4.2 Modified Newton-Raphson Optimization Method

As mentioned earlier the modified Newton-Raphson optimization method was found to be the most suitable to perform the likelihood maximization according to the criteria defined above. This method is used in the case study in Chapter 4 and it will hence be discussed in detail in this chapter.

4.2.1 Data

To simplify the discussion of the optimization technique, the form in which the data should be arranged is defined here with specific reference to vibration monitoring.

Suppose we have n cases of renewal, called histories, in our data and i is used to indicate the history number, i.e. $i=1,2,...,n$. Let T_i denote time to failure or suspension in a particular history and use $c_1, c_2, ..., c_n$ as event indicators such that $c_i = 1$ if T_i is a failure time and $c_i = 0$ in case of suspension. The number of failures present in the data is thus $r = \sum c_i$.

To be able to develop the model for time-dependent covariates we set k_i to be the number of inspections or vibration measurements at moments t_{ij} during a certain history i over the period $(0, T_i]$ for $j = 1,2,...., k_i$ such that:

$$0 = t_{i0} < t_{i1} < t_{i2} < .... < t_{ik_i} = T_i \tag{4.4.}$$

Suppose that covariate vector $\overline{z_j^i} = (z_{j1}^i, z_{j2}^i, ..., z_{jm}^i)$ consisting of m covariates, is measured during history i. For convenience of estimation, it could be assumed that $\overline{z_l^i}(t_{\text{int}}) = \overline{z_l^i}(t_{ij})$, where $t_{ij} \le t_{\text{int}} < t_{i(j+1)}$. The data for history i can be summarized as follows:

Time	*Covariates*			
t_{i0}	z_{01}^i	z_{02}^i	.....	z_{0m}^i
t_{i1}	z_{11}^i	z_{12}^i	.....	z_{1m}^i
:	:	:	.....	:
t_{ik_i}	$z_{k_i1}^i$	$z_{k_i2}^i$	.....	$z_{k_im}^i$

Table 4.1.: *Data summary for particular history*

Some covariates can be time-independent, i.e. vary with history *i*, but not with time *t*, or in mathematical terms $z_s^i = z_0^i$ for any valid value of *s*.

4.2.2 Definition of the Objective Function

The Weibull PHM as introduced in (2.17.) is repeated here for convenience as equation (4.5.):

$$h(t, \overline{z(t)}) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \cdot \exp(\bar{\gamma} \cdot \overline{z(t)}) \tag{4.5.}$$

If the model in (4.5.) is transformed with an auxiliary equation $a = -\beta \ln \eta$, (4.5.) becomes:

$$h(t, \overline{z(t)}) = \beta \cdot t^{\beta-1} \cdot \exp(a + \bar{\gamma} \cdot \overline{z(t)}), \tag{4.6.}$$

which is more convenient for calculation procedures. We will now construct the maximum log-likelihood function as explained in section (2.3.2) with (4.6.) as a function of $\bar{\theta}$, where $\bar{\theta} = (a, \beta, \gamma_1, \gamma_2, ..., \gamma_m)$. Also assume that $c_1 = c_2 = ... = c_r = 1$ and $c_{r+1} = ... = c_n = 0$ for simplification of notation. The log-likelihood will be of the form:

$$l(\bar{\theta}) = v(\bar{\theta}) - u(\bar{\theta}), \tag{4.7.}$$

where

$$\begin{aligned} v(\bar{\theta}) &= \sum_{i=1}^{r} \ln h(t_i, \overline{z(t_i)}) \\ &= ra + r \ln \beta + (\beta - 1) \cdot A + \sum_{b=1}^{m} \gamma_b B_b \end{aligned}, \tag{4.8.}$$

with

$$A = \sum_{i=1}^{r} \ln t_i \; ; \quad \text{and} \quad B_b = \sum_{i=1}^{r} z_{k_i b}^i \tag{4.9.}$$

Similarly for the second part of equation (4.7.):

$$u(\bar{\theta}) = \sum_{i=1}^{n} \int_0^{t_i} h(s, \overline{z_i(s)})\mathrm{d}s$$
$$= \exp(a) \cdot \sum_{i=1}^{n} \sum_{j=0}^{k_i-1} \exp(\sum_{g=1}^{m} \gamma_g z_{jg}^{i}) \cdot (t_{i(j+1)}^{\beta} - t_{ij}^{\beta}) \quad (4.10.)$$

The most appropriate value of $\bar{\theta}$ is found where the objective function $l(\bar{\theta})$ is a maximum, i.e. where all the partial derivatives with respect to $\bar{\theta}$ are zero or $\partial l(\bar{\theta})/\partial\bar{\theta}) = 0$.

4.2.3 Partial Derivatives

The first and second partial derivatives for $l(\bar{\theta}) = v(\bar{\theta}) - u(\bar{\theta})$ are required in the Newton-Raphson method.

First and second partial derivatives of $v(\bar{\theta})$ are:

$$\frac{\partial v(\bar{\theta})}{\partial a} = r, \quad \frac{\partial v(\bar{\theta})}{\partial \beta} = \frac{r}{\beta} + A, \quad \frac{\partial v(\bar{\theta})}{\partial \gamma_b} = B_b \quad (4.11.)$$

$$\frac{\partial^2 v(\bar{\theta})}{\partial a^2} = 0, \quad \frac{\partial^2 v(\bar{\theta})}{\partial \beta^2} = -\frac{r}{\beta^2}, \quad \frac{\partial^2 v(\bar{\theta})}{\partial {\gamma_b}^2} = 0, \quad \frac{\partial^2 v(\bar{\theta})}{\partial \theta_i \theta_j} = 0 \ (i \neq j) \quad (4.12.)$$

First and second partial derivatives of $u(\bar{\theta})$ with respect to a are:

$$\frac{\partial u(\bar{\theta})}{\partial a} = u(\bar{\theta}), \quad \frac{\partial^2 u(\bar{\theta})}{\partial a^2} = u(\bar{\theta}), \quad \frac{\partial^2 u(\bar{\theta})}{\partial a \partial \beta} = \frac{\partial u(\bar{\theta})}{\partial \beta}, \quad \frac{\partial^2 u(\bar{\theta})}{\partial a \partial \gamma_b} = \frac{\partial u(\bar{\theta})}{\partial \gamma_b} \quad (4.13.)$$

First and second partial derivatives of $u(\bar{\theta})$ with respect to β are:

$$\frac{\partial u(\bar{\theta})}{\partial \beta} = \exp(a) \cdot \sum_{i=1}^{n} \sum_{j=0}^{k_i-1} \exp(\sum_{g=1}^{m} \gamma_g z_{jg}^{i}) \cdot (t_{i(j+1)}^{\beta} \ln t_{i(j+1)} - t_{ij}^{\beta} \ln t_{ij})$$
$$\frac{\partial^2 u(\bar{\theta})}{\partial \beta^2} = \exp(a) \cdot \sum_{i=1}^{n} \sum_{j=0}^{k_i-1} \exp(\sum_{g=1}^{m} \gamma_g z_{jg}^{i}) \cdot (t_{i(j+1)}^{\beta} \ln^2 t_{i(j+1)} - t_{ij}^{\beta} \ln^2 t_{ij})$$
$$\frac{\partial^2 u(\bar{\theta})}{\partial \beta \partial \gamma_b} = \exp(a) \cdot \sum_{i=1}^{n} \sum_{j=0}^{k_i-1} \exp(\sum_{g=1}^{m} \gamma_g z_{jg}^{i}) \cdot z_{jb}^{i} (t_{i(j+1)}^{\beta} \ln t_{i(j+1)} - t_{ij}^{\beta} \ln t_{ij}) \quad (4.14.)$$

First and second partial derivatives of $u(\bar{\theta})$ with respect to γ_b are:

$$\frac{\partial u(\bar{\theta})}{\partial \gamma_b} = \exp(a) \cdot \sum_{i=1}^{n} \sum_{j=0}^{k_i-1} \exp(\sum_{g=1}^{m} \gamma_g z_{jg}^i) \cdot z_{jb}^i (t_{i(j+1)}^{\beta} - t_{ij}^{\beta})$$

$$\frac{\partial^2 u(\bar{\theta})}{\partial \gamma_b^{\,2}} = \exp(a) \cdot \sum_{i=1}^{n} \sum_{j=0}^{k_i-1} \exp(\sum_{g=1}^{m} \gamma_g z_{jg}^i) \cdot (z_{jb}^i)^2 \cdot (t_{i(j+1)}^{\beta} - t_{ij}^{\beta})$$

$$\frac{\partial^2 u(\bar{\theta})}{\partial \gamma_b \gamma_s} = \exp(a) \cdot \sum_{i=1}^{n} \sum_{j=0}^{k_i-1} \exp(\sum_{g=1}^{m} \gamma_g z_{jg}^i) \cdot (z_{jb}^i) \cdot (z_{js}^i) \cdot (t_{i(j+1)}^{\beta} - t_{ij}^{\beta}) \qquad (4.15.)$$

4.2.4 Numerical Procedure

The objective of the numerical procedure is to find the value of $\bar{\theta}$ where all the partial derivatives are zero. Let $F(\bar{\theta}) = \partial l(\bar{\theta})/\partial \bar{\theta} = (\partial l/\partial a, \partial l/\partial \beta, \partial l/\partial \gamma_1, ..$ $...., \partial l/\partial \gamma_m)$ and $G(\bar{\theta}) = \partial^2 l/\partial \bar{\theta}^2 = \partial^2 l/\partial \theta_i \partial \theta_j$. (Matrix notation is suppressed for convenience). The following approximation for $F(\bar{\theta})$ can be used: $F(\bar{\theta}) \approx F(\bar{\theta_0}) + G(\bar{\theta_0}) \cdot (\bar{\theta} - \bar{\theta_0})$ where $\bar{\theta_0}$ is an initial estimate. It is required to solve $F(\bar{\theta_0}) + G(\bar{\theta_0}) \cdot (\bar{\theta} - \bar{\theta_0}) = 0$ to determine the optimal value of $\bar{\theta}$.

The conventional Newton-Raphson procedure would solve for $\bar{\theta}$ as follows:

i. Estimate a meaningful initial value for $\bar{\theta}$, i.e. $\bar{\theta_0}$.
ii. Calculate $F(\bar{\theta_0})$ and $G(\bar{\theta_0})$.
iii. Solve for Δ_0 in the system $G(\bar{\theta_0})\bar{\Delta_0} = -F(\bar{\theta_0})$.
iv. Set $\bar{\theta_1} = \bar{\theta_0} + \bar{\Delta_0}$ and repeat the procedure until convergence.

Instead of the conventional Newton-Raphson method, a variable metric method (quasi-Newton method) can be used to overcome some numerical difficulties. In this modified Newton-Raphson method, $G(\bar{\theta})$ is not calculated directly but an approximation of $G(\bar{\theta})$ is used which is chosen to be always positive definite, thereby eliminating the possibility of singular matrices. The approximation of $G(\bar{\theta})$ is explained in detail in reference [80].

To accelerate convergence and increase accuracy of the procedure, the data is transformed to more numerically convenient forms before the iteration process is started. All recorded times, including inspection times, times to failure and times to suspension as well as the scale parameter η are divided by a value C, where:

$$C = (1/N) \sum_{i,j} t_{ij}\,, \qquad (4.16.)$$

and N is the total number of recorded times. With these scaled observations, initial values of $\eta_0 = 3 \cdot C$ and $\beta_0 = 1.5$ were found to be very reliable in the numerical procedure. The covariates are also standardized to have the same relative magnitude with the following relation:

$$z_{jl}^{*i} = \frac{z_{jl}^{i} - z_b^{avg}}{s_b}, \tag{4.17.}$$

where

$$z_b^{avg} = \frac{1}{N_b}\sum_{i,j} z_{jb}^{i} \quad \text{and} \quad s_b^2 = \frac{1}{N_b}\sum_{i,j}(z_{jb}^{i} - z_b^{avg})^2, \tag{4.18.}$$

with N_b the number of recordings of a specific covariate. This standardization of the covariates accelerates convergency considerably with initial values $\overline{\gamma_0} = \overline{0}$.

Methods to vary step sizes of the procedure as well as stopping rule procedures are discussed in Press *et al.*[80].

5 Goodness-of-fit Tests

Goodness-of-fit tests for the PHM are all aimed at evaluating the assumptions (see section (2.3.)) on which the model is based. Methods to evaluate the first assumption, the i.i.d. assumption, were discussed in section (2.1.) of Chapter 1 in considerable detail and are not repeated here.

For the second assumption, graphical methods are usually employed to test whether an influential covariate has been omitted from the model. Plots of estimated cumulative hazard rates versus the number of renewals for different strata should be approximately linear with slope equal to one, if no influential covariate has been omitted[81].

Two approaches can be used to test the validity of the third assumption. Graphical techniques have been used most widely and is considered to be the first approach.[81,82,83] The other approach is either based on hierarchical models or makes use of analytical techniques. In the case of hierarchical models, a time-dependent covariate is introduced into the model and tests are then performed to establish whether the estimate of the effect of this covariate is significantly different from zero.[2,84,85,86] A review of these methods is given by Kay[87].

5.1 Graphical Methods

Graphical methods suitable for testing the assumptions of the PHM can generally be categorized into three groups: cumulative hazard plots, average hazards plots and residual plots. In this discussion the emphasis is on residual plots because of its versatility and enormous level of inherent information.

5.1.1 Cumulative Hazard Plots

Measured values of a certain covariate may often be grouped into different levels, also referred to as *strata*, r. For example, the covariate $z_r(t)$ which occurs on s different levels and for which the proportionality assumption is to be tested is assigned to one of the s strata. Therefore, the hazard rate in this case can be written as:

$$h_r(t, \overline{z_r(t)}) = h_{0r}(t) \cdot \exp\left(\sum_{j=1, j\neq r}^{m} \gamma_j \cdot z_j(t)\right) \qquad (5.1.)$$

To explain cumulative hazard plots further, consider a binary covariate $z_r(t)$ of which the level indicator s has two values, *0* and *1*. This yields the following in terms of total hazard rate:

$$h(t, \overline{z(t)}) = h_0(t) \cdot \exp\left(\sum_{j=1, j\neq r}^{m} \gamma_j \cdot z_j(t)\right) \cdot \exp(\gamma_r); \quad (s=1) \qquad (5.2.)$$

$$h(t, \overline{z(t)}) = h_0(t) \cdot \exp\left(\sum_{j=1, j\neq r}^{m} \gamma_j \cdot z_j(t)\right); \quad (s=0) \qquad (5.3.)$$

Therefore, to satisfy the assumption that $h(t, z_x(t)) \propto h(t, z_y(t))$ we obtain from (5.2.) and (5.3.) that $h_{0r}(t) = c_r h_0(t)$ for $r = 1,2,...,s$ where $c_1, c_2, ..., c_s$ are constants equal to $\exp(\gamma_r \cdot z_r(t))$ for all strata. A similar relation holds for the cumulative hazard rate, i.e. $H_{0r}(t) = c_r H_0(t)$. If plots of the logarithm of the estimated cumulative hazard rates against time are constructed, they will be shifted by an additive constant γ_r, the regression parameter of a specific stratum. Thus, if the proportionality assumption is valid, the two plots should be approximately parallel and separated according to the different values of the covariates. Figure 5.1. below illustrates this concept for a case where the PHM is valid.

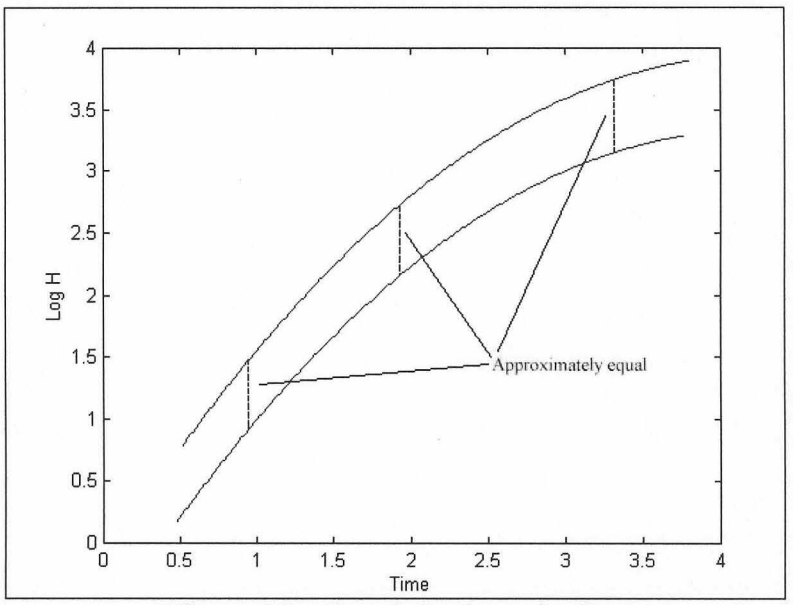

Figure 5.1.: *Cumulative hazards plots*

5.1.2 Average Hazards Plots

Average hazards plots for different strata are based on the assumption that $h_{rj}(t)$ is a slowly varying function that can be approximated by piecewise constant functions, i.e. $h_{rj}(t) = h_{rj}$ in the time interval between inspection number *i(j-1)* and *ij*, where *i* is the history number. Cox[82] defined $S_{rj}^{(i)}$ to be the total estimated operational time at risk in the strata *r* between inspections $i(j-1)^{th}$ and j^{th}. The operational time is the timescale obtained when the contribution at each inspection point associated with a specific covariate is weighted by $\exp(\gamma \cdot \overline{z_i(t)})$. An auxiliary random variable is now defined as:

$$Z_{rj}^{(p)} = \frac{1}{(2i - 1/3)} - \ln\left(\frac{S_{rj}^{(i)}}{i}\right), \qquad (5.4.)$$

which is independent with mean $\ln(h_{rj})$ and variance $1/(i-0.5)$ [82]. If plots of $Z_{rj}^{(i)}$ against the midpoint of the time interval are constructed, the different plots should be parallel and spaced according to the estimated value of the covariate $\overline{z_r(t)}$ defining the strata, if the model fits the data.

5.1.3 Residual Plots

Residual plots are constructed with Cox-generalized residuals for the PHM. Cox-generalized residuals are given by:

$$r_i = \begin{cases} u_i, \text{ if } t_i \text{ is a failure time} \\ u_i + 1, \text{ if } t_i \text{ is a suspension time} \end{cases} \quad (5.5.)$$
$$\text{where } i = 1,2,....,n.$$

In (5.5.) u_i is defined as:

$$u_i = \frac{1}{\eta^\beta} \sum_{j=0}^{k_i - 1} \exp(\overline{\gamma} \cdot \overline{z_j^i}) \cdot \left[t_{i(j+1)}^\beta - t_{ij}^\beta \right], \quad (5.6.)$$

with the same notation used as in section (4.2.). The calculation of u_i can be checked by noting that:

$$\sum_{i=1}^{n} u_i = r; \text{ and } \sum_{i=1}^{n} r_i = n, \quad (5.7.)$$

with r denoting the number of failures in this case. The unknowns in (5.6.) are determined during the model fitting procedure as described in paragraph (4.2.). With these Cox-generalized residuals known, several plots can be constructed to assess the goodness of fit visually.

5.1.3.1 Residuals Against Order of Appearance

Here the residual for every history $i = 1,2,....,n$ is plotted against the corresponding history number, i.e. $(x_i; y_i) = (i; r_i)$.

The residuals should all be scattered around the straight line $y = 1$. Note that the residual values of suspended cases will always be greater than *1*. If an upper limit of $r_i = 3$ (95%) and a lower limit of $r_i = 0.05$ (5%) are chosen, it is expected that at least 90% of the residuals will fall inside these limits if the model fits the data.

5.1.3.2 Ordered Residuals Against Expectation

If the calculated residuals $r_1, r_1, ..., r_n$ are ordered in ascending order we get $r_1^* \le r_2^* \le ..... \le r_n^*$, the ordered residuals. The expected values of the residuals are:

$$E_i = E_{i,n} = \frac{1}{n} + \frac{1}{n-1} + ... + \frac{1}{n-i+1} \quad (5.8.)$$

For a Weibull PHM that fits the data well, the points $(x_i; y_i) = (E_i; r_i^*)$ will be distributed around the line $y = x$. Note however that the difference in consecutive expectations increases and that there will be a concentration of cases below the value *1* (around 50% - 60%). The points on the right side of the $y = x$ line need not necessarily be close to the line to indicate an appropriate model. This is because the variability of the residuals increases with order number. It is possible to improve the situation by using suitable transformations for the residuals.

To transform all points to lie between *0* and *1* with an approximately equally spaced x-axis we could use $x_i = 1 - \exp(-E_i)$ and $y_i = 1 - \exp(-r_i^*)$. It is possible to stabilize the variance by using $x_i = (2/\pi)\sin^{-1}\{\exp[-(E_i/2)]\}$ for the x-axis and $y_i = (2/\pi)\sin^{-1}\{\exp[-(r_i^*/2)]\}$ for the y-axis. All points will again lie between *0* and *1*. For further discussion on this method see reference [88, 89].

5.2 Analytical Goodness-of-Fit Tests

Graphical goodness-of-fit tests are often interpreted totally different by different analysts. For this reason, analytical goodness-of-fit tests have tremendous value since it is totally objective. Several analytical tests have been used on the PHM in the literature, amongst others, the χ^2 test, the log rank test, Z-test (normal distribution test), Kolmogorov-Smirnov test, Wald test, the doubly cumulative hazard function, the likelihood ratio test, score tests and generalized moments specification tests. See references [3,84,90,91,92,93,94,95,96,98]. Three of these tests are discussed below.

5.2.1 *Z*-Test

Before the *Z*-test can be presented, some comments have to be made about transformation rules in statistics. Transformation rules describe the changes in the mean, variance and standard deviation of a distribution when every item in a distribution is either increased or decreased by a constant amount. These rules also describe the changes in the mean, variance and standard deviation of a distribution when every item in the distribution is either multiplied or divided by a constant amount.

- **Rule 1:** Adding a constant to every item in a distribution adds the constant to the mean of the distribution, but it leaves the variance and standard

deviation unchanged.

- **Rule (2):** Multiplying every item in a distribution by a constant multiplies the mean and standard deviation of that distribution by the constant and it multiplies the variance of the distribution by the square of the constant.

The *Z*-test makes use of a special application of the transformation rules, the *Z*-score statistic from which inferences are made. The *Z*-score for an event, indicates how far and in what direction that event deviates from its distribution's mean, expressed in units of its distribution's standard deviation. The mathematics of the *Z*-score transformation are such that if every event in a distribution is converted to its *Z*-score, the transformed scores will necessarily have a mean of zero and a standard deviation of one.

For the PHM, the *Z*-test can be used by letting $r_1^* \leq r_2^* \leq ..... \leq r_n^*$ be the ordered residuals, as before, and define $Z_i = r_i^* / r_n^*$ and $m = n - 1$. The *Z*-score for the PHM is then:

$$Z = \frac{\sum_{i=1}^{m} Z_i - m/2}{\sqrt{m/2}} \tag{5.9.}$$

Inferences about the value of Z can be made by calculating the p-value of Z, using the normal distribution.

5.2.2 Kolmogorov-Smirnov Test

This testing procedure is classified as a frequency test of the degree of agreement between distributions of a sample of generated random values (or a sample of empirically gathered values) and a target distribution. Since it is known that the residuals of the PHM should have an exponential distribution, the Kolmogorov-Smirnov test is performed on the PHM residuals to check the model fit.

The null hypothesis is that the cumulative density function of the PHM residuals is equal to the cumulative density function of an exponential distribution fitted on the residuals. To be able to test the null hypothesis, a test statistic, called the *D-statistic*, is introduced. This statistic is defined as the largest absolute difference between the Weibull PHM residuals and the cumulative exponential distribution and inferences on the goodness-of-fit of a model is made based on this statistic. The procedure seems simple but becomes fairly complicated for the PHM with censored data[97].

As before, assume $r_1^* \leq r_2^* \leq ..... \leq r_n^*$ to be the residuals ordered by magnitude and c_i to be an event indicator as defined in section (4.2.1). Let s_i and a_i be sequences defined by:

$$s_{i+1} = s_i \cdot \left(1 - \frac{1}{n-i}\right)^{c_i} \tag{5.10.}$$

$$a_{i+1} = a_i + \frac{n}{(n-i)\cdot(n-i-1)} \cdot c_i, \tag{5.11.}$$

with $s_0 = 1$, $s_n = \exp(-r_{n-1}^*)$, $a_0 = 0$, $a_n = 0$ and $i = 0,1,2,....,n-2$. The D-statistic is then:

$$D = \max_{0 \leq i \leq n-1} \left\{ \max \left\{ \frac{\left| s_i - \exp(-r_i^*) \right|}{(1+a_i)\exp(-r_i^*)}, \frac{\left| s_{i+1} - \exp(-r_i^*) \right|}{(1+a_{i+1})\exp(-r_i^*)} \right\} \right\} \tag{5.12.}$$

From the *D*-statistic, a *p*-value can be determined to evaluate the quality of the fit.

5.2.3 Wald Test

A test specifically developed for testing the quality of parameter estimations by the method of maximum likelihood, is the Wald test. This test is categorized under likelihood ratio tests and can be used to evaluate the appropriateness of specific coefficients in the estimated regression vector and not only the total goodness-of-fit of a model. This attribute of the Wald test is very useful for the PHM because the contribution of different covariates to the quality of the model can be assessed.

The Wald test statistic for a specific regression coefficient is then given by:

$$W_i = \frac{n \cdot (\theta_i)^2}{\mathrm{Var}(\theta_i)}, \tag{5.13.}$$

where $\mathrm{Var}(\theta_i)$ is the variance of the regression coefficient and n is the sample size. Inference on the Wald test statistic is made by calculating *p*-values from the χ^2 distribution. See [98] for further details.

6 Optimal Decision Making with the Proportional Hazards Model

The PHM supplies us with an accurate estimate of a component's present risk to fail (hazard rate), based on its primary use parameter and the influence of covariates. This educated knowledge of the hazard rate should be utilized to the full to obtain economical benefits, otherwise the PHM estimation exercise is futile.

Economical benefits from a statistical failure analysis can be guaranteed with a high confidence level if the minimum long term life cycle cost (LCC) of a component is determined and pursued, i.e. if renewal always takes place at either the statistical minimum LCC or in the case of failure prior to the minimum LCC. The instant of minimal LCC can easily be specified in terms of time for statistical models where time is the only age parameter but for models including covariates, this makes no sense because the covariates influence survival time.

For optimal decision making with the PHM in reliability, only one model is used in practice, a model specifically developed for the PHM by Makis and Jardine[13,14]. This specifies the optimal renewal policy in terms of an optimal hazard rate which will lead to the minimum LCC. At every inspection the latest hazard rate is calculated and if it exceeds the optimal hazard rate the component is renewed, otherwise operation is continued. If the recommendations of the decision model is obeyed, the LCC will strive to a minimum over the long run.

To be able to determine the hazard rate that will lead to the minimum LCC it is required to predict the behavior of covariates. Makis and Jardine's model does this by assuming the covariate behavior to be stochastic and approximating it by a non-homogeneous Markov chain in a finite state space. The Markov chain leads to transition probabilities of covariates from where the optimal hazard rate is calculated.

An approach of predicting the useful remaining service life of a component and acting preventively on the prediction rather than pursuing a statistical optimum sounds intuitively meritorious but no research on techniques for such an approach in conjunction with the PHM has been published up to date.

6.1 The Long Term Life Cycle Cost Concept

LCC is a concept used widely in statistical failure analysis. Several models to achieve this minimum for repairable systems and renewal situations which depends only on time can be found in the literature. See [15] and [22] for an overview. The

minimum LCC in renewal situations arise from two important quantities in practice namely the cost of unexpected renewal or failure of a component, C_f, and the cost of preventive replacement C_p. It is normally much more expensive to deal with an unexpected failure than it is to renew preventively. A balance has to be obtained between the risk of having to spend C_f and the advantage in the cost difference between C_f and C_p without wasting useful remaining life of a component. The optimum economic preventive renewal time will be at this balance point. LCC's are usually compared when expressed as cost per unit time.

The LCC concept will be illustrated with a simple Weibull model, i.e. a model without covariates. If a component is renewed either preventively after t_p time units or at unexpected failure for every life cycle over the long term, the total expected cost for a life cycle would be:

$$C_t = C_p R(t_p) + C_f[1 - R(t_p)] \tag{6.1.}$$

Since, the LCC is usually expressed in cost per unit time the average life expectancy has to be calculated as well:

$$L_e = (t_p + T_p)R(t_p) + (t_f + T_f)[1 - R(t_p)], \tag{6.2.}$$

where t_f is the expected length of a failure cycle under the condition that failure occurred before t_p and T_p and T_f are the times required for preventive renewal and failure renewal, respectively. When (6.1.) and (6.2.) are expressed in terms of the Weibull reliability functions and divided into each other, the LCC per unit time, if renewed at t_p over the long term, is:

$$C(t_p) = \frac{C_p \cdot e^{-(t_p/\eta)^\beta} + C_f \cdot [1 - e^{-(t_p/\eta)^\beta}]}{(t_p + T_p) \cdot e^{-(t_p/\eta)^\beta} + \int_0^{t_p} \left(t \cdot \frac{\beta}{\eta} \cdot \left(\frac{t}{\eta} \right)^{\beta-1} \cdot e^{-(t/\eta)^\beta} \right) dt + T_f \cdot [1 - e^{-(t_p/\eta)^\beta}]} \tag{6.3.}$$

The preventive renewal time that will lead to the minimum LCC, t_p^*, is found where $D_t[C(t_p)] = 0$. An example of (6.3.) is shown in Figure 6.1. for Weibull parameters of $\beta = 1.80$ and $\eta = 430$ days.

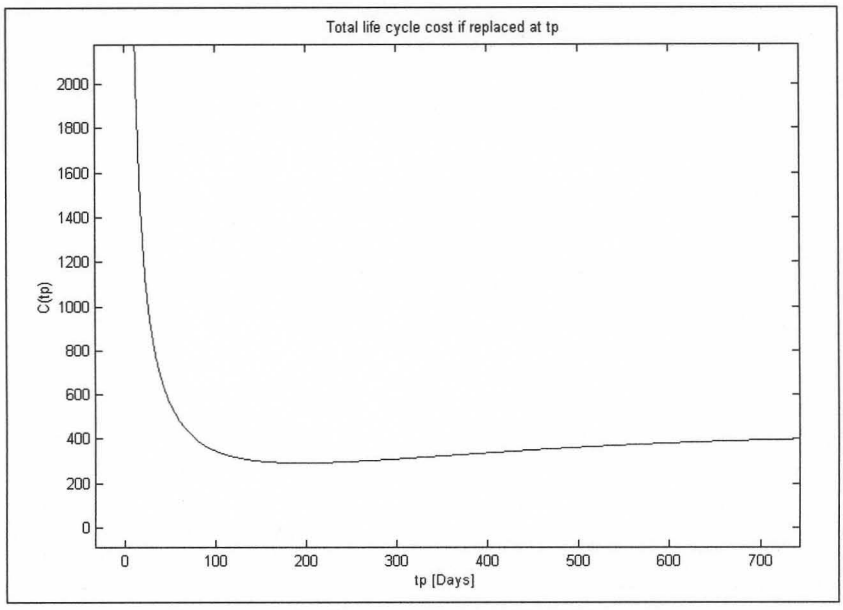

Figure 6.1.: *Example plot of LCC function*

A distinct minimum for $C(t_p)$ in Figure 6.1. exist at 221 days.

6.2 Prediction of Covariate Behavior

Techniques used to predict covariate behavior in Makis and Jardine's model is discussed in this section.

6.2.1 Transition States and Covariate Bands

Transition states have to be defined for the covariates before it can be modeled with a Markov chain. For this reason, every range of covariate values is divided into appropriate intervals or bands and every covariate band is defined as a covariate state. Covariate bands are then used as boundaries for the transition probabilities in the transition probability matrix (TPM). For numerical convenience, 4 or 5 bands are usually selected between upper and lower bands except for the last band which need not have an upper bound.

6.2.2 Markovian Chains and the Transition Probability Matrix

Suppose that $\{X_0, X_1, X_2, ...\}$ is a multidimensional Markov process which makes up a component's renewal history such that $X_k = (z_{k1}(t), z_{k2}(t), .., z_{km}(t)) \in \Re^m$ where m is the number of covariates, and $z_{ki}(t)$ is the k^{th}

observation of variable i before renewal, performed at time $t = k\Delta$, $(k = 0,1,2,....)$ while Δ is a fixed inspection interval. A stochastic process $\{X_0, X_1, X_2, ...\}$ is assumed to be Markovian if, for every $k \geq 0$,

$$P\{X_{k+1} = j \mid X_k = i, X_{k-1} = i_{k-1}, X_{k-2} = i_{k-2}, ..., X_0 = i_0\} = P\{X_{k+1} = j \mid X_k = i\} \quad (6.4.)$$

where $j, i, i_0, i_1, ..., i_{k-1}$ are defined states of the process, in this case the covariate bands.

The transition probability for any covariate in state i to undergo a transition to state j for a given inspection interval Δ is:

$$P_{ij}(k) = P_{ij}(k,\Delta) = P(X_{k+1} = j \mid T > (k+1)\Delta, X_k = i), \quad (6.5.)$$

where T denotes time to renewal as before and i and j denote any two possible states.

Suppose we have a sample $X_{i0}, X_{i1}, X_{i2}, ...$ and let $n_{ij}(k)$ denote the number of transitions from state i to j at k throughout the sample, where the sample may contain many histories:

$$n_{ij}(k) = \#\{X_k = i, X_{k+1} = j\} \quad (6.6.)$$

Similarly, the number of transitions from i at time $k\Delta$ to any other state can be calculated by:

$$n_i(k) = \#\{X_k = i\} = \sum_j n_{ij}(k) \quad (6.7.)$$

It is now possible to estimate the probability of a transition from state i to state j at time $k\Delta$ with the following relationship derived with the maximum likelihood method:

$$\hat{P}_{ij}(k) = \frac{n_{ij}(k)}{n_i(k)}, \quad k = 0,1,2,... \quad (6.8.)$$

If it is assumed that the Markov chain is homogeneous within the interval $a \leq k \leq b$, i.e. $P_{ij}(k) = P_{ij}(a)$, the transition probability can be estimated by:

$$\hat{P}_{ij}(k) = \frac{\sum_{a \le k \le b} n_{ij}(k)}{\sum_{a \le k \le b} n_i(k)}, \quad a \le k \le b \tag{6.9.}$$

It would also be possible to assume that the entire Markov chain is homogeneous, then $P_{ij} = P_{ij}(k)$, for $k = 0,1,2,....$ and hence the transition probabilities are estimated by:

$$\hat{P}_{ij} = \frac{n_{ij}}{n_i}, \text{ where } n_{ij} = \sum_{k \ge 0} n_{ij}(k),\ n_i = \sum_j n_{ij} \tag{6.10.}$$

It is not realistic to assume that the transition probabilities of vibration covariates are independent of time. For this reason continuous time is divided into w intervals, $[0, a_1], (a_1, a_2], ..., (a_w, \infty)$, in which the transition probabilities are considered to be homogeneous. This manipulation simplifies the calculation of the TPM tremendously without loosing much accuracy.

The estimations of the TPM above all assumed that the inspection interval Δ was constant. In practice, this is rarely the case. This would mean that recorded data with inspection intervals different than Δ have to be omitted from TPM calculations, thereby loosing valuable information about the covariates' behavior. To overcome this problem a technique utilizing transition densities (or rates) is used. Assume that the Markov chain is homogeneous for a short interval of time. The probability of transition from $i|_{t=0} \to j|_{t=t}$ is $P_{ij}(t) = P(X(t) = j \mid X(0) = i)$ and the rate at which the transition will take place is $D_t[P_{ij}(t)] = \lambda_{ij}$, $(i \ne j)$. For the case where $i = j$ the transition rate can be derived with the following argument. Suppose the system is in state $i|_{t=0}$ and state $j|_{t=t}$ with r possible states. If the sum over all probabilities over t is taken:

$$\begin{aligned} P_{i0}(t) + P_{i1}(t) + P_{i2}(t) + ... + P_{ir}(t) &= 1 \\ \sum_j P(X(t) = j \mid X(0) = i) &= 1 \\ \text{or } \sum_j P_{ij}(t) &= 1 \end{aligned} \tag{6.11.}$$

If we take the time derivative,

$$\sum_j \frac{\partial}{\partial t}\left[P_{ij}(t)\right] = 0$$

$$\therefore \lambda_{i0} + \lambda_{i1} + \lambda_{ii} + ... + \lambda_{ir} = 0 \qquad (6.12.)$$

$$\lambda_{ii} = -\sum_{i \neq j} \lambda_{ij}$$

The value of any λ_{ij}, $(i \neq j)$ can be approximated by:

$$\hat{\lambda}_{ij} = \frac{n_{ij}}{\Omega_i}, \quad n_{ij} = \sum_k n_{ij}(k) \qquad (6.13.)$$

where, k runs over the given interval of time and Ω_i is the total length of time that a state is occupied in the sample. The calculation of the transition rates can be generalized for the system from any state i to j at any time t with:

$$P'_{ij}(t) = \sum_l P_{il}(t)\lambda_{lj} \qquad (6.14.)$$

Equation (6.14.) provides a system of differential equations that has to be solved to obtain the transition probability matrix. A solution to the system of differential equations solution is:

$$P(t) = \exp(A \cdot t), \qquad (6.15.)$$

where $P(t) = (P_{ij}(t))$ and $A = (\lambda_{ij})$. (Brackets denote matrices). This can be calculated by the series:

$$P(t) = \sum_{n=0}^{\infty} A^n \frac{t^n}{n!}, \qquad (6.16.)$$

which is fast and accurate. Statistical tests (such as χ^2) can be used to confirm the validity of the homogeneity assumption over the given time intervals.

6.2.3 Calculation of the Optimal Decision Policy

Two different renewal possibilities are considered in Makis and Jardine's model: (i) Variant 1, where preventive renewal can take place at any moment; (ii) Variant 2, where preventive replacement can only take place at moments of inspection. Only Variant 1 will be discussed since Variant 2 is only a simplification of Variant 1.

A basic renewal rule is used: if the hazard rate is greater than a certain threshold value, preventive renewal should take place otherwise operations can continue. The objective here is thus to calculate this threshold level while taking working age and covariates into account.

The expected average cost per unit time is a function of the threshold risk level, d, and is given by [13,14]:

$$\Phi(d) = \frac{C_p + KQ(d)}{W(d)}, \tag{6.17.}$$

where $K = C_f - C_p$. $Q(d)$ represents the probability that failure replacement will occur, i.e. $Q(d) = P(T_d \geq T)$ with T_d the preventive renewal time at threshold risk level d or $T_d = \inf\{t \geq 0 : h(t, \overline{Z(t)}) \geq d / K\}$. $W(d)$ is the expected time until replacement, regardless of preventive action or failure, i.e. $W(d) = E(\min\{T_d, T\})$. The optimal threshold risk level, d^*, is determined with fixed point iteration to get:

$$\Phi(d^*) = \min_{d>0} \Phi(d) = d^*, \tag{6.18.}$$

if the hazard function is non-decreasing, e.g. if $\beta \geq 1$ and all covariates are non-decreasing and covariate parameters are positive. If covariates are not monotonic, then the fix point iteration does not work, and $\min_{d>0} \Phi(d)$ should be found by a direct search method. During the calculation of d^* it is necessary to calculate $Q(d)$ and $W(d)$ which is no a trivial procedure. To do this we define the covariate vector $\overline{z(t)} = [z_1(t), z_2(t), \ldots\ldots, z_m(t)]$ as before with $\overline{i(t)} = [i_1(t), i_2(t), \ldots\ldots, i_m(t)]$ the state of every covariate at time t. Thus, for every coordinate l let $X^l(i_l(t))$ be the value of the l^{th} covariate in state $i_l(t)$ (representative of the state) at moment t, and $X(\overline{i(t)}) = \{X^1(i_1(t)), \ldots ., X^m(i_m(t))\}$. We could express the hazard rate now as:

$$h(t, \overline{i(t)}) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \exp(\overline{\gamma} \cdot X(\overline{i(t)})) \tag{6.19.}$$

From (2.20.), the conditional reliability function can be defined as $R(j, i, t) = P(T > j\Delta + t \mid T > j\Delta, \overline{i(t)})$, which becomes after substitution:

$$R(j, \overline{i(t)}, t) = \exp\left\{ - \exp(\overline{\gamma} \cdot \overline{X}) \cdot \left[\left(\frac{j\Delta + t}{\eta} \right)^{\beta} - \left(\frac{j\Delta}{\eta} \right)^{\beta} \right] \right\} \tag{6.20.}$$

with $0 \le t \le \Delta$. If $h(t,\overline{i(t)})$ is a non-decreasing function in t, and if we define $t_i = \inf\{t \ge 0 : h(t,\overline{i(t)}) \ge d/K\}$ and the k_i's as integers such that $(k_i - 1)\Delta \le t_i < k_i\Delta$ we can calculate the mean sojourn time of the system in each state with:

$$\tau(j,\overline{i(t)}) = \begin{cases} 0, & j \ge k_i \\ \tau(j,i,a_i), & j = k_i - 1 \\ \tau(j,i,\Delta), & j < k_i - 1 \end{cases}, \tag{6.21.}$$

where $a_i = t_i - (k_i - 1)\Delta$ and $\tau(j,i,s) = \int_0^s R(j,i,t)dt$. Similarly, the conditional cumulative distribution function for this situation is:

$$F(j,\overline{i(t)}) = \begin{cases} 0, & j \ge k_i \\ 1 - R(j,i,a_i), & j = k_i - 1 \\ 1 - R(j,i,\Delta), & j < k_i - 1 \end{cases} \tag{6.22.}$$

Let for each j, $\tau_j = (\tau(j,i))_i$ and $F_j = (F(j,i))_i$ are column vectors, and $(P_j) = (R(j,i,\Delta)P_{il}(j))_{il}$ is a matrix. From here, the column vectors $W_j = (W(j,i))$ and $Q_j = (Q(j,i))$ are calculated as follows:

$$\begin{aligned} W_j &= \tau_j + P_j W_{j+1} \\ Q_j &= F_j + P_j Q_{j+1} \end{aligned} \tag{6.23.}$$

Then $W = W(0,i_0)$ and $Q = Q(0,i_0)$, where i_0 is an initial state of the covariate process, usually $i_0 = 0$. By starting the calculation with a large value for j, where $W_{j+1} = Q_{j+1} = 0$ and working back to 0, it is possible to solve for W and Q from (6.23.). The above calculation procedure is described in detail in [14]. A forward version of this backward calculation is numerically more convenient and much faster (see [23]), which can be suitably adjusted for non-monotonic hazard functions also.

Thus, once the optimal threshold level is determined we renew the item at the first moment t when:

$$\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \exp\left(\underline{\gamma} \cdot \overline{z(t)}\right) \ge \frac{d^*}{K}, \tag{6.24.}$$

or, which is practically more convenient, when

$$\bar{\gamma} \cdot \overline{z(t)} \geq \delta^* - (\beta - 1)\ln t\,, \qquad (6.25.)$$

where $\delta^* = \ln(\frac{d^* \eta^\beta}{K\beta})$.

A warning level function is defined only in terms of time by:

$$g(t) = \delta^* - (\beta - 1) \cdot \ln(t)\,, \qquad (6.26.)$$

with $g(t)$ strictly decreasing if $\beta > 1$.