A SURVIVAL ANALYSIS ON ENDURANCE TESTS OF CHAINSAWS AND POWER CUTTER COMPONENTS AT HUSQVARNA GROUP AB

JOHANNES LINDVALL

Master's thesis
2022:E48

LUND UNIVERSITY

Faculty of Engineering
Centre for Mathematical Sciences
Mathematical Statistics

CENTRUM SCIENTIARUM MATHEMATICARUM

Abstract

Due to the steadily growing world chainsaw market and the market's relative competitiveness, high performing chainsaws are important for Husqvarna Group AB to remain competitive. In the development phase a large number of endurance tests are run to evaluate the product's performance. Given the data gathered from these tests, the main goal of this thesis was to model the time until failure and evaluate the relationship between the measured predictors and the failure times. To model the data, traditional survival theory was used, which included fitting a Cox proportional hazard model and an extended Cox model for time-varying variables.

The results of the analysis showed that the temperature of the surrounding equipment has a significant effect on the failure times of chainsaws for both test methods. For the first test method the volatility of the revolutions per second(RPS) in the engine also proved to have a significant effect. Component cracks in power cutters seemed to increase with dry weather and low temperatures, confirming what the engineers at Husqvarna Group AB believed to be the case before this thesis was carried out.

Other future analysis methods were discussed together with some identified data collection challenges. When the technical difficulties for collecting the proposed new data sets have been resolved, the models suggested in this thesis may be improved by linking a number of new predictors to the failure times.

Acknowledgements

I would like to thank Nader Tajvidi at Lund University for the support he has given me throughout the work with this thesis. He has given me valuable guidance and encouragement which have been very appreciated. I am very happy to have had him as my supervisor.

I would also like to thank Lars Walfridsson at Husqvarna Test Systems Engineering for the opportunity to write my thesis at his department. To be able to work with data from a real life problem have been most instructive and motivating. I also appreciate all his support in discussing practical problems related to the thesis and getting in contact with the right persons.

Finally, I would like to thank all the other Husqvarna employees that I have been in contact with during the course of this work who have been more than willing to help out in explaining technical aspects and processes related to the problem at hand.

Contents

Chapter 1

Introduction

1.1 Background

According to the research company IMARC(2020) the global chainsaw market grew at an annual rate of 5% (Compound Annual Growth Rate) over the years 2015-2020. According to Precedence Research(2022) the global market was estimated to 3.6$ billion 2021 and will reach 4.92$ billion 2030 with an annual growth rate of 3.5% from this year. This shows that there undoubtedly seems to be an increasing demand for chainsaws worldwide.

The market is quite fragmented with many important players which leads to relatively high competition. Husqvarna group is mentioned as one of the major players by Precedence Research and Husqvarna themselves state that they are number 2 world wide when it comes to handheld devices such as chainsaws. The Husqvarna division responsible for handheld devices, such as chainsaws, provides 62% of the groups net turnover. This highlights the importance of competitive chainsaw models for both the future and the present financial performance of the company(Husqvarna Group AB. 2022. p. 27).

1.2 Test Systems Engineering at Husqvarna

The department Test Systems Engineering at Husqvarna group builds testbeds for mechanical products and hardware objects. These are used to run tests on complete products and components to evaluate their technical performance. During the tests a great number of measurement data as well as component failures are collected. Testing is both time and resource consuming and it is of interest to continuously improve the testing methodology.

As of today there is uncertainty about what conclusions and knowledge can be gained from the existing test data when looking at large amounts of tests together in contrast to evaluating individual tests separately. If the relationship between the measurement data and the test time can be modeled with high precision on a large scale then a great number of aspects around the test procedure could be improved. To do this, there is a need to try out some modern data analysis methods on the existing data to determine what is possible in the existing situation and where the focus should be in the future to improve the data analysis capacity.

Some examples of potentially positive results of being able to model the relationship described above are related to three different aspects.

- Methodological aspect: If the time to failure can be predicted with low margin of error then tests could be stopped close to failure, which would enable dissection of the test object to increase knowledge about the mechanical causes of failure. Greater knowledge about what causes failures would improve the test methodology by measuring the right things and running the right tests.

- Economic aspect: The data collection method could be improved by collecting only the most relevant data and discarding irrelevant data. This could lead to resource savings by saving data storage capacity and investments in data collection equipment. Better understanding of the test environment could also motivate investments in test equipment to be able to control the surrounding conditions of the test environment. Finally, measuring the right things and running the right tests is related to getting more for the invested money.

- Time aspect: Being able to stop tests earlier and, once again, running the most relevant test would save time. This also has potential to increase the product development speed.

1.3 Why survival analysis?

An interesting statistical method used to analyse failure time data is survival analysis. When studying survival theory, examples related to the theory are often related to clinical research(Kent, 2021). Studies in this area could, for example, be the effect of a certain cancer treatment by studying the time until death of the patient and different biological characteristics. This method, however, can be used in a lot of other applications when the aim is to investigate the time until a certain event occurs and what factors affect it as for example a test failure at Husqvarna. As Kent(2021) points out in his article about applications of survival analysis, one of the most important reasons to use survival analysis is that one wants to make decisions before observing all of the data. What he means with that is that some objects of study have not been studied long enough in order to experience the event in question. Take for example the time until a chainsaw fails in one of the tests at Husqvarna, where a test might be stopped before a failure have occurred. The reason for this is that, for practical and economical reasons, there needs to be a limit for when the tests are stopped even if a failure have not occurred. If an object have not failed during this time it will lead to an "incomplete" observation, or so-called censored data. The ability of handling such observations is the strength of survival analysis and the strongest reason that it is used in this study. However, as Kent(2021) points out, the statistical methods of survival analysis only works if there are enough complete, non-censored observations. The trade-off of having enough complete data points and having data that are similar enough to be used in the same analysis will be a recurrent challenge throughout this study.

1.4 Aim of the study

The first objective of the study is to describe and model the time until failure for endurance tests on Husqvarna chainsaws. The relationship between the measured predictors and failures of chainsaw tests will be analysed. The analysis aims to answer the following questions:

- How well does the measured factors explain the component failures in chainsaw tests?
- What predictors have the strongest effect on the time until failure of chainsaws?
- Was the chosen statistical method suitable for the problem and what other methods would be suitable to try out in the future?
- What improvements regarding data collection could be made in order to enable better modelling of the time until failure in the future?

The second objective is to investigate the relationship between the local weather conditions and cracks in components mounted on power cutters. The following question will be answered:

- What weather conditions poses the greatest risk for cracks in components in power cutters?

1.5 Restrictions

We will restrict ourselves to only look at chainsaw data for two different test methods: 'real usage emulated' and 'wear test'. These will be denoted test method 1 and 2. The test methods are explained in Section 3.2.

We will not look at recurrent event analysis, even though there are multiple events registered per test in some cases.

For the analysis of components in power cutters the predictors to analyse were restricted to the air humidity, air temperature and air pressure.

Chapter 2

Theory

2.1 Introduction to survival analysis

The basic problem in survival analysis is studying the time until a certain event. That event can be many different things, typically the death of an individual in biomedicine, the breakout of a disease, relapse in criminal actions, recovery from a disease or breakdown of a component or mechanical product. In this report the latter will be used in the provided examples since it is the event of study in this report. In survival analysis one is often interested in comparing groups of individuals that are exposed to different types of external or internal factors and draw conclusions on how those factors influence the time to event or risk of experiencing the event.

2.1.1 Data and basic terminology

Typically the data consists of a number of comparable individuals or objects which either experienced the event or did not. The time that they have been studied is registered together with the values of possible explanatory variables.

A central concept in survival analysis is the concept of censoring. The so-called censored objects are objects that are included in the study but the exact time when they got the event is not known. The censored data can be divided into, left, right and interval censored data(Kleinbaum and Klein, 2020 p. 7-8). The most relevant for this report is the right censored data, which means that the study of the object is ended before the event occurred. Relating this concept to the data in this study the following explanation is adequate: the mechanical objects in the study are run for a maximum pre-set amount of time and if the object is functioning at that time the test is ended. Therefore the exact time that the objects would have failed if the test was run to failure is not known, hence these objects are right censored.

The concept of censoring provides an advantage for survival analysis in comparison to, for example, linear and logistic regression. To study the time until the event, also objects that did not have the event can be included in the model, which greatly increases the available data in some cases.

2.2 The survivor function and the hazard function

The survivor function, $S(t)$ and the hazard function, $h(t)$ are two main concepts considered in survival analysis. To give some context to the formulas presented below, the random variable describing the survival time of a certain object is denoted T. The survivor function, $S(t)$ describes the probability of survival after a certain time, t, for a certain object. As an example the probability that an object survives longer than 5 time units can be written as:

$$S(5) = P(T > 5) \tag{2.1}$$

The survivor function also has some basic theoretical properties:

1. The survivor function is non-increasing as t increases.
2. $S(0) = 1$, in effect the probability of surviving past time 0 is 1, since no objects have failed at the beginning of the study.
3. $S(\infty) = 0$, meaning that if one would wait sufficiently long all objects would have failed(Kleinbaum and Klein, 2020, p. 9-10).

The hazard function, however, is a measure of the immediate rate of failing at a certain time. It gives the failure rate at the start of a certain time interval, given that the object has survived up to time t and can therefore be called the conditional failure rate(Kleinbaum and Klein, 2020, p. 11-12). The hazard function is defined mathematically as:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} \tag{2.2}$$

Noting that the time interval, Δt in equation (2.2) goes towards zero, the interpretation as the immediate failure rate at the time t is justified. The hazard function, $h(t)$ has the following theoretical properties:

1. $h(t) \geq 0$

2. $h(t)$ has no upper bound(Kleinbaum and Klein, 2020, p. 13).

The way the hazard function changes over time gives an indication of the distribution of the failure process of the underlying object of study. If the hazard function is constant over time we get an exponential surival model. If it increases or decreases over time the process resembles an increasing or decreasing Weibull distribution. Because of these properties the hazard function can be used to identify a specific model form that fits the data(Kleinbaum and Klein, 2020, p. 13-14).

Additionally, if one knows either the survivor function or the hazard function one can derive the other. The relationship between the survivor function and the hazard function looks like in equation (2.3) and equation (2.4).

$$S(t) = \exp\left[-\int_0^t h(u)\,du\right] \tag{2.3}$$

$$h(t) = -\left[\frac{dS(t)/dt}{S(t)}\right] \tag{2.4}$$

2.3 The Kaplan-Meier estimator

When estimating the survival curve of survival data without censored values it is easy to calculate the estimated probability of functioning past different time periods. An example of this can be seen in Table 2.1 where t is the failure time, n_i is the number of objects functioning at the start of each time interval $[t_i, t_{i+1}]$, m_i is the number of objects failing at time t, and q is the number of censored objects at each time t.

Table 2.1: Example table of how to estimate probability of functioning past a certain time(Kleinbaum and Klein, 2020, p. 62).

i	t_i	n_i	m_i	q_i	$\hat{S}(t_i)$
0	0	20	0	0	1
1	1	20	1	0	$19/20 = 0.95$
2	4	19	3	0	$16/20 = 0.8$
3	6	16	5	0	$11/20 = 0.55$
4	7	11	0	0	...
...	...	...	...	...	...

The estimated probability of functioning past a certain time, $\hat{S}(t_i)$ can be obtained by looking at the number of objects functioning at the end of a given time period and divide it by the total number of objects included in the study according to equation (2.5).

$$\hat{S}(t_i) = \frac{n_{i+1}}{n_0} \tag{2.5}$$

When estimating the survival function at different time periods for data that include censored objects an alternative method is needed. This method is called the Kaplan-Meier(KM) method. The advantage of this method in describing the data in relation to typical descriptive measures such as average expected lifetime and median lifetime, is that the survival of groups can be compared at different points in time(Kleinbaum and Klein, 2020, p. 61-63).

The Kaplan-Meier method for estimating the survival function is illustrated with Table 2.2, constructed similarly to Table 2.1 but including censored objects at each point in time.

Table 2.2: Example table of how to estimate probability of functioning past a certain time using the Kaplan-Meier method.

i	t_i	n_i	m_i	q_i	$\hat{S}(t_i)$
0	0	20	0	0	1
1	1	20	1	1	$19/20 = 0.95$
2	4	18	3	3	$0.95 \cdot \frac{15}{18} = 0.79$
3	6	12	2	2	$0.79 \cdot \frac{10}{12} = 0.66$
4	7	8	1	4	$0.66 \cdot \frac{7}{8} = 0.58$
...	...	...	...	...	...

To obtain the probability of functioning after a certain time, t_i, the following conditional probability and formula is used:

$$\hat{S}(t_i) = P(T > t_i | T \geq t_i) = \hat{S}(t_{i-1}) \cdot \left(\frac{n_i - m_i}{n_i} \right) \tag{2.6}$$

Or the full formula including all the conditional probabilities from $t_i = 0$:

$$\hat{S}(t_n) = \prod_{i=1}^{n} P(T > t_i | T \geq t_i). \tag{2.7}$$

Since the estimated survival function at different points in time is empirically based, the resulting function will often look like steps as one of the curves in Figure 2.1.

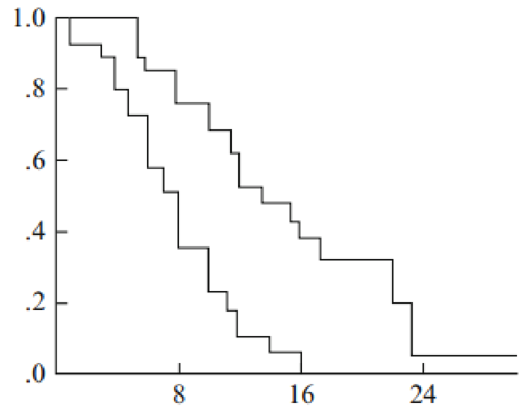

Figure 2.1: An example of two different KM-curves(Kleinbaum and Klein, 2020, p. 67). The y-axis shows the estimated survival probability and the x-axis some given time unit.

2.4 The Log-Rank Test

When one wants to investigate the survival of several different groups the KM-curve of each group can be calculated and plotted in the same figure as in Figure 2.1. To be able to determine if the survival curves of the two different groups are statistically different from each other the log-rank test can be used.

To perform the log-rank test we assume the null hypothesis, H_0, that there is no difference between the survival curves and calculate a test statistic that is chi-square distributed. The log-rank test statistic is calculated according to the following formula:

$$Z_{log-rank} = \frac{\left(O_{1/2} - E_{1/2}\right)^2}{Var\left(O_{1/2} - E_{1/2}\right)} \tag{2.8}$$

where $O_{1/2}$ is the sum of the observed counts of failures over the study period for either the first or the second group. $E_{1/2}$ is the sum of the expected counts of failures over the study period for either one of the groups. It does not matter if one uses the first or the second group in this case but one need to use the same for O and E. The expected failures at time t_i for group 1, are calculated according to equation (2.9) and summed up to get E. The first term in the multiplication refers to the proportion of individuals that is still in the risk set of group 1 at that time, and the second term refers to the sum of the number of actual failures at that time for both groups(Kleinbaum and Klein, 2020, p. 68-70).

$$e_{1t_i} = \left(\frac{n_{1t_i}}{n_{1t_i} + n_{2t_i}} \right) \cdot (m_{1t_i} + m_{2t_i}) \tag{2.9}$$

Since the log-rank test statistic in equation (2.8) is chi-square distributed with one degree of freedom the significance of the test is easily evaluated.

The log-rank test can also be applied to compare the survival curves of more than two groups. In that case the null hypothesis states that all the survival curves are the same.

2.5 Confidence intervals for the KM-curve

Confidence intervals for the estimated probabilities in the KM-curve can be calculated according to equation (2.10) with the help of Greenwood´s formula for the variance according to equation (2.11) (Kleinbaum and Klein, 2020, p. 78).

$$\hat{S}_{KM}(t_i) \pm \lambda_{1-\alpha/2} \sqrt{V\hat{a}r(\hat{S}_{KM}(t))} \tag{2.10}$$

$$V\hat{a}r(\hat{S}_{KM}(t)) = (\hat{S}_{KM}(t))^2 \sum_{i:t_i \leq t} \left(\frac{m_i}{n_i(n_i - m_i)} \right) \tag{2.11}$$

2.6 Cox proportional hazards model

While the KM-curve is good for visualizing the failure process and comparing the overall failure of different groups, it does not provide a way of investigating the effect of a certain factor on the survival probability or hazard rate. To analyse the relationship between explaining factors and time to failure the Cox proportional hazards(PH) model is useful.

2.6.1 Model formulation

The formula for the Cox PH model can be seen in equation (2.12) where $\mathbf{X} = (X_1, ..., X_p)$ is the explanatory variables. $h_0(t)$ is the baseline hazard that depends on time but not on the explanatory variables and $e^{\sum_{i=1}^{p} \beta_i X_i}$ is the part describing the time-independent effect of the explanatory variables.

$$h(t, \mathbf{X}) = h_0(t) e^{\sum_{i=1}^{p} \beta_i X_i} \tag{2.12}$$

If all the explanatory variables are equal to zero the model reduces to the baseline hazard function, $h_0(t)$. The baseline hazard function is unspecified making it a semiparametric model. It is a robust model in the sense that the results of it will closely resemble the results of the true parametric model(Kleinbaum and Klein, 2020, p. 108-110).

2.6.2 Estimating the parameters using the Cox likelihood

The method to estimate the parameters in the Cox PH model is called the Cox likelihood since it is a variation of the maximum likelihood estimation method. The formulation of the Cox likelihood is based on the order of the failure times in the data set and not on the exact distribution of the data, since the semi-parametric character of the model does not enable specifying the exact distribution. The expression for the Cox likelihood with k failure times can be formulated according to equation (2.13). L_i is the contribution of the i:th failure time to the Cox likelihood. This is the hazard for the i:th object divided by the sum of the hazards of the objects still left in the risk set at the time of failure(Kleinbaum and Klein, 2020, p. 127-131). The hazards are calculated according to the hazard function from equation (2.12) in Section 2.6.1. In the expressions for L_i the baseline hazard will be present in each term of the denominator and the numerator and therefore it will cancel. This is fundamental for the estimation of the parameters.

$$L = L_1 \times L_2 \times ... \times L_k = \prod_{i=1}^{k} L_i \tag{2.13}$$

After having formulated the Cox likelihood the parameters in the Cox PH model are estimated as usual by maximizing the log Cox likelihood function with respect to the β-parameters in the model. The solution to the maximization problem is the solution that maximises the probability of getting the order of failure times that was observed from the data.

2.6.3 Interpreting the model by calculating the Hazard Ratio(HR)

Having fitted a Cox PH model the effect of each explanatory variable is illustrated by the estimated β-coefficient related to that variable. The effect of the coefficient on the outcome, the time until event, is often expressed as a hazard ratio (HR). The hazard ratio relates the effect, on the risk of an event occurring, of an object´s exposure variable to an object that does not have the same exposure to that variable. If one has studied logistic regression the interpretation of the hazard ratio is similar to that of the odds ratio. As an example, a hazard ratio of 1 means that the effect of the variable is none. A hazard ratio of 5 means that the effect of the variable is 5 times the hazard of an unexposed object. And a hazard ratio of 1/5 means that the hazard is 1/5 of the unexposed individual.

Given two objects with the set of predictor values $X_1 = x + 1$, $X_2 = x$, the estimated hazard function $\hat{h}(t, X)$ and $\hat{\beta}$-coefficient related to those predictors, the hazard ratio between the two objects can be calculated according to equation (2.14).

$$\hat{HR} = \frac{\hat{h}(t, X_1)}{\hat{h}(t, X_2)} = \frac{\hat{h}_0 e^{\hat{\beta}(x+1)}}{\hat{h}_0 e^{\hat{\beta}x}} = e^{\hat{\beta}} \tag{2.14}$$

The result in equation (2.14) shows the effect of adding 1 unit to the value of predictor x in relation to an object that did not experience this increase. The point of this is to illustrate the effect of a unit increase in the presented Cox PH model. As the number of predictors increase the complexity in the interpretation of predictor changes increases. Since the result in (2.14) is multiplied by the baseline hazard a one unit increase can also be interpreted as increasing or decreasing the baseline hazard with a multiplicative factor of $e^{\hat{\beta}}$.

2.6.4 Adjusted survival curves

Another useful quantity that can be calculated after having fitted the Cox PH model is the adjusted survival curve. When no specific model is fitted to the survival data the survival curve is estimated using the KM-method (Kleinbaum and Klein, 2020, p. 120). Using the estimated Cox PH model it is possible to calculate survival curves that adjust for the explanatory variables. To obtain the estimated survival curves the formula in equation (2.15) with the baseline survival function denoted, $S_0(t)$ is used.

$$S(t, \mathbf{X}) = S_0(t)^{e^{\sum_{i=1}^{p} \beta_i X_i}} \tag{2.15}$$

The parameters in equation (2.15) and $S_0(t)$ are estimated when the Cox PH model is fitted. With this expression a survival probability can be obtained for each value of t.

The adjusted hazard function could of course also be achieved in the same way using the relationship between the two stated in Section 2.2.

2.7 Evaluating the proportional hazards assumption

The proportional hazard assumption states that the difference in hazard between two objects is explained only by multiplying the underlying hazard function with a certain scaling factor(Davidson-Pilon, 2022). This scaling factor is assumed to not vary with time. Three ways of evaluating the proportional hazards assumption are explained by Kleinbaum and Klein(2020, p. 162). These are two graphical approaches, looking at the log-log plots of the survival curves and comparing the observed versus the expected survival curves. The third approach is by performing a goodness of fit test for the relevant predictors. In evaluating the PH assumption the goodness of fit test was used to since it gives a more objective result than the results of the graphical approaches(Kleinbaum and Klein, 2020, p. 181).

2.7.1 Goodness of fit

The idea behind the test is to calculate the Schoenfeld residuals for each predictor being tested. The Schoenfeld residuals can be calculated for each predictor and object that have failed. It is calculated for each failure time and object by taking the observed predictor value minus the weigthed average predictor value for all the other objects still at risk at the failure time in question(Kleinbaum and Klein, 2020, p. 181). The correlation between the Schoenfeld residuals and the order of the failure times is then evaluated. They should be unrelated if the PH assumption is satisfied.

The null hypothesis is that the predictors satisfy the PH assumption and the alternative hypothesis that they do not. When performing the test one predictor at a time is tested given that the other predictors satisfy the PH assumption. In the analysis part of the report the *proportional_hazard_test* provided by the Lifelines Python library (Davidson-Pilon. 2022) was used to evaluate the PH assumption.

2.8 The extended Cox model

If the proportional hazards assumption is not satisfied the Cox model could be extended to include time-varying variables. Another reason to use the extended Cox model is that the variables measured actually is varying with time and the extended Cox model is able to handle this.

2.8.1 Time-varying variables

Three types of time-varying variables can be considered: defined time-variables, internal variables and external variables(Kleinbaum and Klein, 2020, p. 247). Defined time-variables are pre-defined, for example the temperature multiplied by time. "Internal" variables change value because of internal characteristics of the subject of study. This could for example be the temperature measured in the engine, which is affected by internal processes in the engine. "External" variables change value because of external characteristics in the environment. An example of this could be the temperature in the testbed which changes because of many factors, the temperature outside, or the energy emitted from the test object in the testbed.

2.8.2 The counting process format

When looking at time-varying variables the data used for fitting the model will be in a different format than for the regular Cox PH model. This format is called counting process(CP) format. For the regular Cox PH model the data is listed with one observation per study object, containing the failure time, object id and the values of the predictors. In the counting process format there are numerous observations in time per study object, where the value of the predictor can be different at different points in time. This is coded by having each row in the data table represent a time interval for a certain study object. The value of the different predictors at that time is provided together with information on whether the object failed or not during the time interval. An example of this can be seen in Table 2.3 below.

Table 2.3: Example of the counting process data format.

object id	start	stop	event	temperature(°C)	engine RPS
1	0	3	0	20	160
1	3	6	0	19	155
1	6	9	0	21	175
1	9	12	1	22	176
2	0	3	0	20	160
...	...	...	...	...	...

2.8.3 Constructing the extended Cox model

Even though there are different ways of defining the time-varying variables the mathematical structure of the extended Cox model will be the same. The type of time-varying variables focused on in this report will be "internal" variables. The resulting extended Cox model will look like in equation (2.16), which is pretty much the same as in the Cox PH model but now the parameters are fitted to all of the data in the CP format.

$$h(t, \mathbf{X(t)}) = h_0(t)e^{\sum_{i=1}^{p} \beta_i X_i(t)} \tag{2.16}$$

2.8.4 The extended Cox likelihood

The Cox likelihood for the extended model is in large parts the same as for the Cox PH model. Now, however the hazard function for specific objects change with time. This will lead to a time-varying contribution to the total hazard for a certain object in the total hazard of the denominator of L_i in equation (2.13) Section 2.6.2.

Chapter 3

Description of data source

The data used in the study comes from the internal Husqvarna Group database where data from tests performed on products from 7 different product categories is registered: chainsaws, clearing saws, power cutters, lawn movers, batteries, accessories and various handheld and wheeled products.

3.1 Registering failures

A test object consists of several components, for which failures are registered. The failures are registered either automatically, or at the end of test after manual examination. When manually registering failures the exact failure time is not known and since the test managed to run until the end time these are treated as right censored observations according to Section 2.1.1. Additionally, some tests are stopped early, even though no failures are registered. These tests also become right censored. The practical reasons for this could for example be the need of running a more important test in the testbed when no other testbeds are available. Another reason could be that sufficient information related to the performance of the object in question have already been obtained. Therefore the test is stopped ahead of time in order to save time.

3.2 Registering predictor data

Examples of registered predictor data during the tests are temperatures on different parts of the test object, temperatures on surrounding equipment and in the environment and revolutions per second(RPS) in the engine. The measure point names are keyed in manually by the operator, which for example led to a temperature measure having 10 different names in one case, even though it referred to the same predictor. These types of problems was handled in the data gathering and preparation part.

The tests are run according to a pre-specified test method where the purpose is either emulating real usage(denoted test method 1) or general wear(denoted test method 2). The test time is divided into time periods. For each time period, the test object have gone through a number of different cycle steps. For test method 1 the steps vary between full throttle with or without emulated load and for test method 2 between idling and full throttle. Minimum, maximum and mean values from the measure points are registered from each step in the time periods.

3.3 Data from SMHI

The weather data on relative air humidity, air pressure and air temperature was collected from the Swedish Meteorological and Hydrological Institute(SMHI) as hourly measurements from the data station Gothenburg A. The relative air humidity is given in percentages, air temperature in hPa and the temperature in degrees Celsius(SMHI, 2022).

Chapter 4

Method

4.1 Information gathering about the test environment and methodology

Information about the test methodology, test environment and other testing context have been gathered through regular visits and meetings with people from different functions at Husqvarna. References to personal communication relates to these events. The point of these meetings have been to understand the source of the data, interpret the behaviour of it and help with filtering of relevant data for the analysis.

Regular meetings, twice a week, have also been held with Lars Walfridsson, supervisor for the thesis from Husqvarna. Through these meetings support in navigating the database and getting in contact with with relevant people have been provided.

4.2 Software programs

All Husqvarna data were extracted from the external Husqvarna database using Microsoft SQL Server Managment Studio(SSMS) and SQL queries. For the statistical analysis and data cleaning Python programming language was used with the integrated development environment(IDE) Spyder which has a suitable interface for scientific computing. The Python *Pandas* library was used which enables statistical analysis of large amounts of data. The *Lifelines* library(Davidson-Pilon, 2022) contains the implemented statistical models needed for the analysis.

4.3 Data gathering and preparation of the chainsaw data

As mentioned at the end of Section 1.3 as many failures as possible together with possible predictors was desirable. The data set also needed to contain tests that were comparable with each other with respect to a number of aspects. The following aspects were considered when filtering out a final data set for the survival analysis:

1. Product category: The performance of objects from different product categories cannot be directly compared with each other because of their differences. Data from the product category 'chainsaws' was chosen since it contained the largest amount of failure data.

2. Test method: Tests run with different test methods expose the objects to different loads and wear. Therefore, the data was filtered on the type of test method. The test methods analysed were test method 1 and 2.

3. Test program: The test programs are run slightly different depending on the intended use of the products. There were for example 3 different test programs found in the data from test method 1. The difference of the actual products from these tests are the strength of the engine and the bar length on the chainsaw.

4. Product model: Husqvarna produces a large number of different chainsaw models that are intended for different usage. This needs to be taken into consideration when comparing the performance of models in different tests.

5. Predictor names: When inspecting the raw data not all tests contained measurements from all measure points. As described in Section 3.2 there were many different measure point names and to enable analysis without losing too much data, the data set was filtered on test objects that had measurements on certain predictors. This was done by keeping data points of predictors that had more than 100 000 occurrences in the initial raw data.

6. Other disqualified tests: Sometimes tests are run with very specific purposes such as testing components in the testbed or single components in a product. Such test are often very short, not complete with respect to measure points and are part of multiple tests on the

same test objects. Because of these reasons they were discarded in the analysis.

7. Test cycle step: The available data for each test cycle step was investigated. Unfortunately not all test cycle steps contained the same amount of data from the measure points. Therefore the data was filtered on the cycle steps that had the most data points registered. This was the load cycle step for the data with test method 1 and the full throttle cycle step for the data with test method 2.

8. Measurement type: Of the three measurement types that are registered during the tests not all of them contained the same amount of data points, for the same predictors. This means, that if for example the max values from the load cycle at every time period were analysed with a certain set of predictors, the same set of predictors and tests did not also have min values registered. Therefore a new data set was needed to analyse min or mean values instead. Because of this, the focus of the analysis had to be on the measurement type with the largest amount of data points for the relevant chosen predictors, this was the max values for both test methods.

4.3.1 Descriptive analysis of the data

After a first filtering of the data had been made, it was visualized in different ways to get a better overview of the underlying process. The visualization was also important to recognize problematic data such as single outlier values or outlier tests in the analysis.

Plots of the data against time, histograms and KM-curves(see Section 2.3) were used for initial visualization of the data.

4.3.2 Preparation of data for the mean based Cox PH model

In some cases there were multiple failures registered for a test, since failed components are sometimes changed for new ones to be able to finish the test. For the analysis the first failure time was used to define the failure of a test.

For the Cox PH model the data needed to be in the format of one observation per test and therefore the mean values of the max measurements from the cycle steps were calculated for the whole test time. Before this, outlier predictor values were removed, which is discussed more in detail in Section 5.1.4.

4.3.3 Preparation of data for the extended Cox model

To apply the extended cox model with time-varying variables the data needed to be transformed into the counting process format. The details of this can be seen in Section 2.8.2, *The counting process format*.

4.4 Data gathering and preparation of the power cutter component data

The data gathering on cracks in power cutter components was simple since all non failed components in power cutters could be collected together with all failures due to cracks. The date that the test was set up was collected for each component to be able to match it with the right weather conditions that month.

Hourly data from SMHI(2022) was collected from the weather station closest to Jonsered outside Gothenburg. Mean values were then calculated for each month for each predictor and resulted in one observation per failure time. The data from SMHI was in high enough quality and therefore no further preparation was needed.

Chapter 5

Analysis and results

5.1 Tests run with test method 1

Having prepared the data according to the steps described in Section 4.3, the data set for test method 1 on chainsaws is summarized as follows:

- 146 objects
- 59 failures
- Predictors:
    1. RPS
    2. temperature 1 (surrounding environment)
    3. temperature 2 (on test object)
    4. temperature 3 (surrounding equipment)
    5. temperature 4 (surrounding equipment)
    6. Standard deviation of the RPS
    7. Standard deviation of temp. 1
    8. Standard deviation of temp. 2
- Test cycle step: 'load'
- Measurement type: Max-values
- 5 main model groups: denoted 1-5
- 3 different test programs: denoted 1-3

5.1.1 KM-estimator and descriptive measures

First of all, the survival function was estimated using the KM estimator as in Figure 5.1 with the estimated median survival time of 90.7 time periods. The median was obtained using the Kaplan-Meier estimator and is in effect given by the relationship $S(t) = 0.5$.

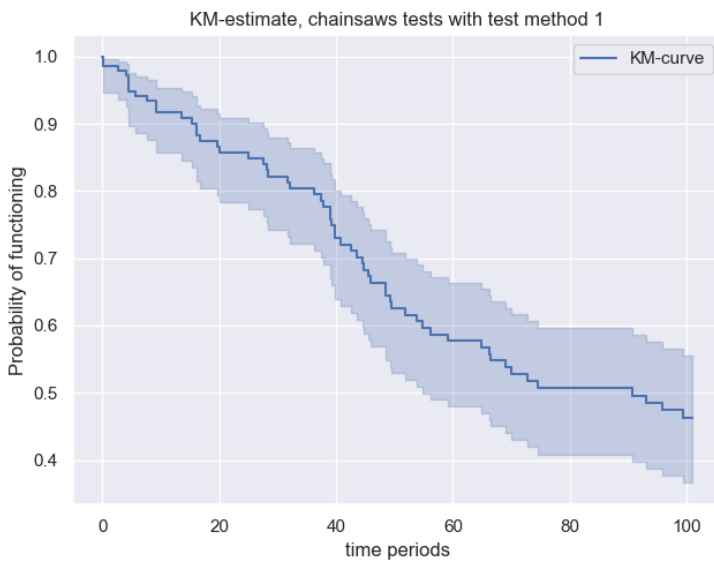

Figure 5.1: KM-curve for all chainsaw tests with test method 1.

Looking at the KM curve for all the tests it can be seen that about 46% of the chainsaws are functioning at the end of the test.

5.1.2 Comparing the survival of different groups

As mentioned in Section 5.1 the chainsaws can be roughly divided into 5 different groups based on the model type. A few chainsaws did not fit in into these groups and did not themselves create a group big enough to be meaningful as a comparison. They were still judged to be similar enough to the rest of the tests to be included in the study. Therefore they were only included when looking at the category 'All'.

One way of comparing the survival of the different groups is by looking at the estimated median survival time. This however, does not give a way of comparing the survival at different points in time. Another way of investigating if there are any differences between the groups is by comparing their estimated survival functions. This was done by estimating the survival function separately for the different groups using the KM-estimator and testing it against the survival function of the objects not included in that group. For this the log-rank test was used.

There were 5 log-rank tests used to test the survival of each group against the rest. A summary of some basic statistics from this analysis can be seen in Table 5.1. A comparison of model group 3 against the rest and model group 5 against the rest is shown in the plots below. These were the groups where the difference in survival was the most evident.

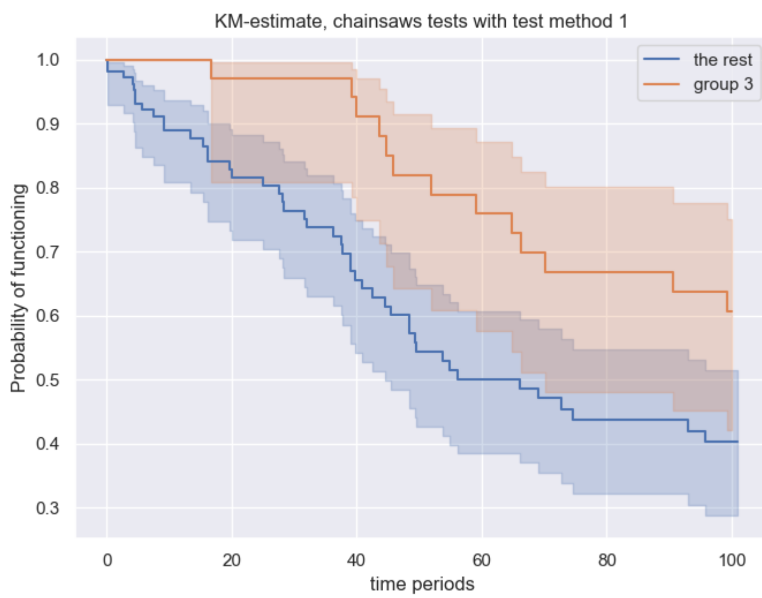

Figure 5.2: KM-curve for model group 3 against the rest.

From Figure 5.2 it seems like the survival of group 3 is better over time than for the rest of the objects. This is also confirmed on a significance level of 0.05 by the p-value of 0.02 when testing the significance between the two groups.

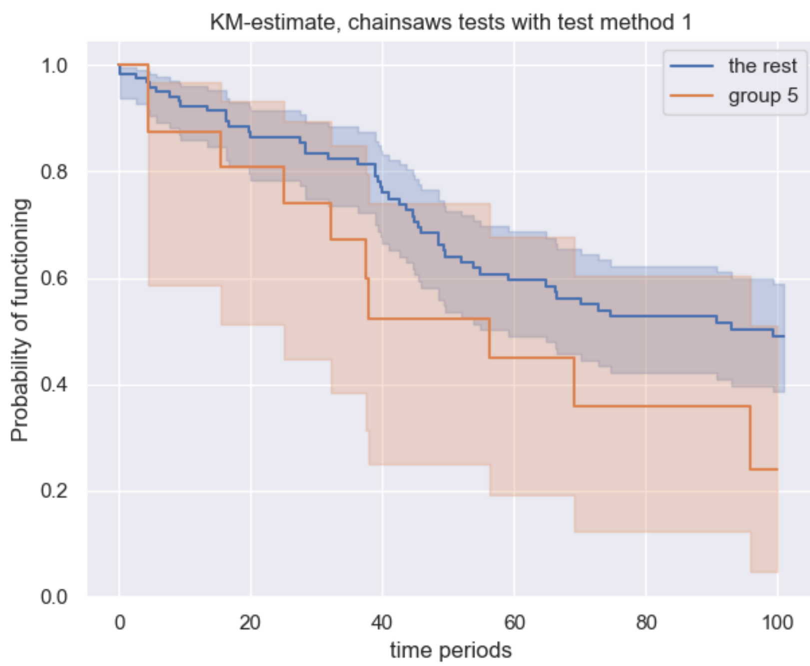

Figure 5.3: KM-curve for model group 5 against the rest.

Looking at Figure 5.3, at first it seems like group 5 performs worse than the rest of the objects since the curve is constantly lower. However, the confidence intervals are overlapping all the time, and the p-value of 0.08 shows that it cannot be concluded that the curves are different from each other.

Furthermore, the log-rank test for the other groups, 1, 2 and 4 did not imply any significant difference in survival as can be seen from the p-values in Table 5.1.

Table 5.1: Summarizing statistics on the different groups and the total population.

Group	n	events	Median survival	p-value(log-rank test)
All	146	59	90.7	-
1	12	3	40.8	0.92
2	42	13	66.2	0.35
3	35	13	-	0.02
4	23	11	-	0.96
5	20	10	56.2	0.08

From Figure 5.1 it can be seen that the median is not stated for group 3 and 4. The reason for this is that there was no data point for $S(t) = 0.5$, when half of the risk set had failed. Therefore the median could not be calculated.

5.1.3 Predictors against time

As a first step in inspecting the predictor data some of the data was plotted against time. Since the data was in time series format and there were about 146 test objects with a total of over 144 000 data points in the data set used for the analysis not all time series data is shown below. An excerpt from the data set of 20 tests that failed are shown for the RPS(Figure 5.4), temperature 1(Figure 5.6), 2(Figure 5.7), 3(Figure 5.8) and 4(Figure 5.9). Each different colour represent the time series for a specific test and the same 20 tests that failed are shown for all predictors.

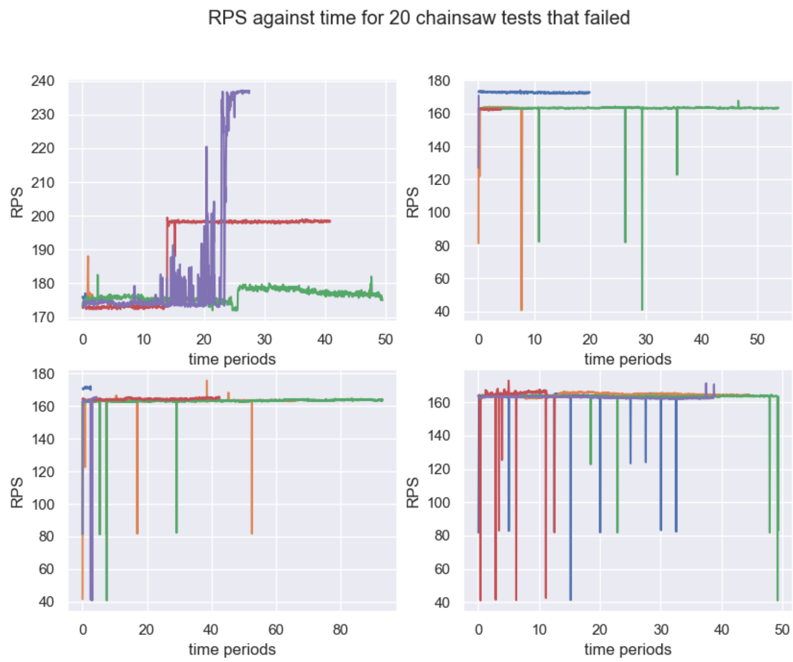

Figure 5.4: An excerpt of the data visualizing the RPS against time for the test from 20 different test objects that failed.

It can be seen that at least one test shows shows erratic behaviour when approaching failure, this is the purple one in the first subplot. This type of behaviour was rather an exception than a rule as can be seen from the picture, but also when looking at other observations not present here. For the most cases the RPS was close to constant over time with short dips created by test stoppages causing the RPS to fall for short time periods. Because of the stoppages the RPS fell to 0 in some sporadic cases.

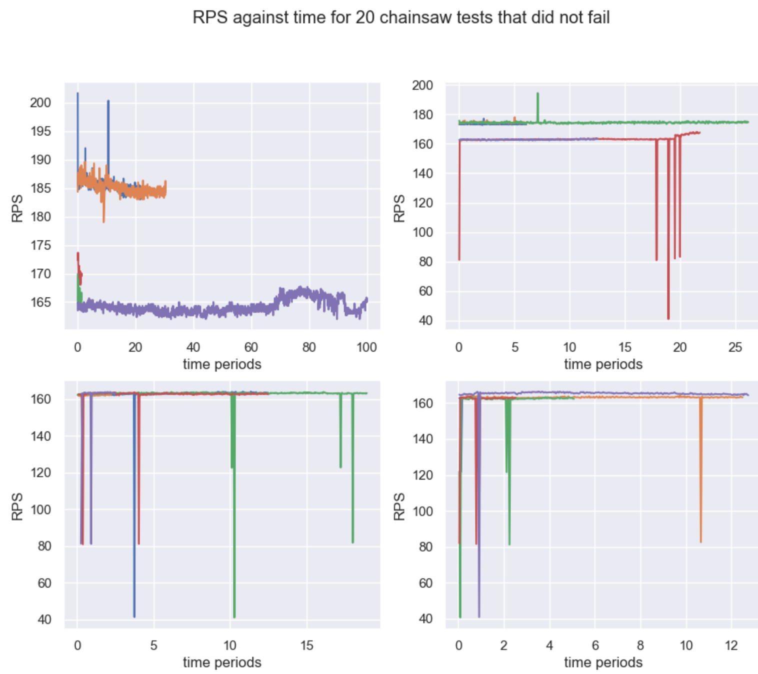

Figure 5.5: An excerpt of the data visualizing the RPS against time for the test from 20 different test objects that did not fail. All but one are censored.

As a contrast to Figure 5.4, Figure 5.5 shows some time series for tests performed that did not fail, most of which were censored. It can be seen that just by inspecting the time series and comparing the figures there were no obvious systematic pattern in the graphs that could tell if an object was going to fail or not.

Furthermore, some plots are shown for time series data on temperature 1 and 2.

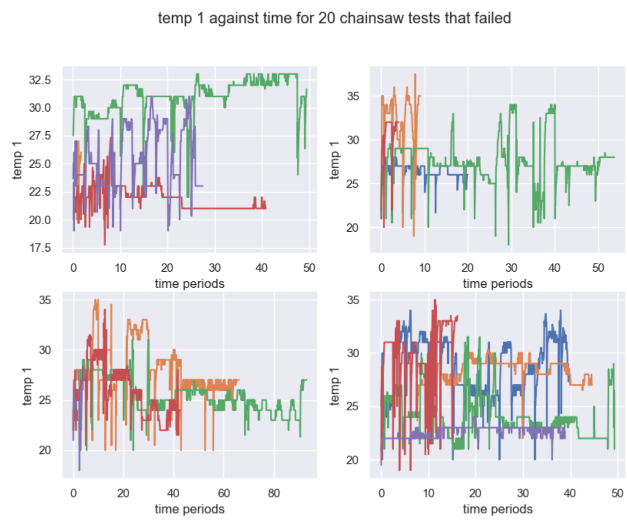

Figure 5.6: An excerpt of the data visualizing temperature 1 against time for test from 20 different test objects that failed.

As can be seen in Figure 5.6 temperature 1 changes in a different way than the RPS over time and there is more of a variation. The temperature 1 is largely affected by the temperature outside which explains varying values. It is also affected by the heat the chainsaw emits to its environment when running.

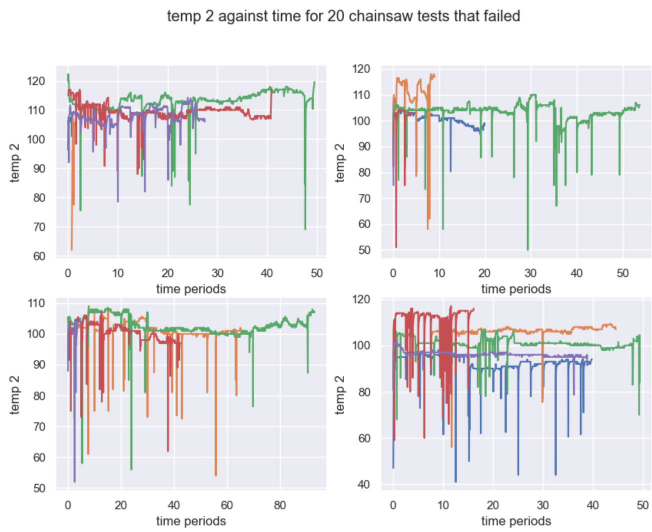

Figure 5.7: An excerpt of the data visualizing temperature 2 against time for test from 20 different test objects that failed.

Temperature 2 seen in Figure 5.7 behaves similar to the RPS over time but here the stoppages have a larger impact since the temperature changes slower than the RPS.

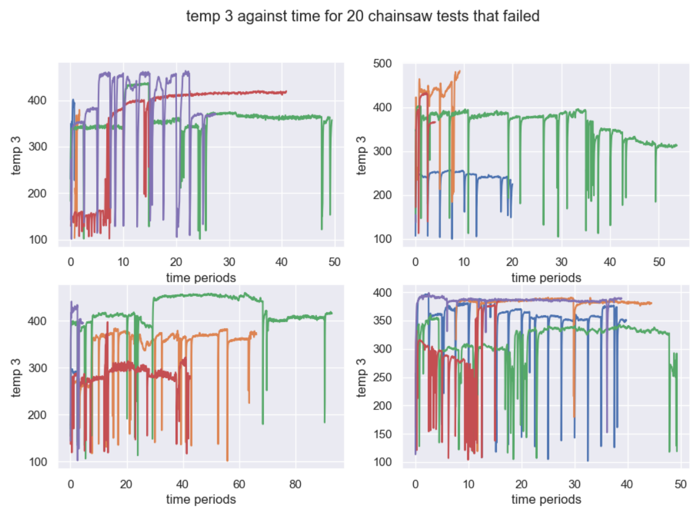

Figure 5.8: An excerpt of the data visualizing temperature 3 against time from 20 different tests that failed.

Temperature 3 can be seen having similar structure to temperature 2 in the way that it also have many short dips due to stoppages. Intuitively it is not clear how this predictor should be related to failures since it is a measurement that comes from surrounding equipment. However, the data on this predictor was extensive and was therefore included in the analysis.

The same reasoning regarding relevance and availability of the data was done for the temperature 4 data seen in Figure 5.9 below.

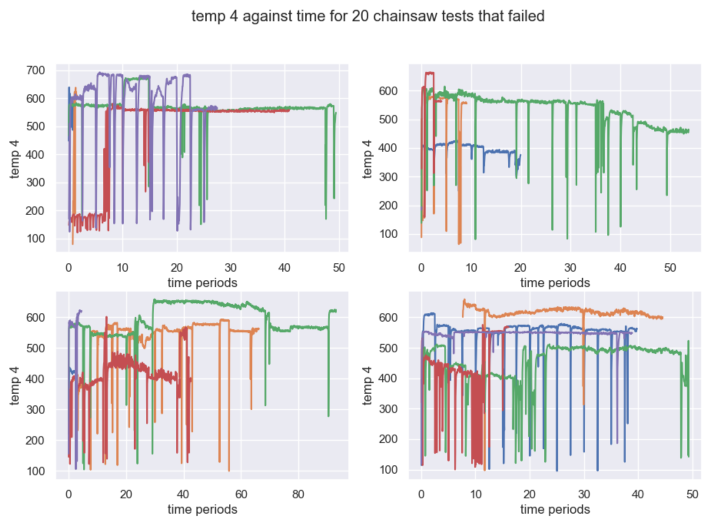

Figure 5.9: An excerpt of the data visualizing temperature 4 against time for test from 20 different test objects that failed.

Just by visual examination of the curves, temperature 3 and 4 seem to follow each other during the duration of the tests. This imply a positive correlation between the predictors and would need to be taken into consideration in the analysis.

5.1.4 Handling outliers

As shown in the figures above there are quite a lot of sporadic dips in all of the predictor data because of the test stoppages. This would certainly affect the value of the estimated parameters in the following model and it can be discussed whether these should be included or not in the analysis.

What for sure should not be included in the analysis and would confound the results are cases where whole series of measure points are wrong. Some such cases were found where, for example, temperature 2 suddenly drops to around 20-25 degrees Celsius and stays there for the rest of the test. The reason for such an error could be due to the measuring equipment coming loose and therefore leading to the temperature in the surrounding environment being registered instead. Such problems were handled in different ways depending on the model used for analysis.

For the mean based Cox PH model single outlier values were not expected to have that big of an impact on the mean of the whole series. Instead observations with unreasonable mean values were removed. Examples of this are mean values of temperature 2 below 40 degrees, which means that someone probably have mixed up this measure point with temperature 1 when entering the name of the measure point. Other outliers could be unreasonably high values of any predictor value or mean values of zero.

For the extended Cox model there were no calculated mean values and instead single data points were removed if unreasonably high or low indicating faulty measurements.

5.1.5 Pairplot of mean values

As a first approach to modeling the relationship between the predictors and the failure times the mean values over the time series were calculated. This was done for several reasons. First of all I wanted to start as simple as possible in modeling the relationship, in order to not miss obvious relationships. Secondly, by inspection of the plots in the previous section and the rest of the tests many time series were relatively stable. They were just on slightly different levels with small spikes or dips that did not seem obviously related to the failures. The idea was that the mean values would capture the differences over time for the predictor values of the whole test implying a different strain on the object. If two objects have about the same values of a certain predictor but one is stopped a lot more or have short spikes then this would also affect the calculated mean and in that sense take this strain on the object into consideration.

Before fitting the Cox PH model a visual inspection of the mean values of the predictors and their relation to each other was appropriate. This was done according to Figure 5.10 and 5.11 below. The main reasons for looking at these plots was to discover any obvious correlation between the predictors that would contribute to the analysis.

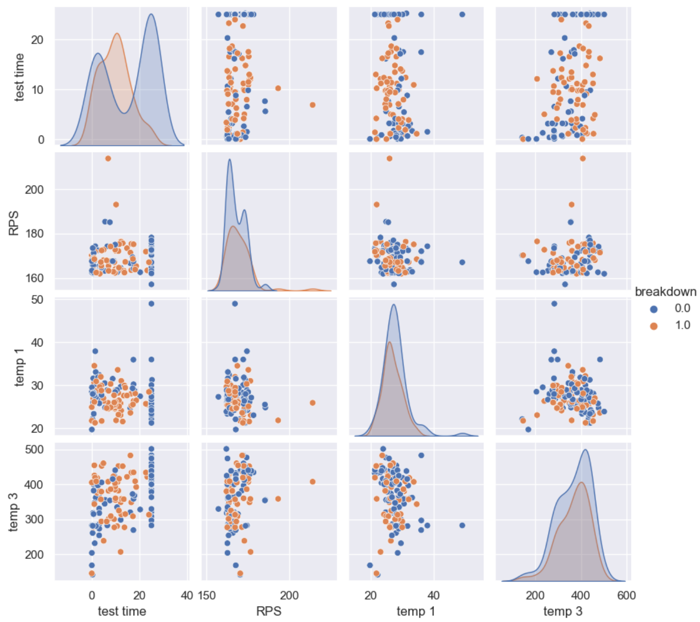

Figure 5.10: Plotting all combinations for three of the predictors against each other. The orange observations represent observations with failures and the blue without.

From Figure 5.10 one could suspect a small correlation between temperature 1 and the temperature 3, but there does not seem to be any strong correlation between the three predictors shown.

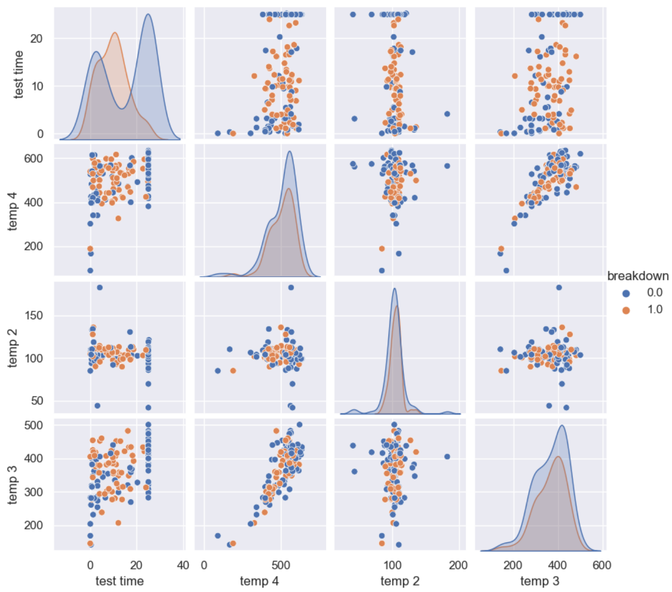

Figure 5.11: Plotting all combinations of the two other predictors together with temperature 3. The orange observations represent observations with failures and the the blue without.

Looking at Figure 5.11 there is a clear positive correlation between temperature 3 and 4. This will be important to keep in mind when looking at what coefficients should be included in the Cox PH model.

5.1.6 Simple mean based Cox PH model

As a start there were measurements for 5 different continuous variables as stated in Section 5.1. The results of the group comparison suggested that group 3 had a different survival from the others, which means that it would be interesting to bring in a categorical variable also to adjust for this difference. Given the nature of the data shown in Section 5.1.3, there was of special interest to incorporate the volatility of the time series in the model. Especially the volatility of the RPS was thought to have an impact on the failure time of a certain test. Therefore the standard deviation of the RPS, temperature 1 and temperature 2 for the whole test were calculated and tested for significance.

When comparing the goodness of fit between different models the main tool used was the partial AIC produced from the Cox PH fitter in the Python *Lifelines* library(Davidson-Pilon, 2022). The 'Partial' AIC refers to the fact that some observations are censored and therefore the partial likelihood is used in the AIC. The significance of the coefficients together with the likelihood-ratio(LR) test were also used in the model selection process.

As a first step, a model was fitted with the 5 directly measured predictors. The coefficients, exponentiated coefficients and their p-values can be seen in Table 5.2. The partial AIC for this model was 510.56.

Table 5.2: Statistical information of the model fitted with the 5 measured predictors, ordered by coefficient values.

Predictor	β_i	e^{β_i}	p-value
RPS	0.03	1.03	0.07
temp. 2	0.02	1.02	0.11
temp. 3	0.00	1.00	0.94
temp. 4	-0.00	1.0	0.15
temp. 1	-0.06	0.94	0.11
group 3	-0.94	0.39	0.03

To get a visual idea of the difference of the predictors Figure 5.12 shows the 95% confidence interval for the estimated coefficients in the first Cox PH model including all measured predictors together with the categorical variable group 3. This figure is useful to look at when one wants to determine the level of significance and the precision in the estimates of a coefficient. However it does not intuitively give an interpretation of the coefficients effect on the survival. For this, it is better to look at the exponentiated coefficients in Table 5.2. There it can be seen that, after adjusting for the other factors, if a test object belongs to group 3 then the hazard is about 61% lower than if it does not.

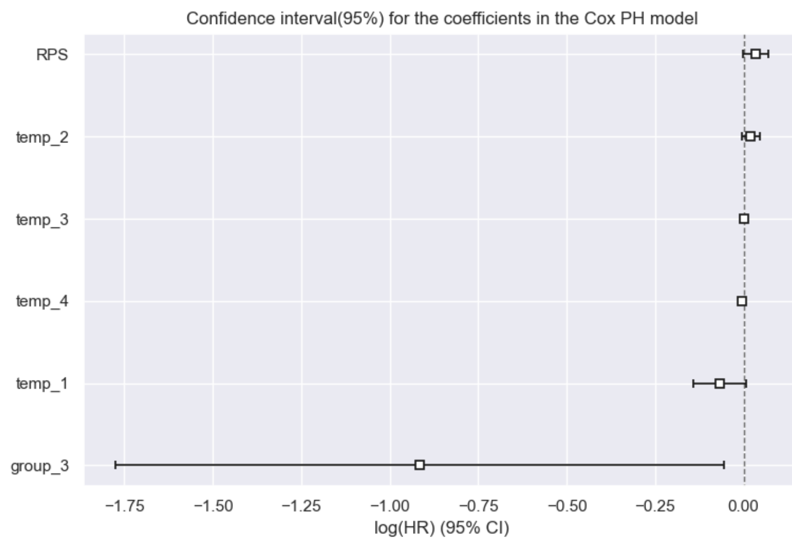

Figure 5.12: The estimated coefficients and their 95% confidence interval.

Figure 5.12 is shown to illustrate the significance of the different predictors when adjusted for the others. The single most important coefficient to determine the time to failure seems to be the categorical variable determining if an object belongs to group 3 or not. However the scale of the effect has great uncertainty in relation to the others, since the interval is so wide. It was of interest to investigate what effect the measured predictors have on the failure times independently of the chainsaw model. Therefore the 'group 3'-coefficients was excluded in further analysis of the data.

When removing the 'group 3'-coefficient the significance of temperature 3 increased and temperature 4 decreased. This together with the strong correlation between those, led to the decision of excluding temperature 4 from the model. Also the RPS coefficient was far from significant and was excluded.

Having a model with the coefficients seen in Table 5.3 with a partial AIC of 506.82, the standard deviations as predictors were included to see if this could improve the model. There was an idea that the volatility of, for example, the RPS could have a wearing effect on the objects in the tests since it puts a certain strain on the engine.

Table 5.3: Statistical information of the model fitted with 3 measured predictors.

Predictor	β_i	e^{β_i}	p-value
temp. 1	-0.07	0.93	0.09
temp. 2	0.03	1.03	0.02
temp. 3	-0.01	0.99	< 0.005

The only standard deviation that significantly improved the model was the one related to the RPS with a partial AIC of 508.76. A LR-test was also performed where a model that included also the standard deviation of temperature 1 was included. This resulted in a test statistic of 1.759 with one degree of freedom, which can be compared to the corresponding quantile form the chi-square distribution. Since $1.759 < 3.841$ it is not significant and the model including the standard deviation of temperature 1 was not preferred. The final model with its estimated coefficients and significance levels can be seen in Figure 5.13 and Table 5.4.

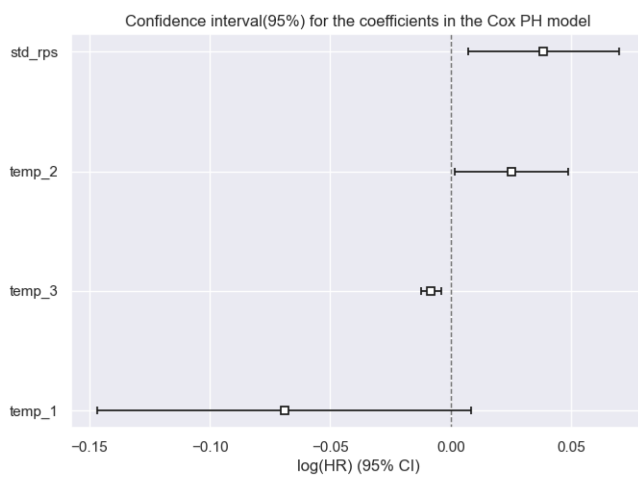

Figure 5.13: The estimated coefficients and their 95% confidence interval for the final model.

Table 5.4: Statistical information of the final model for the data from test method 1.

Predictor	β_i	e^{β_i}	CI(95%) of e^{β_i}	p-value
std. dev. RPS	0.04	1.04	[1.01, 1.07]	0.02
temp. 2	0.02	1.03	[1.00, 1.05]	0.04
temp. 3	-0.01	0.99	[0.99, 1.00]	< 0.005
temp. 1	-0.06	0.94	[0.87, 1.02]	0.12

As a final way of evaluating the models fit to the data and compare it with others there was also a way of predicting the expected survival time in the Python *Lifelines* library(Davidson-Pilon, 2022) given a set of predictors. Predictions was made using the Cox PH model but on the data for which the model was fitted. This was done only on the failure observations and resulted in a RMSE of 37.4. The required time period for a test was 100 periods.

5.1.7 Testing the PH assumption

To determine if it was reasonable to assume proportional hazards for predictors in the Cox PH model a proportional hazard test was performed. None of the predictors violated the proportional hazard assumption and the results of the test can be seen in Table 5.5 below.

Table 5.5: Results from the proportional hazard test for each of the predictors from the final Cox PH model.

Predictor	p-value
std. dev. RPS	0.43
temp. 2	0.61
temp. 3	0.24
temp. 1	0.18

5.1.8 Adjusted survival curves

Having fitted a Cox PH model an adjusted survival curve can be calculated based on the estimated coefficients in the model. To obtain the survival curve the values for the predictors need to be provided. For the adjusted survival curve in Figure 5.14 the mean values of the observed predictors were given.

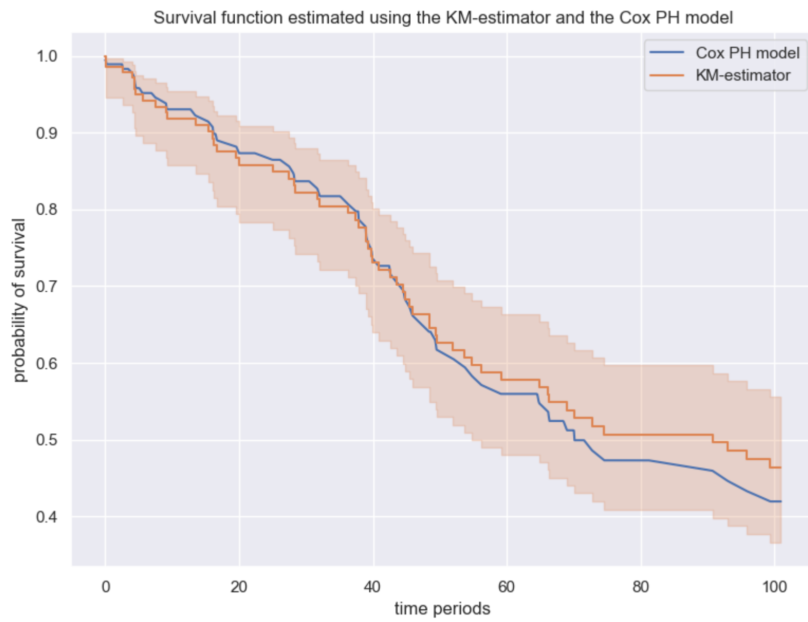

Figure 5.14: The survival curve estimated using the KM-estimator and the Cox PH model.

As a comparison the initial parametric KM-estimate of the survival curve is plotted together with the adjusted survival curve given by the Cox PH model in Figure 5.14. The difference between the models are quite small during the first half of the test time but becomes visible at the second half. Even though a difference is visible it can be concluded that it is quite small when looking at the wide confidence interval of the KM-estimate.

5.1.9 Implications for the survival function

To get a better understanding of the relative risk implied by the Cox PH model one can keep all but one predictor constant and vary the values of the other predictors to see what that implies for the survival over time of an object. In Figure 5.15 and 5.16 the effect on the survival of varying temperature 3 or temperature 2 is illustrated.

By lowering temperature 3 the risk of failure increases and by raising temperature 2 the risk of failure increases. From the figures it seems like the effect of higher temperature 2 or lower temperature 3 is greater at the end of the test. However, this conclusion should not be drawn too fast since the uncertainty is also a lot higher at the end of the test when the amount of observations are fewer. The higher uncertainty is also illustrated by the confidence interval of the KM-estimator in Figure 5.14.

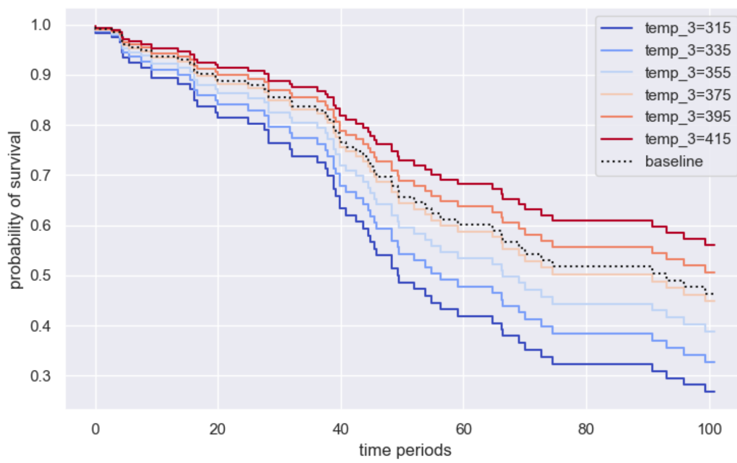

Figure 5.15: The partial effect of varying temperature 3 given that the other predictors do not change.

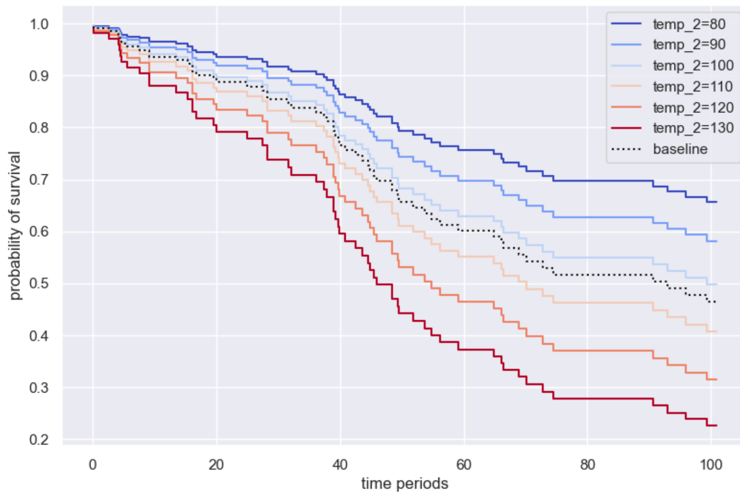

Figure 5.16: The partial effect of varying temperature 2.

5.1.10 The extended Cox model

For the extended Cox model the time-varying character of all of the predictors was examined by putting the data into counting process format as explained in Section 2.8.2. For this data set one observation was lost when removing outlier values. This was a test object with no specifically strange values except for a very early failure. It is judged that this would not make an important difference in the comparison of the models.

The predictors that were included was in large the same as the above with the exception that a rolling standard deviation could now be included. This was done for the RPS, since it was deemed the most relevant.

The first fitted model contained the 5 directly measured predictors from above plus the rolling standard deviation. Different lags for the rolling standard deviation was tested with the most significant based on the previous 5 time periods. The partial AIC was calculated to 475.64.

Table 5.6: The predictor coefficients of the extended Cox model and their significance.

Predictor	β_i	e^{β_i}	p-value
rolling std dev RPS	0.03	1.03	0.14
RPS	0.02	1.03	0.10
temp. 1	0.00	1.00	0.90
temp. 2	0.00	1.00	0.53
temp. 4	-0.00	1.00	0.90
temp. 3	-0.01	0.99	0.01

A visualization of the confidence intervals for the coefficients is once again provided by Figure 5.17.

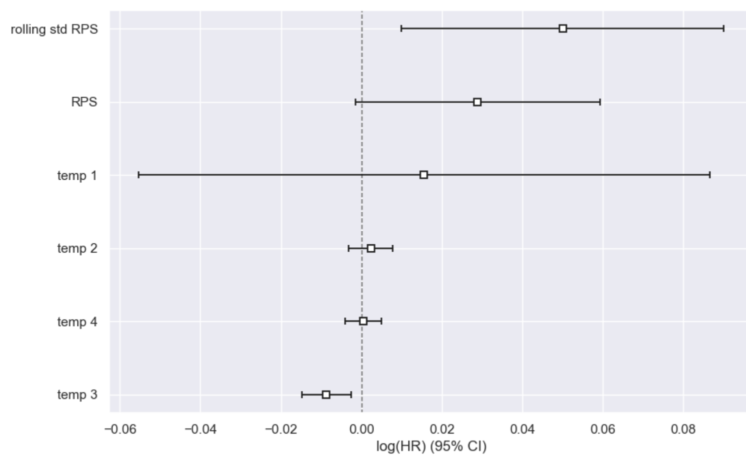

Figure 5.17: Confidence intervals for the coefficients in the extended Cox model with 6 predictors.

Removing temperature 4 and temperature 1 since they were highly insignificant yielded an improvement of the partial AIC to 471.9. However, the coefficient of temperature 2 was still highly insignificant and was therefore also removed. This improved the AIC further to 470.41 and resulted in the best model with only three coefficients as seen in Table 5.7 and Figure 5.18.

Table 5.7: The predictor coefficients of the best extended Cox model and their significance.

Predictor	β_i	e^{β_i}	CI(95%) of e^{β_i}	p-value
RPS	0.03	1.03	[1.00, 1.06]	0.06
rolling std dev RPS	0.03	1.03	[1.01, 1.09]	0.01
temp. 3	-0.01	0.99	[0.99, 1.00]	< 0.005

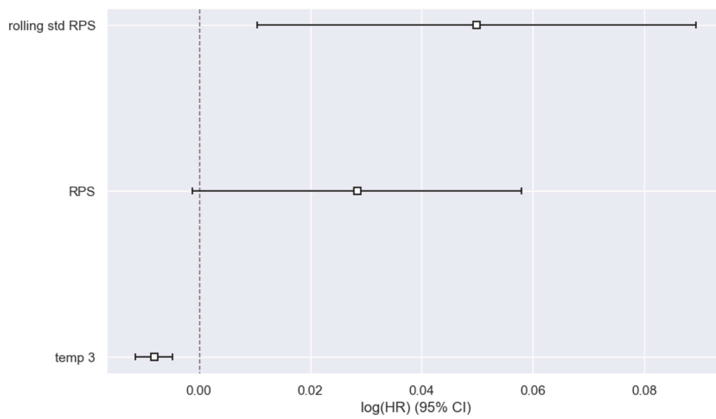

Figure 5.18: Confidence intervals for the coefficients in the best time-varying model.

5.1.11 The partial hazard over time

As way to exploit the advantage of fitting the extended Cox model is that the partial hazard can be calculated for every time step of a test. This could be interpreted as a way of monitoring the strain on the object as an actual value in the calculated hazards. This was performed for 12 tests that failed at the end of the time period. To lower the fluctuations in the presented plots the *log* partial hazard was plotted over time. This can be seen in Figure 5.19 and 5.20.

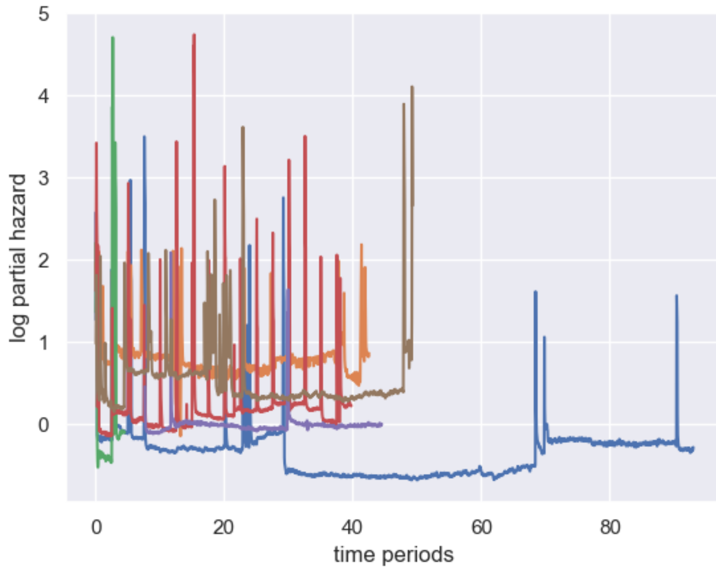

Figure 5.19: The log partial hazard plotted over time for 6 different tests.

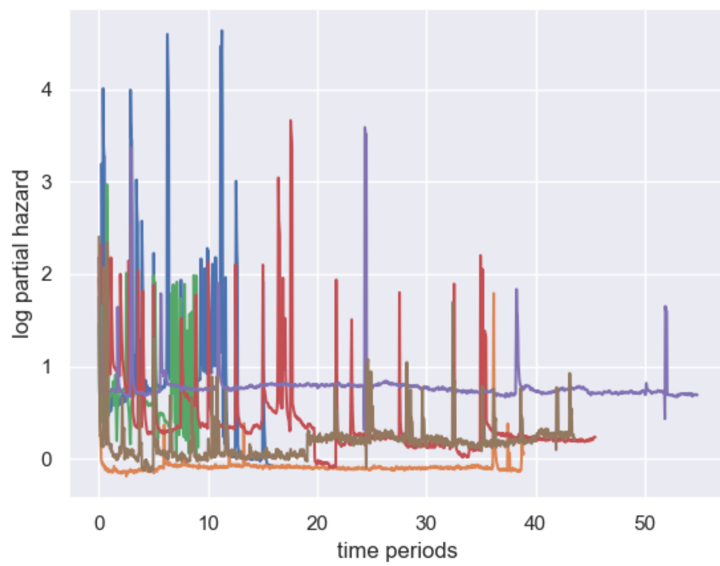

Figure 5.20: The log partial hazard plotted over time for 6 different tests.

5.1.12 Comparing the Cox PH and the extended Cox model

To remind the reader of the formula for the extended Cox model equation (5.1) is provided below. The partial hazard is the parametric part of the hazard function below, in effect the right hand side without the baseline hazard, $h_0(t)$.

$$h(t, \mathbf{X}) = h_0(t)e^{\sum_{i=1}^{p} \beta_i X_i(t)} \tag{5.1}$$

The extended model uses the full time series to estimate the parameters in the Cox model and therefore a partial hazard can be obtained for every time period and object. Since the observations in the Cox PH model was based on mean observations, the mean of the partial hazards from the extended Cox model was calculated as a relevant comparison. The partial hazards produced for the two models was then plotted next to each other against the time in test for the objects that failed. For a model to be a good fit and indication of failure the predicted partial hazards should be high for the objects that failed early and low for the objects failing very late in the tests. The results of plotting the hazards generated from the two different models is visible in Figure 5.21.

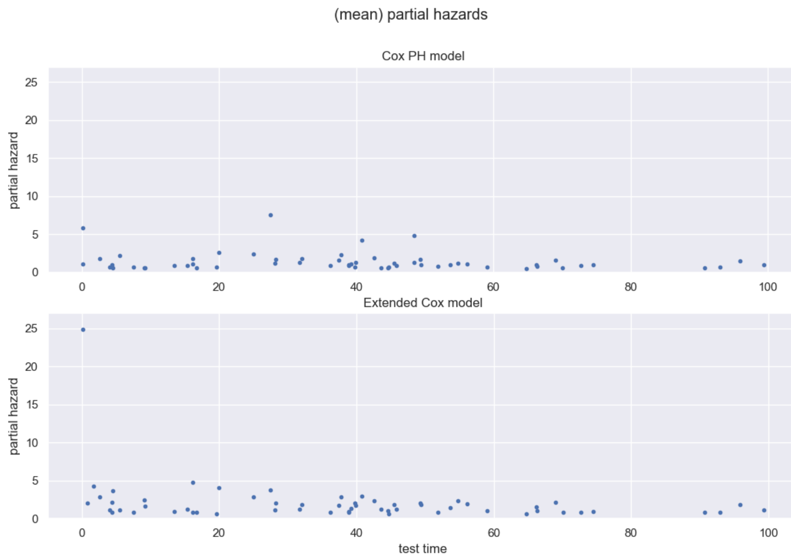

Figure 5.21: The partial hazards plotted for the two models.

From a brief visual inspection the difference in the two models is quite small. It is visible from both of the models that a higher hazard indicates shorter failure times though the relationship is not particularly strong. More data and evaluation is needed to examine which model is a better fit. It is reasonable to think that the model using all of the time series data is giving more robust estimates however the amount of data points used in estimating the parameters are a lot in comparison to the mean based Cox PH model.

5.2 Tests run with test method 2

5.2.1 Summary of the data set

As a summary, the data set for the tests run with test method 2 consisted of:

- 146 observations
- 30 failures
- Possible predictors:
    1. RPS

2. temperature 1
3. temperature 3
4. temperature 4
5. standard deviation of RPS
6. standard deviation of temperature 1

5.2.2 Mean based Cox PH model

In the analysis of the most suitable model first the significance of all of the predictors were evaluated individually. The reason for the different approach in comparison to the last data set was the low amount of failure observations in this data set. It was suspected that bringing in all the coefficients directly would make everything insignificant.

Temperature 3, the RPS and temperature 4 were significant by themselves. However, temperature 4 and 3 was once again strongly correlated and therefore only temperature 3 was included in further analysis.

The model including temperature 3 showed the strongest significance and comparing the partial AIC it resulted in the lowest value. Any other combination led to a higher value and to insignificant coefficients. Therefore it was the best according to the analysis. To compare a model only including the RPS or only including temperature 3 Table 5.8 is provided with relevant information.

Table 5.8: Cox PH model including only a coefficient for the RPS or temperature 3.

Model:	e^{β_i}	CI(95%) of e^{β_i}	p-value	partial AIC	RMSE
only RPS	0.962	[0.927, 0.998]	0.04	267.64	65.7
only temp. 3	0.989	[0.980, 0.999]	0.02	266.31	64.7

5.2.3 The extended Cox model

For the extended Cox model measurements with the same amount of predictors were used, but now data with 115 999 observations were used. This yielded similar results as for the PH model, with the coefficient for temperature 3 being the most significant coefficient. Having only a model with coefficients for this predictor yielded a partial AIC of 267.73. Adding coefficients for the RPS or a rolling standard deviation of the 5 last values for the RPS did not show any significance in the coefficients and worsened the AIC. A coefficient for temperature 1 did improve the AIC slightly to 266.94, though the coefficient was not significant. The information on this model can be seen in Table 5.9 below.

Table 5.9: The predictor coefficients of the extended Cox model for tests with test method 2.

Predictor	β_i	e^{β_i}	CI(95%) of e^{β_i}	p-value
temp. 3	-0.01	0.99	[0.99, 0.99]	< 0.005
temp. 1	0.08	1.08	[0.98, 1.20]	0.10

5.2.4 The partial hazard

In the same way as in Section 5.1.12 the partial hazards of the two models were compared for predictions made on the 30 failures in the data set. The results can be seen in Figure 5.22 below.

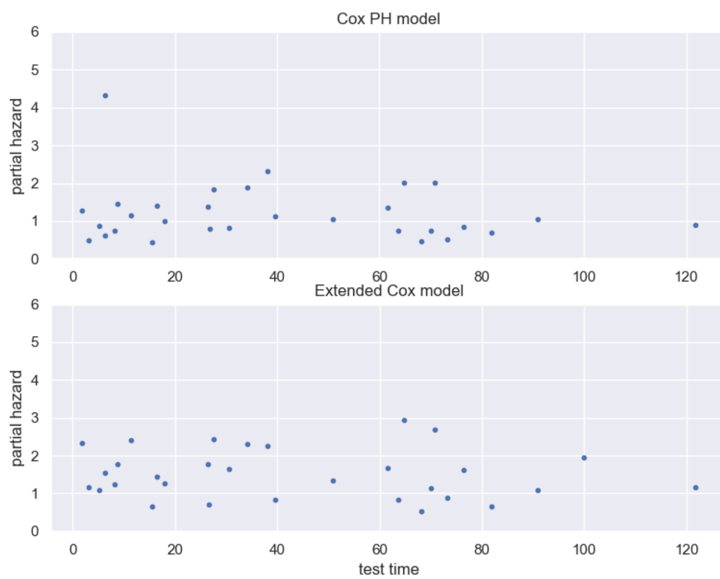

Figure 5.22: The partial hazards plotted for the two models.

From the figure it seems like the Cox PH model implies a slightly better fit because of the higher hazards for lower failure times. The amount of failures in relation to total observations is however quite few. Therefore too strong conclusions from this data set should not be drawn.

5.3 Component failures due to cracks

For the last, a bit more delimited, problem of examining the surrounding weather conditions influence on cracks in components from power cutters the data can be summarized as follows:

- 224 component observations
- 112 failures(due to cracks)
- date of test
- monthly measurements from the SMHI station in Gothenburg:
    1. relative air humidity(%)
    2. air temperature(°C)
    3. air pressure(hPa)

For the model the mean temperature of the month of the test was taken. A Cox PH model was fitted to the data to investigate if the measurements could explain some of the risk related to cracks in the components.

5.3.1 Visualising the data

Starting off, all the predictor data was plotted against the test time and against each other to get a comprehensive view of the data. The observed failures are marked with orange dots as can also be seen in the Figure 5.23 legend. Inspecting the data there were no extreme data points with respect to the different predictors and no extremly clear correlation in the plots shown. A small negative correlation can be suspected between the temperature and the humidity.

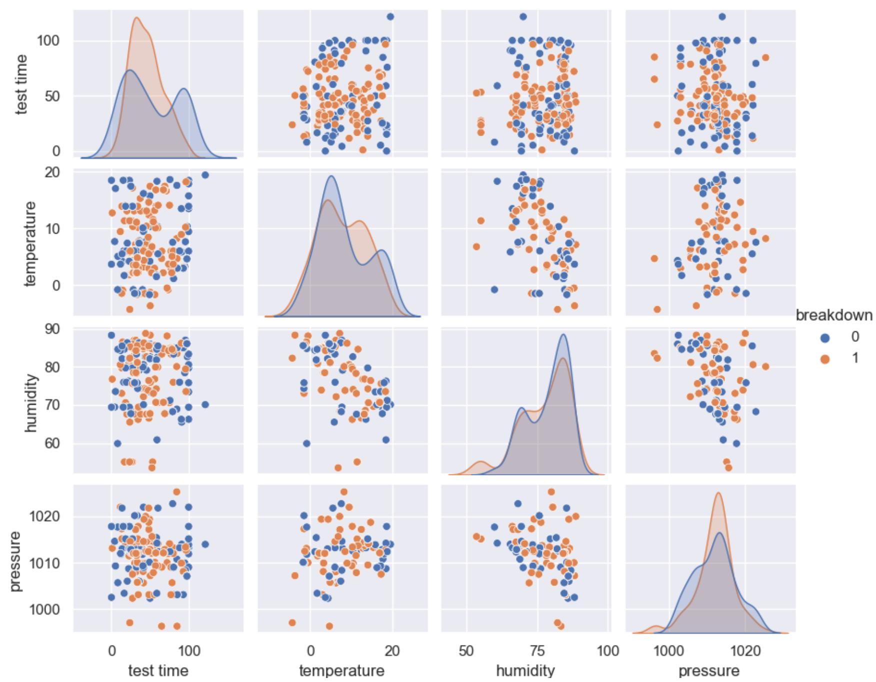

Figure 5.23: The predictor values plotted against each other and against the test time.

Also the KM-estimate was calculated for the dataset. This is seen in Figure 5.24.

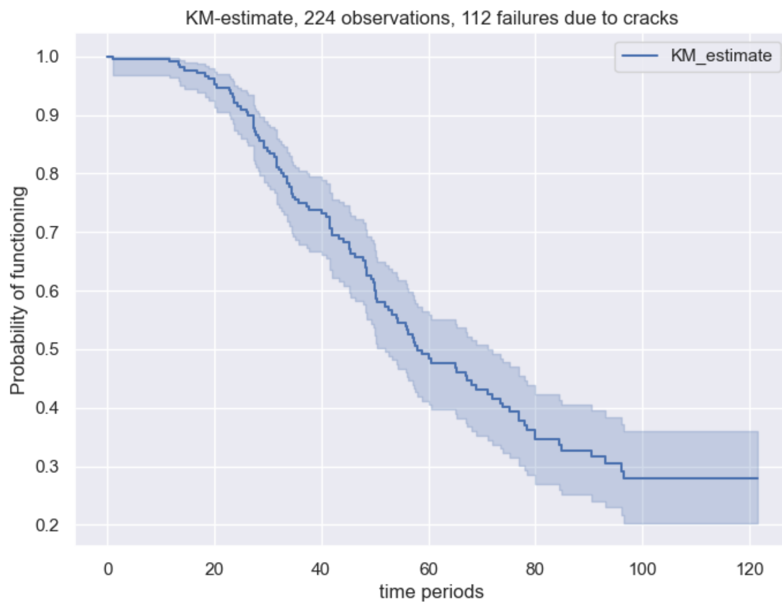

Figure 5.24: Kaplan-Meier estimate for the component failures.

It can be seen that the probability of survival longer than 100 time periods is about 30% when just looking at failures due to cracks in the components.

5.3.2 Model selection

Initially fitting the three weather measurements resulted in confidence intervals(95%) for the coefficients seen in Figure 5.25. From Figure 5.25 is seems like humidity is the only factor to have a significant effect on the probability of failure due to cracks in components. The interpretation here is that higher humidity leads to a lower risk of failure.

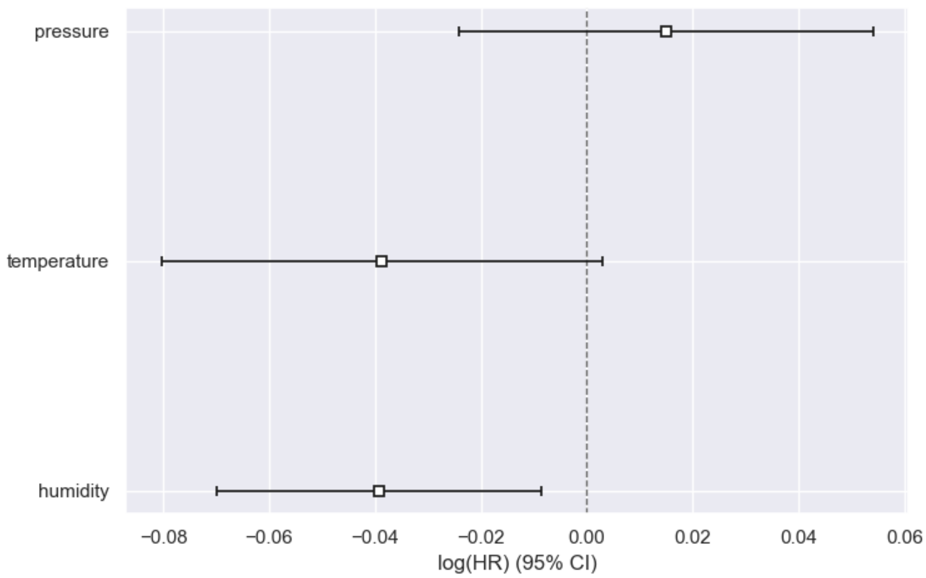

Figure 5.25: Confidence intervals(95%) of the coefficients in the Cox PH model for component failures due to cracks.

A slightly better fit was yielded by removing pressure from the model and the resulting coefficient estimates and their p-values can be seen in Table 5.10.

Table 5.10: The predictor coefficients for the model of the relationship between weather conditions and component failures.

Predictor	β_i	e^{β_i}	CI(95%) of e^{β_i}	p-value
air humidity	-0.04	0.96	[0.93, 0.98]	< 0.005
air temperature	-0.04	0.96	[0.92, 1.00]	0.07

Chapter 6

Discussion and conclusions

6.1 Test method 1

Summarizing the results from the analysis of the data with test method 1, both the mean based model and the extended model indicate that the standard deviation of the RPS and temperature 3 have an effect on the survival time.

Starting with the RPS, there is reason to believe that the volatility of the RPS value during the test affect the survival. The reason for this is that it was only the standard deviation of the RPS in the Cox PH model that showed any significance and not the actual values of the RPS over the test time. Similar results was seen in the extended Cox model where the rolling standard deviation was significant. The interpretation of this could be that a high volatility in the RPS implies instability in the engine and therefore increases the risk of failure.

Moving on to the temperature 3 coefficient, it was significant in both variations of the Cox model. It is not as intuitively clear why this is since the temperature is measured in surrounding equipment. Worth noting is that the value of the estimated coefficient is quite small in relation to the other coefficients in the model. The confidence interval is also small which indicates better precision in the estimate. Therefore it can be said that an increasing temperature 3 lowers the risk of failure even if the effect is quite small in relation to other explaining factors that might exist.

Another aspect to discuss is the effect of the chainsaw model type on the survival time. Section 5.1.2 and Table 5.2 in Section 5.1.6 indicates that the model type is important even though there are large variations within a group that are better explained by the measured factors. It would be interesting to segment the analysis further by just looking at a certain chainsaw model type one at a time. A deeper analysis into this was unfortunately not done due to two reasons. The first being the sharp decrease in observations when segmenting the data set by model type. The second was due to time constraints. The time aspect is explained by the fact that it takes quite a lot of time to extract large amounts of data from the database, clean it and then analyse it. I also wanted to get a chance to look at some data from the other test method.

The biggest reason I wanted to include the log partial hazard over time in Section 5.1.11 was that it could potentially serve as a type of index indicating the strain on the test object. My expectations was that the hazard would increase as the test object approaches failure and therefore one could use the absolute value of the hazard to predict a near failure. By inspection of the plots generated, this did not turn out fruitful. The plots seen in Section 5.1.11 do not imply a more evident pattern towards failure than any of the plots presented for the individual predictors presented in Section 5.1.3.

6.2 Test method 2

This data set was not investigated as thoroughly as the first data set. The reason for this was the much sparser data with only half as many failures, which led to very few failures when dividing the data into smaller groups. Because of this I focused more of the analysis on the first data set and did not dive as deep into this one.

Summarizing the results of the model fitted to the data it is interesting to see that the temperature 3 coefficient proved significant also in this data set. Furthermore the estimated coefficients were almost exactly the same as for test method 1. The difference became clear when fitting only a coefficient with the RPS since the effect of this coefficient turned out to be negative. This means that a higher RPS indicates lower risk of failure. The difference when estimating this coefficient was that there were no adjustments for any other coefficients. When looking at the RMSE of predictions on the 30 failures in the data set, the model with temperature 3 seems to be a slightly better fit.

For the extended Cox model the only difference was an addition of an insignificant coefficient related to temperature 1. As can be seen from the plot in Figure 5.22 with the partial hazards, this difference seem to have very weak practical consequences.

6.3 Component cracks in power cutters

The results from the fitted model partly confirms the thesis of the laboratory engineers working with power cutters. Here the thesis was that dry weather together with cold temperatures would increase the risk of component cracks. From the coefficients of the model including humidity and temperature it can be seen that the lowest risk of failure is when the humidity is high and the temperature is high. However the coefficient for the temperature is not significant and therefore the conclusions regarding the temperature need more investigation.

6.4 Other future methods to explore

Moving on to answer the question of the suitability of the Cox model in describing the relationship between the measured factors and the time until failure. I believe the model is quite suitable in describing the general relationship between the test times and the predictors. A weakness in the model is however that it might not be the most suitable model in handling high resolution time series data. As mentioned in the end of Section 6.1 the hope when starting the thesis was to be able to fit the extended model with many parameters and use it to calculate the partial hazards as a high resolution time series that would resemble a relatively smooth curve. This would then serve as some kind of index to determine how worn down a test object is. Then, for example, a threshold could be set for which the test should be stopped in order to avoid failure. This however did not really work very well and I think one reason for this is that the predictor data available is not very linked to failures. There is no obvious trend that when some of the predictors change value the strain on the object increases significantly. I believe that this could change if other factors are measured in great amounts and with high quality. What factors this should be the technical experts working closely to the chainsaws need to answer.

Other reasons that the idea mentioned above did not work could be that the data was not pre-processed enough and outlier data had to big of an influence on the estimated coefficients.

As a future work I still believe that it would be interesting to look at some time-series model or neural network that weighs the observations differently depending on time. Such a model would be better at capturing erratic behavior close to failure. An example of this could be a Long Short-Term Memory(LSTM) network.

6.5 Data collection challenges

There were some challenges related to the available data that were encountered during the course of the thesis which will be discussed in the following paragraphs.

One problem was inconsistency in what is measured, and incomplete data registration. The first relates mainly to the fact that the test objects that were tested with the same method did not have the same predictors registered. The reason for this can of course be many and one is probably that the test method have changed as the years have passed, which have lead to the data measured looking inconsistent. I also suppose that the test method evolves over time, both in the materials used but also in what cycle steps the chainsaws go through. Ultimately this problem resulted in having to remove large portions of tests with and without failures which in the end probably worsened the estimates of the survival and hazard functions. The incomplete data registrations refers to the fact that not all cycle steps that had max values registered had min and mean values registered. It was the same case for the test cycle steps. As a consequence the combination of cycle step and measurement types with the most registered data points was chosen. It would of course have been interesting to look at the other measurement types but my method and time did not permit me to do it in this thesis.

Another challenge that resulted in loss of useful data was the fact that quite a lot of component failures were registered with the same test time as the required maximum test time, which was 100 time periods, after I had standardized it. Since those failures were registered manually after inspection at the end of test, the times were not reliable. Because the tests were able to run the full time they were considered as having not failed(see Section 3.1). More information that could contribute to the analysis would be available if the real failure times of these were possible to capture.

Finally, another challenge was related to the manually entered predictor names. This creates a headache for the person analysing the data and some valuable information was probably also lost when renaming the predictors and only picking the most common predictor names. The biggest challenge with this is that it is almost impossible to analyse the data before the predictor names have been cleaned and standardized. To mitigate this problem there should only by beforehand determined alternatives to choose as predictor names when setting up the tests.

References

Davidson-Pilon, Cameron. 2022. Python library *lifelines*. https://lifelines.readthedocs.io/en/latest/index.html. (Collected 2022-05-23).

Husqvarna Group AB. 2022. *Husqvarna Groups årsredovisning 2021*. https://www.husqvarnagroup.com/sv/report/husqvarna-groups-arsredovisning-2021. (Collected 2022-05-06).

IMARC. 2020. *Chainsaw Market: Global Industry Trends, Share, Size, Growth, Opportunity and Forecast 2021-2026*. https://www.imarcgroup.com/chainsaw-market. (Collected 2022-05-06).

Kent, Brian. The Crosstab Kite. 2021. *Applications of survival analysis (that aren't clinical research)*. https://www.crosstab.io/articles/survival-analysis-applications. (Collected 2022-05-07)

Kleinbaum, David G; Klein, Mitchel. 2020. *Survival analysis: A Self-Learning Text*. Third edition. New York: Springer.

Precedence Research. 2022. *Chainsaw Market Size To Reach Around US$ 4.92 Bn By 2030*. https://www.precedenceresearch.com/chainsaw-market. (Collected 2022-05-06)

Sveriges meteorologiska och hydrologiska institut(SMHI). 2022. *Ladda ner meteorologiska observationer*. https://www.smhi.se/data/meteorologi/ladda-ner-meteorologiska-observationer #param=airtemperatureInstant,stations=all,stationid=71420. (Collected 2022-03-09)

Master's Theses in Mathematical Sciences 2022:E48
ISSN 1404-6342

LUTFMS-3450-2022

Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lu.se/