate.

For example, a covariate such as failure number is often fitted with the formulation N (where N is the failure number). Better formulations of the covariate, such as $\sqrt{N}$, N^2, ln N etc. may provide greater explanation of the processes, see Davies et al (1987), or lead to the proportionality assumption being fulfilled where previously it had not.

Transformations for covariates are usually incorporated into models as part of the exploratory process of stepping through one model to, hopefully, a better model.

Figure 7.1 summarises the approach which should be taken in applying PHM to field data of a repairable system, in the form of a flow diagram.

7.2 SUMMARY OF USE OF GRAPHICAL DIAGNOSTICS

We have seen clearly the need for assessing the appropriateness and fit of PHM models. In this thesis we have investigated in particular four graphical techniques suggested in the literature.

7.2.1 PROPORTIONALITY PLOTTING

It has been standard in the literature to assess the proportionality assumption for different levels of a covariate by stratifying at each level and plotting the logarithm of the baseline cumulative hazard for each stratum against time on the same graph. However, we have shown that looking for constant vertical separation in this plot (true for proportional hazards) is difficult. We have, therefore, in this thesis additionally plotted the vertical separation at failure points.

Also to give an indication as to the extent we might expect the plot to deviate from constant vertical separation, we have simulated confidence bounds for the plot (assuming Weibull baseline hazards).

7.2.2 SCHÖENFELD PARTIAL RESIDUALS

It has been suggested in the literature to look at the local fit of the proportionality assumption by plotting, for each covariate, the Schöenfeld partial residuals against time. These residuals should have a general scatter about the axis. Additionally to the literature we have, in this thesis, plotted a moving average based on groups of a large number of observations superimposed on to each graph. This has eased visual inspection of the plot to assess local fit of the assumption.

7.2.3 COX AND SNELL TYPE RESIDUALS

The literature has suggested the use of these residual quantities to assess the global fit of the model. The basis of this graphical test is essentially to compare observed residuals to expected values from the model.

Since no indication is given in the literature as to how much we may expect the plot to deviate from the expected 45° line we have investigated possible bounds for the plot by simulation.

7.2.4 INFLUENCE FUNCTIONS

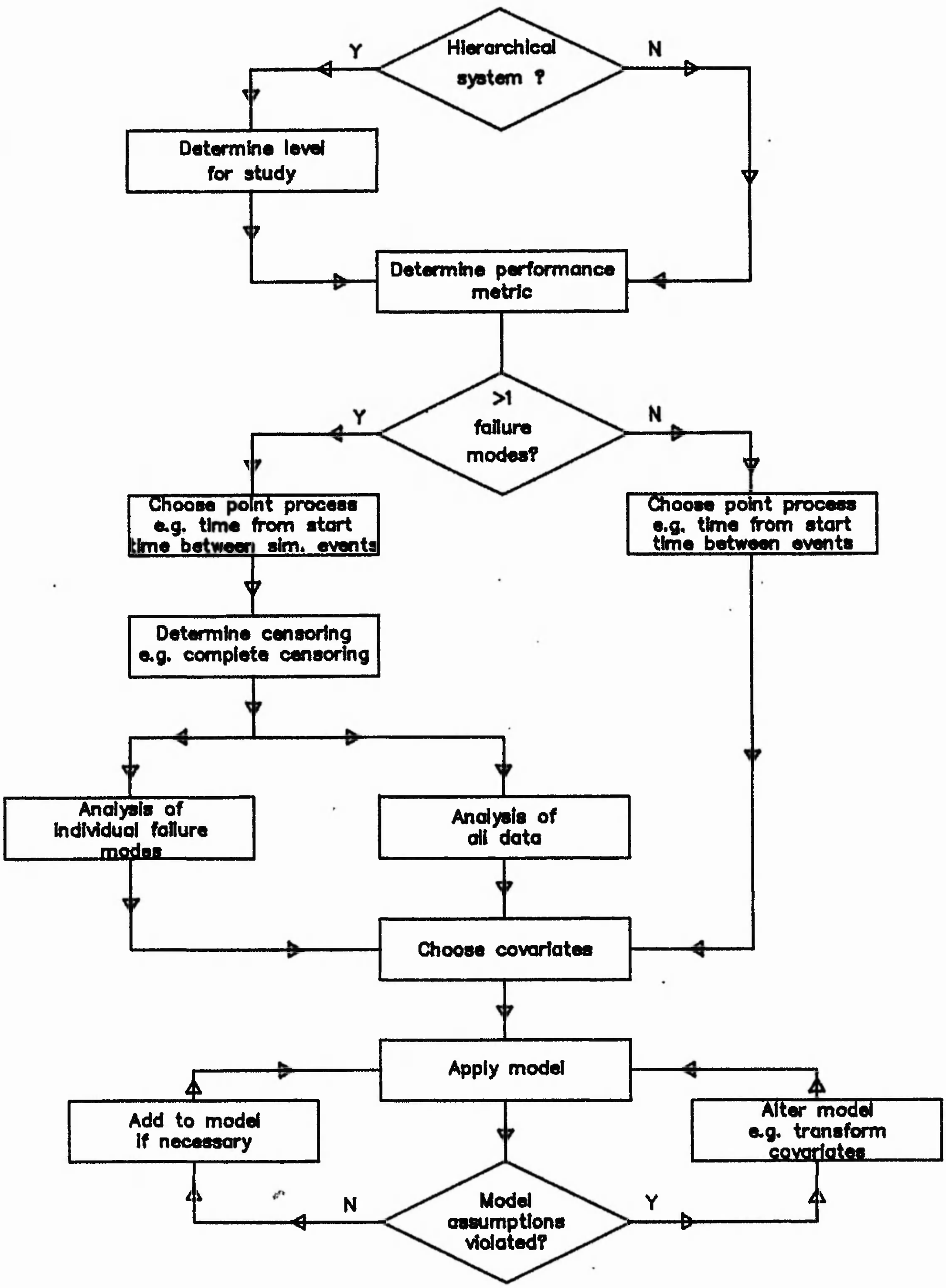
It is common in regression techniques to assess the influence of individual observations on a model by eliminating each observation one at a time and refitting the model.

The influence functions presented in this thesis employ an approximation to the effect of the above technique.

It has been standard in the literature to plot the difference between the coefficient estimate, for each covariate, when the observation is and when it is not included in the model, against the rank of the omitted observation.

However, following the suggestion of Wightman (1987) we plot the z-score for the coefficient obtained when the observation is omitted, against each observation. This enables us to clearly identify observations that may be so influential as to alter the significance of the covariate.

Figure 7.1 Summary of approach to applying PHM to repairable system field data



REFERENCES

AITKIN, M. and CLAYTON, D. (1980). The fitting of exponential, Weibull and extreme value distributions to complex censored survival data using GLIM. Appl. Statist., 29, 156-163.

ANDERSEN, P.K. (1982). Testing goodness of fit of Cox's regression and life model. Biometrics, 38, 67-77.

ARJAS, E. (1987). A graphical method for assessing goodness of fit in Cox's proportional hazards model. J. Am. Statist. Assoc., Vol.83, No.401, 204-212.

ASCHER, H.E. and FEINGOLD, H. (1984). Repairable systems Reliability: Modelling, Inference, Misconceptions and their Causes. Marcel Dekker, New York.

BARTLETT, M.S. (1947). The use of transformations. Biometrics, 39-52.

BRESLOW, N.E. (1974). Covariate analysis of censored survival data. Biometrics, 30, 89-99.

BRYSON, M.C. and JOHNSON, M.E. (1981). The incidence of monotone likelihood in the Cox model. Technometrics Vol. 23, 381-383.

CAIN, K.C. and LANGE, N.T. (1984). Approximate case influence for the proportional hazards regression model with censored data. Biometrics, 40, 493-499.

COX, D.R. (1962). Renewal Theory. Methuen, London.

COX, D.R. (1972). Regression Models and life-tables (with discussion). J.R. Statist. Soc. B, 34, 187-220.

COX, D.R. (1972a). The statistical analysis of dependencies in point processes. In Stochastic Point Processes, P.A.W. Lewis, Ed., Wiley, New York.

COX, D.R. (1975). Partial likelihood. Biometrika, 62, 269-276.

COX, D.R. and LEWIS, P.A.W. (1966). The Statistical Analysis of Series of Events. Methuen, London.

COX, D.R. and MILLER, H.D. (1965). The Theory of Stochastic Processes. Methuen, London.

COX, D.R. and OAKES, D. (1984). Analysis of Survival Data. Chapman and Hall, London.

COX, D.R. and SNELL, E.J. (1968). A general definition of residuals (with discussion). J.R. Statist. Soc. B, 30, 248-275.

DAVIES, N., MARRIOTT, J.M., WIGHTMAN, D.W. and BENDELL, A. (1987). The Musa data revisited: alternative methods and structure in software reliability modelling and analysis. SARSS '87, Elsevier, London.

DRURY, M.R., WALKER, E.V., WIGHTMAN, D.W. and BENDELL, A. (1988). Proportional hazards modelling in the analysis of computer systems reliability. Reliability Engineering and system Safety Journal, 21, 3.

FELLER, W. (1968). An Introduction to Probability Theory and its Applications. Wiley, New York.

GAIL, M.H., SANTER, T.J. and BROWN, C.C. (1980). An analysis of comparitive carcinogenesis experiments based on multiple times to tumor. Biometrics, 36, 255-266.

GILL, R. and SCHUMACHER, M. (1987). A simple test of the proportional hazards assumption. Biometrika, 74, 2, 289-300.

GRAY, C., HARRIS, N., BENDELL, A. and WALKER, E.V. (1988). The reliability analysis of weapon systems. Reliability Engineering and System Safety Journal, 21, 4.

HOLT, J.D. (1978). Competing risk analysis with special reference to matched pair experiments. Biometrika, 65, 159-165.

KALBFLEISCH, J.D. and PRENTICE, R.L. (1973). Marginal likelihoods based on Cox's regression and life model. Biometrika, 60, 267-278.

KALBFLEISCH, J.D. and PRENTICE, R.L. (1980). The Statistical Analysis of Failure Time Data. John Wiley and Sons, Chichester.

KAY, R. (1977). Proportional hazards regression models and the analysis of censored survival data. Appl. Statist. 26, 227-237.

LAGAKOS, S.W. (1981). The graphical evaluation of explanatory variables in proportional hazards regression models. Biometrika, 68, 93-98.

LAWLESS, J.F. (1982). Statistical Models and Methods for Lifetime Data. Wiley, New York.

LAWLESS, J.F. (1983). Statistical methods in reliability (with discussion). Technometrics, Vol. 25, No. 4, 305-335.

LEWIS, P.A.W. (1972): Stochastic Point Processes. Wiley, New York.

LINK, C.L. (1984). Confidence intervals for the survival function using proportional hazards model with covariates. Biometrics, 40, 601-610.

MOREAU,T., O'QUIGLEY, J. and MESBAH, M. (1985). A global goodness-of-fit statistic for the proportional hazards model. Appl. Statist. 34, 212-218.

NOWAKOWSKI, T. (1986). The influence of the operating environment on bus reliability. Reliability Technology - Theory & Applications. J. Moltoft and F. Jensen Ed. Elsevier, Holland.

PETTIT, A.N. and BIN DAUD, I. (1987). Using Schöenfeld's partial residuals to investigate time dependence in Cox's proportional hazards model. Mathematics Report Number 201. Loughborough University of Technology.

PRENTICE, R.L. and GLOECKLER, L.A. (1978). Regression analysis of grouped survival data with application to breast cancer data. Biometrics, 34, 57-67.

REID, N. and CREPEAU, H. (1985). Influence functions for proportional hazards regression. Biometrika, 72, 1-9.

SCHOENFELD, D. (1980). Chi-squared goodness of fit tests for the proportional hazards regression model. Biometrika, 67, 145-153.

SCHOENFELD, D. (1982). Partial residuals for the proportional hazards regression model. Biometrika, 70, 643-649.

STORER, B.E. and CROWLEY, J. (1985). A diagnostic for Cox regression and conditional likelihoods. J. Am. Statist. Assoc. Vol. 80, 389, 139-147.

THOMPSON, W.A. Jr. (1981). On the foundations of reliability. Technometrics, Vol. 23, No. 1, 1-13.

WIGHTMAN, D.W. (1987). The Application of Proportional Hazards Modelling to Reliability Problems. PhD Thesis. Trent Polytechnic, Nottingham.

WIGHTMAN, D.W., WALKER, E.V. and BENDELL, A. (1988). Applications of Proportional Hazards Modelling to hardware and software reliability. To appear: Proc. International Research Conference on Reliability. Missouri, Columbia.

Reliability Engineering and System Safety **21** (1988) 197–214

Proportional Hazards Modelling in the Analysis of Computer Systems Reliability

M. R. Drury

Group Quality, International Computers Ltd, Cavendish Road, Stevenage, Herts SG1 2DY, UK

E. V. Walker,* D. W. Wightman & A. Bendell

Trent Polytechnic, Department of Mathematics, Statistics and Operational Research, Burton Street, Nottingham NG1 4BU, UK

(Received 5 October 1987)

ABSTRACT

The analysis of the reliability of computer systems poses a number of complex problems. With the advent of VLSI and the diversity of usages the modelling of computer reliability has become far from straightforward. The cost of computer breakdowns includes elements for loss of usage, and the cost of repair. To minimise the costs it is essential that computers are as reliable as possible. To achieve reliable designs suitable modelling of reliability has to be undertaken. Proportional Hazards Modelling (PHM) is an efficient technique which can identify the effects of the various explanatory variables which may be associated with variations in the times between faults on a computer system. In this paper we apply PHM to the analysis of fault data on a PCB set from an ICL product.

NOTATION

t — Time in days (time to failure, time between failures, censoring time).

* To whom correspondence should be addressed.

A version of this paper was presented at Reliability '87, 14–16 April 1987, Birmingham, UK, and is reproduced by kind permission of the organisers.

t_i	Time in days between $(i-1)$th and ith events.
$h(t;\, Z_1, Z_2, \ldots, Z_k)$	Hazard function for item of equipment with explanatory variables $Z_1, Z_2, \ldots, Z_k$.
$h_0(t)$	Baseline hazard function.
$H_0(t)$	$\int_0^t h_0(x)\,dx$ cumulative base-line hazard function.
B_i $(i = 1, 2, \ldots, k)$	Parameters of proportional hazards model.
$\hat{B}_i$ $(i = 1, 2, \ldots, k)$	Estimators of parameters of PHM.
Z_i $(i = 1, 2, \ldots, k)$	Explanatory variables in PHM.
$\hat{B}_{ij}$	Estimator of B_i parameter when time j is omitted.
$\hat{r}_{ij}$	Estimator of Schoenfeld partial residual for time j on covariate i.
E	Expectation operator.
SE(x)	Standard error of x.

INTRODUCTION

Reliability in ICL

ICL, part of the STC group, is a leading European Information Systems supplier. ICL supplies a range of complete Information Systems to carefully selected target market-places. This includes system integration and development, the supply of office systems, maintenance, training, professional services and, of course, the design and manufacture of the world's most advanced range of mainframes.

The Reliability of ICL's products is of paramount importance to the Company. Failures of computer equipment are costly to both the users and ICL (who maintain it). The user suffers the costs associated with the lost business and staff being idle waiting for the equipment to be repaired. ICL suffers the cost of excessive failures in terms of man and part costs to fix. It is thus in everyone's interest to make the products as reliable as possible.

ICL collects reliability data at system, unit and component level for its computer products. All failures rectified by the engineers are logged on a computerised database, and weekly summary information for all large mainframe systems is collected to monitor system reliability. So that predictions for future products are as accurate as possible, the results of analysing the system, unit and component reliability data collected enables the prediction models and databases to be continually enhanced. However, with the advent of VLSI and the diversity of usages, the analysis of field returns, and the subsequent enhancement of reliability models is far from straightforward.

Application of proportional hazards modelling

To improve the reliability of future products, the modelling of reliability has to be improved to allow for the large number of variables that potentially affect (or appear to affect) the reliability of equipment. However, firstly these have to be identified. From experience, large differences in the reliability of identical hardware products in different environments and circumstances have been seen. The various methods and systems for testing and burning-in products prior to shipment can have different effects on the resultant reliability seen by a user. It is this type of variable that needs explaining if reliability models are to lead to the improvements looked for.

There is a large literature on the reliability of computer hardware (Dhillon and Ugwu[1]).

In the continuing search for new techniques to analyse the field data collected, and get a handle on the variables affecting reliability, Proportional Hazards Modelling (PHM) presented an exciting new approach to try. The actual reliability of the highly successful ICL 2957, 2958, 2966 and 2988 family of processors was modelled using PHM. Extensive data on the failures of the Printed Circuit Boards comprising the processors, as well as a number of other variables was readily available from ICL's databases.

Four variants of the processor, in both single and dual modes over a total of 364 customer systems were considered. For each system the following was known:

Serial numbers and installation dates of processor(s).
Which variant (or system type).
Operational pattern of system.
Details of each processor hardware PCB failure to component level.
Weekly actual operating hours.
The QC system in place at time of manufacture of the processor(s).

Only failures of PCBs common to all the processors have been considered in the analysis. Options (fitted to only some), cabling and the power supply have not been counted as part of the processor. Failures of processors for systems with more than one could not be identified to the actual failing processor, and so the 'system' is the level at which processor failures are considered in the modelling of the reliability.

PROPORTIONAL HAZARDS MODELLING

Proportional hazards modelling is a technique which can identify the effects of explanatory factors which may be associated with the life length of equipment. PHM can be used to study repairable as well as non-repairable systems. Data may be censored or uncensored. No underlying distribution need be assumed for the structure of the data, making PHM a powerful tool for reliability analysis.

The technique is a method for decomposing the variation in life lengths into orthogonal factors, identifying the significant ones, and reconstituting the model for prediction purposes.

The fundamental equation on which PHM is based is an assumed decomposition of the hazard function for an item of equipment into the product of a base-line hazard function and an exponential term incorporating the effect of a number of explanatory variables varying between items. That is:

$$h(t;\, Z_1, Z_2, \ldots, Z_k) = h_0(t)\exp(B_1 Z_1 + B_2 Z_2 + \cdots + B_k Z_k)$$

where the B_i's are the unknown parameters of the model defining the effects of each of the explanatory variables; the Z_i's are the values of the explanatory variables.

A Z_i can be either a naturally measured variable such as age, or an indicator variable, indicating for example the presence or absence of a change in design.

The explanatory variables are assumed to act multiplicatively on the base-line hazard functions. Thus, for different values of the explanatory variables the hazard functions are proportional to each other over all time t. The base-line hazard function $h_0(t)$ represents the hazard function that the equipment would experience if the covariates all take the base-line value zero.

The B_i's are estimated from the data and tested to see whether each explanatory variable really has an effect in explaining the variation in observed failure times.

In this paper we employ the usual distribution-free approach. The detail of the methodology is not developed here, but the interested reader is referred to Kalbfleisch and Prentice.[2] The method first iteratively estimates the effects of the covariates $B_1, B_2, \ldots, B_k$ using the so-called method of scoring based upon a Taylor Series expansion for each step in the iteration, starting with initial values of zero. Once the estimates converge, tests of whether each explanatory variable has any significant effect are based upon the asymptotic normality of the estimators. A backwards stepwise procedure is incorporated whereby non-significant explanatory factors are excluded one at a time and the model re-run until all factors are significant.

Upon finding estimates for the B_i's, the method then obtains the base-line hazard function based upon discrete hazard contributions at each of the times at which failures were recorded. This can then be compared to standard distributional forms by usual hazard plots.

DATA

Information was available about the processor failures occurring, within the three month time window 01/04/84–30/06/84, at 364 customer systems.

The records included information on the system identifier, date of installation, the usage the system experiences, the number of processors in the system, and details of the processor failures occurring.

APPLICATION

There are many possible model formulations that could be used in applying a proportional hazards study to these systems. For example;

(a) The basic time metric could be taken as the time between processor failures of any type occurring on the same system with covariates included to describe the type of failure found.

(b) The basic time metric could be taken as the time between processor failures on the system of the same type; again covariates could be included to describe the failure type occurring at each event.

For any model there are usually a number of censoring possibilities which could be appropriate.

A good knowledge of the repair procedure would be useful to identify the most appropriate model and censoring structure.

This particular data structure corresponds to a point process observed within a time window. This is a familiar structure in reliability data (see Fig. 1). Such data presents problems since it includes both left and right truncation. The right-hand truncation can be dealt with within PHM fairly easily by defining a censored event time from the last failure viewed to the end of the time window. However, the time from the start of the time window to the first failure viewed needs to be treated differently. Here we assume that the left-truncated time to the first event follows the same distribution as subsequent time between failures but with a different hazard. Such a procedure is less wasteful of data than ignoring such times and was introduced for initial exploration of the data. It is, of course, theoretically sound if the times between failure (TBF's) are exponential.

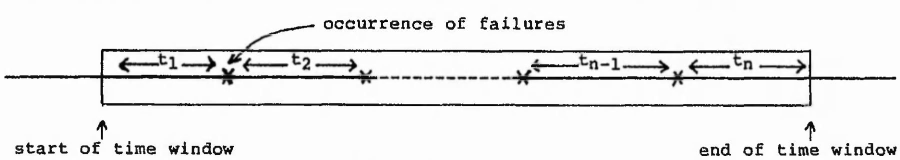


Fig. 1. Time window.

DATA STRUCTURE

The basic time metric, t, is taken to be the time, in days, between failures. Since the systems are repairable, a reasonable starting assumption is that a system is repaired to 'as-good-as-new', irrespective of the particular failure type that occurred. Thus we assume complete censoring. That is, should a failure occur of one particular type, censored events are also assumed at this time for the other failure types.

From the coding of failure types two large groups of events can be identified in the data corresponding to failure types No Faults Found or NFF (92 events), and component type 272 (32 events). All other failure types are grouped together, in this analysis, to form a third failure type OTHER (46 events). It follows from the type of failure mechanism assumed that a failure to a system identified as type 272, also creates censored events for the failure types NFF and OTHER.

Explanatory variables employed in the PHM model were:

Z_1	Z_2	Z_3	Z_4, Z_5	Z_6, Z_7, Z_8	Z_9, Z_{10}	Z_{11}, Z_{12}
Event	Age	Av. h/wk	Failure dummies	System dummies	Number of processors	Quality control

Event. This variable indicates the presence '0' or absence '1' of left truncation.

Age. The age at the time of event is calculated, in months, from the first of the month of installation.

Av. h/wk. The average number of hours use per week each system experiences in the time window is taken from actual field returns.

Failure dummies. These dummy variables (in Table 1) compare different processor failure types.

System dummies. These dummy variables indicate the system type the processor(s) are in. There are four types of system (Table 2). There are 234 type A, 64 type B, 34 type C and 32 type F.

Number of processors. These dummy variables compare the effects of the number of processors without assuming linear effects. Systems can have different numbers of processors depending whether they are single or dual (Table 3).

TABLE 1
Coding of Failure Type

Failure type	*Covariates*	
	Z_4	Z_5
'272' (base)	0	0
NFF	0	1
OTHER	1	0

TABLE 2
Coding of System Type

System type	*Covariates*		
	Z_6	Z_7	Z_8
A (base)	0	0	0
B	0	0	1
C	0	1	0
F	1	0	0

TABLE 3
Number of Processors within System Type, and Covariate Coding

System type		*Number of processors* *1*	*2*	*4*
A		X	X	
B		X		
C		X		
F			X	X
Covariates	(Z_9, Z_{10})	(0, 0)	(0, 1)	(1, 0)

TABLE 4
Coding of Covariates for quality Control

Quality control	*Covariates*	
	Z_{11}	Z_{12}
Pre-change (base)	0	0
Post-change	0	1
Combination	1	0

Quality control. Since the start of manufacture a major change in Quality Control was introduced. Table 4 shows the coding of covariates to model the QC to which the processors within the System have been subject.

SELECTION OF PHM MODEL

The model described above evolved by applying PHM in an exploratory manner. In this process a number of different structures were considered. Some of these had to be neglected, or rather further adjusted, since they gave rise to technical problems such as multicollinearity and monotonic likelihood (Bryson and Johnson[3]). For example, multicollinearity between the previous use and age of the system implied that previous use was neglected, whilst power was found to be an exact linear combination of the dummy covariates representing the system type, so that power was eliminated.

In other cases the specification of some variables was also changed to provide greater information. For example, in an early exploratory formulation the covariate 'age' was taken as the age of the system at the start of the time window. However, due to the presence of a few 'young' systems the respecification to age of system at processor failure gives more physical explanation.

ANALYSIS

Table 5 shows that disaggregation of the 1602 events into the binary covariate categories.

Results

The results after the backwards usual stepwise elimination procedure, based on two-tailed 5% tests, are given in Table 6. The p-value indicates the probability of obtaining such an extreme estimate for the B's just due to chance, if there were no real effect for that covariate. The likelihood ratio statistic, in the last column, is seen to exceed the tabulated upper 5% critical value for a χ^2 distribution with nine degrees of freedom (16·92). This indicates that the model is highly significant as compared to assuming that the covariates have no effect and that the data is homogeneous.

Significant covariates

Z_1—*event.* The positive estimate $\hat{B}_1$ suggests that the times to the first failure tend to be longer than the times between failures, with the t.b.f.'s

TABLE 5
Disaggregation of Events into the Binary Covariate Categories

Covariate	*Covariate = 0*		*Covariate = 1*	
	Number of failures	*Number of censorings*	*Number of failures*	*Number of censorings*
Z_1	103	947	67	485
Z_4	123	945	47	487
Z_5	79	989	91	443
Z_6	148	1 292	22	140
Z_7	157	1 304	13	128
Z_8	157	1 214	13	218
Z_9	157	1 376	13	56
Z_{10}	140	1 261	30	171
Z_{11}	165	1 410	5	22
Z_{12}	133	1 046	37	386

Total number of events = 1 602.
Total number of failures = 170.
Total number of censored events = 1 432.

having a hazard rate which is greater than that for the time to the first event by a multiplicative factor of $\exp(0{\cdot}6384) = 1{\cdot}89$. This, perhaps, counter-intuitive result can be explained in three ways. Firstly, it is apparent, from the raw data, that the majority (71%) of the complete set of systems did not have a processor failure at all during the time window. Thus, these systems have a long censored time to first processor failure compared to the necessarily shorter observed t.b.f.'s. This model bias arises due to the time window

TABLE 6
Final Significant Model after Backwards Stepwise Elimination

Significant covariates	$\hat{B}_i$	*p-Value*	*Likelihood ratio statistic*
Z_1—event	0·638 4	0·000 1	
Z_2—age	−0·033 6	0·001 9	
Z_3—av. h/wk	0·006 4	0·003 0	99·450
Z_5—NFF	0·834 5	0·000 0	(5% critical
Z_6—system F	−0·874 8	0·015 5	value from
Z_8—system B	−0·707 9	0·008 3	tables: 16·92)
Z_9—4 processors	1·507 4	0·001 0	
Z_{10}—2 processors	0·641 7	0·005 0	
Z_{12}—post change	−0·662 2	0·001 5	

nature of the data. Secondly, as we show below, the underlying distributional form for t.b.f.'s turns out to be decreasing hazard rate, so that such left-truncation is expected to increase residual life length. Finally, bunching of failures (e.g. due to misdiagnosis) may be associated with the longer times to first failure.

Z_2—*age*. The sign of $\hat{B}_2$ implies that the older the system the lower the hazard, so that there is reliability growth as the system ages.

Z_3—*av. h/wk*. The sign of $\hat{B}_3$ indicates that the more use the system experiences per week, the higher the hazard.

Fault types. The covariate Z_4 was eliminated since its effect was not significant, thus there is no significant difference between the fault types OTHER and '272'. The covariate Z_5—NFF which compares the fault type NFF with '272' and OTHER fault types, is significant with a positive coefficient implying that the NFF fault type has higher hazard and shorter times to failure than the '272' and OTHER fault types. This is not of great physical significance, and is to be expected since NFF faults are very prevalent.

System. Z_7 which compares system C with the base system A is not significant. Z_6 which compares system F with systems type A and C is significant with a negative coefficient implying that system type F is subject to a smaller hazard than those of systems type A and C. Similarly, systems of type B experience a smaller hazard (about half) compared to systems type A and C.

Number of processors. From the significance of Z_9 it can be seen that systems with four processors experience a higher hazard than those with one. The crude multiplicative factor is approximately 4·94. Similarly, Z_{10} shows that systems with two processors experience a higher hazard than those with one. The crude multiplicative factor is approximately 1·90.

System/number of processors. Based on our model the magnitude of the hazard variation between systems with different numbers of processors can be calculated from the significant B estimates, since this is of operating importance. Table 7 shows these differences.

Quality control. The covariate Z_{11} was eliminated since its effect was not significant. The covariate Z_{12} which compares systems with processors which are covered by the new quality control (only) with the other systems is

TABLE 7
Multiples of Base-line Hazard for Various System/ Number of Processor Combinations

System type	*Number of processors*		
	1	*2*	*4*
A	1·00	1·90	—
B	0·49	—	—
C	1·00	—	—
F	—	0·79	2·06

significant with a negative coefficient implying that the new tighter quality control procedures imply a lower hazard.

GRAPHICAL ANALYSIS

A number of graphical techniques can be employed to examine the appropriateness and fit of the PHM model. Some of these graphs for the current data can be seen in Figs 2–9.

Figure 2 shows a *Weibull hazard plot* for the base-line hazard obtained from the PHM model. The plot is reasonably straight indicating the Weibull as a reasonable distribution for the t.b.f. (King[4]). The shape and scale

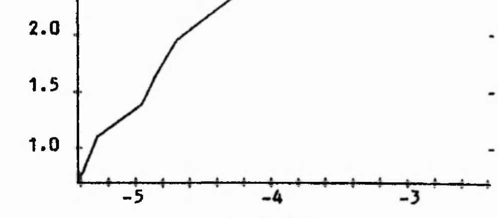


Fig. 2. Weibull baseline hazard plot $[H_0(t) = (t/\theta)^B]$.

parameters are estimated at approximately 0·84 and 1425 days respectively. Since the shape parameter is less than 1 the base-line shows decreasing hazard.

Cox and Snell 'crude' residuals can be calculated and plotted against their expected order statistics (Cox and Snell[5]) to provide a graphical goodness-of-fit test for the model. If the model is a good fit to the data we would expect the residuals to lie on a straight line of gradient 1, passing through the origin (Kay[6]). As yet, however, little guidance is available as to how far the plot can stray from this expected line before the model is no longer considered a good fit. A problem also arises since we are unable to estimate the hazard for censored times greater than the largest t.b.f. or time to first failure. For our purposes, we assign these a hazard equal to that for the greatest time to a failure. It is thus probable that such plots will be left-shifted at their tail end. Another drawback to this plot is that the variance increases as the residuals increase (Aitkin and Clayton[1]). Arcsine transforms can be taken to plot a *variance-stabilised* form of the Cox and Snell residuals and the resulting plot can be seen in Fig. 3. We can see from Fig. 3 that the model appears to be a fairly good fit.

Figures 4 and 5 are *proportionality plots* for two of the significant binary covariates. In the plots the data is stratified on each significant binary covariate, and the model run separately for each stratum. Plotting $\ln H_0(t)$ against t for each stratum on the same graph, should produce plots with constant vertical separation for all t, if the assumption of proportional hazards holds.[6] Figure 4 shows stratification on the dummy variable Z_5. The

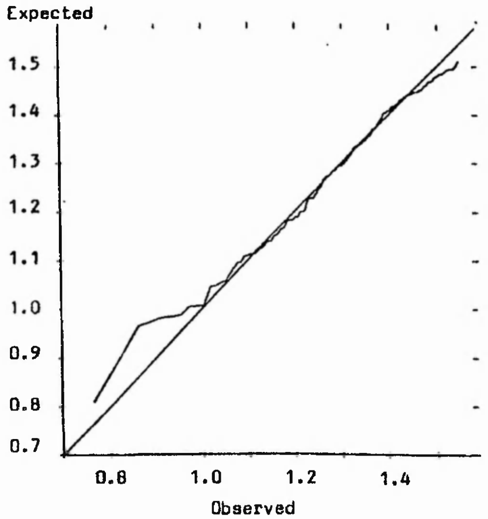


Fig. 3. Variance stabilised Cox and Snell residuals.

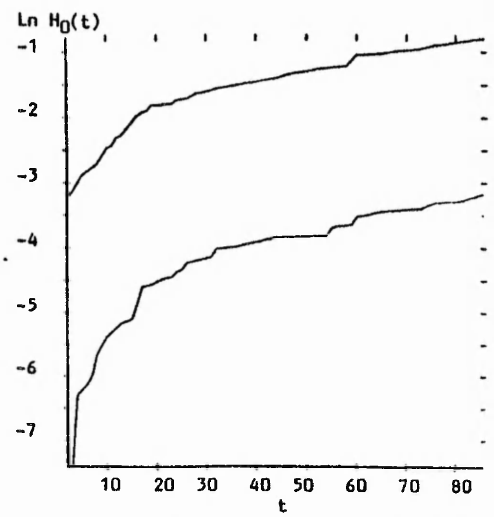


Fig. 4. Proportionality plot for covariate 'NFF'.

two plots are reasonably parallel after about $t = 10$, which implies that the proportional hazards assumption holds for this variable. Proportionality plots for the other significant binary variables included in the model; with the exception of that for Z_1 are similar to Fig. 4, indicating good fit.

Figure 5 shows stratification on the covariate Z_1. For small t, the plots are not constantly vertically separated indicating that this covariate violates the

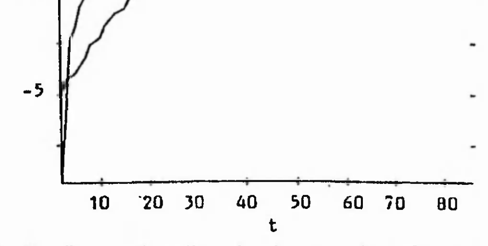


Fig. 5. Proportionality plot for covariate 'first event'.

proportionality assumption. If necessary it is possible to extend the proportional hazards model to include strata, to accommodate such variables that violate proportionality (Kalbfleisch and Prentice[2]). As indicated above the assumption of proportionality for Z_1 was only a tentative first hypothesis. Figure 5 indicates that stratification might instead be worthwhile.

Figures 6 and 7 show *Schoenfeld partial residuals* plotted against time for two of the significant covariates (Schoenfeld[8]). If proportional hazards holds, $E(\hat{r}_{ij}) \simeq 0$, where $\hat{r}_{ij}$ is the estimate of the Schoenfeld partial residual for event j on covariate i, and a plot of $\hat{r}_{ij}$ against t will be centred about zero for all t. Since, as often in reliability problems, there are a large number of

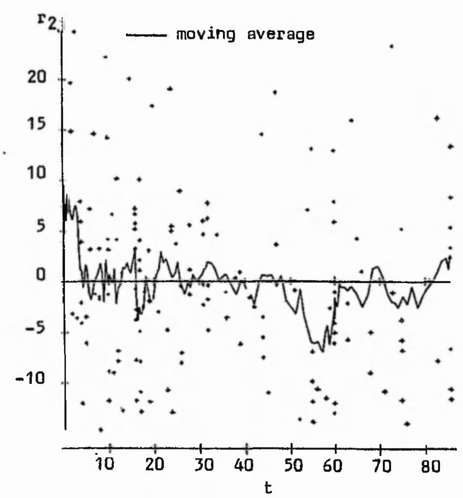


Fig. 6. Schoenfeld residuals for covariate 'age'.

tied t.b.f.'s the density of partial residuals at many of the points cannot be clearly seen on the plots, particularly for binary covariates. To ease visual inspection, therefore, a moving average it shown on each plot based on intervals of 10 t.b.f.'s. If proportionality holds for the covariate identified, the moving average should centre about zero for all t.

Figure 6 shows the Schoenfeld partial residuals plotted for the covariate age. This plot is typical for a non-binary covariate. For all t the plot has a reasonable scatter about zero; this is borne out by the oscillation of the moving average about the axis. Thus, the proportional hazards assumption does not appear violated by this covariate. It is also noticeable that there are no obvious outliers, which might have indicated times to failure which need further examination.

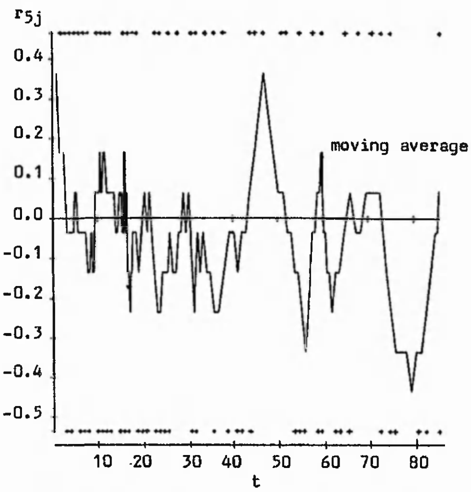


Fig. 7. Schoenfeld residuals for covariate 'NFF'.

Figure 7 shows the Schoenfeld partial residuals plotted for the covariate Z_6. This plot is typical of those for binary covariates.[8] There is a split of residuals above and below the axis corresponding to the two categories of the covariate. The plot largely bears out the assumption that the residuals have scatter about zero. Thus the proportional hazards assumption does not appear violated by this covariate.

Schoenfeld residual plots for the other significant binary variables are typically similar to Fig. 7, with the exception of that for the covariate Z_1. The plot for Z_1 suggests that there is a trend in the residuals. In agreement with the interpretation of Fig. 5, this may suggest that the covariate violates the assumption of proportionality.

Figures 8 and 9 show standardised plots of the *influence* of individual times to failure on the $\hat{B}_i$ parameter estimates for some of the significant covariates (Cain and Lange,[9] Reid and Crepeau[10]). The form we present here shows the estimated normal deviate for the covariate coefficient when each single time to failure and censoring time is excluded from the model, one at a time. This is plotted against the order by magnitude of the times on the horizontal axis. We can thus examine which times to failure have times most influential on the observed significance of the covariates, and which if any, would if deleted remove the significance of the covariate. For the 5% two-tailed tests these points correspond to estimated normal deviates in the range $(-1{\cdot}96, +1{\cdot}96)$.

The influences of an event is the difference between the $\hat{B}_i$ estimates with and without the event included; $I_{ij} = \hat{B}_i - \hat{B}_{ij}$, where $\hat{B}_{ij}$ is the estimate of B_i

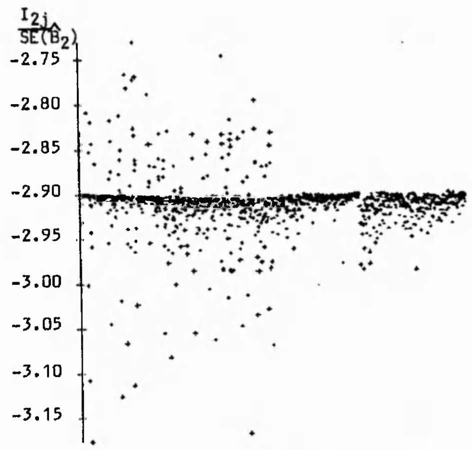


Fig. 8. Influence of events for covariate 'age'.

made when time j is excluded. The figures presented here are based on the simplifying approximation of treating the standard error (SE) of the B_i estimator as unaltered after elimination of the single data point. This is likely to be valid for large data sets, such as here. However, since it is an approximation we should examine the actual standard error for points on the influence plots close to the $\pm 1{\cdot}96$ limits.

Figure 8 shows the influence of times on the covariate age. We can see that there are no times or censorings that are likely to alter the significance of the covariate (at 5% level) if omitted. The highest density of points is close to the

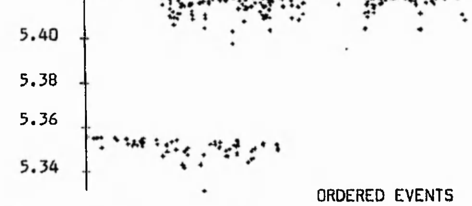


Fig. 9. Influence of events for covariate 'NFF'.

normal deviate for this covariate in the full data set; most of these correspond to censoring points.

Figure 9 shows the influence of times to failure on the dummy variable NFF. Again there are no times which alter the significance of the variable when omitted. The structure of this plot is typical of those for binary covariates. There are three distinct groupings; the central group comprises mainly censored events, and the groups above and below are split on the value of the covariate in question for each time to failure event.

Plots obtained for the other significant covariates are similar to Figs 8 and 9. Whilst not standard practice, it may on occasion be interesting to also look at such influence plots for non-significant covariates, to see if there are any events which when omitted would move the covariate into significance. For the current case the $\hat{B}$ estimates and their significance for the non-significant covariates, at the stage when they were eliminated from the model in the backwards stepwise procedure, can be seen for each covariate in Table 8. In all cases (except perhaps for Z_4) they are largely non-significant and there is little argument to undertake such an approach.

TABLE 8
Significance Level at Which Covariates Were Eliminated from the Model

Covariates	B_i	*p-Value*
Z_{11}—combination	−0·2472	0·3351
Z_7—system C	0·3714	0·1430
Z_4—OTHERS	0·3844	0·0467

CONCLUSIONS

The results of the initial study reported here indicate that the PHM model described fits the structure of the data quite well. The treatment of time to first failure by use of a covariate, to avoid the problem of left truncation, was least successful and there may well be a need here for further work.

The PHM model has enabled us to identify variables which have a significant effect in explaining the times between failures for the systems as well as the direction and magnitude of these effects.

We have identified the effects of configuration parameters and have been able to make comparisons between the failure types. The model has also enabled us to identify the underlying structure of the point process, and the distributional form for the base line distribution. Unlike other techniques we did not have to make distributional assumptions prior to the analysis.

Work is continuing in collaboration between ICL and Trent Polytechnic to clarify outstanding issues and further understanding of the processes involved.

ACKNOWLEDGEMENTS

Mrs Walker's work for this paper was funded by a grant from the NAB. Mr Wightman received partial funding from UKAEA; contract number IR55527, and from the Alvey grant on Software Reliability Modelling.

REFERENCES

1. Dhillon, B. S. and Ugwu, K. I. Bibliography of literature on computer hardware reliability, *Microelectron. Rel.*, **26** (1986), pp. 131–53.
2. Kalbfleisch, J. F. and Prentice, R. L. *The Statistical Analysis of Failure Time Data*, Wiley, New York, 1980.
3. Bryson, M. C. and Johnson, M. E. The incidence of monotone likelihood in the Cox model, *Technometrics*, **23** (1981), pp. 381–3.
4. King, J. R. *Probability Charts for Decision Making*, Industrial Press Inc., New York, 1971.
5. Cox, D. R. and Snell, E. J. A general definition of residuals, *J. R. Stat. Soc.*, **B30** (1968), pp. 248–75.
6. Kay, R. Proportional hazard regression model and the analysis of censored survival data, *Applied Statistics*, **26**(3) (1977), pp. 227–37.
7. Aitkin, M. and Clayton, D. The fitting of exponential, Weibull, and extreme value distributions to complex consored survival data using GLIM, *Applied Statistics*, **29**(2) (1980), pp. 156–63.
8. Schoenfeld, D. Partial residuals for the proportional hazards regression model, *Biometrika*, **69** (1982), pp. 239–41.
9. Cain, K. C. and Lange, N. T. Approximate case influence for the proportional hazards regression model with censored data, *Biometrics*, **40** (1984), pp. 493–9.
10. Reid, N. and Crepeau, H. Influence functions for proportional hazards regression, *Biometrika*, **72** (1985), pp. 1–9.

APPENDIX B

Reliability Engineering and System Safety **21** (1988) 245–269

The Reliability Analysis of Weapon Systems

C. Gray, N. Harris

British Aerospace Army Weapons Division, Seven Hills Way, Stevenage, Herts SG1 2DA, UK

A. Bendell & E. V. Walker*

Trent Polytechnic, Department of Mathematics, Statistics and Operational Research, Burton Street, Nottingham NG1 4BU, UK

(Received 5 October 1987)

ABSTRACT

The complexity of modern weapon systems presents great difficulty for the reliability analyst. In many analytical contexts, failures of the equipment may take place at various levels in a hierarchy. Thus failures may be attributed at the lowest level to components, or alternatively at higher levels to circuit boards, sub-assemblies, assemblies or modules. The logistic support for such systems is highly complicated. The failure data faced by the reliability analyst records failures at various levels of equipment aggregation, hence standard reliability analysis methods, as well as new techniques such as Proportional Hazards Modelling, need adaption in order to facilitate the complex point processes underlying the data structure.

In this paper we consider the analysis of early field data for a major weapon system in current military use. The problems of data extraction and manipulation are discussed, and the adaption of methodologies to the current data structure highlighted. Emphasis is placed upon the exploration of the data structure and the categorisation of mechanisms.

* To whom correspondence should be addressed.
A version of this paper was presented at Reliability '87, 14–16 April 1987, Birmingham, UK, and is reproduced by kind permission of the organisers.

Reliability Engineering and System Safety 0951-8320/88/$03·50

NOTATION

t	time (time between faults, censoring time)
t_i	ith TBF
$h(t; z_1, z_2, \ldots, z_k)$	hazard function for item of equipment with covariates $z_1, z_2, \ldots, z_k$
$h_0(t)$	baseline hazard function
$H_0(t)$	cumulative baseline hazard function $= \int_0^t h_0(x)\,\mathrm{d}x$
β_i $(i = 1, 2, \ldots, k)$	parameters of Proportional Hazards model denoting effects of covariates $z_1, z_2, \ldots, z_k$
$\hat{\beta}_i$ $(i = 1, 2, \ldots, k)$	estimates of parameter of Proportional Hazards model
z_i $(i = 1, 2, \ldots, k)$	covariates in Proportional Hazards model
I_{ij}	influence of the jth TBF on the ith covariate
$\hat{\beta}_{ij}$	estimate of β_i with jth TBF excluded
N.D. $(\hat{\beta}_{ij})$	normal deviate of $\hat{\beta}_{ij}$

INTRODUCTION

The successful reliability engineering of complex systems can benefit from inputs from the statistician as well as the engineer. The image of the statistician's role in reliability is that he/she organises and analyses failure data for the purpose of measuring, modelling and predicting reliability. There is however, an emerging view that the expertise of the statistician can also be productively exploited in addressing the reliability engineer's primary objectives: those of achieving and improving reliability.

Whilst the engineer and the statistician use different methodologies, they have a common objective: that of achieving a reliable product. We can identify two roles that the statistician can play in the design process. The first is passive in which the statisticians involve themselves with the organising and analysis of data, so as to provide the designer (decision maker) with relevant information. The second role is a more active one, in which the statistician goes beyond being a mere provider of information or guidance counsellor, but becomes a full participant in the design process, and is thus entitled to make recommendations which are supported by his analysis as to the actual form a design will take. To achieve this requires a substantial reorientation in the statistician's approach to design, he/she must regard themselves as capable of more than just modelling a design, but just as, if not more importantly, having the potential of impacting a design. This has considerable implications as to the types of models and analysis that will be pursued. A new relationship between the designer and statistician is called for, which now emphasises an iterative and interactive approach.

Whilst in the 1960s and 1970s, the statistician could have argued a lack of technical support for this role, there have been significant recent theoretical and computational developments. If a distinction has to be made between then and now, it is that in the past emphasis has been on developing methods and models that operate on quantitative (—cardinally measured) information, which whilst important, all too often required the statistician to ignore soft qualitative (—categorically measured or rule based) information, which in engineering is often more important. Recent trends are redressing this imbalance.

One development of significance is the general emergence of the theory of generalised linear models (GLM). This has unified many areas of statistics which have considerable potential for analysing reliability data; e.g. probit analysis which concerns itself with quantifying the relation between a stimulus and its response:[1] contingency table analysis in association with Log-Linear models, which concerns itself with the analysis of Cross-Classified Data;[2,3] and (although only very tenuously GLM) 'Proportional Hazards Modelling' which concerns itself with developing failure models based on explanatory variables.[4] Whilst Probit Analysis and log-linear models have only recently been introduced into reliability,[5,6] Proportional Hazards has on occasions been applied with considerable success.[7,8] The development of the software package GLIM (available from NAG) offers the competent professional statistician the facility of computerising much of this type of analysis.

Almost irrespective of the analytical framework and the modelling approach, in order to suggest particular model formulations these methods should be used in conjunction with the exploratory approach to data known as EDA (Exploratory Data Analysis).[9] When used in conjunction with a particular model structure, EDA proceeds very much in line with conventional modelling theory, namely that of Model Identification, Model Parameterisation, Model Validation and Iteration.

The reliability analysis of weapon systems reflects all of these considerations, but is complicated by the features of particular systems. Modern weapon systems, their deployment and maintenance, and not least of all their record keeping, are all inherently complicated. Whilst the reliability of such systems may be very good, their reliability analysis is usually limited and difficult. For these and other reasons statistics has largely failed to make a major impact on the reliability of weapon systems.

This may not be the statistician's fault, for example, the form and the quality of the data available from the field is often poor, usually dominated by operational requirements, product of development and security considerations, and imbedded by the historical data analysis methodology current at the time of development years earlier. Indeed, it is not atypical for the data system to be recording at the same time too much irrelevant data and too little informative data, as well as being full of errors. Concepts of data validation and data entry are yet to permeate this field.[5]

BACKGROUND

The mature weapons system considered in this paper has an established reputation for reliability. It has been in operation for over a decade and has been successfully deployed worldwide in a variety of environments from hot-wet to extreme cold. It has been transported over many types of rough terrain.

The system is of modular design as shown in Fig. 1. A complete system

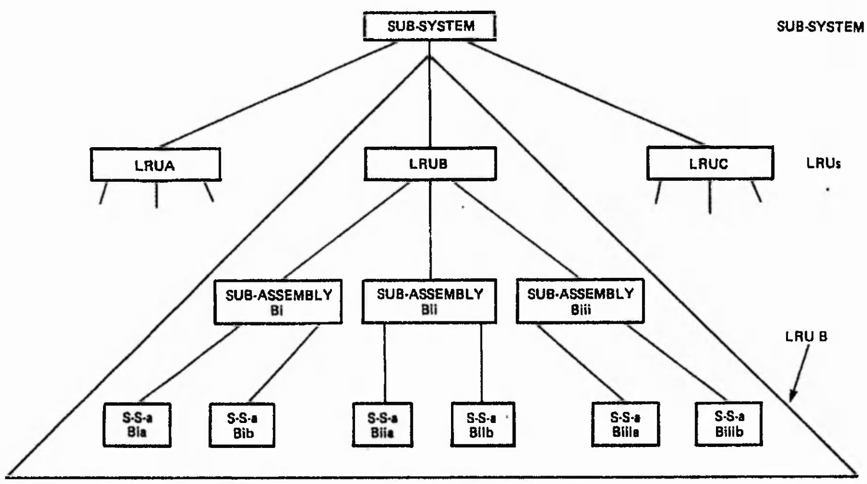


Fig. 1. Hierarchical nature of sub-systems.

comprises a missile launcher, optical tracker, and generator, and for Blind Fire Systems an additional radar tracker. The analysis in this paper only discusses three types of subsystem, which are referred to as sub-systems (1), (2) and (3). These sub-systems are interchangeable and also of modular design, the principle modules being designed as Line Replacement Units (LRUs). The system has been designed to be maintained by combat troops in a battlefield environment. First line maintenance consists of changing the LRU whilst at second line the fault will be traced to the sub-assembly (and possibly down to component) level. The philosophy of repair by replacement is supported by automated performance testing, thereby ensuring maximum operational availability. The modular design has allowed individual units to be upgraded in line with technological developments, e.g. introducing Built in Test Equipment (BITE). This has led to enhanced performance and a high achieved availability. It is a successful military system with a proven track record which has and is being continually upgraded to meet and surpass the demands of its role on the battlefield.

Data source

The source of reliability data on this weapons system is the computerised database built up by the REME, Radar Branch using the FORWARD (Feedback of Repair Workshops and Reliability Data) reporting system. The function of this database was to provide to management logistic information necessary to the management of equipments, spares, and resources of a complex weapons system. It also served to monitor reliability and provide engineering information to improve reliability.

The database, by today's standards, can be considered of conventional design. The repair technician was required to complete a descriptive jobcard outlining relevant information, such as date, serial number, elapsed time indicator readings (ETIs), and fault classification code. This information is then transcribed and stored in the computer database, thereby in principle building up a complete historical record of reliability repair and maintenance data.

It has been recognised that there were several factors which detrimentally affected the integrity of the FORWARD database. Omissions, inconsistencies and errors were introduced due to the amount of human effort required, the need for human judgement and interpretation at various stages in the data gathering process and the amount of data transcription required. Consequently it is estimated that the FORWARD databases are about 70% accurate.

Four databases detailing schedules, environmental/deployment data, ETI readings and defect data were available. A number of features in the database made data extraction difficult and complicated. For example there was inconsistent formatting within the data bases which precluded file merging, also certain fields of the deployment database were free format and thus difficult to extract information from. Because of these problems only the defect data base, which contains records about faults found, was investigated.

Structure of the data in the defect database

Each record of the defect database refers to a fault. The particular unit for which data is recorded is identified by its sub-system type (i.e. launching unit, optical unit, etc.) and serial number. Each record contains fields detailing:

(1) The date that the fault was identified.
(2) The holding group of the unit at the time of failure (User identifier i.e. squadron or battery). In this analysis there were six different holding groups.
(3) The type and serial number of the LRU that was removed and replaced. If more than one LRU was removed at one time a separate record was generated for each.
(4) The level in the hierarchy (see Fig. 1) to which the fault was subsequently traced. A separate record was generated for each identified fault.
(5) ETI reading when the fault occurred. The ETI is a four digit counter (i.e. 0000–9999) which measures the time the unit has spent in normal mode. (A launching unit has two modes: alert and normal. A radar unit has three modes: high alert, low alert, and normal. An optical unit has just the one: normal mode.)

Note: The ETIs are themselves known to be extremely unreliable; often sticking. The counters are 'throw-away', so should one appear to be 'misbehaving' it may be replaced. However, the replacement counter often does not start at zero, nor is it set to the last reading of the discarded counter. The value of information from the ETIs is therefore suspect.

A generalized representation of the fault structure

A reliability model is a means of representing the failure events of the system in a useful form. Since there is no such thing as a universal reliability model, model formulation must be directed by knowledge of the system and the data available. Initial considerations of the system lead to a generalised representation of the fault structure from which realistic and detailed models can be developed.

Using Fig. 1 which outlines the system structure and from the knowledge of the data base structure and the data available we can construct the representation shown in Fig. 2. Faults at the lowest level are component faults. Faults at the next highest level are sub-assembly faults which are the superposition of component faults plus other faults such as interconnection problems. Faults at the next level up are LRU faults which are a further superposition of sub-assembly faults conjugated with other faults that cannot be attributed to sub-assembly or component fault. The highest level is the sub-systems level which represents a total aggregation by superposition of all events reported as faults.

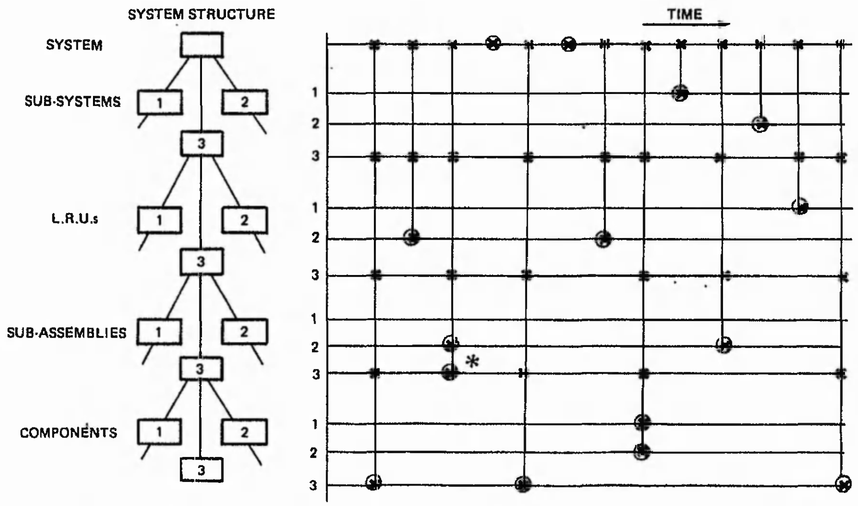


Fig. 2. A generalised representation of the fault structure. ○, lowest part of system to which the fault was traceable; *, example of fault traceable to 2 sub-assemblies.

POSSIBLE ANALYSIS METHODS

A number of analysis methods that are associated with reliability were considered, some of which were discarded as being inadequate or potentially misleading. e.g.

(i) *Multivariate analysis*
This approach[6,10] was not ideal because a lifetime characteristic could be identified within the data structure.

(ii) *Homogenizing the data and distribution fitting*
This was not appropriate since among other reasons the data were sparse in many of the possible combinations of factors. The known hierarchical structure also complicates such an approach.

(iii) *Time series analysis*
Although potentially promising, due to time considerations it was decided not to consider this approach fully. It can be used to study and describe a sequence of observations which depend on time, space or an index. In reliability such a time series approach may identify some structure in the data.[11] An indexed set of observations can be constructed here by counting the number of fault records in the database in some ordered time interval.

Figure 3 shows the number of records per month appearing in the database for sub-systems (2) in holding group I.

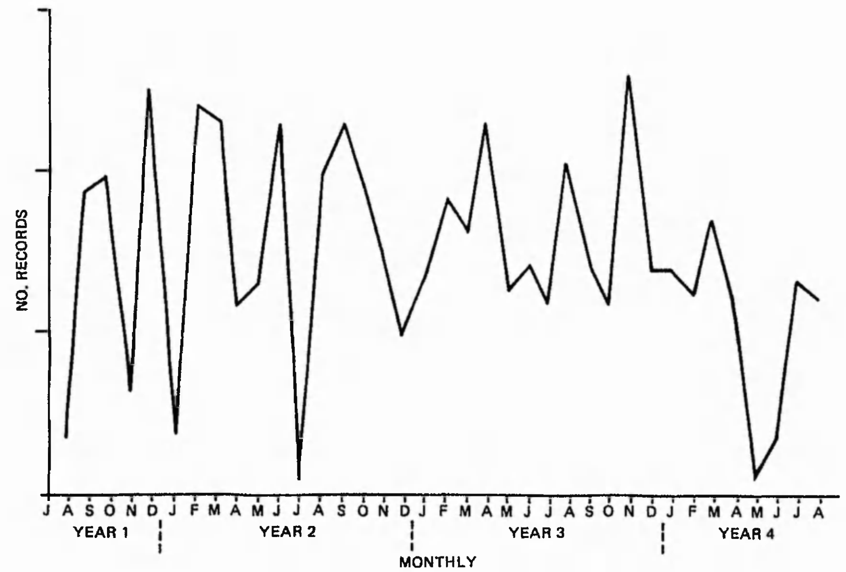


Fig. 3. Number of fault records, occurring within each month, for sub-systems (2) in holding group I.

There neither appears to be notable trend nor periodic behaviour. Plots for other sub-system types and holding groups are similar. The series, are however, shorter for holding groups II–VI, since fault records did not begin for these until some months after the beginning of the reporting for holding group I. The holding groups are known to have received their weapon systems at different times, although the exact dates each unit received them are not known.

(iv) *Proportional hazards modelling (PHM)*

Because of the considerable potential of PHM in this line of analysis, it was decided to pursue it further. An initial model is that times between faults (TBF's) may form a modified renewal process, where the TBF's are independent but non-identically distributed due to the effects of certain covariates upon the probability of a fault occurring. As there are a number of explanatory variables for TBF's of the LRUs, Proportional Hazards Modelling (PHM) is known to be a useful technique.[8,12,13] It identifies significant explanatory factors for TBF by an orthogonal decomposition of the lifetime variation. The model may then be reconstituted for prediction purposes.

Two important advantages of PHM in this application are:

(a) There is no need, within PHM, to specify a particular distribution a priori.

(b) Data may be censored or uncensored.
PHM is based on an assumed decomposition of the hazard function for an item of equipment into the product of a base-line hazard and an exponential term which incorporates the effect of explanatory factors varying between items.

$$h(z_1, z_2, \ldots, z_k) = h_0(t) \exp(\beta_1 z_1 + \beta_2 z_2 + \cdots + \beta_k z_k) \qquad t > 0$$

The β_i's are unknown parameters of the model defining the effects of each of the explanatory factors. The z_i's are the values of these explanatory factors; they can be either a naturally measured variable, or an indicator (dummy) variable indicating the presence or absence of a factor.

The base-line hazard function $h_0(t)$ represents the hazard function that the equipment would experience if the covariates all took the base-line value zero. From these considerations it was decided to use the Proportional Hazards Modelling as a basis for this analysis.

MODEL FORMULATION

There is a multitude of formulations of the PHM model which could be employed in the analysis of the data, the different models having different time variables, censoring structures, and covariates. Selection between them is iterative, based on experience and engineering information. The aim is to explore the data, in order to maximise explanation.

Types of model

Initially four types of simple but physically plausible PHM model were identified. General covariates which could be included in all the models are; the holding group that the particular faulty unit was in, the season in which the fault occurred, the ETI reading at the time of the fault, and a time trend.

Model A
Here the series of events occurring for a *particular serial numbered LRU* is followed, see Fig. 4. The time metric here is the time between the LRU being entered into a sub-system, and its being removed from that same sub-system when found faulty. The next TBF will be for that same LRU within another sub-system. Covariates can include which particular sub-system the LRU was in when the fault occurred.

For this model to be formulated, it is necessary to know when the main assembly block was returned to service, so that its repair time is not included in the TBF calculation.

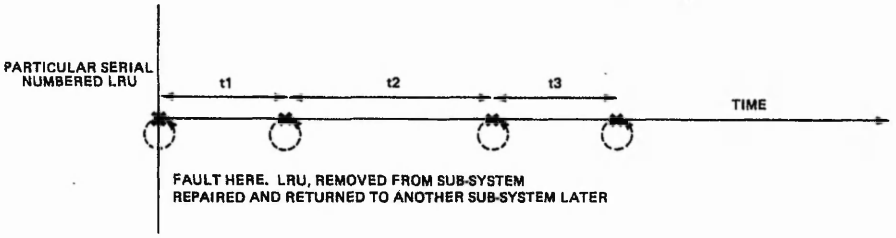


Fig. 4. Series of events to a particular LRU.

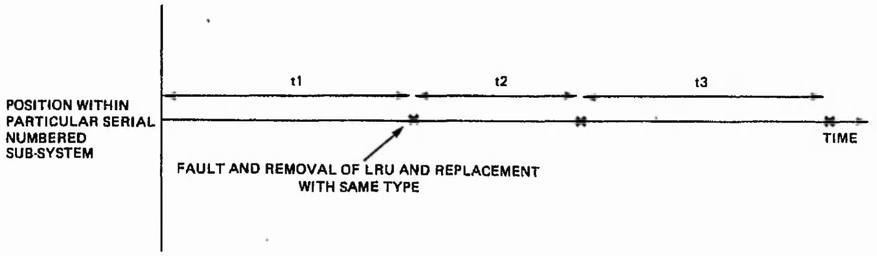


Fig. 5. Series of events to the position of a certain type of LRU within a particular sub-system.

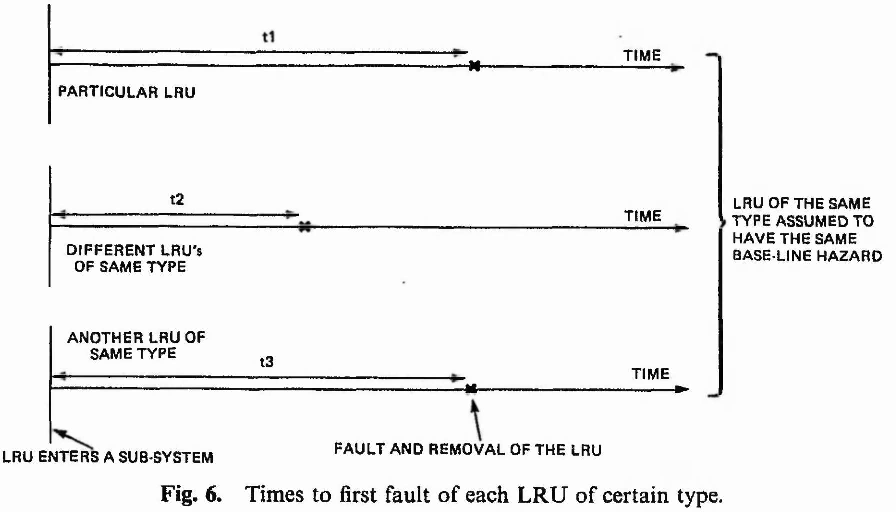


Fig. 6. Times to first fault of each LRU of certain type.

Model B
Here the series of events to the *position of a certain type of LRU, within a particular serial numbered sub-system* is followed. The time metric here is the time between faults to the same type of LRU in the fixed position within the sub-system. Figure 5 illustrates this series of events. Covariates can include which serial numbered LRU was removed at each fault.

Model C
The types of model here, look at *times to first fault of the LRUs of a given type*. Figure 6 illustrates this. Covariates could include which particular serial numbered LRU was faulty, and which particular serial numbered sub-system it was in when the fault occurred.

Model D
This model type looks at the events at the sub-system level. The *series of events on a particular serial numbered sub-system* is followed. The time metric here is the time between faults, irrespective of which LRUs are faulty, occurring on a particular unit. Figure 7 illustrates this series of events. Covariates can include which type of LRU was faulty.

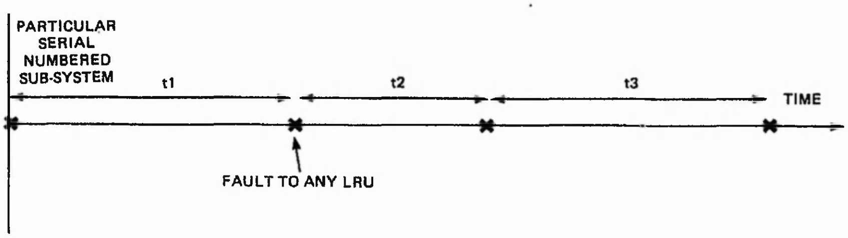


Fig. 7. Series of events to a particular sub-system.

CENSORING STRUCTURES

In dealing with the lifetimes of sub-assemblies or lower level units it is necessary to introduce censoring concepts. This is because a non-faulty sub-assembly is removed from the field when its parent LRU is replaced. Many censoring structures can be identified by interpreting various fault mechanisms in the hierarchical levels. Many of these structures can result in considerable complexity. The examples here are chosen because they seem reasonable in terms of what is known about the methods of maintenance and fault recording used.

In the simplest case consider Fig. 8 which depicts an LRU which has just

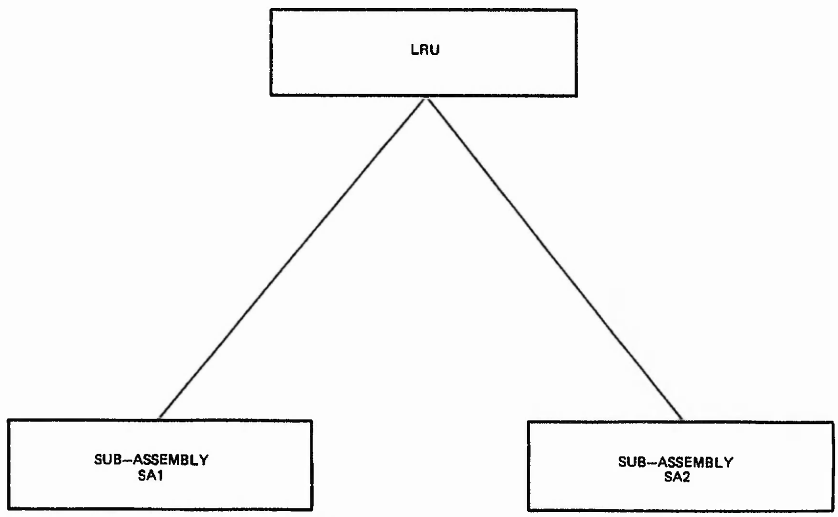


Fig. 8. LRU with two sub-assemblies.

two levels in its hierarchy, and two sub-assemblies. Four failure modes are possible for such an LRU, which could be recorded in the data.

$(\overline{\mathrm{SA1}}, \overline{\mathrm{SA2}})$—Fault recorded to LRU but to neither sub-assembly.
$(\mathrm{SA1}, \overline{\mathrm{SA2}})$—Sub-assembly 1 only has fault.
$(\overline{\mathrm{SA1}}, \mathrm{SA2})$—Sub-assembly 2 only has fault.
(SA1, SA2)—Both sub-assemblies have faults.

Four censoring structures that might be applicable are given below, and summarised in Table 1:

(i) *Two Failure Mechanism structure:* Two failure mechanisms (F1) and (F2) only are causing failure.

(F1)—sub-assembly 1 only fails, e.g. due to wearout.
(F2)—sub-assembly 2 only fails, e.g. due to wearout.

Each failure mechanism censors the other, i.e. 2 competing risks (see Fig. 9).

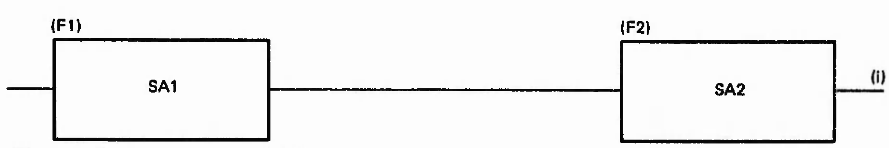


Fig. 9. Representation of failure mechanisms for censoring structure (i). The blocks show the parts of the system which may fail and the (F –), the failure mechanism which causes that part of the system to fail.

TABLE 1
Effects of Various Censoring Structures

Failure mode	*Censoring Structure* (*i*)		(*ii*)		(*iii*)		(*iv*)	
	Failure mechanism	*Effect on analysis*	*Failure mechanism*	*Effect on analysis*	*Failure mechanism*	*Effect on analysis*	*Failure mechanism*	*Effect on analysis*
($\overline{\text{SA1}}$, $\overline{\text{SA2}}$)	—	Ignored	(F0)	Censors (F1) & (F2)	(F0)	Censors (F1), (F2), (F12)	(F0)	Censors (F1), (F2)
(SA1, $\overline{\text{SA2}}$)	(F1)	Censors (F2)	(F1)	Censors (F0) & (F2)	(F1)	Censors (F0), (F2), (F12)	(F1)	Censors (F0), (F2)
($\overline{\text{SA1}}$, SA2)	(F2)	Censors (F1)	(F2)	Censors (F0) & (F1)	(F2)	Censors (F0), (F1), (F12)	(F2)	Censors (F0), (F1)
(SA1, SA2)	—	*	—	*	(F12)	Censors (F0), (F1), F2)	(F12)'	Censors nothing

* Treated as two independent faults, of types (SA1, $\overline{\text{SA2}}$) and ($\overline{\text{SA1}}$, SA2).

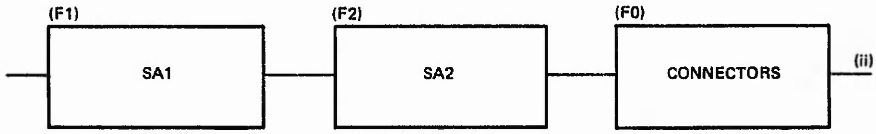


Fig. 10. Representation of failure mechanisms for censoring structure (ii).

($\overline{\text{SA1}}$, $\overline{\text{SA2}}$) type failures are ignored as being recording errors.
(SA1, SA2) type failures are treated as being two independent simultaneous failures due to chance. Consequently they are treated as (SA1, $\overline{\text{SA2}}$) type failure and a ($\overline{\text{SA1}}$, SA2) type failure.

(ii) *Three Failure Mechanism structure:* As well as the 2 failure mechanisms from (i) an additional failure mechanism, (F0), exists

(F0)—The LRU fails without either sub-assembly failing, e.g. due to the connectors failing.

Each failure mechanism censors the others, i.e. 3 competing risks (see Fig. 10).

(SA1, SA2) type failures are treated the same as in (i).

(iii) *Four Failure Mechanism structure:* In addition to the three failure mechanisms from (ii) a fourth failure mechanism (F12) exists.

(F12)—Both sub-assemblies fail together due to a common external cause, e.g. a power surge.

Each failure mechanism censors the other, i.e. 4 competing risks (see Fig. 11). This accounts for the four possible failure modes in the data.

(iv) *Modified Four Failure Mechanism structures:* In addition to the three failure mechanisms from (ii) a fourth failure mechanism F(12)′ exists:

(F12)′—a failure mechanism which accounts for simultaneous failure.

This failure mechanism is included because simultaneous occurrence of faults occurs more often than expected according to the

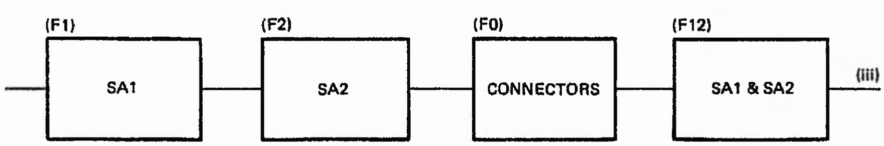


Fig. 11. Representation of failure mechanisms for censoring structure (iii).

independent failure model, thus we allow for an increased tendency for simultaneous failures. However this has not been reflected in the censoring structure, i.e. this failure mechanism neither censors any other failure mechanism or is censored by any other failure mechanism.

ANALYSIS

In the current application it is not possible to apply a model of type A since the data on when an LRU returns to the field is not available. Here, out of the remaining model types we consider the applications of some simple models of type D and C.

D models

For these models, the time metric was taken as the time in days between faults on a particular serial numbered sub-system. However, to avoid many TBF's of zero, multiple records on the same day were treated as a single fault event. In order to model the TBF's together, each sub-system was assumed to have the same baseline hazard. Times to first failure were ignored since the date of entry into service was unknown. Similarly the times since last failure were ignored since the time when reports ceased was unknown. Because D models operate at the subsystem level no censoring was required.

Ten explanatory variables were employed in the PHM model:

z_1, z_2, z_3, z_4, z_5 holding group dummies	z_6, z_7, z_8 season dummies	z_9 ETI	z_{10} time trend

(a) holding group dummies—The base group was selected to be the holding group that had the longest period of reported events. The dummy variables as defined in Table 2 compare the hazard rate for each holding group to that for the base group.
(b) season dummies—These dummy variables, compare the hazards during each of spring, summer and autumn respectively to the baseline season winter.
(c) ETI—This covariate is the actual ETI reading value as recorded at the time of the fault. (Irrespective of any apparent error)
(d) Time trend—This covariate, time in days since an arbitrary start date, allows for the possibility of a time trend affecting the TBF hazard.

TABLE 2
Coding of Holding Group Covariates

		z_1	z_2	z_3	z_4	z_5
(base)	holding group I	0	0	0	0	0
	holding group II	1	0	0	0	0
	holding group III	0	1	0	0	0
	holding group IV	0	0	1	0	0
	holding group V	0	0	0	1	0
	holding group VI	0	0	0	0	1

Because of the large number of types of LRUs present in a sub-system the types of LRU were not included as covariates.

Results for sub-system (1)

The results after the usual backwards stepwise elimination procedure based on two-tailed 5% tests, are given in Table 3. The p-values indicate the probability of obtaining such extreme estimates for the β_i's just due to chance, if there were no real effect for the covariate.

Discussion

(a) The likelihood ratio statistic: This, is seen to exceed the tabulated upper 5% critical value for a chi-squared distribution with four degrees of freedom. This indicates that the fitted model provides significantly more explanation than the model that the covariates have no effect and that the data is homogeneous.

(b) Significant Covariates: The positive estimates of β_1 and β_5 imply that sub-systems (1) from holding groups II, and more particularly VI, experience a higher hazard than those in the other groups (by factors of 45% and 75% respectively).

TABLE 3
Model for Sub-system 1 after Backwards Stepwise Elimination

Significant covariates	β_i	*p-Value (less than)*	*Likelihood Ratio Stat.*
z_1—holding group II	0·373 214	0·000 05	68·695
z_5—holding group VI	0·562 168	0·000 05	(5% critical
z_6—spring	0·126 810	0·020 55	from tables
z_{10}—time trend	−0·000 604	0·000 05	= 9·488)

The positive estimate of β_6 indicates that sub-systems (1) experience approximately a 13% higher hazard in the spring than during the other seasons, possibly reflecting increased exposure due to exercises. This covariate is however only marginally significant, there being approximately a 2% chance that the effect is purely spurious.

The negative estimate of β_{10} indicates that there is decreasing hazard as time passes (at the rate of approximately a 20% reduction per year). The number of sub-systems (1) entering the field was known to be rising over the period. Assuming that times to first fault are not shortening with calendar time, the sub-systems (1) are becoming more reliable.

(c) Non-Significant Covariates: The holding units III–V do not appear to have a significantly different hazard rate, for their sub-systems (1), to the base holding group. The hazards during summer and autumn do not appear to be significantly different to that during winter.

Interestingly, the ETI reading was found to be non-significant. The elapsed time in normal mode was initially considered as important in explaining the fault rates of the units. The finding that this is non-significant, in this model is probably mostly due to the unreliability of the ETI's themselves.

GRAPHICAL VALIDATION TECHNIQUES

A number of geographical techniques are available to examine the appropriateness and goodness-of-fit of the PHM model.

(1) *Baseline hazard plot*

Figure 12 shows a Weibull hazard plot for the baseline hazard obtained from the PHM model. The plot is reasonably straight indicating that, apart from the effects of the covariates, the Weibull is a reasonable distribution for the times between faults (e.g. Ref. 14). The shape and scale parameters are estimated at approximately 0·84 and 11·02 days respectively. Since the shape parameter is less than 1 the TBF's exhibit a decreasing hazard rate.

(2) *Cox and Snell residuals*

A graphical goodness-of-fit test for the whole model is provided by plotting Cox and Snell 'crude' residuals[15] against their expected order statistics. If the model is a good fit we expect the residuals to lie on a straight line, of gradient 1, passing through the origin.[16] Figure 13 shows such a plot.

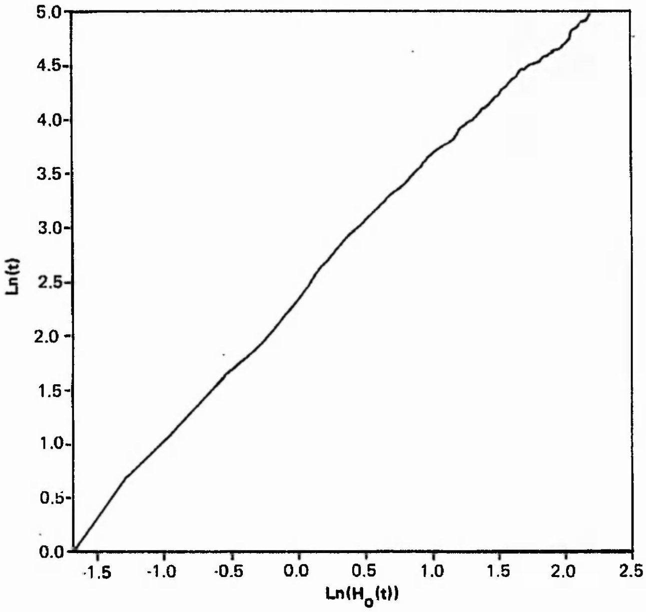


Fig. 12. Baseline Weibull hazard plot.

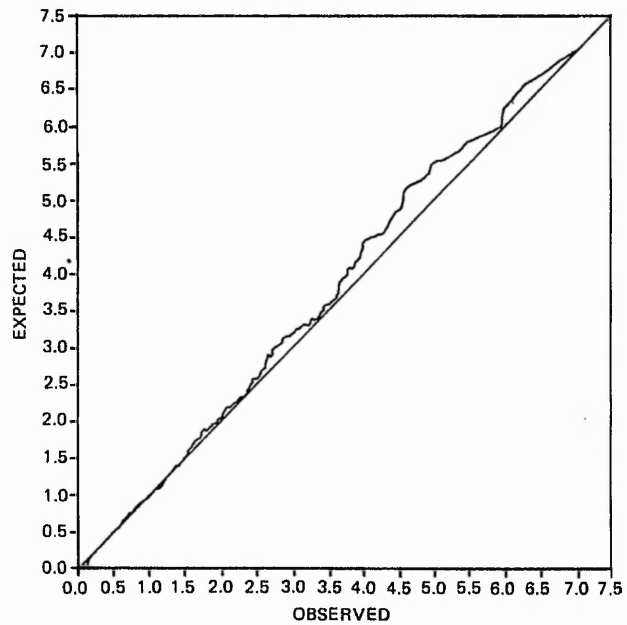


Fig. 13. Cox and and Snell residuals.

(3) *Proportionality plots*

The PHM model assumes that all covariates affect the baseline hazard proportionally. The validity of this assumption needs to be verified from the data for each covariate, so that if it is unjustified, the model can be adapted where necessary. Figures 14–16 are proportionality plots for the significant binary covariates. In the plots the data is stratified on each covariate, and the model is then fitted separately to each stratum. If the assumption of proportionality holds, plotting the $\log_e$ baseline cumulative hazard $\ln H_0(t)$ against t for each stratum on the same graph, should produce plots with constant vertical separation for all t.[16] Figure 14 shows such a plot for the stratification on the covariate z_1, holding group II, for which the proportionality assumption, whilst not perfect, appears approximately valid. Figure 15 stratifies on the covariate z_5, holding group VI. As in Fig. 14, the vertical separation changes towards the end of the graphs for which there is relatively little data (hence the crossing point). The curves remain however, reasonably well separated and the proportionality assumption again appears plausible for the majority of the data which corresponds to Tbf's

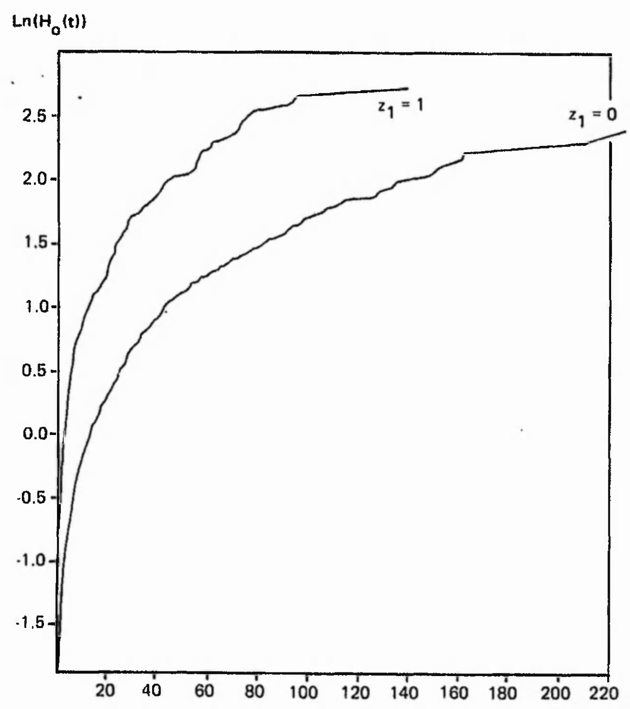


Fig. 14. Proportionality plot for covariate z_1.

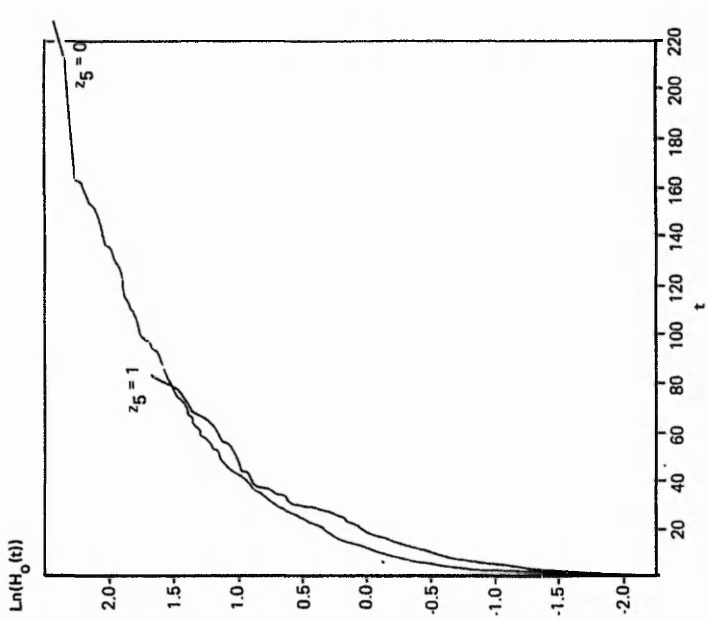


Fig. 15. Proportionality plot for covariate z_5.

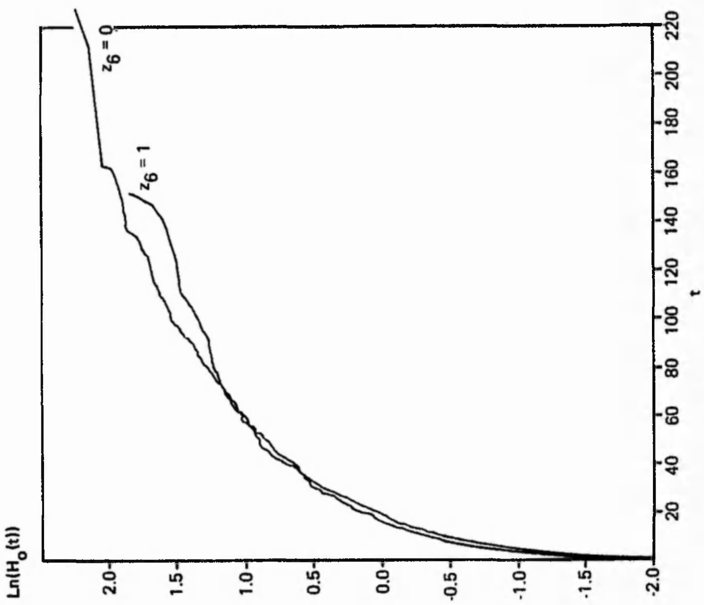


Fig. 16. Proportionality plot for covariate z_6.

under 20 days. Figure 16 stratifies on the covariate z_6, spring. The plots cross in numerous places, reflecting the fact that there is little difference between the estimated hazards for the two groups, as is apparent from the marginality of the significance of z_6 in Table 3.

(4) *Influence functions*

Figure 17 shows a standardised plot of the influence of individual TBF's on the β_6 parameter estimates for the covariate spring.[17,18] The plot shows the estimated normal deviate for the covariate coefficient when each single TBF and censoring time is excluded from the model, one at a time. This is plotted against the order of magnitude of the TBF's on the horizontal axis. We can thus examine which TBF's have times most influential on the observed significance of the covariate, and which if any would, if deleted, remove the significance of the covariate. For the 5% two-tailed tests these points correspond to estimated normal deviates in the range $(-1{\cdot}96, +1{\cdot}96)$. The influence of an event is the difference between the β_i estimate with and without the TBF included: $I_{ij} = \hat{\beta}_i - \hat{\beta}_{ij}$, where $\hat{\beta}_{ij}$ is the estimate of β_i made when TBF j is excluded. The figure presented here is based on the simplifying approximation of treating the standard error of the β_i estimator as unaltered after elimination of the single data point. The letters on the plot represent the

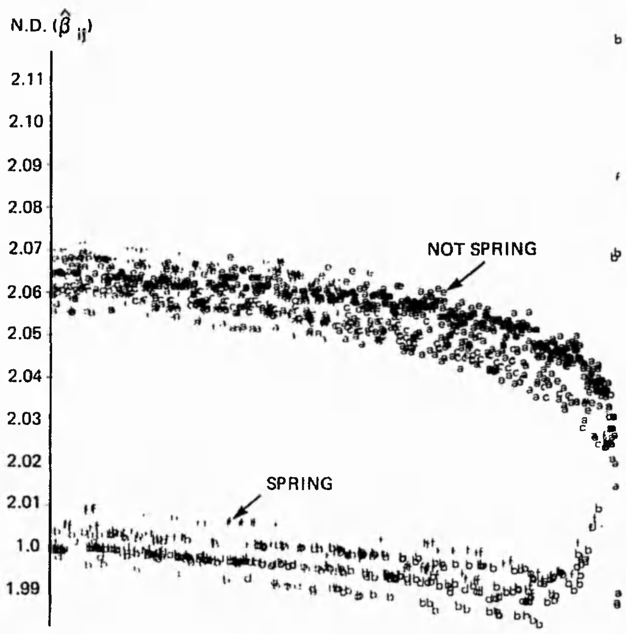


Fig. 17. Influence plot for covariate, Spring.

combinations of the significant covariates associated with each TBF. The plot splits into two distinct groups, one associated with TBF's in spring, and one with TBF's not in spring. This split is common for binary covariates. Although the covariate is marginal there do not appear to be any TBF's that would change the significance of the covariate if omitted.

Results for sub-systems (2)

Commencing from the same set of initial covariates as for sub-systems (1), and applying the same model D structure, the results for sub-systems (2) after backwards elimination as shown in Table 4.

TABLE 4
Model for Sub-systems (2) after Backwards Stepwise Elimination

Significant covariates	$\hat{\beta}_i$	*p-Value (less than)*	*Likelihood Ratio Stat.*
z_1—holding group II	0·400 882	0·000 05	33·71
z_2—holding group III	0·256 648	0·014 05	(5% critical
z_5—holding group VI	0·532 513	0·000 05	from tables
z_{10}—time trend	−0·000 627	0·000 15	= 9·488)

Discussion
As may be expected, the results are broadly similar to those obtained for sub-systems (1). Now, however the spring is no longer significant, and holding group III is marginally significant. The results for z_1, z_5, and z_{10} are highly significant as before, and are of the same sign and order of magnitude as in the model for the sub-systems (1).

It may not be surprising that the results for two types of sub-system are so similar, since the holding groups operate their units as complete systems. Thus if holding group I exercised their sub-system (1) more than did holding group II, then they necessarily also exercised their sub-systems (2) more.

C MODELS

For illustrative purposes here, we present the analysis of a PHM model of type C, for a particular type of LRU, present in sub-systems (3), selected because it has only two sub-assemblies, and two hierarchical levels. The censoring structure (iv) described in the previous section was applied.

The covariates z_1–z_{10} were again used to describe holding group, season,

TABLE 5
Coding of Covariates for Level of Fault

	z_{11}	z_{12}	z_{13}
sub-assembly 1 faulty (base)	0	0	0
sub-assembly 2 faulty	0	0	1
both sub-assemblies faulty simultaneously	1	0	0
neither sub-assembly faulty	0	1	0

ETI, and time trend and z_{11}–z_{13} were introduced for the censoring procedure for the hierarchical system. Table 5 shows the coding of the covariates z_{11}–z_{13}.

Results

After backwards stepwise elimination only one covariate, z_{11} (recording simultaneous faults), remained significant; with $\beta_{11} = 1{\cdot}871\,28$ and a p-value of less than 0·000 05.

Discussion

The positive estimate of β_{11} indicates that there are shorter times to the simulaneous faults on the sub-assemblies than times to faults due to the other mechanisms.

No differences between the holding groups or the seasons was identified in this model.

Once again the ETI reading is found to be non-significant, and in this model there is no evidence that a time trend is affecting the times to first fault of this type of LRU.

CONCLUSIONS

1. Whilst reliability for weapon systems may be good, reliability analysis for weapons systems has been poor and thus has limited the role of the statistician effectively contributing to the design process.
2. Poor database design, recording and quality hampers analysis.
3. Reliability analysis methods need modification to be applied satisfactorily to weapon systems.
4. Proportional Hazards Modelling is an effective method for the analysis of weapon systems data with its wealth of auxiliary (covariate) information.

5. Various model formulations, corresponding to different physical interpretations are possible. The reliability analyst's role is to explore these in the search for structure and categorisation of mechanisms. In such an approach on-line graphical aids as partly reported here are of great value.
6. This paper has presented an initial review of parts of the reliability analysis of a weapons system currently being undertaken by British Aerospace and Trent Polytechnic personnel. Work in this area is continuing.

ACKNOWLEDGEMENT

Mrs Walker's contribution to this paper was funded under a NAB grant.

REFERENCES

1. Finney, D. J., *Probit Analysis*, Cambridge University Press, Cambridge, UK, 1971.
2. Goodman, L. A., *Log-Linear Models and Latent Structure Analysis*, University Press of America, 1984.
3. Goodman, L. A., *Analysing Cross Classified Data having Ordered Categories*, Harvard University Press, Cambridge, MA, 1984.
4. Kalbfleisch, J. D. and Prentice, R. L., *The Statistical Analysis of Failure Time Data*, John Wiley, New York, 1980.
5. Bendell, A., Collecting, analysing and applying reliability data in the process industry, an overview, *Proc. Seminar on Data Collection and Analysis for Reliability Assessment*, I. Mech. E., 1986.
6. Bendell, A., New methods in reliability analysis, in *Reliability Technology Theory and Applications*, eds J. Møltoft and F. Jensen, Elsevier Science Publishers B.V., Amsterdam, 1986, pp. 351–7.
7. Bendell, A., Walley, M., Wightman, D. W. and Wood, L. M., Proportional hazards modelling in reliability analysis—An application of brake discs on high speed trains, *Quality and Reliability*, **2**(1) (1986), pp. 45–52.
8. Wightman, D. W. and Bendell, A., The practical application of proportional hazards modelling, *Reliab. Engng*, **15** (1986), pp. 29–53.
9. Tukey, J. W., *Exploratory Data Analysis*, Addison-Wesley, London, 1977.
10. Libberton, G. P., Bendell, A., Walls, L. A. and Cannon, A. G., Reliability data collection and analysis for automatic fire detection systems on a large industrial site, *Proc. Seminar on Data Collection and Analysis for Reliability Assessment*, I. Mech. E., 1986.
11. Walls, L. A. and Bendell, A., Time series methods in reliability, *Reliab. Engng*, **18** (1987), pp. 239–65.
12. Cox, D. R., Regression models and life-tables, *J.R. Stat. Soc. B*, **34** (1972), pp. 184–220.

13. Ascher, H., Reliability models for repairable systems, in *Reliability Technology Theory and Applications*, eds J. Møltoft and F. Jensen, Elsevier Science Publishers, Amsterdam, 1986, pp. 177–85.
14. King, J. R., *Probability Charts for Decision Making*, Industrial Press Inc., New York, 1971.
15. Cox, D. R. and Snell, E. J., A general definition of residuals, *J. R. Stat. Soc. B*, **30** (1968), pp. 248–75.
16. Kay, R., Proportional Hazard regression models and the analysis of censored survival data, *Applied Stats.*, **26**(3) (1977), pp. 227–37.
17. Cain, K. C. and Lange, N. T., Approximate case influence for the proportional hazard model with censored data, *Biometrics*, **40** (1984), pp. 493–9.
18. Reid, N. and Crépeau, H., Influence functions for proportional hazards regression, *Biometrica* **72** (1985), pp. 1–9.

APPENDIX C

APPLICATIONS OF PROPORTIONAL HAZARDS MODELLING TO HARDWARE AND SOFTWARE RELIABILITY

Dr Dave Wightman, Mrs. Estelle Walker and Professor Tony Bendell

TRENT POLYTECHNIC

Nottingham, England

The paper describes work at Trent Polytechnic in extending uses of Proportional Hazards Modelling (PHM) within both hardware and software applications. The Reliability Group, led by Professor Tony Bendell, has experience of applying PHM to a number of different industries and technologies. As well as discussing the adaptation of PHM to diverse reliability data structures, we discuss developments in diagnostic aids employed to investigate model structure. Illustrative material is drawn from applications in railway engineering, computer hardware, weapon systems, electricity supply, computer software and electronic components.

INTRODUCTION: PHM

The origin of PHM is the seminal paper by Professor D R Cox which was presented to the Royal Statistical Society in March 1972 (Cox 1972).

PHM combines concepts from biostatistics and reliability theory; incorporating regression-like arguments for explanatory variables into life-table analysis. It is a technique whereby identification of independent effects of variables thought to influence the life-length of equipment is possible without the necessity of specifying the distributional form of life *a-priori*.

The model is structured on the assumed decomposition of the hazard function into the product of a baseline or generic hazard function and (usually) an exponential term incorporating the effects of variables. The fundamental equation is:

$$h(t;z_1,z_2,\ldots,z_k) = h_o(t).\exp(\beta_1 z_1 + \beta_2 z_2 + \ldots + \beta_k z_k) \qquad (1)$$

where the β's are the unknown parameters of the model defining the effects of each of the explanatory variables; the z_i's are the values of the explanatory variables; $h_o(t)$ is the baseline hazard function (usually distribution free).

From (1) the effects of the z_i's are seen to act multiplicatively on the baseline hazard $h_o(t)$, so that for different values of an explanatory variable their respective hazard functions are proportional over all time.

The unknown parameters β are estimated through the maximisation of Cox's partial likelihood, (Cox, 1972). Optimisation procedures available include the expectation-maximisation (EM) algorithm, (Cox and Oakes, 1984). However, for the material illustrated in this paper we employ a method whereby we firstly take the natural logarithm of the partial likelihood and then obtain the first and second partial differentials with respect to β_j, $j=1,2,\ldots,k$. The parameters $\beta_1,\beta_2,\ldots,\beta_k$ are estimated iteratively based on a Taylor Series expansion for each step of the iteration, starting with initial values of zero (the method of scoring). Tests of whether each explanatory variable has any significant effect are based on the asymptotic normality of the estimators. A backwards stepwise procedure is incorporated whereby non-significant factors are excluded one at a time and the model rerun until all the factors are significant.

Having obtained a set of significant explanatory variables we then obtain an estimate of $h_0(t)$, using the approach in Kalbfleisch and Prentice (1980).

PHM RELIABILITY APPLICATIONS

The reliability literature mainly concentrates on data from repairable systems. The reported data is diverse, including moterettes (Dale, 1983), marine gas turbines and ship sonars (Ascher, 1983), valves in light water nuclear generating plants (Booker et al, 1980), aircraft engines (Jardine and Anderson, 1984), sodium sulphur batteries (Ansell and Ansell, 1986), transmission equipment (Argent et al, 1986), (Manning et al, 1987), weapon systems (Gray et al, 1987) and subsurface safety valves (Lindquist et al, 1988). However, in some analyses the form of the models employed are in general basic and are not developed to take specific account of the complexities arising in reliability data.

Many of the PHM analyses of repairable systems in reliability employ one (of four) of the formulations proposed for a Leukaemia study by Prentice, Williams and Peterson (1981)(PWP). The paper by PWP(1981) concentrates upon data arising from a large number of study subjects with a small number of failures on each subject. Experience at Trent has shown that data structures arising in reliability studies are considerably more diverse. Such structures reflect the data collection processes and procedures as well as the field deployment and failure phenomena.

The question of appropriate structure for modelling and analysis centres about what are the appropriate point processes to describe failures of repairable systems. In this respect it is essential when analysing these complex data sets that PHM is employed as an exploratory data analysis tool, searching for appropriate pattern and structure.

Some commonly occurring reliability data structures are illustrated in Figure 1. Within the constraints of this paper we are of course unable to discuss in much detail the structure and analysis of all the data structures identified in Figure 1. However, we briefly describe each in the next section, and in particular concentrate on PHM for competing risks

SINGLE OBSERVATION ON EACH ITEM

The application of PHM to a single failure observation on each item of equipment (of which non-repairable items are a special case) has the advantage that much of the complexity and necessary exploration associated with its application to multiple observations on repairable systems is unnecessary. However, the application still requires the adoption of a flexible approach in which different modelling formulations are considered, each of which contributes to the understanding of the data set under study.

REPAIRABLE ITEMS/ITEMS UNIDENTIFIED

A commonly encountered process is where we have information on failures of a group of items within which failures to individual items are not labelled. Thus, we do not have any information on the past failure history of individual items in the group. Often in this situation, t, in equation (1) is taken as the series of subsequent failure times for the group. However, the possibilities are heavily dependent upon the context in which the data arises.

REPAIRABLE ITEMS/ITEMS IDENTIFIED

Prentice, Williams and Peterson (1981) introduce a variation on the basic PHM in which the structure of the model considered is essentially one that allows for reliability growth or decay in subsequent inter-failure periods on the same item. The concept behind this formulation is that items move through strata upon failure, so that prior to the first failure they are in stratum 1, after the first failure and prior to the second they are in stratum 2, etc. PWP (1981) suggested four alternative approaches to analysis.

DATA FROM ONE SYSTEM

Often reliability data takes the form of a sequence of events on one system. Frequently in these data sets the underlying failure mechanism is based upon the time between subsequent failures. However, more than one cause of failure may be present, so that for example; t in equation (1) may be based on the time between subsequent failures due to the same cause.

NESTED/HIERARCHICAL DATA STRUCTURES

The nested/hierarchical nature of some systems leads to the identification of a number of different point processes on which to base any PHM analysis. These point processes correspond to different levels of the hierarchy; system, sub-system, assembly, sub-assembly, etc. In analysing such complex data it is necessary not only to consider carefully the choice of t in equation (1) but also the censoring structures that may be appropriate. Gray et al (1987) found that many censoring 'structures' could be identified by interpreting various fault mechanisms in the hierarchical levels.

APPLICATION OF PHM TO COMPETING RISKS

Competing risks formulations within the PHM techniques exist. Two models were introduced by Holt (1978); one which has the same baseline for each cause with cause specific β coefficients, and the other with different baseline and β coefficients for each cause. These are discussed in Kalbfleisch and Prentice (1980).

In this paper we consider a model for competing risks whereby the basic time metric, t, in equation (1), is taken as the time between consecutive failures. Given that there are l + 1 failure modes, l binary dummy variables are introduced into the covariate set to represent them. A censoring event is generated at each failure time for each of the other failure modes.

We now show that the partial likelihood for this model construction factorises into two terms, one containing only information from failure mode variables and the other being the usual partial likelihood as shown below.

The partial likelihood for a model without competing risks is:

$$Lp = \prod_{i=1}^{n} \frac{\exp(\beta_1 z_{1i} + \beta_2 z_{2i} + \dots + \beta_k z_{ki})}{\sum_{m\varepsilon R_i} \exp(\beta_1 z_{1m} + \beta_2 z_{2m} + \dots + \beta_{km})}$$

where n = total number of failures

k = number of covariates

Under competing risks each event in the model (2) above now becomes 1 + 1 events. Let γ_j represent the associated parameters for the indicator variable x_j for the j'th failure mode. The partial likelihood now becomes

$$L_e = \prod_{i=1}^{n} \frac{\exp(\beta_1 Z_{1i} + \beta_2 Z_{2i} + \cdots + \beta_k z_i + \gamma_{1 1i} + \gamma_2 x_{2i} + \cdots + \gamma_l x_{li})}{\sum_{M\varepsilon R_i} \begin{array}{l} \exp(\beta_1 Z_{1m} + \beta_2 Z_{2m} + \cdots + \beta_k Z_{km} + 0) + \\ \exp(\beta_1 Z_{1m} + \beta_2 Z_{2m} + \cdots + \beta_k z_{km} + \gamma_1) + \\ \quad \vdots \\ \exp(\beta_1 z_{1m} + \beta_2 z_{2m} + \cdots + \beta_k z_{km} + \gamma_l) \end{array}} \qquad (3)$$

hence,

$$Le = \prod_{i=1}^{n} \frac{\exp(\gamma_1 x_{1i} + \gamma_2 x_{2i} + \dots + \gamma_l x_{li}) \ \exp(\beta_1 z_{1i} + \beta_2 z_{2i} + \beta_k z_{ki})}{\sum_{M\varepsilon R_i} [1 + \exp(\gamma_1) + \dots + \exp(\gamma_l)] \exp(\beta_1 z_{1m} + \beta_2 z_{2m} + \dots + \beta_k z_{km})}$$

$$= \prod_{i=1}^{n} \frac{\exp(\gamma_1 x_{1i} + \dots + \gamma_l x_{li})}{[1+\exp(\gamma_1)+\dots+\exp(\gamma_l)]} \times \prod_{i=1}^{n} \frac{\exp(\beta_1 z_{1i} + \dots + \beta_k z_{ki})}{\sum_{M\varepsilon R_i} \exp(\beta_1 z_{1m} + \dots + \beta_k z_{km})}$$

$$L_e = L_f \times L_p \qquad (4)$$

where

$$L_f = \prod_{i=1}^{n} \frac{\exp(\gamma_1 x_{1i} + \dots + \gamma_l x_{li})}{[1+\exp(\gamma_1)+\dots+\exp(\gamma_l)]} \qquad (5)$$

Maximising the partial likelihood L_e is equivalent to maximising L_p and the factor involving the dummy variables independently. Hence the estimates of the β's from the competing risks model will be the same as those from the initial model.

Now maximising the factor L_f:

$$\frac{d\log L_f}{d\gamma_j} = \sum_{i=1}^{n} x_{ji} - \frac{\exp(\gamma_j)}{[1+\exp(\gamma_1)+\ldots+\exp(\gamma_l)]} = 0$$

$$j=1,2,\ldots,l$$

let n_j = number of the n events for which $x_{ji} = 1$ then

$$n_j - \frac{n.\exp(\gamma_j)}{(1 + \exp\gamma_1)+\ldots+\exp(\gamma_l\ ))} = 0 \qquad (6)$$

$$j=1,2,\ldots,l$$

From (6) a set of l simulations linear equations can be generated which can be solved to give the parameter estimates γ for the dummy variables. The parameter estimates γ are in terms only of the number of 'deaths' within each failure mode.

COVARIATE FORMULATION

From the investigations of the Central Electricity Generating Board's (CEGB) transmission failure data, see Argent et al (1986), Manning et al (1987) and Manning (1988), one aspect of interest to the CEGB was the non-significance of the overhead line length as a covariate. A possible explanation of the non-significance of the line length is that there is little information in the covariate; the line length figure was updated at the end of each year for the first 7 years of data and remained constant for the remainder of the period.

Given the line length values, of particular interest to the CEGB was that by increasing the line length by a multiple, whether there was evidence that the hazard was increased by the same multiple. Multiplying the line length by k and investigating whether the hazard is multiplied by k can be achieved by taking the logarithm of the line length (say x), viz

$$e^{\beta \ln kx} = ke^{\beta \ln x}$$

$$e^{\beta \ln k} . e^{\beta \ln x} = ke^{\beta \ln x}$$

$$e^{\beta \ln k} = k$$

$$k^{\beta} = k$$

$$\beta = 1$$

where $e^{\beta \ln x}$:- hazard for length of line x. So that a $\beta \sim 1$ would indicate the property of increasing the length of the line k times would increase the hazard k times. However, running the model with the logarithm of line length found this formulation of the covariate non-significant and a β coefficient markedly different from 1.

It is usual practice in the PHM applications in reliability to include covariate information as recorded. However, other formulations for the covariate information may be more appropriate. For example, in Davies et al (1987) who analysed the 16 software failure data sets for Musa (1980) different formulations of the failure number covariate were considered - these were (where N is the failure number) N, $\sqrt{N}$, N^2, $1/N$ and $\ln N$. Investigation of the form of the failure number information was undertaken since in many conventional software reliability models failure number is an integral part of the model. From Davies et al (1987) there is the suggestion that the inclusion of failure number information (although not statistically significant in the majority of cases) as N or $\sqrt{N}$ are the most appropriate formulations.

The above examples were introduced to investigate specific questions within particular data sets. However, in a more general sense, work at Trent Polytechnic has recently focused on whether *a-priori* it is possible to choose/obtain covariate formulations that lead to the proportionality assumption being fulfilled.

EXAMPLE OF ADAPTION OF PHM FOR A LARGE DATA SET

Due to the number of events and covariates the problem with computer space constraints has occurred in some of the analyses undertaken by the reliability group at Trent. A recent problem of this nature occurred in considering the use of PHM in the investigation of the data held in the electronic component data base at Loughborough University, Loughborough. Investigation focused on component types which had a 'reasonable' number of failures for analysis purposes. However, because of the number of (same) components of the circuit board, the number of boards in a unit and the number of units in the field, substantial censoring information was generated (in one case of the order 500,000 censoring observations). This number of observations creates computer program problems, e.g. in terms of dimension statements and also storage and data handling/manipulation problems.

Background knowledge supported by inspection of the data revealed that the vast majority of the censoring observations had a censored time larger than the greatest failure time, so that in PHM these observations are given a rank equal to that of the last failure. Inspection also revealed that for large groups of observations the covariate values for each group and the number of observations in the group. The computing routines employed for PHM at Trent were altered to include reading of the summarised data file for the censored observations and to amend the calculations that are associated with the largest rank.

DIAGNOSTICS FOR PHM

A number of graphical techniques can be employed to examine the appropriateness and fit of the PHM model.

A method of testing the crucial assumption of proportionality between different levels of a covariate is provided by Kay (1977). Cox (1979) introduce two techniques similar to this, one which employs the logarithm of the hazard, the other the survivor distribution.

Kay (1977) suggests an approach to testing the appropriateness of the model whereby 'residual' quantities as defined by Cox and Snell (1968) are obtained. These quantities should exhibit approximately the properties of a random sample from a censored exponential sample with failure rate 1, if the model is appropriate. Hence, survival estimates based on the residuals should when plotted on a logarithmic scale, yield approximately a straight line with slope - 1. Self (1981) extends the technique to include time dependent covariates. Aitken and Clayton (1980) describe a variance-stabilised version of the plot based upon an Arc-sin transformation.

Schoenfeld (1982) defines partial residuals for each significant covariate which can be used to look for local deviations from the proportional hazards model.

Lagakos (1981) defines residual scores for each individual from consideration of the partial likelihood score function and the cumulative hazard. Lagakos then proceeds to adjust the observed ranks based upon this information. If the model is appropriate, then the re-adjustment should account for the effects of the covariates.

Cain and Lange (1984) and Reid and Crepeau (1985) obtain influence functions for the proportional hazards model. These influence functions, for each explanatory factor, approximate the effect of individual cases upon the estimate of the associated coefficient. Strorer and Crowley (1985) also discuss a diagnostic for estimating the changes in β due to the deletion of a single observation.

The recent paper by Barlow and Prentice (1988) represents residuals for relative risk regression as an estimator of a stochastic integral with respect to the martingale arising form a subject's failure time counting process. Previously proposed residuals for individual study subjects and for specific time points are shown to be special cases of this definition.

Despite a great deal of work having been undertaken in this area the use of graphical diagnostics is still currently crude. This is, therefore, an area for development.

The strengths and the limitations of three of these diagnostics; proportionality plots, Schoenfeld residuals and influence functions, are discussed below.

PROPORTIONALITY ASSUMPTION

The assumption of proportionality in PHM is that different values of the covariate have hazard functions that are proportional to each other over all time.

The most commonly applied method to test whether a covariate follows the proportionality assumption is to stratify upon the covariate of interest and for each stratum (level) of the covariate plot the logarithm of the cumulative baseline hazard against time, see Kay (1977). That is, for a binary covariate, z_k say, we have for the hazard function

$$h(t;\underline{z},z_k) = \begin{cases} h_o(t)\exp(\underline{\beta z})\exp(\beta_k z_k), & z_k = 1 \\ h_o(t)\exp(\underline{\beta z}), & z_k = 0 \end{cases}$$

and for the logarithm of the cumulative hazard

$$\ln[\Sigma h_o\ (t)] + \beta z + \beta_k \qquad , \qquad z_k = 1$$

$$\ln[\Sigma h_o\ (t)] + \beta z \qquad , \qquad z_k = 0$$

So that, if the assumption is appropriate then plotting the logarithm of the cumulative hazard for each stratum against time should result in a constant vertical separation. Figure 2 illustrates this graphically.

The problem with this procedure is that it contains a highly subjective element. As a first step to minimize the subjectivity the vertical separation of the plots at various points along the time scale can be plotted separately below the graph, this is illustrated by Figure 3 taken from the analysis of ICL hardware for the covariate indicating 'No fault found' (See Drury et al 1987). Although, we should expect this to reveal a straight horizontal line if proportionality holds it does not, however, give any real indication as to whether there is a reasonable constant vertical separation present. It would therefore be useful to construct confidence bounds around the estimates. Initial bounds for the log baseline cumulative hazards were produced, from a transformation of the confidence interval around the baseline survivor function, which was constructed form an asymptotic estimate of the variance for the survivor function (Link 1984). These asymptotic bounds were very wide and offered little information, an example can be seen in Figure 4. The +'s are the 95% bounds around the top plot, and the x's those for the lower plot.

Simulated bounds were considered in an attempt to narrow the bounds. Assuming a Weibull distributed baseline with parameters estimated from a hazard plot for each of the strata, 90% limits were constructed by simulating twenty groups of fifty failure times from the estimated Weibull distribution. The problem with these, although the bounds are a little narrower, is that they are not always defined at small t, since the first simulated failure time can be relatively large, an example can be seen in Figure 5.

From the bounds in Figure 5, despite their shortcomings as described above, we are able to see that the two strata are clearly separated (neither set of bounds overlap) indicating a real difference between the levels of the covariate. The bounds generally narrow as t increases indicating less variance in the latter part of the plots.

Work is continuing at Trent in the area of minimizing the level of subjectivity in this diagnostic.

SCHOENFELD RESIDUALS

The Schoenfeld residuals (Schoenfeld, 1982) are known as 'partial residuals' since a set is obtained for each covariate.

For the j'th covariate at failure time t_i the partial residual r_{ji} is defined:

$$r_{ji} = z_{ji} - E[z_{ji} \mid R_i]$$

where z_{ji} is the value of the j'th covariate at failure time t_i, and

$$E[Z_{ji} \mid R_i] = \frac{\sum_{R_i} Z_{ji} \exp(\beta z)}{\sum_{Ri} \exp(\beta z)}$$

the partial residuals are obtained from elements of the score vector (the vector of first differentials of the log likelihood with respect to β_j).

If proportional hazards holds $E[r_{ji}]$ 0, and a plot of r_{ji} versus t_i will be centred about 0.

For a binary covariate there is a split of residuals, above and below the axis, corresponding to the two categories of the covariate. Since there are often in reliability problems a large number of tied failure times the density of partial residuals at many of the points cannot be seen. Visual inspection is also hindered since the two bands are rarely equidistant from the axis. To ease visual inspection, therefore, we have added to the plot a moving average based on intervals of 20 failure times. The moving average can then be looked at for local fit of the model, an example can be seen in Figure 6. The +'s on Figure 6 represent single failures and the x's tied failure points.

Experience at Trent tends to show that the appearance of these residual plots is highly affected by the pattern of censoring observations. To this end we now consider plotting censoring observations on the same figure. As an example we show Figure 7 the Schoenfeld residuals and censoring observations for the binary covariate 'route' in the analysis of the brake discs on high speed trains, see Bendell et al (1986) and Wightman (1987). The x's on the plot are the partial residuals as defined by Schoenfeld (1982), whilst the +'s are the partial residuals obtained using a slight approximation in the presence of tied failure points, see Wightman (1987), the o's are the censoring events plotted at their covariate value and censoring time.

The plots of the o's gives an indication of the distribution of the censoring points for each covariate and may reveal outliers in the censoring observations or provide evidence of a relationship between censoring and the covariate.

The residuals on Figure 7 form two bounds reflecting the binary nature of the covariate. There are more positive than negative residuals which reflects there being more failures on the West route than the East. The predominance of failures on the West route is balanced, as is necessary since the sum of residuals for each covariate is zero, by the East route residuals being greater in magnitude. In Figure 7 the patterns of censoring points for the East route (o's along the mileage axis at residual value one) show no clear difference, however, the censoring observation at approximately 470,000 miles on the East route is particularly unusual.

This outlying point would not be detected in the usual Schoenfeld plots and would in general go unnoticed because of the observation being censored and many of the diagnostics and plots not explicitly using 'time to failure' information, but rather rank information.

ESTIMATED INFLUENCE FUNCTIONS

Cain and Lange (1984) and Reid and Crepeau (1985) present essentially the same technique for approximating the influence of individual items upon each of the β coefficients obtained from PHM. These empirical influence functions can then be used in an informal manner to identify influential observations which may greatly affect statistical inferences regarding the covariates.

The influence of each observation on the β coefficients can be obtained in an exact manner by dropping each observation in turn and refitting the model. However, this is not feasible in most practical applications because the number of observations and covariates implies prohibitive computer time.

Based on a Taylor series expansion Cain and Lange (1984) employ a first order approximation to

$$\hat{\beta} - \hat{\beta}_{(j)}$$

where:- $\hat{\beta}$ estimation of β with all observations

$\hat{\beta}_{(j)}$ estimation of β with j'th observation missing

$\hat{\beta} - \hat{\beta}_{(j)}$ is the influence of the j'th observation on $\hat{\beta}$

This representation of the influence is shown to consist of the partial residual of Schoenfeld (1982) (discussed above) and a component which is the effect that an item has on the β coefficient via all the risk sets that the item is a member of.

Cain and Lange (1984) and Reid and Crepeau (1985) represent the estimated influence function graphically; Cain and Lange plotting the standardised influence against the rank of survival time, and Reid and Crepeau plot influence against covariate value. However, we plot the estimated change in the z-score for when each observation is omitted against the rank of the observation. This can be achieved if it is assumed that the variance-covariance matrix does not change fundamentally when one observation is omitted (this assumption is already made in the calculation of the estimated influence function). Now for each coefficient the estimated influence of the j'th observation ($I_{(j)}$) is

$$\hat{I}_{(j)} \simeq \hat{\beta} - \hat{\beta}_{(j)}$$

Dividing by the standard deviation of $\hat{\beta}$($St(\hat{\beta})$) obtained from the variance-covariance matrix.

$$\frac{\hat{I}_{(j)}}{St(\hat{\beta})} = \frac{\hat{\beta}}{St(\hat{\beta})} - \frac{\hat{\beta}_{(j)}}{St(\hat{\beta})}$$

$$\frac{\hat{\beta}_j}{St(\hat{\beta})} = \frac{\hat{\beta}}{St(\hat{\beta})} - \frac{\hat{\beta}_{(j)}}{St(\hat{\beta})}$$

$$z'_j = z - \frac{\hat{I}_{(j)}}{St(\hat{\beta})}$$

where:

z'_j is the estimated z-score when the j'th observation is ommitted from the analysis.

z is the z-score obtained when all the observations are analysed.

$I_{(j)}$ the calculated influence of the j'th observation.

To illustrate our plot we consider the route covariate from the British Railways High Speed Train brake disc analysis, see Bendell et al (1986) and Wightman (1987). With 3 significant covariates, 2 of which are two-level covariates and the other a three-level covariate, there are 12 possible combinations of covariate values. On the 'influence function' plot, Figure 8, each possible combination of covariate values is represented by a letter, failures by upper-case letters and censoring events by lower-case letters. Figure 8 shows that the removal of any observation does not change the high statistical significance of the coefficient since for all observations the estimated z-score exceeds 1.96.

The plot also shows that failures have in general a larger effect than censoring observations.

To illustrate the accuracy and the stability of the estimated influence (z'-score) we consider observation [A] on Figure 8; Table 1 summarises the effect of removing this observation.

Table 1. Effect of Removing Observation [A]

		Covariate		
		Route	Material/ Bolting Interaction	Braking System/ Bolting Interaction
z-score	Actual after Elimination	3.5617	4.5129	4.4819
	Estimated (z'-score)	3.5913	4.5139	4.4820
Standard Deviation	All Data	0.1944	0.2094	0.1034
	Less Observation [A]	0.1854	0.2095	0.1033

Comparing actual and estimated z-score values in Table 1 there is seen to be good agreement, particularly for covariates material/bolting position and braking system/bolting position. Also, the use of the standard deviation, based upon the whole data set, appears reasonable.

CONCLUSIONS

There is a need to avoid a black-box application of the PHM methodology, particularly in light of the unrefereed nature of much of the early literature. PHM is, however, a good tool for use in an exploratory manner.

A lot of experience is associated with identifying and modelling structure in the data set, and also in using and interpreting the diagnostics.

The major need is for the improvement of graphical diagnostics. Work is being undertaken at Trent to systematically develop these.

ACNOWLEDGEMENTS

Parts of this work were undertaken under a National Advisory Board grant for work on Hardware and Software Reliability, and as part of the Alvey Software Reliability Modelling project.

REFERERNCES

AITKIN, M. and CLAYTON, D. (1980). The fitting of exponential, Weibull and extreme value distributions to complex censored survival data using GLIM. Appl. Statist., 29, 156-163.

ANSELL, R.O. and ANSELL, J.I. (1986). Modelling the reliability of sodium sulphur cells. Proc. of 9th Advances in Reliability Technology Symposium, Bradford, B1/R/1-5.

ARGENT, S.J., BENDELL, A., MANNING, P.T., MARSHALL, J., RYAN, S.G. and WIGHTMAN, D.W., (1986). Proportional hazard modelling in the analysis of transmission failure statistics. Proc. of 5th Euredata Conference, Heidelberg.

ASCHER, H.E. (1983). "Regression Analysis of Repairable Systems Reliability", in Electronic Systems Effectiveness and Life Cycle Costing, J.K. Skwirzynski, Ed., Springer-Verlag, Berlin, 119-133.

BARLOW, W.E. and PRENTICE, R.L. (1988). Residuals for relative risk regression. Biometrika, 75, 65-74.

BELSLEY, D.A., KUH, E. and WELSCH, R.E. (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. Wiley, New York.

BENDELL, A., WALLEY, M., WIGHTMAN, D.W. and WOOD, L.M. (1986). Proportional hazards modelling in reliability analysis - an application of brake disks on high speed trains. Quality and Reliability Engineering Intl., 2, 45-52.

BOOKER, J., CAMPBELL, K., GOLDMAN, A.G., JOHNSON, M.E. and BRYSON, M.C. (1980). Application of Cox's proportional hazards model to light water reactors. Los Almos Scientific Laboratory, Report LA-8834-SR.

CAIN, K.C. and LANGE, N.T. (1984). Approximate case influence for the proportional hazards regression model with censored data. Biometrics, 40, 493-499.

COOK, R.D. (1986). Assessment of local influence. J.R. Statist. Soc., B, 48, 133-169.

COX, D.R. (1972). Regression models and life-tables (with discussion). J.R. Statist. SOC., b, 30, 248-275.

COX, D.R. (1979). A note on graphical analysis of survival data. Biometrika, 66, 188-190.

COX, D.R. and OAKES, D. (1984). Analysis of Survival Data. Chapman and Hall, London.

COX, D.R. and SNELL, E.J. (1968). A general definition of residuals (with discussion). J.R. Statis. Soc., B, 30, 248-275.

DALE, C. (1983). Application of the proportional hazards model in the reliability field. Proc. of the 4th National Reliability Conference, 5B/1/1-9.

DAVIES, N., MARRIOTT, J.M., WIGHTMAN, D.W. and BENDELL, A. (1987). The Musa data revisited: alternative methods and structure in software reliability modelling and analysis. SARSS '87, Elsevier, London.

DRURY, M., WALKER, E.V., WIGHTMAN, D.W. and BENDELL, A. (1987). Proportional hazards modelling in the analysis of computer systems reliability. Proc. Reliability '87, Birmingham. To appear; Reliability Engineering and System Safety Journal, (1988), 21.

GRAY, C., HARRIS, N., BENDELL, A. and WALKER, E. (1987). The reliability analysis of weapon systems. Proc. Reliability '87, Birmingham. To appear; Reliability Engineering and System Safety Journal (1988), 21.

HOLT, J.D. (1978). Competing risk analysis with special reference to matched pair experiments. Biometrika, 65, 159-165.

JARDINE, A.K.S and ANDERSON, M. (1984). Use of concomitant variables for estimations and setting component replacement policies. Proc. of the 8th Advances in Reliability Technology Symposium, B3/2/1-6.

KALBFLEISCH, J.D. and PRENTICE, R.L. (1980). The Statistical Analysis of Failure Time Data. John Wiley and Sons, Chichester.

KAY, R. (1977). Proportional hazards regression models and the analysis of censored survival data. Appl. Statist. 26, 227-237.

LINDQVIST, B., MOLNES, E. and RAUSAND, M. (1988). Analysis of SCSSV performance data. Reliability Engineering and System Safety. Vol. 20, No. 1, 3-18.

LAGAKOS, S.W. (1981). The graphical evaluation of explanatory variables in proportional hazards regression models. Biometrika, 68, 93-98.

LINK, C.L. (1984). Confidence intervals for the survival function using Cox's proportional hazards with covariates. Biometrics, 40, 601-610.

MANNING, P.T., RYAN, S.G., BAXTER, M.J. and BENDELL, A. (1987). Proportional hazards modelling of transmission equipment failures. Proc. Reliability '87, Birmingham. To appear; Reliability Engineering and System Safety Journal, (1988).

MANNING, P.T. (1988). The compatibility of reliability models with the underlying data structure. Proc. 10th Advances in Reliability Symposium, Elsevier, London.

MUSA, J.D. (1980). Software Reliability Data Submitted to the DCAS. Bell Telephone Laboratories.

PREGIBON, D. (1981). Logistic regression diagnostics. Am. Statist., 9, 705-724.

PRENTICE, R.L., WILLIAMS, B.J. and PETERSON, A.V. (1981). On the regression analysis of multivariate failure time data. Biometrika, 68, 373-379.

REID, N. and CREPEAU, H. (1985). Influence functions for proportional hazards regression. Biometrika, 67, 145-153.

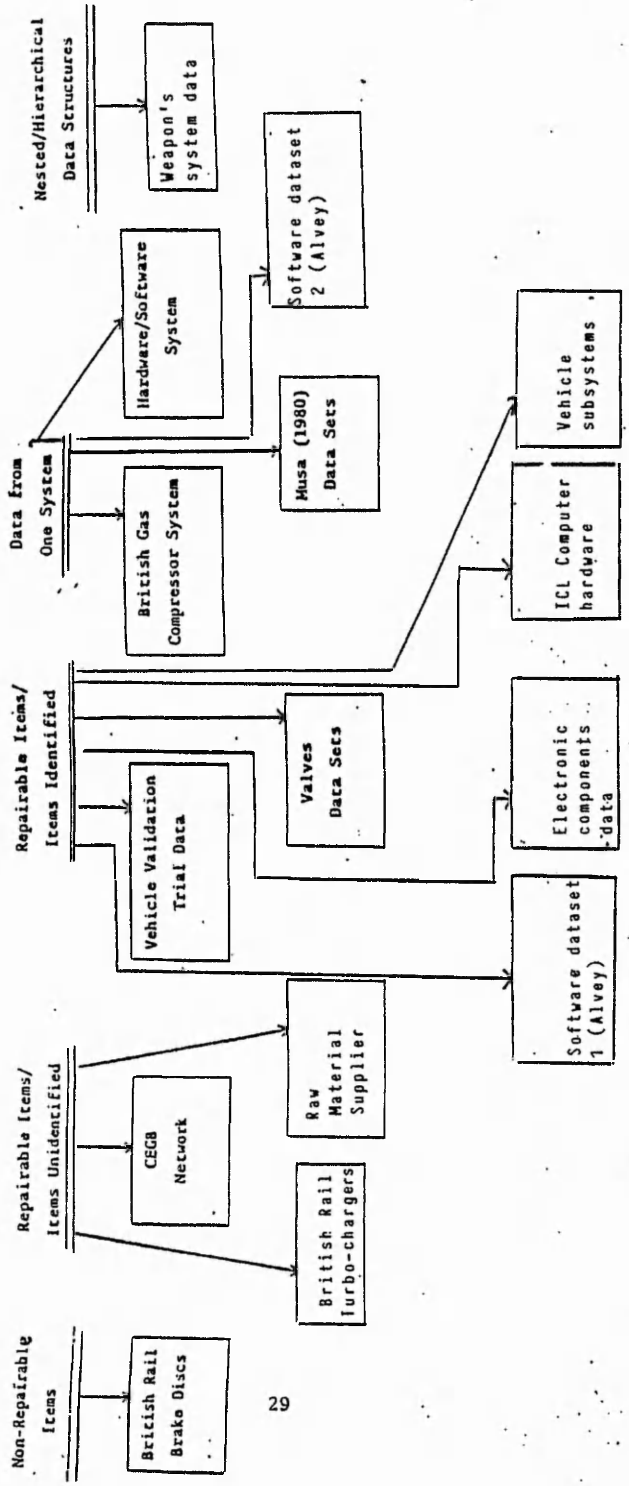
SCHOENFELD, D. (1982). Partial residuals for the proportional hazards regression model. Biometrika, 69, 239-241.

SELF, S.G., (1981). Proportional hazards model with time-dependent covariates: Large sample theory and related topics. PhD. Thesis, Department of Biostatist., Univ. of Washington.

STORER, B.E. and CROWLEY, J. (1985). A diagnostic for Cox regression and conditional likelihoods. J. Am. Stat. Assoc., Vol. 80, 389, 139-147.

WIGHTMAN, D.W. (1987). The application of proportional hazards modelling to reliability problems. PhD Thesis. Trent Polytechnic, Nottingham.

FIGURE . 1 Common Data Structures in Reliability and Data Sets we have Studied within these Structures



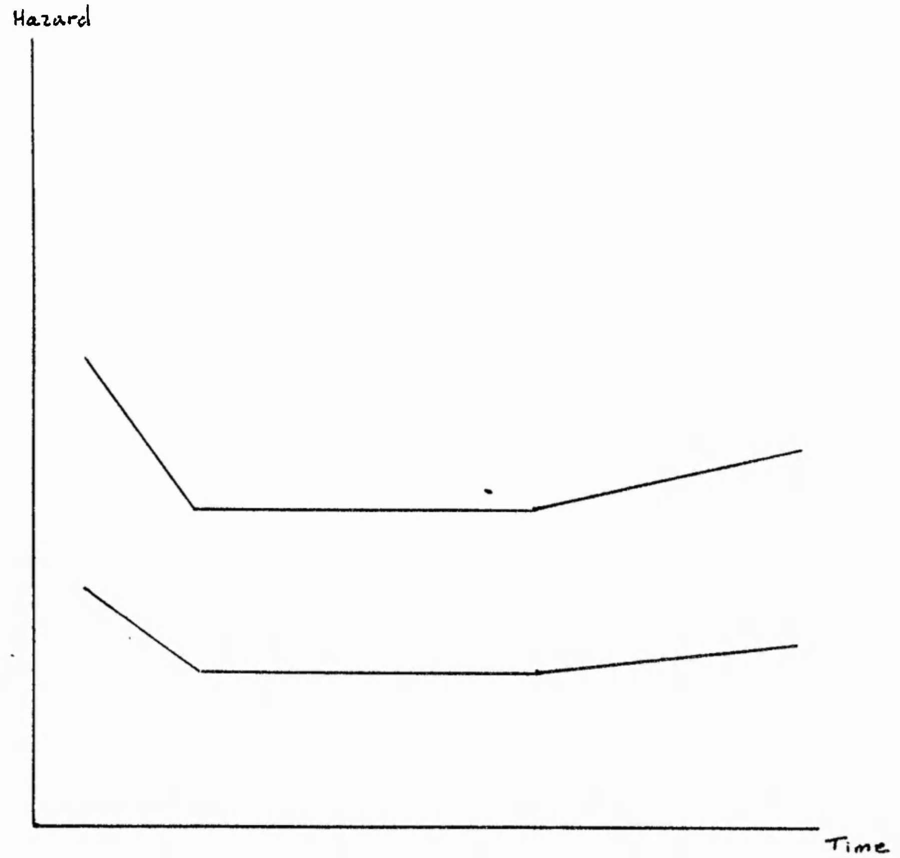


Figure 2. Hazard function for different covariate value

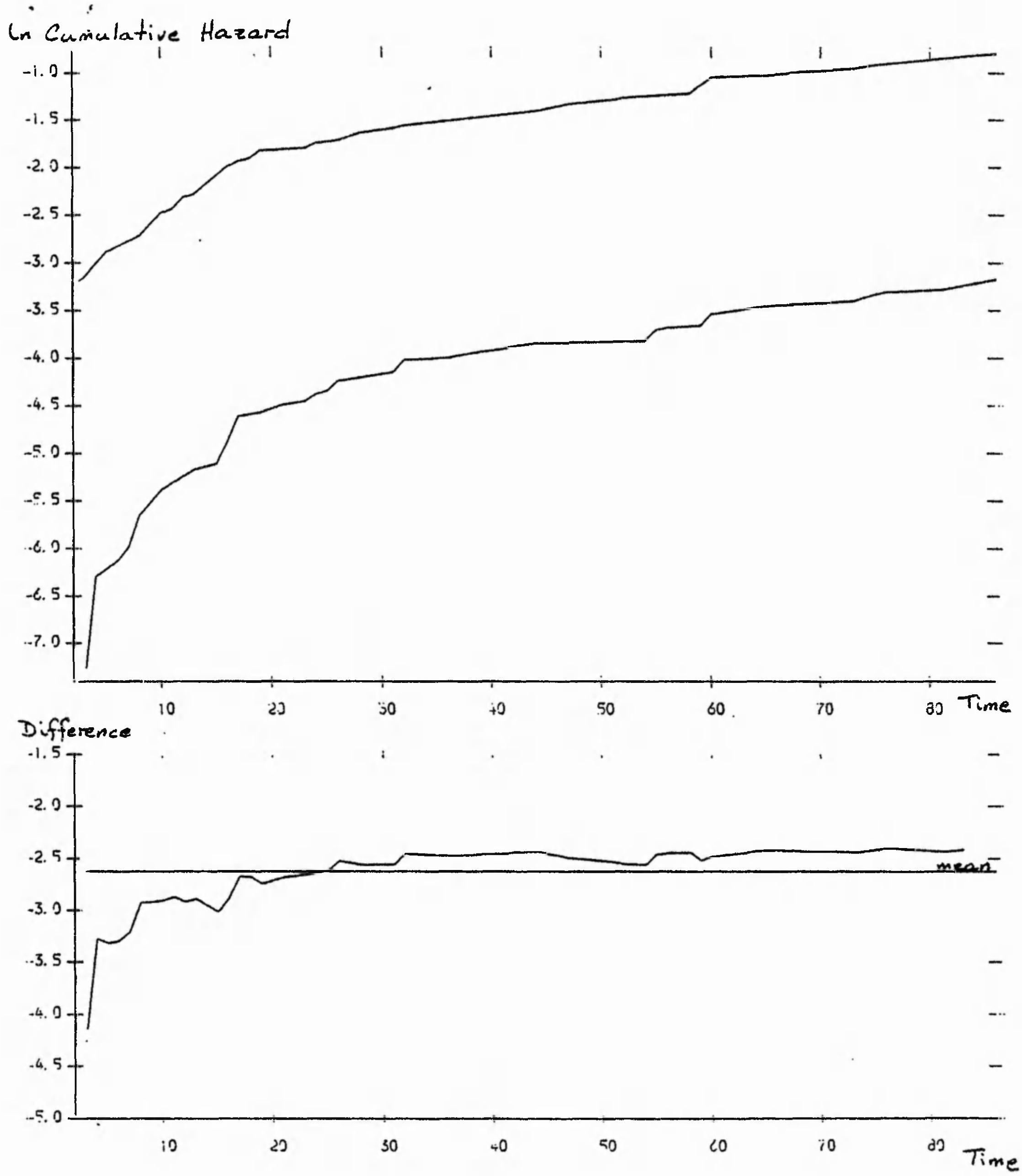


Figure 3. Proportionality plot

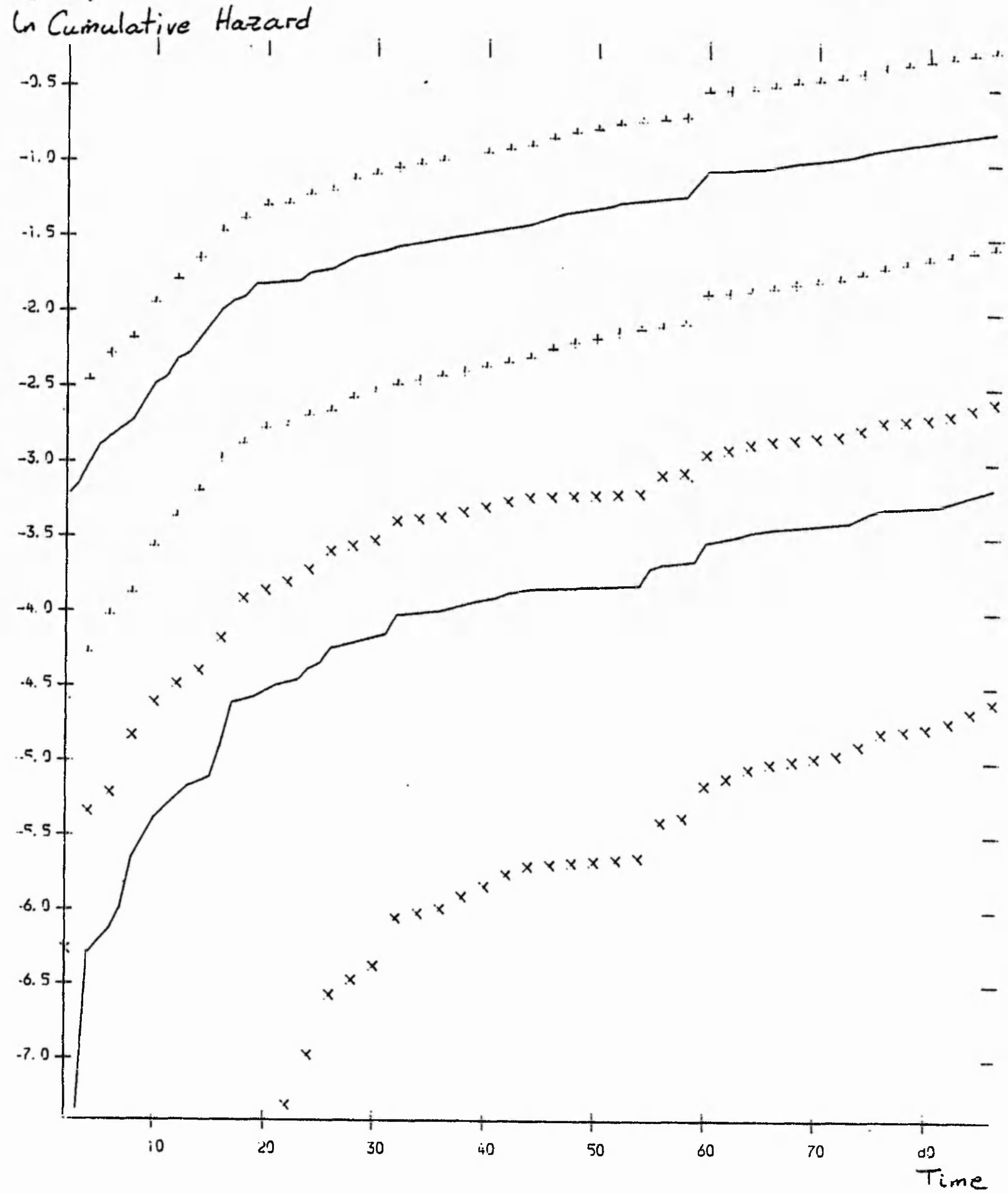


Figure 4. Proportionality plot

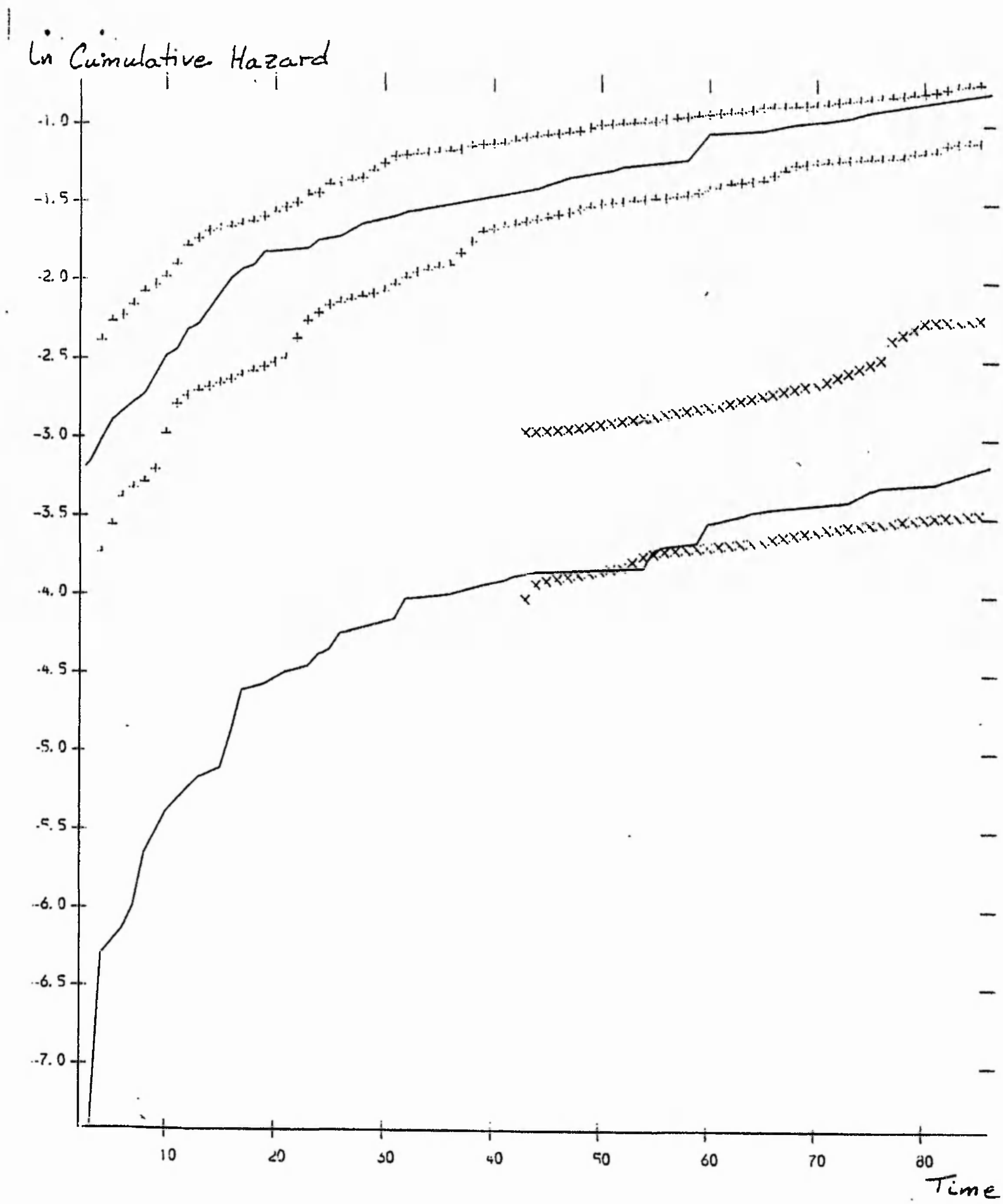


Figure 5. Proportionality plot

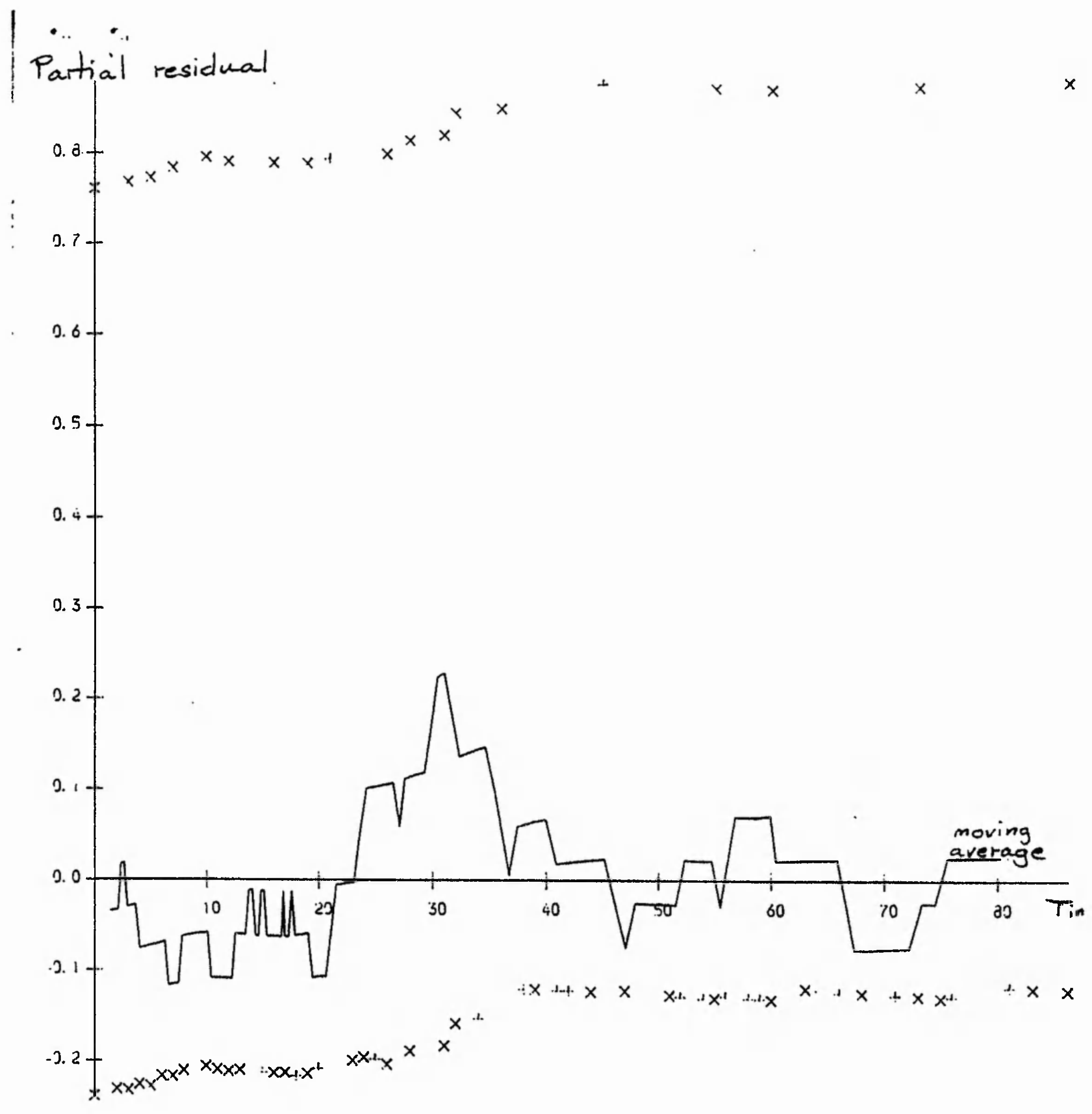


Figure 6 Schöenfeld residuals for '2 procs' covariate

FIGURE 7. Schoenfeld Residuals for Route Covariate

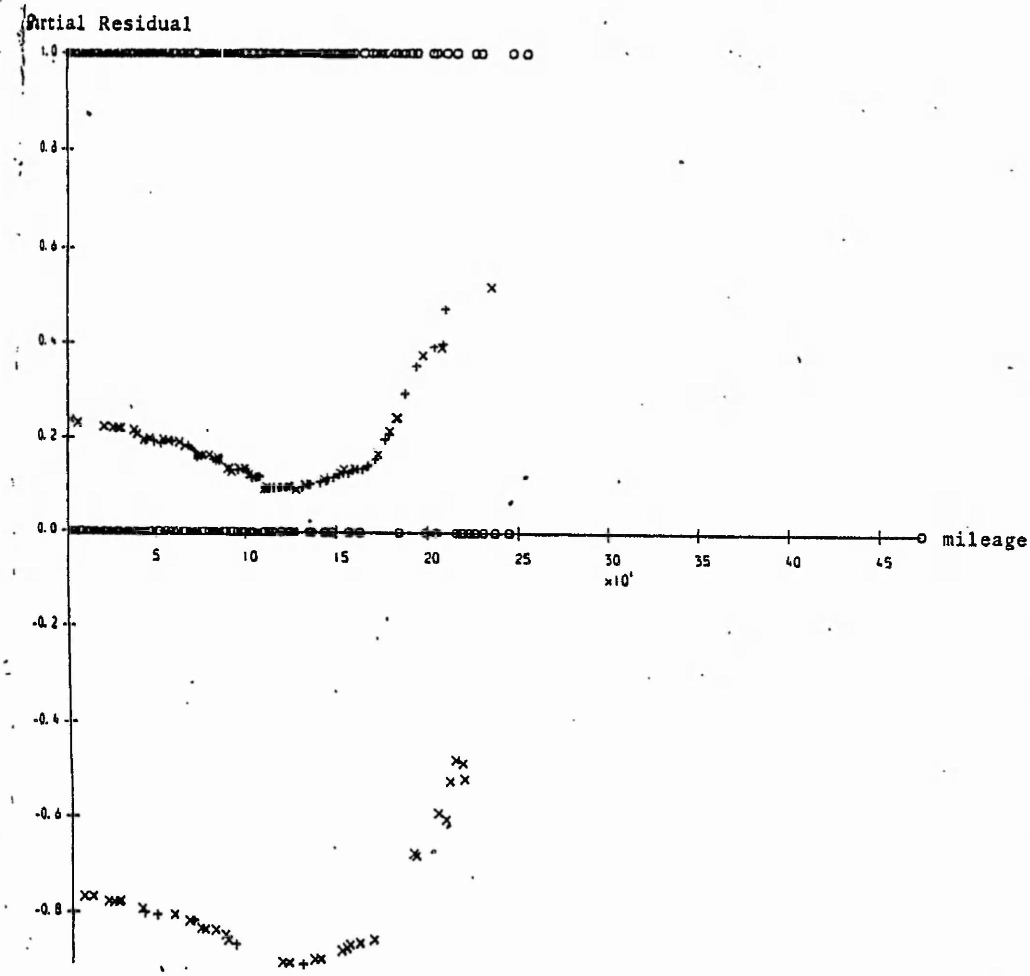


FIGURE 8 Estimated Influence on z-score for Route Observations