Chapter 3

Proportional Hazards Modelling

1 Introduction

In the preceding chapters, the problem that motivated this research was described, possible solutions to the problem were considered and the Proportional Hazards Model (PHM) was selected as the most logical route to the solution of the problem. In Chapter 3, the PHM is considered in detail.

Cox proposed the original PHM in 1972, initially intended for biomedical applications[2]. The model was immediately considered to be a revolution in life data analysis and it is still applied on a wide variety of survival data today. It only started to become popular amongst reliability modelers in the early 1980's, especially because of the model's ability to model the hazard rate without making assumptions about its functional form (if used in its semi-parametric form). The PHM have ever since the 1980's been applied in diverse reliability applications, for example, component failures in a light water reactor plant[61], marine gas turbine and ship sonar[8], motorrettes[7], aircraft engines[50], high speed train brake discs[51], sodium sulfur cells[52], surface controlled subsurface safety valves[53], machine tools[54], diesel engines[5], aircraft cargo doors[55], rolling mills[56,57], power transmission cables[58,59] and components of a mine loader[60].

It is impossible to include all the theory, which was developed over the years of the model's existence in this dissertation. For this reason, the discussion is limited to the Proportional Hazards theory required to apply the model in practical situations although some attention is given to the original PHM for the sake of completeness. An optimal renewal decision making technique developed specifically for use with the model is considered as well.

Numerical methods required to implement the model in real life situations are also described in a fair amount of detail.

2 The Proportional Hazards Model

Before introducing the PHM, the probabilistic hazard rate $h(t)$ as derived in Chapter 1 is repeated here for convenience as equation (2.1):

$$h(t) = \frac{f(t)}{R(t)} \tag{2.1}$$

The probabilistic hazard rate is a function of time only, a property which seriously limits the function's abilities in reliability modelling as discussed previously. Cox addressed this problem in the PHM by assuming that the hazard rate of a component can be determined by the product of an arbitrary and unspecified baseline hazard rate, $h_0(t)$ and a functional term $\lambda(\overline{z(t)})$. The baseline hazard rate is a function of time only and the functional term is a function of time and covariates (concomitant or explanatory variables). (If the covariates are independent of time, the functional term is only a function of the covariates, i.e. $\lambda(\bar{z})$)[1].

$$h(t, \bar{z}) = h_0(t) \cdot \lambda(\overline{z(t)}), \tag{2.2}$$

where $\bar{\gamma}$ is a regression vector estimated during model fitting procedures.

There are several possible forms for the functional term $\lambda(\overline{z(t)})$. Some are: the exponential form, $\exp(\bar{\gamma} \cdot \overline{z(t)})$; the logarithmic form, $\log(1 + \exp(\bar{\gamma} \cdot \overline{z(t)}))$; the inverse linear form, $1/(1 + \bar{\gamma} \cdot \overline{z(t)})$; or the linear form, $1 + \bar{\gamma} \cdot \overline{z(t)}$. The exponential form of the functional term has been used most often in reliability applications (and is the only form considered in this dissertation) and then equation (2.2) becomes:

$$h(t, \overline{z(t)}) = h_0(t) \cdot \exp(\bar{\gamma} \cdot \overline{z(t)}) \tag{2.3}$$

All theory on PHM presented in this dissertation will allow for time-dependent covariates, i.e. $\overline{z(t)}$, for the sake of generality.

This discussion deals with two forms of the PHM, namely (a) the semi-parametric PHM; and (b) the fully parametric PHM. When using the semi-parametric form, no assumption needs to be made about the shape of the baseline hazard rate when estimating the

[1] Although time, *t*, is used throughout as the unit of measure in this dissertation, any other suitable use parameter could be used instead, such as mileage or tons processed.

regression coefficients, although this only yields relative risks. To determine absolute risks, the baseline hazard rate has to be estimated first. This feature is considered to be a huge advantage even though some numerical difficulties are often encountered. The fully parameterized model makes use of a continuous distribution, most often the Weibull distribution because of its flexibility, for the baseline hazard rate, which makes it much more numerically tractable.

2.1 Assumptions of the PHM

The assumptions on which the PHM is postulated are best illustrated for the model with time-independent covariates:

a) Renewal times are independent and identically distributed.
b) All influential covariates are included in the model.
c) The ratio of any two hazard rates as determined by any two sets of covariates $\overline{z_1}$ and $\overline{z_2}$ associated with a particular component has to be constant with respect to time, i.e. $h(t,\overline{z_1}) \propto h(t,\overline{z_2})$. (This assumption implies that the covariates acts multiplicatively on the hazard rate of the component).

Assumption (c) is illustrated graphically in Figure 2.1 below:

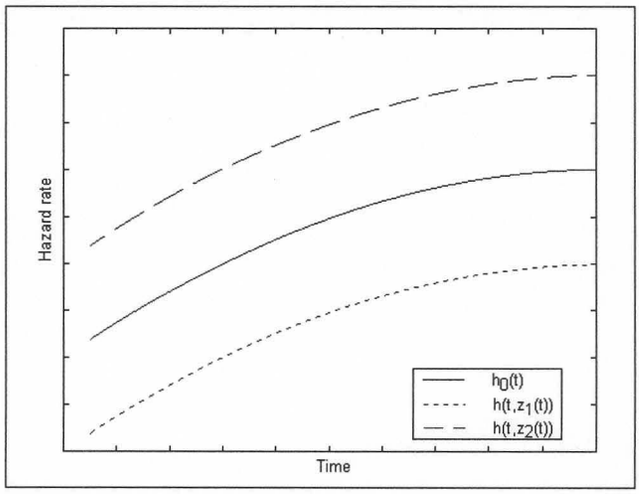

Figure 2.1: *Graphical illustration of PHM assumption (c)*

Figure 2.1 shows that the two hazard rates, $h(t,\overline{z_1})$ and $h(t,\overline{z_2})$, calculated from different covariate vectors associated with a specific component are proportional to each other with respect to time.

2.2 The Semi-Parametric PHM

The semi-parametric PHM has a very valuable attribute in that no assumption needs to be made about the baseline hazard rate of the model when estimating the regression coefficients. This means that relative risks of the item under consideration can be estimated without any knowledge about the time dependent failure behavior of the item, as is contained in the baseline hazard rate. The reason for the existence of this advantageous property becomes clear in the explanation of the estimation technique proposed by Cox with the proposal of the PHM, called the *method of partial likelihood*.

As an introduction to the method of partial likelihood, the *method of maximum likelihood* is discussed in general terms.

2.2.1 The Method of Maximum Likelihood

The method of maximum likelihood is a well known method for, amongst other uses, estimating regression coefficients and is used widely in the literature. It is important to describe this method before introducing partial likelihood since these methods are closely related and the method of maximum likelihood is used to determine the fully parametric PHM coefficients.

Likelihood refers to the hypothetical probability that an event, which has already occurred, would yield a specific outcome. This concept differs from that of probability in that probability refers to the occurrence of future events, while a likelihood refers to past events with known outcomes.

The method of maximum likelihood starts off by specifying the joint probability distribution function for events, in this case, failures. (For the moment the possibility of suspensions is omitted). We assume a sample of size n drawn from some probability distribution. Let $\overline{\boldsymbol{T}} = [\boldsymbol{T}_1, \boldsymbol{T}_2, ..., \boldsymbol{T}_n]$ denote failure times as random variables. The probability that $\boldsymbol{T}_i = T_i$ can be expressed informally as the probability density at T_i, i.e. $P\{\boldsymbol{T}_i = T_i\} = f(\boldsymbol{T}_i, \overline{\gamma}, \overline{z_i(t)})$. By using the law of total probability, the joint probability of all the random variables is given by:

$$P\{\boldsymbol{T}_1 = T_1, \boldsymbol{T}_2 = T_2, ..., \boldsymbol{T}_n = T_n\} = \prod_{i=1}^{n} f(\boldsymbol{T}_i, \overline{\gamma}, \overline{z_i(t)}) \qquad (2.4.)$$

Once the random variables have been observed, (2.4.) can be used to calculate

the likelihood of events with all observed values fixed and $\bar{\gamma}$ the only unknown as shown in (2.5.) below:

$$L(\bar{\gamma}) = \prod_{i=1}^{n} f(T_i, \bar{\gamma}, \overline{z_i(t)}) \tag{2.5.}$$

Equation (2.5.) can be extended to include suspended observations (see reference [3]):

$$L(\bar{\gamma}) = \prod_{i=1}^{n} f(T_i, \bar{\gamma}, \overline{z_i(t)})^{c_i} \cdot R(T_i, \bar{\gamma}, \overline{z_i(t)})^{1-c_i} , \tag{2.6.}$$

for a data set with n renewals each at time T_i with $c_i = 1$ in case of failure and $c_i = 0$ for suspended observations. The value of $\bar{\gamma}$ that maximizes L is the most appropriate regression vector for the model since it maximizes the probability of occurrence of the observed data set. Numerical methods suitable to estimate the regression vector are discussed later in Chapter 3.

Maximum likelihood estimation is suitable for estimating both the semi-parametric and the parametric PHM although numerical difficulties are often experienced with this technique and the semi-parametric model.

2.2.2 Method of Partial Likelihood

In his original paper[2], Cox suggested parameter estimation for the semi-parametric PHM by maximizing an expression which he called *conditional likelihood*. This term gave rise to much discussion in the literature where the validity of referring to the term as conditional likelihood was argued by critics. For the estimation technique of Cox to be a conditional likelihood function it had to be a likelihood function based on the conditional distribution of data, given some statistic, which it was certainly not[62]. It was neither a marginal likelihood function because then it had to be based on the marginal distribution of some reduction of the data. In reply[63] Cox showed (somewhat informally[3]) that his technique was accurate and consistent with considerably less numerical difficulties. He also eliminated all confusion in terminology by renaming the technique to *partial likelihood* because of the fact that it was not a likelihood function in the usual sense.

Suppose a vector of random variables denoting failure times (for the moment suspensions are not considered) is observed, $\bar{T} = [T_1, T_2, ..., T_n]$, which comes

from a probability density $f(\overline{\boldsymbol{T}}, \overline{z_i(t)}, \overline{\gamma}, h_0(t))$. The baseline hazard function $h_0(t)$ is considered to be a nuisance function in this case. If a one on one transformation on the data in $\overline{\boldsymbol{T_i}}$ is performed with auxiliary variables $\boldsymbol{A_1}, \boldsymbol{B_1}, ..., \boldsymbol{A_n}, \boldsymbol{B_n}$ such that $\boldsymbol{A^{(k)}} = [\boldsymbol{A_1}, ..., \boldsymbol{A_k}]$ and $\boldsymbol{B^{(k)}} = [\boldsymbol{B_1}, ..., \boldsymbol{B_k}]$ the likelihood of $\boldsymbol{A^{(k)}}$ and $\boldsymbol{B^{(k)}}$ is:

$$\prod_{k=1}^{m} f(b_k \mid b^{(k-1)}, a^{(k-1)}, \overline{\gamma}, h_0(t)) \cdot \prod_{k=1}^{m} f(a_k \mid b^{(k)}, a^{(k-1)}, \overline{\gamma}) \qquad (2.7.)$$

Cox[63] defined the second product on the right hand side of (2.7.) as the partial likelihood function since only a part of the joint probability density function is considered and the nuisance function $h_0(t)$ is eliminated. The mathematical proof of the validity of partial likelihood is discussed in reference [63]. See also reference [64] for a thorough discussion on the topic of partial likelihood.

For the PHM, the partial likelihood can be constructed as follows: as before, suppose a certain number of similar items have been renewed on k occasions of which n were failures at $T_1, T_2, ..., T_n$ with corresponding covariate vectors $\overline{z_1(t)}, \overline{z_2(t)}, ..., \overline{z_n(t)}$. Define the order statistic to be $O(t) = [T_{(1)}, T_{(2)}, ..., T_{(n)}]$ and the rank statistic to be $\overline{r(t)} = [(1),(2),....,(n)]$. The order statistic refers to the $T_{(i)}$'s ordered from smallest to largest and the notation (i) in the rank statistic refers to the label attached to the i^{th} order statistic. Consider a set $R(t_{(i)})$ of items at risk at time $T_{(i)}^{-}$. The partial probability that item (i) fails at $T_{(i)}$ given that the items $R(t_{(i)})$ are at risk and that exactly one failure occurs at $T_{(i)}$ is:

$$\frac{h(T_{(i)}, \overline{z_{(i)}(t)})}{\sum_{l \in R(T_{(i)})} h(T_{(i)}, \overline{z_l(t)})} = \frac{\exp(\overline{\gamma} \cdot \overline{z_{(i)}(t)})}{\sum_{l \in R(T_{(i)})} \exp(\overline{\gamma} \cdot \overline{z_l(t)})}, \qquad (2.8.)$$

where $i = 1,2,....,k$. Equation (2.8.) shows that the baseline hazard rate has no effect on the joint probability and hence no effect on the estimated values of $\overline{\gamma}$. The partial likelihood can now be calculated by taking the product over all the failure points:

$$PL(\overline{\gamma}) = \prod_{i=1}^{k} \frac{\exp(\overline{\gamma} \cdot \overline{z_{(i)}(t)})}{\sum_{l \in R(T_{(i)})} \exp(\overline{\gamma} \cdot \overline{z_l(t)})} \qquad (2.9.)$$

To account for the possibility of ties at a specific $T_{(i)}$, i.e. the occurrence of more than one failure at a specific $T_{(i)}$, Breslow[65] has derived the following

approximation of the partial likelihood (for numerical tractability):

$$PL(\bar{\gamma}) = \prod_{i=1}^{k} \frac{\exp(\bar{\gamma} \cdot \overline{s_i(t)})}{\left[\sum_{l \in R(T_{(i)})} \exp(\bar{\gamma} \cdot \overline{z_l(t)})\right]^{d_i}}, \qquad (2.10.)$$

where $s_i = \sum_j \overline{z_{ij}(t)}$ is the sum of the covariates of the d_i items observed to fail exactly at $T_{(i)}$.

The value of $\bar{\gamma}$ that maximizes $PL(\bar{\gamma})$ is the most appropriate for the semi-parametric PHM. At this optimal point, the partial derivatives of $PL(\bar{\gamma})$ with respect to all the *m* measured covariates should be zero, i.e.:

$$U_j(\bar{\gamma}) = \frac{\partial \log PL(\bar{\gamma})}{\partial \gamma_j} = \sum_{i=1}^{k} [s_{ji} - d_i A_{ji}(\bar{\gamma})] = 0 \quad (j = 1,2,\ldots,m), \qquad (2.11.)$$

where s_{ji} is the j^{th} element in the vector s_i and

$$A_{ji}(\bar{\gamma}) = \frac{\sum_{l \in R(T_{(i)})} z_{jl} \exp(\bar{\gamma} \cdot \overline{z_l(t)})}{\sum_{l \in R(T_{(i)})} \exp(\bar{\gamma} \cdot \overline{z_l(t)})} \qquad (2.12.)$$

A suitable optimization technique is required to perform the maximization such as Newton-Raphson iteration. This technique and others are described later in Chapter 3.

2.2.3 Efficiency of Partial Likelihood Estimation

After the introduction of partial likelihood, its efficiency was measured and compared by many researchers, for example Kalbsfleich and Prentice[3], and Effron[68].

Kalbsfleich and Prentice investigated the efficiency of the partial likelihood with the following two questions in mind: (1) Can the estimated $\bar{\gamma}$ be improved at all for the case where $h_0(t)$ is unspecified?; and (2) How does the partial likelihood estimate of $\bar{\gamma}$ compare to a maximum likelihood estimate? Two separate investigations were done, one for time-independent covariates and one for time-dependent covariates.

It was found for time-independent covariates that the partial likelihood

estimation was 'reasonably' efficient. The details concerning the investigation are not discussed, see reference [3] pp. 103-113 for a comprehensive explanation. Cooper and Darch[69] agree with this statement after doing a study on armoured vehicles. For time-dependent covariates it is difficult to predict the partial likelihood's efficiency and it could in some cases even be very low. See reference [3], pp. 140-141.

2.2.4 Estimation of the Baseline Hazard Rate

The baseline hazard rate function represents the hazard rate that an item would experience if covariates had no effect on the item. No assumption needs to be made about its functional form for the semi-parametric PHM provided that the regression vector $\bar{\gamma}$ is known. Several authors have developed techniques with which the baseline hazard rate can be estimated, including Cox[2], Kalbsfleish and Prentice[3], Breslow[65,66] and Link[67]. The method of Breslow is presented here.

Suppose that $h_0(t)$ is a step function which jumps just before the occurrence of a failure and is constant between times to failure, i.e.:

$$h_0(t) = h_{0_i}, \quad T_{(i-1)} < t \leq T_{(i)}, \quad i = 1,2,....,n \tag{2.13.}$$

With $h_0(t)$ defined as in (2.13.), an expression for the joint distribution can be derived and reduced to Cox's partial likelihood, exactly as explained in 2.2.2 which results in:

$$h_{0_i} = \frac{d_i}{(T_{(i)} - T_{(i-1)}) \sum_{l \in R(T_{(i)})} \exp(\bar{\gamma} \cdot \overline{z_l(t)})}, \tag{2.14.}$$

with h_{0_i} completely distribution-free. The distribution-free baseline hazard rate is often used to check the appropriateness of continuous distributions used for the baseline hazard rate.

2.3 The Fully Parametric PHM

By assuming a continuous distribution for the form of the baseline hazard rate, the PHM is completely parameterized. The very versatile Weibull distribution (and the only one considered in this research project) is most often used for the parameterization because of its flexibility. It is impossible to estimate the baseline

hazard rate independently for the fully parametric PHM[1] and the distribution- and regression parameters have to be estimated simultaneously. Considerably less numerical problems arise for the Weibull PHM and its is much more numerically convenient.

2.3.1 Statistical Model

The time-dependent Weibull distribution is given by:

$$f(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \cdot \exp(-(t/\eta)^{\beta}), \tag{2.15.}$$

with its corresponding hazard rate function:

$$h(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}, \tag{2.16.}$$

where β and η are shape and scale parameters of the distribution, respectively.

If the Weibull distribution is used as the baseline hazard rate of the PHM as presented in (2.3.), the model becomes:

$$h(t, \overline{z(t)}) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \cdot \exp(\overline{\gamma} \cdot \overline{z(t)}), \tag{2.17.}$$

which is a fully parametric model.

From reliability theory we know that the reliability, R, of a component under the influence of ageing only, just before renewal at T_i is:

$$R(T_i) = \exp\left(-\int_0^{T_i} h(t)\mathrm{d}t\right) = \exp(-(T_i/\eta)^{\beta}) \tag{2.18.}$$

If $U_i = (T_i/\eta)^{\beta}$, then U_i has a unit negative exponential distribution. Similar to (2.18.), the reliability at a time T_i for a component under the influence of time-independent covariates according to the PHM can be estimated by:

[1] Also referred to as the Weibull Proportional Hazards Model in this dissertation.

$$R(t,\bar{z}) = \exp\left[-\int_0^{T_i} \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \mathrm{d}t \exp(\bar{\gamma}\cdot\bar{z})\right], \qquad (2.19.)$$
$$= \exp\left[-(T_i/\eta)^{\beta}\exp(\bar{\gamma}\cdot\bar{z_i})\right]$$

with $U_i = (T_i/\eta)^{\beta}\exp(\bar{\gamma}\cdot\bar{z_i})$, again with unit exponential distribution. For the case of time-dependent covariates, the covariates have to be included in the integration to estimate the reliability of a component at time T_i:

$$R(t,\overline{z(t)}) = \exp\left[-\int_0^{T_i} \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \exp(\bar{\gamma}\cdot\overline{z_i(t)})\mathrm{d}t\right], \qquad (2.20.)$$
$$= \exp\left[-\int_0^{T_i} \exp(\bar{\gamma}\cdot\overline{z_i(t)})\mathrm{d}(t/\eta)^{\beta}\right]$$

with $U_i = \int_0^{T_i} \exp(\bar{\gamma}\cdot\overline{z_i(t)})\mathrm{d}(t/\eta)^{\beta}$, also with unit negative exponential distribution. In practice, (2.20.) is often approximated by:

$$R(t,\overline{z_i(t)}) = \exp\left\{\sum_{k=1}^{i}\exp(\bar{\gamma}\cdot\overline{z_i^*(t_k)})\cdot\left[\left(\frac{t_{k+1}}{\eta}\right)^{\beta} - \left(\frac{t_k}{\eta}\right)^{\beta}\right]\right\}, \qquad (2.21.)$$

where $0 = t_0 < t_1 < .... < T_i$ are inspection points where covariate measurements were taken and $\overline{z_i^*(t_k)} = 0.5\cdot(\overline{z_i(t_k)} + \overline{z_i(t_{k+1})})$.

2.3.2 Parameter Estimation

The method of maximum likelihood is used to estimate the model's parameters. The full likelihood is obtained by:

$$L(\beta,\eta,\bar{\gamma}) = \prod_i h(T_i,\overline{z_i(T_i)})\cdot\prod_j R(T_j,\overline{z_j(t)}) \qquad (2.22.)$$

where i indexes failure times and $j=1,2,...,n$ indexes failure and suspension times. If (2.17.) and (2.20) are substituted in (2.22.), the full likelihood becomes:

$$L(\beta,\eta,\bar{\gamma}) = \prod_i \frac{\beta}{\eta}\left(\frac{T_i}{\eta}\right)^{\beta-1} \exp\left(\bar{\gamma}\cdot\overline{z_i(T_i)}\right)\cdot \prod_j \exp\left[-\int_0^{T_j}\exp(\bar{\gamma}\cdot\overline{z_j(t)})\mathrm{d}(t/\eta)^{\beta}\right] \quad (2.23.)$$

The same values for β, η and $\bar{\gamma}$ that maximize (2.23.) will also maximize $\log(L(\beta,\eta,\bar{\gamma}))$ or $l(\beta,\eta,\bar{\gamma})$, the log-likelihood. It is numerically much more attractive to maximize $l(\beta,\eta,\bar{\gamma})$ given by:

$$l(\beta,\eta,\bar{\gamma}) = r\ln(\beta/\eta) + \sum_i \ln[(T_i/\eta)^{\beta-1}] + \sum_i \bar{\gamma}\cdot\overline{z_i(T_i)} - \sum_j \int_0^{T_j}\exp(\bar{\gamma}\cdot\overline{z_j(t)})\mathrm{d}(t/\eta)^{\beta}, \quad (2.24.)$$

with r being the number of failure renewals.

Several maximization techniques were tested on (2.24.) with success. This includes (a) a Nelder-Mead type simplex search method as is commonly found in the literature, (b) a BFGS Quasi-Newton method with a mixed quadratic and cubic line search procedure, (c) Snyman's dynamic trajectory optimization method[20,21] and (d) a modified Newton-Raphson procedure that gives fast convergence.

3 Covariates

The ability of the PHM to include covariates in its estimations and predictions is a very attractive attribute of the model. Covariate behavior and their effects on the PHM are not trivial issues however and a proper background knowledge of covariates is required before the model can be used with confidence. Some comments on these issues are presented in this section.

3.1 Effects of Interaction and Omission of Covariates

Interaction (dependency) of covariates can influence parameter estimation significantly and therefore the presence of this phenomenon should be checked. The easiest, most practical way to test for interaction of covariates is by introducing a new temporary covariate. (This is only to get a feel for the behavior of the data and is not very scientific). The temporary covariate is simply the product of the

covariates under discussion and the new covariate's effect on the model is then tested. (Usually the interaction of only two covariates will be checked at a time). If interaction is present, the new covariate will be statistically significant in the model. The results of statistical tests for interaction, i.e. testing for significance of the new covariate in the model, should also be practically justified if possible, to identify possible inconsistencies in the data.

Bendell *et al.*[51] investigated covariate interaction by dividing data into groups (strata) based on major differences in the data and then estimating the regression coefficients for each group. It was shown that a suitable test statistic can be defined to test the effect of a particular covariate on different groups. See also reference [70] for a discussion on 'stratum-covariate' interactions.

Omission of influential covariates from the PHM also effects the regression coefficients of the model. Suppose z_1 and z_2 are two significant covariates with corresponding regression coefficients γ_1 and γ_2. If only z_1 is considered in the model and its coefficient is γ_1, then $|\gamma_1| < |\gamma_2|$. The estimates for γ_1 are asymptotically biased towards γ_2 and have smaller asymptotic variance than γ_2 [71]. The magnitude of the bias depends on the relative importance of the omitted covariate to that of the included covariate. Omission of influential covariates could also lead to overestimation or underestimation of the baseline hazard rate[72].

3.2 Effects of Measurement Error and Misspecification of Covariates

Significant errors in the estimation of the regression vector $\bar{\gamma}$ are possible if an inappropriate parametric form of the baseline hazard rate is specified or if errors in covariate measurement are distributed in an dismal manner. The regression vector is influenced in the same manner as in the of case of linear regression when measurement errors in the covariates are present. It is possible to test the significance of the effects of a covariate, in spite of its measurement error and misclassification, provided that sufficient information or assumptions are available relating to the covariate error or misclassification distribution[73]. Lagakos[74] determined that the efficiency of the partial likelihood estimator may be very low compared to the correct model if covariates are misspecified.

3.3 Effects of Monotonicity and Multicolinearity of Covariates

The 'usual' difficulties of regression analysis like multicolinearity, monotonicity and large covariate values are often encountered in the PHM as well. In such cases, maximization procedures used for parameter estimation often fail to converge.

Monotone increasing or decreasing covariate values in a data set, when ordered according to the magnitude of times to failure, is the biggest cause for divergence of maximization procedures. Bryson and Johnson[75] suggest seven steps to avoid the problems associated with monotonicity of covariates. In a censored data set, it may occur that the covariate at each failure time is either the largest or the smallest of all covariates in the risk set at that time. In such cases the regression vector estimate is often infinite.

It is important to formulate the covariates in such a manner that colinearity is avoided during the estimation of the regression vector since results may be very inaccurate. The data set could be analyzed in different groups based on a trail-and-error method to address this problem. Peduzzi *et al.*[76] published a general procedure for the selection of covariates in a nonlinear regression analysis to avoid colinearity. This is useful when there is a large number of covariates and it is difficult to determine the priority of selection of covariates for the model to their confusing effects on the times to renewal[59].

3.4 Time-dependent Covariates

Numerous papers have been published on the theory of time-dependent covariates in the PHM. This include topics like efficiency of estimation techniques for regression coefficients of time-dependent covariates[3], a two-step PHM model to accommodate time-dependent covariates more accurately[77], techniques to detect time-dependent effects of fixed covariates[78] and graphical techniques based on partial residuals suitable to detect time varying effects of covariates[79]. The detailed theory and analysis of time-dependent covariates are not important for this dissertation but rather its practical use and therefore this discussion is limited to some practical calculation issues.

Covariates in the Weibull PHM in (2.20.) are allowed to be time-dependent and assumed to be known for all values of time, t. This assumption is not entirely valid for the case of vibration covariates because vibration inspections are normally done on a discrete periodic basis. It is thus necessary to estimate covariate values between inspections for the model in (2.20.) to hold. Experience has shown that

there is no fixed rule for this estimation (especially not for vibration covariates) and every situation should be considered separately. In (2.21.) an estimation technique was presented where the covariate values between any two inspections were taken to be the average of the covariate values between the two particular inspections. Jardine *et al.*[4] describe a similar method that was used with success on aircraft and marine engine failure data.

In some situations where covariates have a monotonic behavior, conventional interpolation techniques can be used with confidence such as linear, hyperbolic, parabolic, geometric or exponential interpolation, depending on the particular situation. It can be much more difficult to predict the values of other, non-monotonic, situations between inspections. In these cases the most sensible option is usually to consider the covariate behavior as a continuous right jumps process, where covariate values only increase or decrease at inspections and remain constant between inspections.

4 Numerical Model Fitting Procedures

Four optimization techniques were implemented successfully to fit the Weibull PHM with the method of maximum likelihood, i.e. converged to the point where all the objective function's partial derivatives were zero, namely:

i. A Nelder-Mead type simplex search method.
ii. A Standard BFGS Quasi-Newton method with a mixed quadratic and cubic line search procedure.
iii. Snyman's dynamic trajectory optimization method[20,21]
iv. A modified Newton-Raphson procedure.

The performance of each one of the methods was measured according to their economy (number of iterations needed before convergence, number of objective function evaluations and number of partial derivative evaluations) and robustness (the accuracy of initial values required for convergence and its ability to handle steep valleys and discontinuities in the objective function). Methods (1) and (2) maximized the likelihood function successfully but performed fairly mediocre. Snyman's method was found to be somewhat expensive but extremely robust which is a very valuable attribute. The modified Newton-Raphson method proved to be by far the most economical and fairly robust as well. Of the four above mentioned techniques, this technique is certainly the most suitable for optimization of the maximum likelihood function.

For the above mentioned reasons, only Snyman's method and the modified Newton-Raphson method are considered in this discussion on numerical model fitting

procedures. Snyman's method is presented in fairly general terms to illustrate its robustness, but the Newton-Raphson method is described in detail since this method is used in the case study in Chapter 4.

4.1 Snyman's Dynamic Trajectory Optimization Method

Snyman's method models a conservative force field in m-dimensions (the number of variables in the objective function) with the objective function and then monitors the trajectory of a particle of unit mass (released from rest) as it 'rolls' down the objective function to the point of least potential energy, which is the minimum of the objective function.

In this general presentation of Snyman's technique, the objective function is $l(\bar{\theta})$, the maximum log-likelihood function as presented in (2.24.), where $\bar{\theta} = [\beta, \eta, \bar{\gamma}]$.

4.1.1 Characteristics

The attributes of Snyman's technique can be summarized as follows:

i. It uses only gradient information, i.e. $\bar{\nabla}[l(\bar{\theta})]$.
ii. No explicit line searches are performed.
iii. It is extremely robust and handles steep valleys and discontinuities in the objective function or gradient with ease.
iv. This algorithm seeks low local minimum and it can be used as a basic component in a methodology for global optimization.
v. The method is not as efficient on smooth and near quadratic functions as classical methods.

4.1.2 Basic Dynamic Model

Assume a particle of unit mass in a m-dimensional conservative force field with potential energy at $\bar{\theta}$ given by $l(\bar{\theta})$, then the force experienced by the particle at $\bar{\theta}$ is given by:

$$m\bar{a} = \bar{\ddot{\theta}} = -\bar{\nabla}[l(\bar{\theta})], \tag{4.1.}$$

from which it follows that for the time interval $[0, t]$:

$$\frac{1}{2}\left\|\dot{\bar{\theta}}(t)\right\|^2 - \frac{1}{2}\left\|\dot{\bar{\theta}}(0)\right\|^2 = l(\bar{\theta}(0)) - l(\bar{\theta}(t)) \qquad (4.2.)$$

Equation (4.2.) can be simplified by expressing it in terms of kinetic energy as:

$$T(t) - T(0) = l(0) - l(t) \qquad (4.3.)$$

From (4.3.) it evident that $l(t) + T(t) = \text{constant}$, which indicates conservation of energy in the conservative force field. It should also be noted that $\Delta l = -\Delta T$, therefore as long as T increases, l decreases, which is the basis of the dynamic algorithm.

4.1.3 Basic Algorithm

Suppose $l(\bar{\theta})$ has to be minimized from a starting point $\bar{\theta}(0) = \bar{\theta}_0$, then the dynamic algorithm is as follows:

i. Compute the dynamic trajectory by solving the initial value problem, $\ddot{\bar{\theta}}(t) = -\bar{\nabla}[l(\bar{\theta}(t))]$, $\dot{\bar{\theta}}(0) = 0$ and $\bar{\theta}(0) = \theta_0$. In practice the numerical integration of the initial value problem is often done by the 'leap-frog' method. Compute for $k = 0,1,2,....$ and time step Δt, the following: $\bar{\theta}^{k+1} = \bar{\theta}^k + \dot{\bar{\theta}}^k \Delta t$ and $\dot{\bar{\theta}}^{k+1} = \dot{\bar{\theta}}^k + \ddot{\bar{\theta}}^k \Delta t$, where $\ddot{\bar{\theta}}^k = -\bar{\nabla}[l(\bar{\theta}^k)]$ and $\dot{\bar{\theta}}_0 = (1/2)\ddot{\bar{\theta}}_0 \Delta t$.

ii. Monitor $\dot{\bar{\theta}}(t)$, the velocity of the particle. As long as the kinetic energy $T = \frac{1}{2}\left\|\dot{\bar{\theta}}(t)\right\|$ increases, the potential energy decreases, i.e. $l(\bar{\theta})$ decreases.

iii. As soon as T decreases, the particle is moving uphill and the objective function is increasing, i.e. $\|\dot{\bar{\theta}}^{k+1}\| \leq \|\dot{\bar{\theta}}^k\|$. Some interfering strategy should be applied to extract energy from the particle to increase the likelihood of decent. A typical interfering strategy is to let $\dot{\bar{\theta}}^k = (1/4)(\dot{\bar{\theta}}^{k+1} + \dot{\bar{\theta}}^k)$ and $\bar{\theta}^{k+1} = (1/2)(\bar{\theta}^{k+1} + \bar{\theta}^k)$ after which a new $\bar{\theta}^{k+1}$ is calculated and the algorithm is continued.

iv. To accelerate convergence of the method, the algorithm should allow for magnification and reduction of the stepsize, Δt, depending on the particle's position.