

# Accelerated testing with application in finance

by

Anel Oppel

Submitted in partial fulfilment of the requirements for the degree  
Master of Science (Mathematical statistics)  
in the Faculty of Natural and Agricultural sciences  
University of Pretoria, Pretoria

November 2016

**Publication data:**

Anel Opper. Accelerated testing with application in finance. Master's Mini-dissertation, University of Pretoria, Department of Statistics, Pretoria, South Africa, November 2016.

Electronic, hyperlinked versions of this dissertation are available online, as Adobe PDF files, at:

<http://upetd.up.ac.za/UPeTD.htm>

## Accelerated testing with application in finance

by

Anel Opper

E-mail: [anel.opper@gmail.com](mailto:anel.opper@gmail.com)

### Abstract

The event of a default for low-default portfolios, such as sovereign debt or banks, have received much attention as a result of the increasing instabilities in financial markets. The lack of sufficient default information on low-default portfolios complicates the protection of such portfolios. Default protections have typically, in the past, relied on extreme value theory and reporting the value at risk. The focus here, is the application of an engineering concept, accelerated test techniques, to the problem of insufficient data on low-default portfolios. In the application, high-default portfolios serve as stressed cases of low-default portfolios. Since high-default portfolios have more data available, viewing it as a stressed case of a low-default portfolio enables us to extrapolate the data to the low-default portfolio environment, and do estimation such as estimating the default probability for a low-default portfolio. The flexible framework through which the above is achieved, is provided.

**Keywords:** Extreme value theory, Accelerated test, Binary data, Life-stress link function, Market capitalization, one-shot device testing, single stress factor.

**Supervisors** : M. T. Loots

Dr. F. J. C. Beyers

**Institution** : University of Pretoria Department of Statistics

**Degree** : Master of Science

## Declaration

I, *Anel Oppel*, declare that this mini-dissertation, which I hereby submit for the degree *Master of Science (Mathematical statistics)* at the University of Pretoria, is my own work and has not been previously submitted by me for a degree at this or any other tertiary institution.



---

*Anel Oppel*

25/11/2016

## Acknowledgements

I would like to thank the following persons and entities who's support has been instrumental in the completion of the mini-dissertation (and my continued survival through it's completion):

- My supervisors Mr. M. T. Loots and Dr. F. J. C. Beyers for their guidance and support, numerous lively discussions, frequent comments and suggestions which made this work possible
- The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.
- The financial assistance of the Barclays Africa chair in Actuarial science towards this research is hereby acknowledged.
- My husband Mr. T. W. Leenen, my family, friends and fellow students.

# Contents

<b>List of Figures</b>	<b>iii</b>
<b>List of Graphs</b>	<b>iv</b>
<b>List of Tables</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
1.1 A review of accelerated tests: Basic concepts . . . . .	5
1.1.1 Life distribution . . . . .	7
1.1.2 Life-stress relationships . . . . .	8
1.1.3 Parameter estimation methods . . . . .	10
1.1.4 Reliability information . . . . .	10
1.1.5 Censored data . . . . .	11
1.1.6 Introductory examples . . . . .	17
1.1.7 Additional considerations . . . . .	27
1.2 Objectives of the study . . . . .	28
1.3 Contributions . . . . .	28
1.4 Dissertation outline . . . . .	28
<b>2 Exploring the EM algorithm</b>	<b>30</b>
2.1 Introduction to the EM algorithm . . . . .	30
2.2 The EM algorithm for Log-normal, left- and right censored data . . . . .	41
2.3 Summary . . . . .	45

<b>3</b>	<b>Simulation exploration</b>	<b>46</b>
3.1	The log-returns of a company . . . . .	46
3.2	Simulation results . . . . .	55
<b>4</b>	<b>Concluding comments</b>	<b>73</b>
4.1	Summary . . . . .	73
4.2	Future work . . . . .	74
	<b>Bibliography</b>	<b>75</b>
<b>A</b>	<b>The coding Appendix for Chapter 2</b>	<b>80</b>
A.1	SAS code for Example 3 from Chapter 2 . . . . .	80
A.2	SAS code for Example 4 from Chapter 2 . . . . .	82
<b>B</b>	<b>The coding Appendix for Chapter 3</b>	<b>85</b>
B.1	SAS code for Section 3.1 from Chapter 3 . . . . .	85
B.2	SAS code for Section 3.2 from Chapter 3 . . . . .	96
B.2.1	This code generates Figure 3.5 . . . . .	96
B.2.2	This code generates Figure 3.6 . . . . .	101
B.2.3	This code generates Table 3.16–3.18 . . . . .	104
B.2.4	This code generates Figure 3.7 . . . . .	107
B.2.5	This code generates Figure 3.11– 3.14 . . . . .	110
<b>C</b>	<b>Acronyms</b>	<b>119</b>
<b>D</b>	<b>Derived publications</b>	<b>121</b>

# List of Figures

3.1	Return of big- vs. small CAP companies . . . . .	51
3.2	Sigma vs. market CAP . . . . .	52
3.3	Log sigma vs. market CAP . . . . .	53
3.4	Non-linear model of log sigma vs. market CAP . . . . .	55
3.5	Failure data . . . . .	58
3.6	Empirical lifetime distribution . . . . .	59
3.7	Lifetime distribution for different companies . . . . .	63
3.8	Failure data . . . . .	64
3.9	Empirical lifetime distribution . . . . .	65
3.10	Lifetime distribution for different companies . . . . .	67
3.11	Probability of a default for different companies (wide range) . . . . .	69
3.12	Lifetime distribution for different companies (smaller range) . . . . .	70
3.13	Probability of a default for different companies-wide range . . . . .	71
3.14	Lifetime distribution for different companies-smaller range . . . . .	72



# List of Graphs

1.1	Exploratory Example 1 relation plot . . . . .	22
1.2	Exploratory Example 2 relation plot . . . . .	26

# List of Tables

1.1	Degradation data. . . . .	18
1.3	Exploratory Example 1 . . . . .	20
1.8	Time-to-failure data. . . . .	23
1.9	Exploratory Example 2 . . . . .	24
2.1	EM algorithm Example 3 . . . . .	35
2.3	EM algorithm Example 4 . . . . .	39
2.5	EM algorithm Example 4 . . . . .	40
3.1	Stock data . . . . .	48
3.3	Normality hypothesis . . . . .	49
3.7	Market CAP of two companies . . . . .	51
3.9	A non-linear model . . . . .	54
3.16	Linear link function estimates . . . . .	61
3.18	Power function estimates . . . . .	62
3.21	Linear link function estimates . . . . .	66
3.23	Power function estimates . . . . .	66

# Chapter 1

## Introduction

*Simplicity is prerequisite for reliability.*

–Edsger Dijkstra

Recent instabilities in financial markets along with increases in the number of trade assets in portfolios, have led to questions about the effectiveness of existing risk modelling techniques. In answer, a great deal of research has been done towards finding more appropriate techniques or improving existing methods, [Kiefer \(2009\)](#). Determining the likelihood of rare or extreme events is a key concern in the risk management of financial portfolios. Rare events will not be within the range of the available observations. Low-default portfolios, such as banks and sovereign, are characterized by the lack of data for the event of a default, an extreme event for such portfolios, which lie outside the range of available observations. Financial institutions are required by regulatory bodies to perform regular back testing on their rating models. Back testing involves cross validation on time data and is problematic for low-default portfolios due to insufficient data. For low-default portfolios the capital reserves needed for losses that might be incurred, could be underestimated because of such difficulties.

Frequently, determining the likelihood of extreme events involves extreme quantile estimation such as Value at Risk (VaR), expected shortfall and return level. These quantities form part of a collective referred to as risk measures.

**VaR** could loosely be defined as the capital reserves needed for losses that might be incurred over a holding period of a fixed number of days. Suppose  $X \geq 0$  is a random variable

(rv) which models losses experienced by a portfolio over a certain fixed time, with a continuous cumulative density function (CDF)  $F_X$ .  $VaR_p(X)$  can be defined as

$$VaR_p(X) = F^{-1}(1 - p)$$

where  $F$  will be referred to as the loss distribution. More formally

$$VaR_p(X) = \sup\{x \in \mathfrak{R} : P(X \geq x) < 1 - p\}.$$

VaR quantifies the unconditional tail risk of a portfolio and is not a coherent measure of risk as it is not sub-additive. This means the VaR of a combined portfolio can be larger than the sum of the VaRs of its components.

**Expected shortfall** ( $ES_p$ ) is defined as the expected size of a loss which exceeds  $VaR_p$ . Also sometimes called the conditional VaR or mean excess (in an insurance setting)

$$ES_p(X) = E(X|X > VaR_p).$$

This measure of the tail risk has better theoretical properties than that of VaR.

Lastly, **return level** in layman's terms is the maximum of maximum losses. If maxima observed over successive non-overlapping periods of equal length  $n$ , have distribution  $G$  then the return level is defined as

$$R_n^k = G^{-1}\left(1 - \frac{1}{k}\right).$$

The main challenge in implementing one of the above risk measures will be in estimating the tail of  $F_X$ , the loss distribution. Extreme value theory is applied as a possible solution to this challenge, which is attractive because of the parametric form of the tail distribution.

1. Block maxima designs which considers the maxima obtained in successive period.
2. Excess over threshold defines a maximum as a value greater then a certain, high set, threshold.

Some key concepts regarding these two methods are very briefly discussed.

1. The Block maxima's limiting distribution will follow one of the three extreme value distributions, Frechet, Weibull or Gumbel distribution, [Embrechts et al. \(1999\)](#). These distributions can be re-parametrized and represented as the general extreme value distribution. In practice some norming constants of the limiting distribution are unknown and an un-normalized version of the limiting distribution is also available. The return levels of the limiting distribution are of interest.
2. The excess over threshold method is concerned with the conditional excess distribution. Estimating this distribution becomes difficult because few observations are made above the threshold level. Extreme value theory has offered an approximation to the distribution, the generalized Pareto distribution. The distribution can then be used to calculate VaR or expected shortfall.

In both methods, Maximum Likelihood (ML) is used to estimate the parameters of the generalized distributions. [Gilli and Kellezi \(2006\)](#) applied both the block maxima's limiting distribution and the conditional excess distribution to calculate quantities such as the return level and the expected shortfall (defined in terms of VaR) of seven major stock market indices. In their model the excess over threshold proved superior. [Embrechts et al. \(1999\)](#) applied the methods to fire insurance data. [McNeil and Frey \(2000\)](#) studied the tail estimation of a financial return series, stated to be stationary with a stochastic volatility structure leading to dependent returns over time. They proposed an extreme value theory based method, using excess over threshold, which yields a VaR reflecting the current volatility background. This was achieved by applying extreme value theory to the residual series after performing estimation with models often encountered in econometrics.

[Van der Burgt \(2008\)](#) introduced a different method, to that which has been discussed above, for calibrating low-default portfolios. This is done by differentiating the function for the concave shape of the Cumulative Accuracy Profile curve, or the Lorentz curve as it is frequently referred to in economics. The curve tracks the cumulative percentage of defaults against the cumulative percentage of debtors. This function, measuring the shape of the curve contains the parameter  $k$ , called the concavity which is used as a measure of the discriminatory power of the rating model i.e. the ability to distinguish between defaults and non-defaults. More frequently the accuracy ratio, or Gini coefficient, measures discriminatory power. A

relationship exist between  $k$  and the accuracy ratio and is derived by [Van der Burgt \(2008\)](#). The whole method is then applied to a portfolio with exposure to 87 sovereigns: their foreign currency ratings.

[Kiefer \(2009\)](#) proposed a different method for modelling the number of defaults on a low default portfolio, that does not solely rely on historical data. Using Bayes rule, uncertainty about defaults are modelled in the same way that defaults themselves are modelled, with the help of a probability distribution. Starting by considering the problem in the simplest way, as one of independent Bernoulli sampling where in any given year there can either be a default or not. Also stating that assumptions such as independence and identical distributions must be subject to supervisory review. A prior distribution in the Bayesian approach marries the experience of the portfolio manager about aspects of the default rate with what is observed in the data.

Another process which relies on accurately estimating the probability of a default is that of credit-scoring. An applicant will be granted credit based on their credit score. [Stepanova and Thomas \(2002\)](#) showed how survival analysis techniques, which are used in reliability and maintenance modelling, allow credit-scoring models to be built, by extending the Cox's proportional hazard model. The model was then applied to personal loan data. The main advantage of survival analysis is that censored data may be included into the model. Survival analysis models also enables not only the likelihood of a default to be determined, but also how long it will be until the default occurs. [Banasik et al. \(1999\)](#) also applied proportional hazard models to 50000 loans accepted between June 1994 and March 1997 with performance descriptions up to July 1997 and showed that meaningful results can be obtained. The models were compared with logistic regression results and shown to be competitive under certain circumstances.

It is clear that an array of established methods, theories and models are available to deal with the estimation of default probability for low-default portfolios. These methods are constantly being improved upon while research is also conducted on possible new methods.

Towards introducing the proposed method of this paper for dealing with probability estimation on low-default portfolios, consider briefly the following two notions.

**High-default portfolio** Portfolios for which the events under consideration are easily observed. Whether it be loan defaults or severe losses on stock traded. It is straightforward

to apply statistical methods to such data as censoring is not present in the data, the data is consistent and large samples are available.

**Accelerated testing** Employed in manufacturing industries where it is necessary to collect sufficient lifetime information on extremely reliable products. This is problematic as only a few of these highly reliable units will fail (within the limited time available) under normal operating conditions. accelerated test (AT)s are commonly used to induce more failures so as to obtain more lifetime information that will be useful in reliability analysis. To this end, certain controllable factors of these highly reliable products are adjusted to higher-than-normal levels. Such factors might be stress, temperature or pressure (multiple factors might also be used). The use of the word “accelerated” is motivated by obtaining reliability information more rapidly. It is then possible to quantify the lifetime characteristics of the product at normal-at-use conditions.

Would it be possible to use the data available on high-default portfolios, to draw inference on low-default portfolios as is done in accelerated testing? Could some link be established between the high frequency of observations for the high-default portfolios and the low frequency, censored observations of the low-default portfolios? In attempting to answer this question, research was first conducted on accelerated testing.

## 1.1 A review of accelerated tests: Basic concepts

It is necessary to extrapolate failure data collected from the AT, from high levels of stress to normal operating conditions. The extrapolation typically follows from a physical justification which motivates the model choice. A mixture of empirical model fitting with adequate experience in testing similar units could also justify the model chosen for extrapolation. Inference on such characteristics as the reliability and the mean lifetime of devices, when under at use conditions, are done. Stated differently: based on an assumed distributional model and a life-stress relationship, the reliability at a specific mission time, under at use conditions, can be predicted from data collected at higher stress levels. Statisticians have made significant contributions to the development of appropriate models for AT data, [Escobar and Meeker \(2006\)](#). Although we will focus on the application of ATs in the reliability industry, the theory also applies to

other fields of study such as survival analysis, [Ren \(2004\)](#), or even shelf life in bioproduct development, [Rauk et al. \(2014\)](#).

At first glance the mechanics of ATs seem to involve nothing more than regression analysis, perhaps on complicated data such as censored data. ATs require us to extrapolate not only in the accelerating factors but also time. Model choice is thus of critical importance as common practices such as fitting curves to data presents inadequate results.

ATs are partitioned into qualitative and quantitative ATs. In qualitative accelerated testing, the main interest is identifying failures and the modes which are causing these failures. Qualitative ATs don't attempt to make any predictions as to the product's life under normal-at-use conditions. [Freels et al. \(2015\)](#) and [Silverman \(1998\)](#) state that once such failure modes are identified, corrective action is required and the design flaw needs to be removed, so the product is improved upon. [Ma et al. \(2014\)](#) performed a highly accelerated life test to identify 3 failure mechanisms for thin films applied on glass. In quantitative ATs, the interest is in predicting the life of a product, or more specifically, its life characteristics (Mean time to failure (MTTF) or the estimated time at which the probability of failure will reach  $X\%$ ) at normal-at-use conditions, from data obtained in the AT.

The focus here is on quantitative ATs. Intuitively, as the stress used in the AT becomes higher, the duration of the test becomes shorter (failures are observed quicker). However, this poses an additional challenge as uncertainty in the extrapolation increases as the deviation from normal-at-use conditions are increased. Confidence intervals assist in providing a measure of this uncertainty in extrapolation.

It is beneficial to differentiate among ATs based on the type of response variable. Differing models and analyses are then performed depending on this response, [Escobar and Meeker \(2006\)](#).

1. Accelerated binary test (ABT): Binary response variable. One-shot device test are common examples of these types of ABTs. The only information available is whether the product has failed or not.
2. Accelerated life test (ALT): The test is directly related to the lifetime of the product. Here the data is often right- or interval censored.
3. Accelerated repeated measures degradation test (ARMDT): The degradation, at different



points in time, on a sample of units is measured. Instead of failure time data which can be difficult to obtain. Degradation data provides more information than failure time data, as units will not be censored, [Wang et al. \(2015\)](#).

4. Accelerated destructive degradation test (ADDT): These tests are similar to ARMDTs, except that the measurements are destructive, so one can obtain only one observation per test unit.

Further classification of ATs is done by considering the type of stress under which the test operates, [Nelson \(2009\)](#).

1. Constant-stress: Each test item is subjected to a constant stress level, but stress level may differ from item to item. The stress on each item is time independent and does not change with time.
2. Step-stress: Each test item undergoes a pattern of increasing stress levels for a fixed period of time. A test item initially starts at a pre-specified constant stress level for a specified length of time. The stress levels are then increased and maintained step by step at some pre-specified points of time during the experiment. The stress is time dependent and varies as time progresses.
3. Progressive-stress: Each test item is subjected to continuously increasing stress levels over time.

Since the actual modelling of the acceleration is challenging, the prediction of the lifetime under normal operating conditions, based on observations from accelerated levels, becomes even more so. Constant-stress tests are relatively simple to execute and model for reliability estimation, and hence is preferable over the step-stress and progressive-stress tests. The focus here will therefore be the constant-stress test. As an example, in an ALTs of white organic light-emitting diode displays, both constant-stress and step-stress acceleration were applied, [Zhang et al. \(2014\)](#).

### 1.1.1 Life distribution

When performing an accelerated life data analysis the point of departure should be to choose a suitable life distribution. The term life distribution is used for describing the collection of

statistical probability distributions that are often used in reliability engineering and survival analysis. Life data could for instance include a time the product operated successfully (measured in hours, cycles or another relevant metric). Some distributions that better represent life data are commonly called lifetime distributions. Popular choices include the Exponential, Weibull, Log-normal and Gamma distribution.

Other distributions have been used in the literature such as the family of Exponentiated distributions, [Abdel Ghaly et al. \(2015\)](#), because it is richer in that different hazard rates can be accommodated. A mixture of distributions has also been applied to describe different failure modes in products, [Kim and Bai \(2002\)](#).

The lifetime distribution is a function which describes the product at different stress levels. Other important functions related hereto, are the probability density function (PDF) and CDF, the latter used for calculating important reliability results and plots. These basic relationships are now illustrated.

### 1.1.2 Life-stress relationships

After an appropriate life distribution has been explored, the failure rate is required in relating stress factors, such that measurements taken during an experiment can be extrapolated back, to the expected performance under normal operating conditions. The way in which the life distribution changes across various stress levels, is quantified as a curve representing the relationship between the rate of failure and the stress factors, such as a simple linear model; however, this may not suffice to adequately describe the lifetime of products against the stress factors. Various relationships between the lifetime of products and stress factors have been proposed in the literature. Five common relationships are stated here, [Wang and Kececioglu \(2000\)](#).

- **Inverse power law relationship**

The parameter  $\theta(T)$  relates to  $T$  in the form

$$\theta(T) = \frac{1}{\gamma_0 T^{\gamma_1}}.$$

- **Arrhenius accelerated relationship**

Based on the Arrhenius Law for simple chemical-reaction rates, it is one of the com-

monly used acceleration models that predicts the failure time varying with temperature. The parameter  $\theta(T)$  relates to temperature  $T$  in degree Celsius,

$$\theta(T) = \exp \left[ \gamma_0 + \gamma_1 \left( \frac{11605}{T + 273.15} \right) \right].$$

The Arrhenius life-stress model is formulated by assuming that the inverse reaction rate of the process is proportional to life, thus the Arrhenius life-stress relationship is given by

$$\theta(T) = \exp \left[ \gamma_0 + \frac{\gamma_1}{T} \right].$$

Different forms of the Arrhenius life-stress relationship is present in the literature under the same name. The designer needs to be sure which form of the relationship is required, [Rauk et al. \(2014\)](#). A life-stress relationship that is closely related to the above relationship is the Eyring relationship

- **Eyring accelerated relationship**

Just as the Arrhenius relationship, Eyrings relationship is generally applied when thermal stress is the acceleration variable. It relates the stress varying parameter to the stress factor as follows

$$\theta(T) = \frac{1}{T} \exp \left[ -\left( \gamma_0 - \frac{\gamma_1}{T} \right) \right].$$

It is clear that the difference between Eyrings relationship and the Arrhenius relationship is the variable  $\frac{1}{T}$ .

- **Log-linear relationship**

This relationship is commonly used in practice due to its mathematical convenience. It shows the relative importance of stress factors in influencing the failure behaviour, regardless of whether the model is correct or not. The parameter  $\theta(T)$  relates to a stress factor  $T$  in this case in the form

$$\theta(T) = \exp[\gamma_0 + \gamma_1 T].$$

Note that when the transformation of the acceleration variable to  $T = \frac{1}{V}$  is applied, the Arrhenius life-stress relationship is obtained, or when  $T = \exp V$ , the inverse-power relationship is obtained.

A final relationship to mention which might not always sufficiently describe the underlying relationship, is the Linear relationship.

- **The Linear relationship**

This simple relationship is given by

$$\theta(T) = \gamma_0 + \gamma_1 T.$$

This life-stress relationship was assumed in a study of life test data under two failure modes, [Kim and Bai \(2002\)](#), and a comparisons of test plans, [Meeker \(1984\)](#).

### 1.1.3 Parameter estimation methods

The chosen life distribution and life-stress relationship is applied to the sample data in order to estimate the parameters of the distribution. Some of the more sophisticated methods, popular in the literature, are:

- Least Squares (LS) methods
- Weighted Least Squares (WLS) methods
- ML methods

The ML method is commonly adopted in analysing reliability data due to its well-known optimal properties. Depending on the data structure this method can be extended to include the Expectation Maximization (EM) algorithm.

The mathematical details of the EM algorithm will be given in [Chapter 2](#).

### 1.1.4 Reliability information

The most important information obtained from the AT is the reliability function. It is derived using the CDF. The CDF gives the probability of an event (failure) occurring at a specific time, and so, subtracting this probability from 1 yields the reliability function.

$$\begin{aligned}R(t) &= 1 - F(t) \\ &= 1 - \int_0^t f(s)ds \\ &= \int_t^\infty f(s)ds.\end{aligned}\tag{1.1}$$

From this function other reliability results are calculated.

### The conditional reliability function

This is defined as the probability of a product lasting the next test phase, given it successfully lasted/survived the previous test phase

$$R(T, t) = \frac{R(T + t)}{R(T)}.\tag{1.2}$$

### The failure rate function

The failure rate determines the number of failures occurring per unit of time. It gives the instantaneous failure rate and is also known as the hazard function

$$\lambda(t) = \frac{f(t)}{R(t)}.\tag{1.3}$$

### The mean life

This life characteristic was referred to in Section 1.1 (MTTF). The mean life function gives a measure of the average time of operation to failure

$$\bar{T} = m = \int_0^\infty t.f(t)dt.\tag{1.4}$$

#### 1.1.5 Censored data

Often in reliability studies the time to an event is studied; however, problems arise because of censoring within the data. When the time to an event is unknown for a subset of the data, it

is said that the data is censored. This most often happens because the trial/testing period ends before units have failed. In this section extensive use is made of concepts and examples discussed in Chapter 3 of the book: "Survival Analysis, Techniques For Censored and Truncated Data" by [Klein and Moeschberger \(2005\)](#), in which time-to-event data is analysed. These data appear in different ways which poses particular challenges when analysis on the data is performed. As mentioned, censoring might be present within the life data, which imply that some lifetimes are known exactly, and others only to have occurred within certain intervals. Data may exhibit 4 characteristics within this framework. It may be

- Observed, truncated or censored.
- Right, left or interval censored.
- Type I, Type II or randomly right censored.
- Single or multiple censored.

These different types of censoring schemes lead to different likelihood functions. In order to effectively work with censoring in analysis, careful examination of the design, which was used when the survival data was obtained, is required.

### **Observed, truncated or censored**

*Observed data* is the usual data which results from a typical experiment or process. In this, the lifetime of every subject included in the sample is known and used as observed or recorded. If this is the case, the estimate of the survival function  $S(t)$  reduces to the empirical survival function. *Censoring* implies that time to event is unknown for a subset of the data. [Yildirim \(2008\)](#) proposed a mixture model in dealing with commercial mortgage backed securities where a substantial proportion of loans are deemed long-term survivors of the study meaning they will never default. These long-term survivor loans will fall in the censored grouped along with the loans which will eventually default. Thus, the assumption of standard survival models that the censoring is independent of the event is incorrect in this instance and the mixture model may overcome this. Only 4.05% of the loans in the study experienced a default, thus heavy censoring was present. Another feature of time to event studies, not to be confused with censoring, is *truncation*. This is where only the lifetimes of subjects which lie within a certain

observational window are observed. If it does not lie within this window no information on the subjects' lifetime is available. Different to censoring, where at least partial information is available, truncation takes place when the subject is not included in the study. So the number of censored observations are known whereas the number of truncated observations are not known. [Klein and Moeschberger \(2005\)](#) present Example 3.7 as an example of left truncation. Data on ages at death of 462 individuals, who resided in a retirement home during a period of 11 years, have been reported. To enter the old age home, one would have to be a certain age. If an individual died before this age, he/she could not have entered the retirement centre and could thus not have been included in the study. [Boyes et al. \(1989\)](#) commented that often when banks do credit-scoring, not all applicants who applied for credit are included in the sample used to estimate credit assessment functions. Using a choice sample of applicants who received credit and either defaulted or repaid, leads to a sample that is truncated as the applicants who did not receive credit was not included in the sample. For the remainder of this section the focus falls on censored data.

### Right, left or interval censored

If there are patients included in a study with time to event greater than a certain value (the fixed censoring time) the data is *right censored*. The following notation is defined in this regard.  $X$ , the exact lifetime of a specific individual in the study and  $C_r$ , the fixed censoring time (where  $r$  is for *right* censored). The  $X$ 's are assumed to be independent and identically distributed (IID) with PDF  $s(x)$  and survival function  $S(x)$ .  $X$  is rarely observed and instead data from right censored experiments can conveniently be noted by pairs of rvs  $(T, \delta)$ . The value of  $\delta$  (an indicator of whether an observation was observed,  $\delta = 1$ , or was censored,  $\delta = 0$ ) is based on the value of  $T$ . The exact lifetime  $X$  will be known if and only if  $X \leq C_r$  in this case  $\delta = 1$ . Thus

$$\delta = \begin{cases} 1 & \text{if } T = X, \\ 0 & \text{if } T = C_r. \end{cases} \quad (1.5)$$

The following is an equivalent definition of  $T$

$$T = \min(X; C_r)$$

with  $\delta$  equalling 1 or 0 as  $T$  equals  $X$  or  $C_r$ . It is important to note that  $C_r$  is considered fixed and that  $(T, \delta)$  is a pair of rvs.

It is now simple to extend these ideas to the case where left censored data is present. *Left censored* data occurs when time to an event is less than a certain value. The event of interest has occurred before the study began. The previous notation is now extended to the left censored survival problem. Data from left censored experiments can be presented by pairs of rvs  $(T, \epsilon)$ .  $\epsilon$  is now the indicator of whether an observation is censored or not and the exact lifetime,  $X$ , will be known if and only if  $X \geq C_l$ .  $C_l$  is the fixed censoring time ( $l$  is for *left* censored). Thus

$$\epsilon = \begin{cases} 1 & \text{if } T = X, \\ 0 & \text{if } C_l, \end{cases} \quad (1.6)$$

or

$$T = \max(X; C_l)$$

with  $\epsilon$  equalling 0 or 1 as  $T$  equals  $C_l$  or  $X$ .

[Cheng et al. \(2008\)](#) addressed a left censored data problem. They define intellectual capital as being the difference between the market value and the book value of a company. Reporting that the 62% of invested stocks that were spent on tangible assets in 1982 decreased to 16% by 1999. So, they identify the need for companies to sustain growth by continually investing in intellectual capital. The purpose of their study was to investigate which factors drive the value of a company's intellectual capital. The intellectual capital of a company is distributed over positive values and so the creation value is seen as being left censored. The Tobit model is employed as a ML estimator of parameters for left censored data. The financial data of health care firms listed on the S&P500 were used to assemble the Tobit model.

Data can also be *interval censored* and this occurs when subjects have periodic follow-ups. For example: A study is conducted where patients are monitored for the occurrence of a disease in three month intervals. Patients who are disease free after three months but not after six months are interval censored as exact occurrence of the disease is not known. It is only known to have occurred between three and six months after commencement of the study. Interval censoring is a generalized case of left and right censoring. This can be seen if the left end-point of the interval in which the event to occur, is set to  $C_r$  and the right end-point to



infinity. In the reliability context one-shot device testing is an example of interval censored data. During one-shot device testing binary data arises. Methods for analysing such data can be found in the literature, [Balakrishnan and Ling \(2012\)](#). In Chapter 3 we will model interval censored data.

Next, three types of right censoring is considered, which can easily be extended to left censoring.

### **Type I, Type II or competing risk right censored data**

*Type I right censoring* refers to situations where the event is observed if it occurs prior to some pre-specified time. Here the censoring levels are known in advance but the number of censored observations,  $c$  and the number of exact observations,  $r$  are random outcomes. When subjects have different fixed censored times, this is known as *progressive Type I censoring*. This can, for example, occur when two fixed censored times are given and certain subjects are kept in the study beyond the first fixed censored time to be observed, until the second fixed censored time.

Another Type I censoring that occurs, due to subjects entering the study at different times, is called *generalized Type I censoring*. All subjects are still observed under the same fixed censoring time  $C_r$ , but subjects have their own specific observed, fixed censoring times  $C_r - t$ . For example, a person who entered the study at time  $t = 0$ , who has not experienced the event by the end of the study, will have  $T = C_r$ . A person entering the study after it has commenced, say at time  $t = 4$ , who has also not experienced the event by the end of the study will have  $T = C_r - 4$ . These two patients have their own specific observed, fixed censoring times but are observed under the same fixed censoring time,  $C_r$ . Note that in all instances of Type I censoring, the censoring times are fixed.

*Type II censoring* refers to the case where the study continues until the first  $r$  subjects have experienced the event. Here  $r$ , the number of exact observations, is pre-determined and the censoring time,  $T(r)$  the  $r$ th ordered lifetime, is random. Type II censoring can be extended to progressive Type II censoring as with Type I censoring. After observing the first  $r_1$  events, a portion of the surviving subjects are removed from the study, and then the next  $r_2$  events are recorded for the remaining subjects. It is important to note that for Type I censoring a fixed censored time is present, while for Type II, a fixed number of censored subjects (or fixed

number of exact observations) are present. This distinction is important when one constructs the likelihood function of the data. A final type of right censoring of interest here, is *competing risk censoring*. A particular case of competing risk censoring, is random censoring. This occurs when the marginal distribution of some event is considered, but due to competing risk, such as loss to follow up or withdrawal from study, some subjects are removed from the study. The time to event is not available for these subjects and random right censoring occurs. It has been shown in the literature that in order to be able to identify the marginal distribution from randomly censored data, the time to event,  $X$ , and censoring times  $C$ , are required to be independent. Therefore, competing risks are random observations when the probability of the event occurring in a setting, where all other risk cannot occur, is estimated. The additional assumption of time to event being independent of the competing events, are also required. [Yildirim \(2008\)](#) discussed the limitations of these independence assumptions and overcame them by applying the mixture model. It is interesting to note that when an observation is censored, information about the exact lifetime, cannot be obtained. However, when it is observed exactly, information about the observation limit of fixed censoring time, can be obtained. If observation limits are random (as is the case here) this information cannot be obtained. Many right censored medical studies are a combination of random censoring and Type I censoring. For example, subjects may be observed for a fixed censored time of  $C_r$ , some of which will be censored when this time ends, others may leave the study due to competing risks.

[Phillips et al. \(1996\)](#) analysed both fixed rate mortgage and adjustable rate mortgage terminations using a national individual loan data set for 1986 to 1992. The mortgages that originated between 1982 and 1989 were either terminated or censored between January 1986 and June 1992. At the end of the observance periods loans were either still current or removed due to sale of servicing contracts (possibly switching to a different lender). This results in a combination of random right censoring and Type I right censoring. [Stepanova and Thomas \(2002\)](#) and [Banasik et al. \(1999\)](#) mentioned previously, also had competing risk in their loan data.

### Single or multiple censoring

When there is one censoring level, this is known as single censoring. This is for example, Generalized Type I censoring or Type II censoring. When multiple censoring levels are present

this is known as multiple censoring. This is Type I or Type II progressive censoring.

### 1.1.6 Introductory examples

As part of this introduction into ATs a few basic introductory examples are presented so as to ease the reader into the concepts discussed thus far. These examples were constructed using the Reliability procedure in SAS.

**Example 1** *A company has been collecting degradation data over a period of 4 years with the purpose of calculating reliability after 5 years. With the degradation measurement decreasing with time, failure is defined as a measurement of 150 or below. The following table gives the degradation measurements over 4 years*

**Table 1.1:** Degradation data.

Year 1	Year 2	Year 3	Year 4
437	412	246	125
446	420	324	208
497	451	330	229
503	454	426	242
705	554	499	273
737	580	546	297
748	608	554	311
788	610	559	318
818	727	625	393
860	825		403
875	925		470
934			
1124			
1250			
1350			

A linear life-stress relationship was assumed together with a 2-parameter Weibull distribution. The PDF and CDF for the Weibull distribution, are respectively given below

$$f(x) = \begin{cases} \left(\frac{\beta}{\eta}\right)\left(\frac{x}{\eta}\right)^{\beta-1} \exp^{-\left(\frac{x}{\eta}\right)^{\beta}} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases} \quad (1.7)$$

$$F(x) = \begin{cases} 1 - \exp^{-\left(\frac{x}{\eta}\right)^\beta} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases} \quad (1.8)$$

The SAS reliability procedure relates the stress factors to lifetime of products through the Weibull scale parameter, by default, in the following way

$$\eta = \exp(\mathbf{x}'\hat{\beta})$$

where  $\mathbf{x}'\hat{\beta}$  is a regression model containing an intercept term. That is a model of the form

$$\mu_i = \beta_0 + \beta_1 x_{i1} + \dots$$

The parameters were estimated to be:

- $\hat{\beta} = 3.618336$  the Weibull shape parameter as seen in the Weibull PDF and CDF, Equation 1.7 and Equation 1.8 respectively.
- $\gamma_0 = 7.154643$  and  $\gamma_1 = -0.331695$  the regression parameters of the linear life-stress relationship as listed in Section 1.1.2.

Note that SAS uses, by default, the ML estimation method in estimating the above obtained parameters.

An excerpt of the output obtained from the reliability procedure is given below.

**Table 1.3:** Exploratory Example 1: Weibull parameter estimates and group Weibull percentile estimates.

Model Information		
Input Data Set	WORK.CLASSB	
Analysis Variable	deg	Degradation
Distribution	Weibull	

The Weibull life distribution was used

Weibull Parameter Estimates				
Parameter	Estimate	Standard Error	Asymptotic Normal	
			95% Confidence Limits	
			Lower	Upper
Intercept	7.1546	0.0946	6.9691	7.3402
year	-0.3317	0.0357	-0.4016	-0.2618
EV Scale	0.2764	0.0316	0.2208	0.3459
Weibull Shape	3.6183	0.4142	2.8911	4.5285

The Weibull parameter estimates obtained as previously stated

Weibull Parameter Estimates					
Parameter	Estimate	Standard Error	Asymptotic Normal		Group
			95% Confidence Limits		
			Lower	Upper	
EV Location	6.8229	0.0649	6.6957	6.9502	1
EV Scale	0.2764	0.0316	0.2208	0.3459	1
Weibull Scale	918.6901	59.6663	808.8834	1043.4033	1
Weibull Shape	3.6183	0.4142	2.8911	4.5285	1

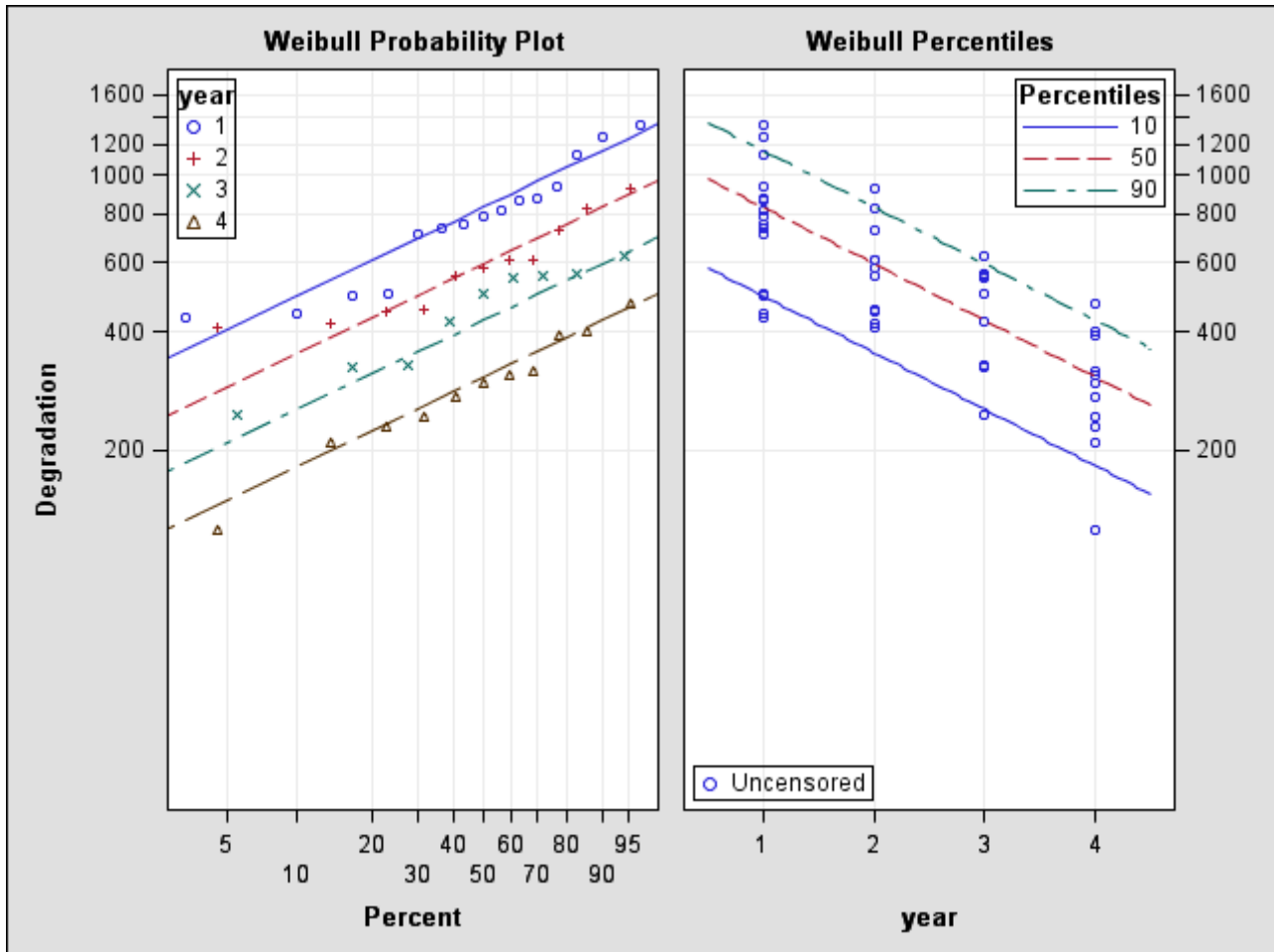
The Weibull parameter estimates obtained from the first years life data. Also given for years two to four, but not shown here

<i>Weibull Percentile Estimates</i>					
<i>Percent</i>	<i>Estimate</i>	<i>Standard Error</i>	<i>Asymptotic Normal</i>		<i>Group</i>
			<i>95% Confidence Limits</i>		
			<i>Lower</i>	<i>Upper</i>	
0.1	136.182599	33.0465551	84.6381212	219.117578	1
0.2	164.959416	36.5456496	106.856054	254.65669	1
0.5	212.586666	41.2163401	145.380084	310.861634	1
1	257.651917	44.6284918	183.48255	361.802856	1
2	312.490286	47.7628903	231.597448	421.637541	1
5	404.268201	51.1688781	315.450798	518.092773	1
10	493.250302	52.9839489	399.606932	608.837938	1
20	606.928947	54.1860843	509.498986	722.990147	1
30	690.925154	54.8986767	591.285575	807.355343	1
40	763.035269	55.768513	661.199059	880.556035	1
50	830.191178	57.0316114	725.609825	949.845728	1
60	896.759984	58.8949952	788.448426	1019.95063	1
70	967.05018	61.6711393	853.425902	1095.80228	1
80	1047.8218	66.010599	926.111971	1185.52676	1
90	1156.84265	73.8935791	1020.7129	1311.12765	1
95	1244.11286	81.823373	1093.64803	1415.27874	1
99	1401.10182	99.3190598	1219.35759	1609.93487	1
99.9	1567.24138	121.63514	1346.08754	1824.72943	1

*The Weibull distribution percentiles obtained for the first years life data. Also given for*

years two to four, but not shown here

A probability plot showing the Weibull distribution being fitted to each group is given below. Also shown is a plot of some of the estimated percentiles for each group. Note that for each year the failure times seem to follow a straight line, indicating that the Weibull distribution does a good job in fitting the data for all years. The lines are also parallel, indicating that the Weibull distribution is a good fit, [Ramirez et al. \(2001\)](#).



Graph 1.1: Exploratory Example 1 relation plot: 2-parameter Weibull life data and linear stress-life relationship.

*It is now easy to estimate the reliability after 5 years or in other words the probability*



that the degradation measurement will not be less than 150, the critical degradation level, in 5 years

$$\begin{aligned}
 R(5) &= 1 - F(150) \\
 &= 1 - P(x(5) \leq 150) \\
 &= 1 - (1 - \exp^{-\left(\frac{x}{\eta}\right)^\beta}) \text{ from Equation 1.8} \\
 &= \exp\left(-\left(\frac{150}{\exp(\gamma_0 + \gamma_1 5)}\right)^\beta\right) \\
 &= \exp\left(-\left(\frac{150}{\exp(7.154643 - 0.331695(5))}\right)^{3.618336}\right) \\
 &= 0.841481.
 \end{aligned} \tag{1.9}$$

**Example 2** Consider the following times-to-failure data at 3 different stress levels.

**Table 1.8:** Time-to-failure data.

Stress	393	408	423
Time failed (hrs)	3850	3300	2750
	4340	3720	3100
	4760	4080	3400
	5320	4560	3800
	5740	4920	4100
	6160	5280	4400
	6580	5640	4700
	7140	6120	5100
	7980	6840	5700
	8960	7680	6400

An Arrhenius life-stress relationship was assumed together with a 2-parameter Weibull distribution. Similar output to that of Example 1 is given below.

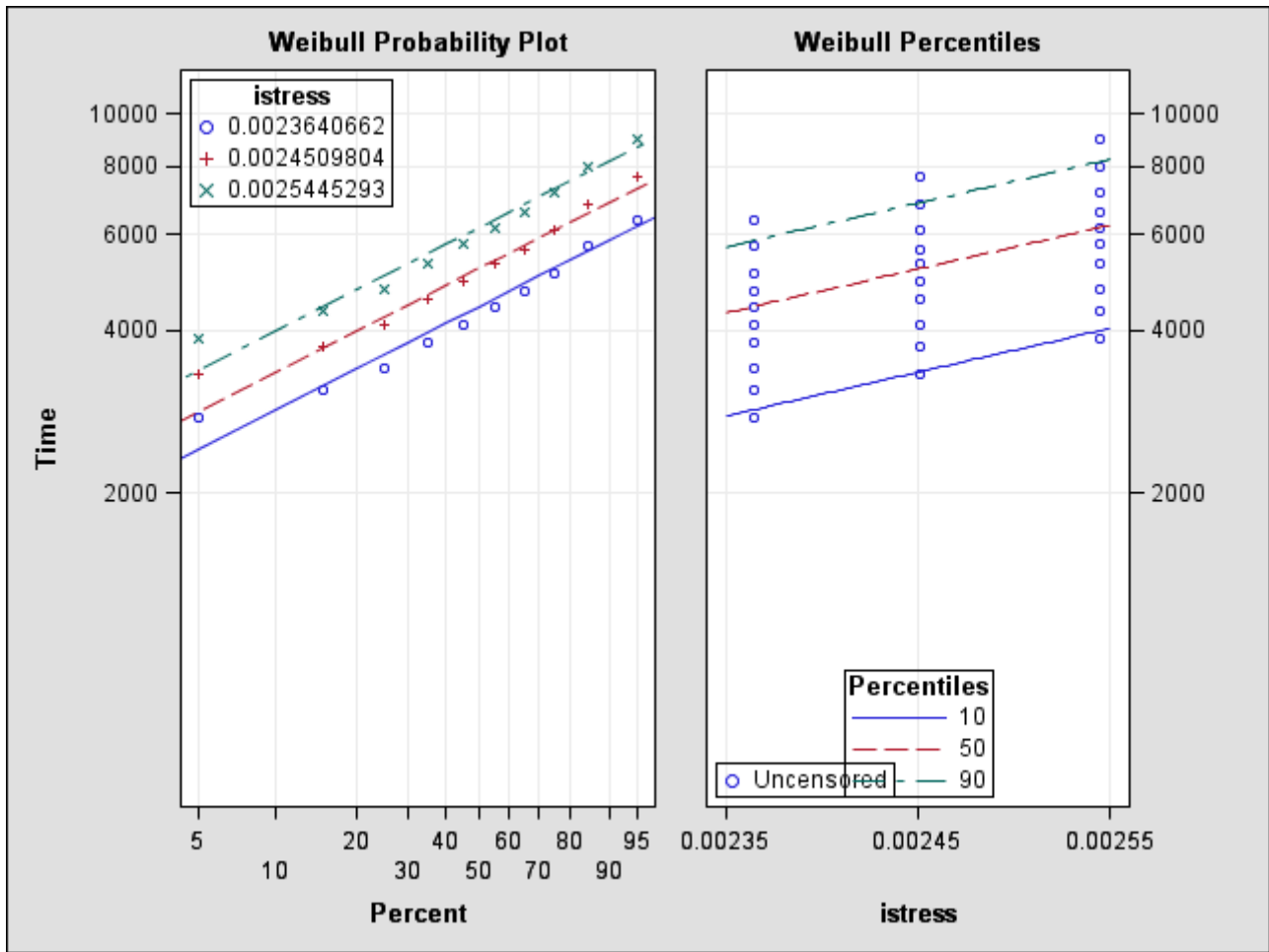
**Table 1.9:** Exploratory Example 2: Weibull parameter estimates and group Weibull percentile estimates.

Model Information		
Input Data Set	WORK.CLASSB	
Analysis Variable	time	Time
Distribution	Weibull	

Weibull Parameter Estimates				
Parameter	Estimate	Standard Error	Asymptotic Normal	
			95% Confidence Limits	
			Lower	Upper
Intercept	4.0773	1.4382	1.2584	6.8961
istress	1861.6187	585.9728	713.1331	3010.1042
EV Scale	0.2330	0.0326	0.1772	0.3064
Weibull Shape	4.2916	0.5996	3.2636	5.6434

Weibull Parameter Estimates					
Parameter	Estimate	Standard Error	Asymptotic Normal		Group
			95% Confidence Limits		
			Lower	Upper	
EV Location	8.4783	0.0690	8.3431	8.6135	0.0023640662
EV Scale	0.2330	0.0326	0.1772	0.3064	0.0023640662
Weibull Scale	4809.1263	331.7239	4200.9936	5505.2920	0.0023640662
Weibull Shape	4.2916	0.5996	3.2636	5.6434	0.0023640662

<i>Weibull Percentile Estimates</i>					
<i>Percent</i>	<i>Estimate</i>	<i>Standard Error</i>	<i>Asymptotic Normal</i>		<i>Group</i>
			<i>95% Confidence Limits</i>		
			<i>Lower</i>	<i>Upper</i>	
0.1	961.77605	239.482183	590.370735	1566.83439	0.0023640662
0.2	1130.49568	257.009882	724.026556	1765.15693	0.0023640662
0.5	1400.06066	278.616153	947.884159	2067.94242	0.0023640662
1	1646.43931	292.832493	1161.85699	2333.12914	0.0023640662
2	1937.32469	304.271277	1424.016	2635.66348	0.0023640662
5	2407.08752	313.661301	1864.54997	3107.48997	0.0023640662
10	2846.66378	315.726457	2290.49065	3537.88596	0.0023640662
20	3390.60372	314.301922	2827.30343	4066.13365	0.0023640662
30	3782.15487	313.744415	3214.61406	4449.89515	0.0023640662
40	4112.34511	315.494568	3538.23206	4779.61367	0.0023640662
50	4415.46296	320.07636	3830.65246	5089.55416	0.0023640662
60	4712.15321	328.184057	4110.89478	5401.35155	0.0023640662
70	5021.70301	341.141857	4395.67855	5736.88473	0.0023640662
80	5373.09147	361.946269	4708.52575	6131.45463	0.0023640662
90	5840.73653	399.933625	5107.20353	6679.62477	0.0023640662
95	6210.09425	437.918421	5408.46281	7130.5419	0.0023640662
99	6864.54735	520.654978	5916.30921	7964.76462	0.0023640662
99.9	7544.72967	624.31688	6415.16495	8873.18506	0.0023640662



Graph 1.2: Exploratory Example 2 relation plot: 2-parameter Weibull life data and Arrhenius life-stress relationship.

*An explanation as to why the inverse stress was used in the model, is now given. As mentioned in Example 1, the SAS reliability procedure relates the stress factors to the lifetime of products through the Weibull scale parameter, by default, in the following way*

$$\eta = \exp(\mathbf{x}'\hat{\beta})$$

*where  $\mathbf{x}'\hat{\beta}$  is a regression model containing an intercept term. That is a model of the form*

$$\mu_i = \beta_0 + \beta_1 x_{i1} + \dots \tag{1.10}$$

*Specifying the Arrhenius life-stress relationship in the SAS procedure will lead to  $x$  (stress) being transformed in Equation 1.10, in the following way*

$$\mu_i = \beta_0 + \beta_1 \left( \frac{1000}{x + 273.15} \right)_{i1} + \dots,$$

*which would lead to a Weibull scale parameter*

$$\eta = \exp\left(\beta_0 + \beta_1 \left( \frac{1000}{x + 273.15} \right)_{i1} + \dots\right).$$

*In order to make use of the Arrhenius life-stress model as stated in Section 1.1.2, the Weibull scale parameter should be defined as*

$$\eta = \exp\left(\beta_0 + \left(\frac{\beta_1}{x}\right)_{i1} + \dots\right)$$

*which requires a regression equation of the form*

$$\mu_i = \beta_0 + \left(\frac{\beta_1}{x}\right)_{i1} + \dots$$

*It is thus easier to use the default linear life-stress relationship, and regress using the inverse stress.*

### 1.1.7 Additional considerations

ATs have been extensively modelled using a Bayesian approach to estimate and predict model parameters, [Sinha et al. \(2003\)](#), [Ling \(2012\)](#), instead of the classical ML, LS or WLS methods mentioned in Section 1.1.3.

Various test plans have been designed where levels of stress, allocation of units to stress, time of inspection and number of units inspected are some of the factors considered and chosen so as to satisfy optimal criteria, [Meeker \(1984\)](#), [Morris \(1987\)](#). These test plans are also designed under different censoring schemes, [Nelson and Kielpinski \(1976\)](#).

Most devices can fail due to more than one failure mode. These multiple failure modes need to be carefully considered as they effect the lifetime distribution at normal-at-use condition. The dominant failure mode at high stress may not be the dominant mode at low stress, [Nelson \(1975\)](#), [Kim and Bai \(2002\)](#).

## 1.2 Objectives of the study

- Conduct a broad review of the literature on reliability analysis, with specific focus on censoring schemes, and ATs.
- Present examples of ATs, analysed in SAS.
- Apply the theory of accelerated testing to stock market data.

## 1.3 Contributions

- Apply the Log-normal distribution to one-shot device testing problems, formulated using the EM algorithm.
- Formulate a low probability event in a stock market in terms of an AT scenario.
- Apply a one-shot device AT to simulated stock market data.

## 1.4 Dissertation outline

The rest of the study is structured as follows

- **Chapter 2**, as mentioned in Section 1.1.3, focusses on the mathematical details of ML methods for parameter estimation, as well as particular examples.
- **Chapter 3** introduces, intuitively, how one might go about applying ATs to financial data by way of a few simulation example. Thus, fully introducing the proposed method of this paper
- **Chapter 4** gives a summary and conclusion of the study. Recommendations for further research are also mentioned.

All SAS code used for the generation of examples and illustrations are provided in the attached appendices:

- **Appendix A** provides the SAS code for the EM algorithm used in Chapter 2.

- **Appendix B** provides the SAS code for the application presented in Chapter 3.
- **Appendix C** provides the acronyms used in the course of this work, along with their respective definitions.
- **Appendix D** lists the proposed publication derived from this work.

## Chapter 2

# Exploring the EM algorithm

In this chapter, the mathematical workings of the EM algorithm, will briefly be explored. The EM algorithm is commonly implemented in statistical estimation problems involving incomplete data, or in problems which can be posed in a similar form, such as mixture estimation, [Friedman et al. \(2001\)](#). The algorithm was first explicitly introduced by [Hartley \(1958\)](#) while [Dempster et al. \(1977\)](#) described the method and presented useful examples. The EM algorithm has since been used in many problems concerning the estimation of parameters in incomplete data under various censoring schemes. Such schemes include progressive Type-I interval censored data [Chen and Lio \(2010\)](#), right-censored data [Scheike and Sun \(2007\)](#), random censored data [Kundu and Dey \(2009\)](#) and both left- and right censored data [Balakrishnan and Ling \(2012\)](#), to name but a few. The focus of this chapter will be the application of the algorithm to both left- and right censored data.

The chapter is structured as follows; Section [2.1](#) gives, at the hand of two examples, a brief introduction of the algorithm. Section [2.2](#) will define the algorithm in terms of a problem which will be presented more thoroughly in Chapter [3](#).

### 2.1 Introduction to the EM algorithm

The EM algorithm is an iterative method of ML estimation for incomplete data. ML estimates are computed from the complete data likelihood instead of the observed likelihood, because the observed likelihood might be numerically infeasible to maximise. The areas of application



include

- Image analysis in medicine such as the ML reconstruction of emission tomography data, as done by [Green \(1990\)](#) and [Kay \(1997\)](#).
- Engineering as presented throughout this paper.
- Genetics including human pedigree analysis as done by [Ott \(1977\)](#) or determining the distribution of one part of an allele by applying the EM to genotype data, such as [Kitamura et al. \(2002\)](#).

In layman's terms the EM algorithm consists of the following steps:

1. Fill in missing values by estimating them
2. Estimate the parameters of the complete dataset
3. Use the estimates to re-estimate the missing data
4. Re-estimate the parameters from the updated complete data step

Step2–4 should be re-iterated until the parameters achieve the desired level of convergence.

Step 1–3 is known as the Expectation step (E-step), while step 4 defines the Maximisation step (M-step) of the process, thus the EM algorithm. Before the above algorithm is presented in formal mathematical notation, two elementary examples are given.

**Example 3** *Simple example given by [Dempster et al. \(1977\)](#).*

*Consider a 4 outcome multinomial distribution*

$$p(x_1, x_2, x_3, x_4) = \frac{n!}{x_1!x_2!x_3!x_4!} \pi_1^{x_1} \pi_2^{x_2} \pi_3^{x_3} \pi_4^{x_4} \quad (2.1)$$

*with*

$$n = x_1 + x_2 + x_3 + x_4 \text{ and } 1 = \pi_1 + \pi_2 + \pi_3 + \pi_4. \quad (2.2)$$

Suppose the probabilities are related by a single parameter  $0 \leq \theta \leq 1$  such that

$$\pi_1 = \frac{1}{2} + \frac{1}{4}\theta \quad (2.3)$$

$$\pi_2 = \frac{1}{4} - \frac{1}{4}\theta \quad (2.4)$$

$$\pi_3 = \frac{1}{4} - \frac{1}{4}\theta \quad (2.5)$$

$$\pi_4 = \frac{1}{4}\theta. \quad (2.6)$$

For an observation  $(x_1, x_2, x_3, x_4)$ , the log-likelihood function is given by

$$\begin{aligned} l(\theta) &= \ln(n!) - \ln(x_1!) - \ln(x_2!) - \ln(x_3!) - \ln(x_4!) + x_1 \ln(\pi_1) + x_2 \ln(\pi_2) + x_3 \ln(\pi_3) + x_4 \ln(\pi_4) \\ &= x_1 \ln\left(\frac{1}{2} + \frac{1}{4}\theta\right) + x_2 \ln\left(\frac{1}{4} - \frac{1}{4}\theta\right) + x_3 \ln\left(\frac{1}{4} - \frac{1}{4}\theta\right) + x_4 \ln\left(\frac{1}{4}\theta\right) + \text{constant} \\ &= x_1 \ln\left(\frac{1}{4}(2 + \theta)\right) + (x_2 + x_3) \ln\left(\frac{1}{4}(1 - \theta)\right) + x_4 \ln\left(\frac{1}{4}\theta\right) + \text{constant} \\ &= x_1 \ln(2 + \theta) + (x_2 + x_3) \ln(1 - \theta) + x_4 \ln(\theta) + \text{constant}. \end{aligned}$$

Differentiation with respect to  $\theta$  yields

$$\frac{\partial l(\theta)}{\partial \theta} = \frac{x_1}{2 + \theta} - \frac{x_2 + x_3}{1 - \theta} + \frac{x_4}{\theta}.$$

The unknown parameter could be determined by solving the above, simple equation. However, for the purpose of this example the EM algorithm is employed. Suppose a five class multinomial distribution is formed by splitting the first class into two classes, with associated probabilities  $\frac{1}{2}$  and  $\frac{1}{4}\theta$  respectively. The original variable  $x_1$  is now the sum of two components  $v_1$  and  $v_2$ .

$$p(v_1, v_2, x_2, x_3, x_4) = \frac{n!}{v_1!v_2!x_2!x_3!x_4!} \pi_{1,1}^{v_1} \pi_{1,2}^{v_2} \pi_2^{x_2} \pi_3^{x_3} \pi_4^{x_4}$$

with

$$n = v_1 + v_2 + x_2 + x_3 + x_4 \text{ and } 1 = \pi_{1,1} + \pi_{1,2} + \pi_2 + \pi_3 + \pi_4$$

and

$$\begin{aligned}\pi_{1,1} &= \frac{1}{2} \\ \pi_{1,2} &= \frac{1}{4}\theta \\ \pi_2 &= \frac{1}{4} - \frac{1}{4}\theta \\ \pi_3 &= \frac{1}{4} - \frac{1}{4}\theta \\ \pi_4 &= \frac{1}{4}\theta.\end{aligned}$$

Thus ,

$$\begin{aligned}l(\theta) &= \ln(n!) - \ln(v_1!) - \ln(v_2!) - \ln(x_2!) - \ln(x_3!) - \ln(x_4!) + v_1 \ln(\pi_{1,1}) + v_2 \ln(\pi_{1,2}) \\ &\quad + x_2 \ln(\pi_2) + x_3 \ln(\pi_3) + x_4 \ln(\pi_4) \\ &= v_1 \ln\left(\frac{1}{2}\right) + v_2 \ln\left(\frac{1}{4}\theta\right) + x_2 \ln\left(\frac{1}{4} - \frac{1}{4}\theta\right) + x_3 \ln\left(\frac{1}{4} - \frac{1}{4}\theta\right) + x_4 \ln\left(\frac{1}{4}\theta\right) + \text{constant} \\ &= (v_2 + x_4) \ln\left(\frac{1}{4}\theta\right) + (x_2 + x_3) \ln\left(\frac{1}{4}(1 - \theta)\right) + \text{constant} \\ &= (v_2 + x_4) \ln(\theta) + (x_2 + x_3) \ln(1 - \theta) + \text{constant}.\end{aligned}$$

The above equation resembles the likelihood function of a binomial distribution with parameters  $n = v_2 + x_4 + x_2 + x_3$  and  $\pi = \theta$ .

The algorithm proceeds with the complete log-likelihood of a five class multinomial with two missing observations  $v_1$  and  $v_2$

$$l_c(\theta) = (v_2 + x_4) \ln(\theta) + (x_2 + x_3) \ln((1 - \theta)) + \text{constant}$$

setting the derivative equal to zero, as previously done, yields

$$\begin{aligned}\frac{\partial l(\theta)}{\partial \theta} &\stackrel{\text{set}}{=} 0 \\ -\frac{x_2 + x_3}{1 - \theta} + \frac{v_2 + x_4}{\theta} &= 0\end{aligned}$$

so that the ML estimate for  $\theta$  is

$$\frac{v_2 + x_4}{v_2 + x_2 + x_3 + x_4}.$$

Conditioning on the current value of the parameter  $\theta^{(k)}$  and the observed data  $x_1$ , the E-step of the iterative EM algorithm fills in the unobservable value with this expected value. It is seen that  $v_2$  is a binomial variable as part of  $x_1$ . Thus  $v_2$  is binomially distributed with  $n = x_1$  and  $\pi = \frac{0.25\theta}{0.5+0.25\theta}$ .

Let  $\theta^{(k)}$  be the value of  $\theta$  after  $k$  iterations, so that

$$\begin{aligned} E[v_1^{(k)} | \theta^{(k)}, x_1] &= x_1 \left( \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{4}\theta^{(k)}} \right) & E[v_2^{(k)} | \theta^{(k)}, x_1] &= x_1 \left( \frac{\frac{1}{4}\theta^{(k)}}{\frac{1}{2} + \frac{1}{4}\theta^{(k)}} \right) \\ &= v_1^{(k)} & &= v_2^{(k)} \end{aligned} \quad (2.7)$$

since  $v_1 + v_2 = x_1$  and  $\theta = \theta^{(k)}$ .

The M-step proceeds by maximizing  $E_{\theta}^{(k)}(l_c(\theta))$  and because  $l_c(\theta)$  is a linear equation,

$$E_{\theta}^{(k)}(l_c(\theta)) = (E_{\theta}^{(k)}[v_2] + x_4) \ln(\theta) + (x_2 + x_3) \ln(1 - \theta)$$

so that the maximum occurs at

$$\theta^{(k+1)} = \frac{v_2^{(k)} + x_4}{v_2^{(k)} + x_2 + x_3 + x_4}. \quad (2.8)$$

The above theory is now practically applied by distributing 197 entities into 4 categories

$$\begin{aligned} \mathbf{x} &= (x_1, x_2, x_3, x_4) \\ &= (125, 18, 20, 34) \end{aligned}$$

where  $\mathbf{x}$  is distributed exactly as in Equation 2.1 – 2.3. The initial estimate of  $\theta$  is chosen to be  $\theta^{(0)} = 0.1$  and  $E[v_2^{(0)} | \theta^{(0)}, x_1]$  is calculated from Equation 2.7 to be

$$\begin{aligned} E[v_2^{(0)} | \theta^{(0)}, x_1] &= x_1 \left( \frac{\frac{1}{4}\theta^{(0)}}{\frac{1}{2} + \frac{1}{4}\theta^{(0)}} \right) \\ &= 125 \left( \frac{\frac{1}{4}(0.1)}{\frac{1}{2} + \frac{1}{4}(0.1)} \right) \\ &= 5.952381. \end{aligned}$$

**Table 2.1:** Convergence of the EM algorithm for Example 3.

v	k	theta
5.952381	0	0.1
5.952381	1	0.5125229
25.49842	1	0.5125229
25.49842	2	0.6102501
29.223737	2	0.6102501
29.223737	3	0.624594
29.747172	3	0.624594
29.747172	4	0.6265252
29.81721	4	0.6265252
29.81721	5	0.6267822
29.82652	5	0.6267822
29.82652	6	0.6268163

The E-step of the algorithm has been performed to provide as initial guess for  $v_2$ , it's conditional expected value. The M-step now treats this estimate,  $v_2$ , as real data and improves the estimate for  $\theta$  using Equation 2.8.

$$\begin{aligned}
 \theta^{(1)} &= \frac{v_2^{(0)} + x_4}{v_2^{(0)} + x_2 + x_3 + x_4} \\
 &= \frac{5.952381 + 34}{25.49842 + 18 + 20 + 34} \\
 &= 0.5125229 .
 \end{aligned}$$

Continuing in this fashion, the value of  $\theta$  stabilizes to less than 0.01% percent change at 0.6268163 in 6 iterations. As can be noted from Table 2.1.

The programming details of this example can be found in Appendix A.

**Example 4** Exponential left- and right censored data *Flury and Zoppè (2000)*

It is assumed that the lifetime of a product follows an exponential distribution with parameter  $\eta$  having PDF

$$f(x, \eta) = \frac{1}{\eta} e^{-\frac{x}{\eta}} \text{ where } x \in (0, \infty); \eta > 0$$

and CDF

$$F(x, \eta) = 1 - \exp -\frac{x}{\eta}.$$

If a rv  $z$  has a upper-truncated exponential distribution with parameter  $\gamma$ , truncated at  $q$ . Then  $z$  has a PDF

$$g(z, \gamma) = \frac{f(z, \gamma)}{F(q, \gamma)}. \quad (2.9)$$

If  $z$  has a lower-truncated exponential distribution then

$$g(z, \gamma) = \frac{f(z, \gamma)}{1 - F(q, \gamma)}. \quad (2.10)$$

Now, consider  $n$  of these exponentially distributed products that are tested until they all fail. The failure times are recorded by the variable  $y$ . In a separate experiment  $m$  products are tested, but only the number of products that fail at time  $t$  will be recorded. The number will be denoted as  $r$ .

The missing data are the failure times of the products in the second experiment, denoted by the variable  $v$ . Based on the complete data, the following likelihood function is obtained

$$\begin{aligned} l_c(\eta; y, v) &= \ln \left\{ \prod_{i=1}^n [f(y_i, \eta)] \prod_{j=1}^m [f(v_j, \eta)] \right\} \\ &= \ln \left\{ \prod_{i=1}^n \left[ \frac{1}{\eta} e^{-\frac{y_i}{\eta}} \right] \prod_{j=1}^m \left[ \frac{1}{\eta} e^{-\frac{v_j}{\eta}} \right] \right\} \\ &= \sum_{i=1}^n \ln \left\{ \frac{1}{\eta} e^{-\frac{y_i}{\eta}} \right\} + \sum_{j=1}^m \ln \left\{ \frac{1}{\eta} e^{-\frac{v_j}{\eta}} \right\} \\ &= -n \ln \eta - \frac{\sum_{i=1}^n y_i}{\eta} - m \ln \eta - \frac{\sum_{j=1}^m v_j}{\eta} \\ &= -(n + m) \ln \eta - \frac{n\bar{y}}{\eta} - \frac{\sum_{j=1}^m v_j}{\eta}. \end{aligned} \quad (2.11)$$

Given a current value of the parameter  $\eta^{(k)}$  and the observed data, the E-step of the iterative EM algorithm, fills in the unobservable value with this expected value. The expected value is obtained by viewing the conditional density of  $v$  as a truncated density, truncated at  $t$  (Equation 2.9 and 2.10).

$$\sum_{j=1}^m E[v_j^{(k)} | \eta^{(k)} = \eta, m, r, t] = \left( \frac{r \int_0^t v^{(k)} \frac{1}{\eta} e^{-\frac{v^{(k)}}{\eta}} dv}{F(t; \eta)} \right) + \left( \frac{(m-r) \int_t^\infty v^{(k)} \frac{1}{\eta} e^{-\frac{v^{(k)}}{\eta}} dv}{F(t; \eta)} \right) \quad (2.12)$$

where

$$\begin{aligned} \int_0^t v^{(k)} \frac{1}{\eta} e^{-\frac{v^{(k)}}{\eta}} dv &= -v^{(k)} e^{-\frac{v^{(k)}}{\eta}} \Big|_0^t + \int_0^t e^{-\frac{v^{(k)}}{\eta}} dv \\ &= -te^{-\frac{t}{\eta}} - \eta e^{-\frac{v^{(k)}}{\eta}} \Big|_0^t \\ &= -te^{-\frac{t}{\eta}} - \eta e^{-\frac{t}{\eta}} + \eta \end{aligned} \quad (2.13)$$

and

$$\begin{aligned} \int_t^\infty v^{(k)} \frac{1}{\eta} e^{-\frac{v^{(k)}}{\eta}} dv &= -v^{(k)} e^{-\frac{v^{(k)}}{\eta}} \Big|_t^\infty + \int_t^\infty e^{-\frac{v^{(k)}}{\eta}} dv \\ &= te^{-\frac{t}{\eta}} - \eta e^{-\frac{v^{(k)}}{\eta}} \Big|_t^\infty \\ &= te^{-\frac{t}{\eta}} + \eta e^{-\frac{t}{\eta}} . \end{aligned} \quad (2.14)$$

Using Equation 2.13 and Equation 2.14 above, Equation 2.12 is further simplified

$$\begin{aligned} \sum_{j=1}^m E[v_j^{(k)} | \eta^{(k)} = \eta, r, m, t] &= \left( \frac{r \int_0^t v^{(k)} \frac{1}{\eta} e^{-\frac{v^{(k)}}{\eta}} dv}{F(t; \eta)} \right) + \left( \frac{(m-r) \int_t^\infty v^{(k)} \frac{1}{\eta} e^{-\frac{v^{(k)}}{\eta}} dv}{1 - F(t; \eta)} \right) \\ &= \left( \frac{r(-te^{-\frac{t}{\eta}} - \eta e^{-\frac{t}{\eta}} + \eta)}{1 - e^{-\frac{1}{\eta}t}} \right) + \left( \frac{(m-r)(te^{-\frac{t}{\eta}} + \eta e^{-\frac{t}{\eta}})}{e^{-\frac{t}{\eta}}} \right) \quad (2.15) \\ &= \left( \frac{-rte^{-\frac{t}{\eta}} + r\eta(-e^{-\frac{t}{\eta}} + 1)}{1 - e^{-\frac{1}{\eta}t}} \right) + \left( \frac{(m-r)(t + \eta)(e^{-\frac{t}{\eta}})}{e^{-\frac{t}{\eta}}} \right) \\ &= (m-r)(t + \eta) + r\eta + \left( \frac{-rte^{-\frac{t}{\eta}}}{1 - e^{-\frac{t}{\eta}}} \right) \\ &= v^{(k)} . \end{aligned} \quad (2.16)$$

From Equation 2.11

$$l_c(\eta; y, v) = -(n + m) \ln \eta - \frac{n\bar{y}}{\eta} - \frac{\sum_{j=1}^m v_j}{\eta},$$

the following is now obtained by applying the results of Equation 2.15

$$\begin{aligned} E_{v^{(k)}|r,m,t,\eta^{(k)}}(l_c) &= -(n + m) \ln \eta - \frac{n\bar{y}}{\eta} - \left(\frac{1}{\eta}\right)v^{(k)} \\ &= -(n + m) \ln \eta - \frac{n\bar{y}}{\eta} - \left(\frac{1}{\eta}\right)\left[(m - r)(t + \eta^{(k)}) + r\eta^{(k)} + \left(\frac{-rte^{-\frac{t}{\eta^{(k)}}}}{1 - e^{-\frac{t}{\eta^{(k)}}}}\right)\right] \\ &= -(n + m) \ln \eta - \left(\frac{1}{\eta}\right)\left[n\bar{y} + (m - r)(t + \eta^{(k)}) + r\eta^{(k)} + \left(\frac{-rte^{-\frac{t}{\eta^{(k)}}}}{1 - e^{-\frac{t}{\eta^{(k)}}}}\right)\right] \\ &= -(n + m) \ln \eta - \left(\frac{1}{\eta}\right)[n\bar{y} + v^{(k)}] \end{aligned}$$

The  $k$ th  $M$  step determines the maximum with respect to the variable  $\eta$ , which given  $\eta^{(k)}$ , occurs at

$$\begin{aligned} \frac{\partial E_{v^{(k)}|r,m,t,\eta^{(k)}}(l_c)}{\partial \eta} &\stackrel{\text{set}}{=} 0 \\ \frac{-(n + m)}{\eta} + \frac{n\bar{y} + v^{(k)}}{\eta^2} &= 0 \\ \frac{n\bar{y} + v^{(k)}}{\eta^2} &= \frac{n + m}{\eta} \quad (2.17) \\ \frac{n\bar{y} + (m - r)(t + \eta^{(k)}) + r\eta^{(k)} + \left(\frac{-rte^{-\frac{t}{\eta^{(k)}}}}{1 - e^{-\frac{t}{\eta^{(k)}}}}\right)}{n + m} &= \eta^{(k+1)}. \end{aligned}$$

Starting with a positive number  $\eta = \eta^{(0)}$ , the algorithm is iterated until convergence.

The above results are further explored at the hand of a practical example. The lifetime of  $n = 15$  devices are simulated from the exponential distribution with  $\eta = 8$ . In keeping with notation the lifetimes of these devices are denoted with  $y$ . Also simulated from this distribution are the lifetimes,  $v$ , of  $m = 20$  devices. These devices are then left- and right censored at  $t = 3$ ,



**Table 2.3:** Number of the failures for Example 4.

The number of failures	y-bar
5	7.9464866

so that  $v$  is unobserved. The simulated data reveals, in Table 2.3, that  $r = 5$  of the  $m = 20$  devices failed before  $t = 3$  and that the average lifetime of the  $y$  devices are 7.9464866.

The initial estimate of  $\eta$  is chosen to be  $\eta^{(0)} = 2$  and  $\sum_{j=1}^m E[v_j^{(k)} | \eta^{(k)}, r, m, t]$  is calculated from Equation 2.15

$$\begin{aligned}
 & \sum_{j=1}^m E[v_j^{(0)} | \eta^{(0)}, r, m, t] \\
 &= \sum_{j=1}^{20} E[v_j^{(0)} | 2, 5, 20, 3] \\
 &= (20 - 5)(3 + 2) + 5(2) + \left( \frac{-5(3)e^{-\frac{3}{2}}}{1 - e^{-\frac{3}{2}}} \right) \\
 &= 80.691746.
 \end{aligned}$$

Here the E-step of the algorithm has been performed and a initial guess for the unobserved data  $v$ , is it's conditional expected value. In the M-step, this estimate is treated as real data and the estimate for  $\theta$  is improved upon, using Equation 2.17.

$$\begin{aligned}
 \eta^{(1)} &= \frac{n\bar{y} + \sum_{j=1}^m E[v_j^{(0)} | \eta^{(0)}, r, m, t]}{n + m} \\
 &= \frac{15(7.9464866) + 80.691746}{15 + 20} \\
 &= 5.7111156.
 \end{aligned}$$

Continuing in this fashion, it is seen from Table 2.5 that the value of  $\eta$  stabilizes to less than 0.1% percent change at 8.5623948 in 11 iterations.

The programming details of this example can be found in Appendix A.

The steps of the EM algorithm can now be given in mathematical notation.

### The EM algorithm

Given a likelihood function  $L_c(\eta; y, v)$  where  $y$  is the observed data and  $v$  the unobserved data, the ML estimates are determined by the marginal likelihood of the observed data  $L_c(\eta; y)$ .

**Table 2.5:** Convergence of the EM algorithm for Example 4.

v	k	theta
80.691746	0	2
80.691746	1	5.7111156
137.51312	1	5.7111156
137.51312	2	7.3345834
162.00889	2	7.3345834
162.00889	3	8.0344626
172.55128	3	8.0344626
172.55128	4	8.3356737
177.0862	4	8.3356737
177.0862	5	8.4652428
179.03658	5	8.4652428
179.03658	6	8.5209679
179.87533	6	8.5209679
179.87533	7	8.5449324
180.23603	7	8.5449324
180.23603	8	8.5552379
180.39114	8	8.5552379
180.39114	9	8.5596696
180.45784	9	8.5596696
180.45784	10	8.5615753
180.48652	10	8.5615753
180.48652	11	8.5623948

1. Calculate  $E_{v|y,\eta^{(k)}}(l_c)$ , the expected value of the log-likelihood function with respect to the conditional distribution of  $v$ , given  $y$  and the current estimate of  $\eta^{(k)}$
2. Find the next iterate of the parameter  $\eta^{(k+1)}$  by maximizing this quantity

The algorithm then iterates between the above two steps until the specified convergence is reached in terms of the difference between  $\eta^{(k)}$  and  $\eta^{(k+1)}$ .

Example 4 is a classical example in the literature on an AT with a binary outcome; know as one-shot device testing. In the next section the same logic is applied to lifetimes that are Log-normally distributed.

## 2.2 The EM algorithm for Log-normal, left- and right censored data

[Nelson and Kielpinski \(1976\)](#) presented theory for ALTs where a linear link-function is used for Log-normal data in a right-censored scheme. [Meeker \(1984\)](#) focused on planning ALTs for Weibull and Log-normal distributions in a Type I censoring scheme. [Li et al. \(2015\)](#) described a sampling plan for ALTs for Log-normal data that is right censored, when the acceleration factors are both known and unknown. [Balakrishnan and Ling \(2014\)](#) implemented the EM algorithm for Gamma distributed lifetimes. [Balakrishnan and Ling \(2012\)](#) and [Balakrishnan and Ling \(2013\)](#) for the Exponential and Weibull distributed lifetimes, respectively. [Vernic et al. \(2009\)](#) implemented the EM algorithm to a two component Log-normal mixture model and a two components mixture model having truncated Log-normal mixture models. The results given in last named article can be used in one-shot device testing where the Log-normal lifetimes are both left- and right censored.

Consider  $M$  devices placed under a single stress factor  $x_i$  and tested at time  $T_i$ . The combination of  $x_i$  (different levels of the single stress) and  $T_i$  (different inspection times) define certain test conditions for current group  $i$  where  $i = 1, 2, \dots, I$ . In each test condition there are  $M$  devices, leading to  $MI$  devices being tested. If the device survives past the inspection time, the device is denoted as a success. If the device failed before the inspection time, it will be denoted as a failure. For each stress  $x_i$ , the number of failures  $r_i$  (out of  $M$ ) is recorded at each inspection time  $T_i$ . It is assumed that the true, but unobserved, lifetimes

$v_{ij}$  where  $i = 1, 2, \dots, I$  and  $j = 1, 2, \dots, M$  are independent and identically Log-normal distributed variables with PDF

$$f(v) = \frac{1}{v\sigma\sqrt{2\pi}} e^{-\left[\frac{(\ln v - \mu)^2}{2\sigma^2}\right]}, \quad v > 0;$$

where the location parameter is  $\mu \in \mathbb{R}$  and the scale parameter  $\sigma > 0$ .

The stress factor can now be related to one or both of the parameters, through a link function, as described in Chapter 1. In this instance the location parameter is related to the stress factor through a linear link-function

$$\mu_i = \beta_0 + \beta_1 x_i.$$

The EM algorithm will estimate the parameters  $\beta_0$  and  $\beta_1$  in the M-step while the expected lifetimes will be approximated in the E-step.

The log-likelihood function based on the complete data is given by

$$\begin{aligned} l_c(\beta_0, \beta_1) &= \sum_{i=1}^I \sum_{j=1}^M \ln \left( \frac{1}{v_{ij}\sigma\sqrt{2\pi}} e^{-\frac{(\ln v_{ij} - \mu)^2}{2\sigma^2}} \right) \\ &= \sum_{i=1}^I \sum_{j=1}^M \left[ -\ln(v_{ij}) - \ln(\sigma\sqrt{2\pi}) - \frac{(\ln(v_{ij}) - \beta_0 - \beta_1 x_i)^2}{2\sigma^2} \right]. \end{aligned} \quad (2.18)$$

The parameter estimates are obtained through the following equations

$$\begin{aligned} \frac{\partial l_c(\beta_0, \beta_1)}{\partial \beta_0} &\stackrel{set}{=} 0 \\ \sum_{i=1}^I \sum_{j=1}^M \left[ \frac{\ln(v_{ij}) - \beta_0 - \beta_1 x_i}{\sigma^2} \right] &= 0 \\ \frac{\sum_{i=1}^I \sum_{j=1}^M [\ln(v_{ij}) - \beta_1 x_i]}{IM} &= \beta_0. \end{aligned} \quad (2.19)$$

For the solution of  $\beta_1$  an iterative formula is adopted

$$\begin{aligned}
 \frac{\partial l_c(\beta_0, \beta_1)}{\partial \beta_1} &\stackrel{set}{=} 0 \\
 \sum_{i=1}^I \sum_{j=1}^M \left[ \frac{(\ln(v_{ij}) - \beta_0 - \beta_1 x_i) x_i}{\sigma^2} \right] &= 0 \\
 \frac{\sum_{i=1}^I x_i \sum_{j=1}^M [\ln(v_{ij}) - \beta_0^{(k)}]}{\sum_{i=1}^I M x_i^2} &= \beta_1^{(k+1)}.
 \end{aligned} \tag{2.20}$$

Recall that if  $v \sim \ln N(\mu; \sigma)$  then  $\ln(v) \sim N(\mu; \sigma)$ . Thus, making the necessary substitution  $u = \ln v$ ,  $u$  becomes a normally distributed rv. From Equation 2.19 and Equation 2.20 it is seen that  $E[\ln(v_{ik}) | \mu_i, T_i, M, r_i, x_i]$  will be required in the expectation step. It is calculate as follows,

$$\begin{aligned}
 &\sum_{j=1}^M E[\ln(v_{ij}) | \mu_i, T_i, M, r_i, x_i] \\
 = &\frac{r_i \int_0^{T_i} \ln(v_i) \frac{1}{v_i \sigma \sqrt{2\pi}} e^{-\frac{(\ln(v_i) - \mu_i)^2}{2\sigma^2}} dv}{\int_0^{T_i} \frac{1}{v_i \sigma \sqrt{2\pi}} e^{-\frac{(\ln(v_i) - \mu_i)^2}{2\sigma^2}} dv} + \frac{(M - r_i) \int_{T_i}^{\infty} \ln(v_i) \frac{1}{v_i \sigma \sqrt{2\pi}} e^{-\frac{(\ln(v_i) - \mu_i)^2}{2\sigma^2}} dv}{\int_{T_i}^{\infty} \frac{1}{v_i \sigma \sqrt{2\pi}} e^{-\frac{(\ln(v_i) - \mu_i)^2}{2\sigma^2}} dv}
 \end{aligned} \tag{2.21}$$

Now apply the transformation  $u = \ln(v)$  so that  $v du = dv$

$$\begin{aligned}
 = &\frac{r_i \int_{-\infty}^{\ln(T_i)} u_i \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(u_i - \mu_i)^2}{2\sigma^2}} du}{\int_{-\infty}^{\ln(T_i)} \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(u_i - \mu_i)^2}{2\sigma^2}} du} + \frac{(M - r_i) \int_{\ln(T_i)}^{\infty} u_i \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(u_i - \mu_i)^2}{2\sigma^2}} du}{\int_{\ln(T_i)}^{\infty} \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(u_i - \mu_i)^2}{2\sigma^2}} du} \\
 &= \text{term1} + \text{term2}.
 \end{aligned}$$

Now let  $w = \frac{u - \mu_i}{\sigma}$  so that  $dw = \frac{1}{\sigma} du$  and  $u = w\sigma + \mu_i$

$$\begin{aligned}
 term1 &= \frac{r_i \int_{-\infty}^{\frac{\ln(T_i) - \mu_i}{\sigma}} (w_i \sigma + \mu_i) \frac{1}{\sqrt{2\pi}} e^{-\frac{w_i^2}{2}} dw}{\int_{-\infty}^{\frac{\ln(T_i) - \mu_i}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{w_i^2}{2}} dw} \\
 &= \frac{r_i \int_{-\infty}^{\frac{\ln(T_i) - \mu_i}{\sigma}} w_i \sigma \frac{1}{\sqrt{2\pi}} e^{-\frac{w_i^2}{2}} dw}{\phi\left(\frac{\ln(T_i) - \mu_i}{\sigma}\right)} + r_i \mu_i \\
 &= \frac{r_i \left[ \frac{-\sigma}{\sqrt{2\pi}} e^{-\frac{w_i^2}{2}} \right]_{-\infty}^{\frac{\ln(T_i) - \mu_i}{\sigma}}}{\phi\left(\frac{\ln(T_i) - \mu_i}{\sigma}\right)} + r_i \mu_i \\
 &= \frac{r_i \left[ \frac{-\sigma}{\sqrt{2\pi}} \left( e^{-\frac{-(\ln(T_i) - \mu_i)^2}{2\sigma^2}} \right) \right]}{\phi\left(\frac{\ln(T_i) - \mu_i}{\sigma}\right)} + r_i \mu_i
 \end{aligned} \tag{2.22}$$

$$\begin{aligned}
 term2 &= \frac{(M - r_i) \int_{\frac{\ln(T_i) - \mu_i}{\sigma}}^{\infty} (w\sigma + \mu_i) \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} dw}{\int_{\frac{\ln(T_i) - \mu_i}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} dw} \\
 &= (M - r_i) \mu_i + \frac{(M - r_i) \int_{\frac{\ln(T_i) - \mu_i}{\sigma}}^{\infty} (w\sigma) \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} dw}{1 - \Phi\left(\frac{\ln(T_i) - \mu_i}{\sigma}\right)} \\
 &= (M - r_i) \mu_i + \frac{(M - r_i) \left[ \frac{-\sigma}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} \right]_{\frac{\ln(T_i) - \mu_i}{\sigma}}^{\infty}}{1 - \Phi\left(\frac{\ln(T_i) - \mu_i}{\sigma}\right)} \\
 &= (M - r_i) \mu_i + \frac{(M - r_i) \left[ \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{-(\ln(T_i) - \mu_i)^2}{2\sigma^2}} \right]}{1 - \Phi\left(\frac{\ln(T_i) - \mu_i}{\sigma}\right)}.
 \end{aligned} \tag{2.23}$$

Using the above Equation 2.22 and Equation 2.23, Equation 2.21 is completed

$$\begin{aligned}
 &\sum_{j=1}^M E[\ln(v_{ij}) | \mu_i, T_i, M, r_i, x_i] \\
 &\frac{r_i \left[ \frac{-\sigma}{\sqrt{2\pi}} \left( e^{-\frac{-(\ln(T_i) - \mu_i)^2}{2\sigma^2}} \right) \right]}{\phi\left(\frac{\ln(T_i) - \mu_i}{\sigma}\right)} + r_i \mu_i + (M - r_i) \mu_i + \frac{(M - r_i) \left[ \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{-(\ln(T_i) - \mu_i)^2}{2\sigma^2}} \right]}{1 - \Phi\left(\frac{\ln(T_i) - \mu_i}{\sigma}\right)}.
 \end{aligned} \tag{2.24}$$

If the values of  $\beta_0$  and  $\beta_1$  at the  $k$ th-step are  $\beta_0^{(k)}$  and  $\beta_1^{(k)}$ , then the  $(k + 1)$ -step of the EM algorithm proceeds by:

1. Computing the missing data in Equation (2.24) by using  $\beta_0^{(k)}$  and  $\beta_1^{(k)}$  for  $i = 1, 2, \dots, I$
2. Starting with  $\beta_0^{(k)}$  for fixed  $E[\ln(v_{ij})|\mu_i, T_i, M, r_i, x_i]$   $i = 1, 2, \dots, I$  find  $\beta_1^{(k+1)}$  iteratively from Equation (2.20)
3. Given  $\beta_1^{(k+1)}$ , compute  $\beta_0^{(k+1)}$  from Equation (2.19)
4. Repeat steps 1–3 until convergence occurs.

## 2.3 Summary

A brief glimpse into one of the estimation methods for a binary response variable AT has now been given. It is seen that the EM algorithm provides flexibility with regards to the type of life-time distribution, link-function, number of stress factors and estimable parameters assumed. It is easy to see where the method may be extended for the above mentioned, although the derivations will become more complex and additional methods such as the Newton Raphson algorithm will be needed [Balakrishnan and Ling \(2013\)](#). The same estimation methods applied in this Chapter can be in the next chapter where the data will be exponentially distributed, a left-and right censored scheme will be assumed for an AT where a single stress factor will be used and a single parameter will be related to this stress factor via a power link function.

## Chapter 3

# Simulation exploration

Now that a thorough understanding of ATs have been obtained, and more specifically of one-shot device testing, the question presented in the introduction can be returned to. Would it be possible to use the data available on high-default portfolios, to draw inference on low-default portfolios as is done in accelerated testing? Could some link be established between the high frequency of observations for the high-default portfolios and the low frequency, censored observations of the low-default portfolios? Could this be achieved by applying the ATs framework to the problem, thus presenting a different method in dealing with insufficient data for low-default portfolios?

These questions are explored in this chapter by way of simulation studies. Section 3.1, introduces the Log-normal stock price of a company and the outline of the proposed model. Section 3.2 implements the model on simulated data and reports the results. Note that all programming information can be found in Appendix B.

### 3.1 The log-returns of a company

Consider the percentage return of a company, and assume that the percentage returns are Normally distributed

$$r_t \sim N(\mu, \sigma) \tag{3.1}$$

the Normal PDF given in (3.2)



$$f(r) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left[\frac{(r-\mu)^2}{2\sigma^2}\right]}, \quad (3.2)$$

where the location parameter is  $\mu \in \mathbb{R}$  and the scale parameter  $\sigma > 0$ .

Following from (3.1) the price of said stock is then Log-normally distributed

$$\begin{aligned} r_t &= \ln \frac{p_t}{p_{t-1}} \\ \therefore p_t &= p_{t-1} e^{r_t} \\ &\sim \ln N(\mu, \sigma). \end{aligned} \quad (3.3)$$

To see if the above assumptions are valid, a test of Normality is done on actual stock data. The 6 month stock data of 431 companies trading in the health sector was obtained along with the market capitalization (CAP) of each company. An excerpt of the data is given in Table 3.1.

**Table 3.1:** An excerpt of the 6 month stock data for 431 companies with their market CAP given in the first row.

Obs	Company	AAC	ABBV	ABC
1	.	376613466.30	92484920568.00	16311214010.00
2	13MAY2016	20.35	62.15	74.71
3	16MAY2016	20.49	61.91	73.76
4	17MAY2016	20.61	62.44	74.40
5	18MAY2016	19.70	59.87	74.20
6	19MAY2016	19.82	60.06	74.46
7	20MAY2016	19.60	59.63	75.14
8	23MAY2016	19.81	59.75	75.00
9	24MAY2016	20.00	59.73	74.37
10	25MAY2016	21.58	61.08	74.59
11	26MAY2016	21.28	61.00	74.04
12	27MAY2016	20.60	61.99	73.78
13	31MAY2016	21.52	62.71	74.20
14	01JUN2016	21.08	62.93	74.76
15	02JUN2016	22.15	62.87	75.81
16	03JUN2016	22.13	64.76	76.70
17	06JUN2016	21.39	62.13	76.30
18	07JUN2016	21.71	62.59	76.39
19	08JUN2016	22.21	62.21	76.31

The test for Normality was performed on the returns of five of the 431 companies. The

p-values of the Shapiro-Wilk Normality tests are given in Table 3.3. It can be seen that based on this p-value 2 of the 5 reject for Normality. For these two companies, the full output is given below. Upon closer inspection it is seen that they do not reject for the Kolmogorov-Smirnov test and the model continuous under the assumption of Normal returns.

**Table 3.3:** The p-value of the Shapiro-Wilk Normality tests.

Obs	JNJ	AGRX	MNOV	NDRM	TLGT
1	0.022427	0.50030	0.10968	0.40769	.004073165

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.975816	Pr < W	0.0224
Kolmogorov-Smirnov	D	0.06471	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.10132	Pr > W-Sq	0.1088
Anderson-Darling	A-Sq	0.621744	Pr > A-Sq	0.1039

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.967848	Pr < W	0.0041
Kolmogorov-Smirnov	D	0.069769	Pr > D	0.1320
Cramer-von Mises	W-Sq	0.163666	Pr > W-Sq	0.0168
Anderson-Darling	A-Sq	0.965512	Pr > A-Sq	0.0159

In order to investigate and apply reliability theory methods to stock price data, the following needs to be considered:

- **Stress factor**

In a single stress model, as will be implemented in this chapter, the life-stress relationship needs to, in essence, regress the parameter(s) of a distribution on a stress factor. Within the frame work of the current problem, market CAP for each company is chosen as the factor displaying how much "stress" each company is operating under. With the understanding that companies with larger market CAP perform more stable in said market, thus are operating under lower stress.

- **The parameters under consideration**

Intuitively it is expected that a small company (smaller market CAP), being the stressed version of a large company, should exhibit higher volatility due to market fluctuations. This will be exhibited through  $\sigma$ .  $\sigma$  Will be chosen larger or smaller when simulating data for small and large companies respectively .

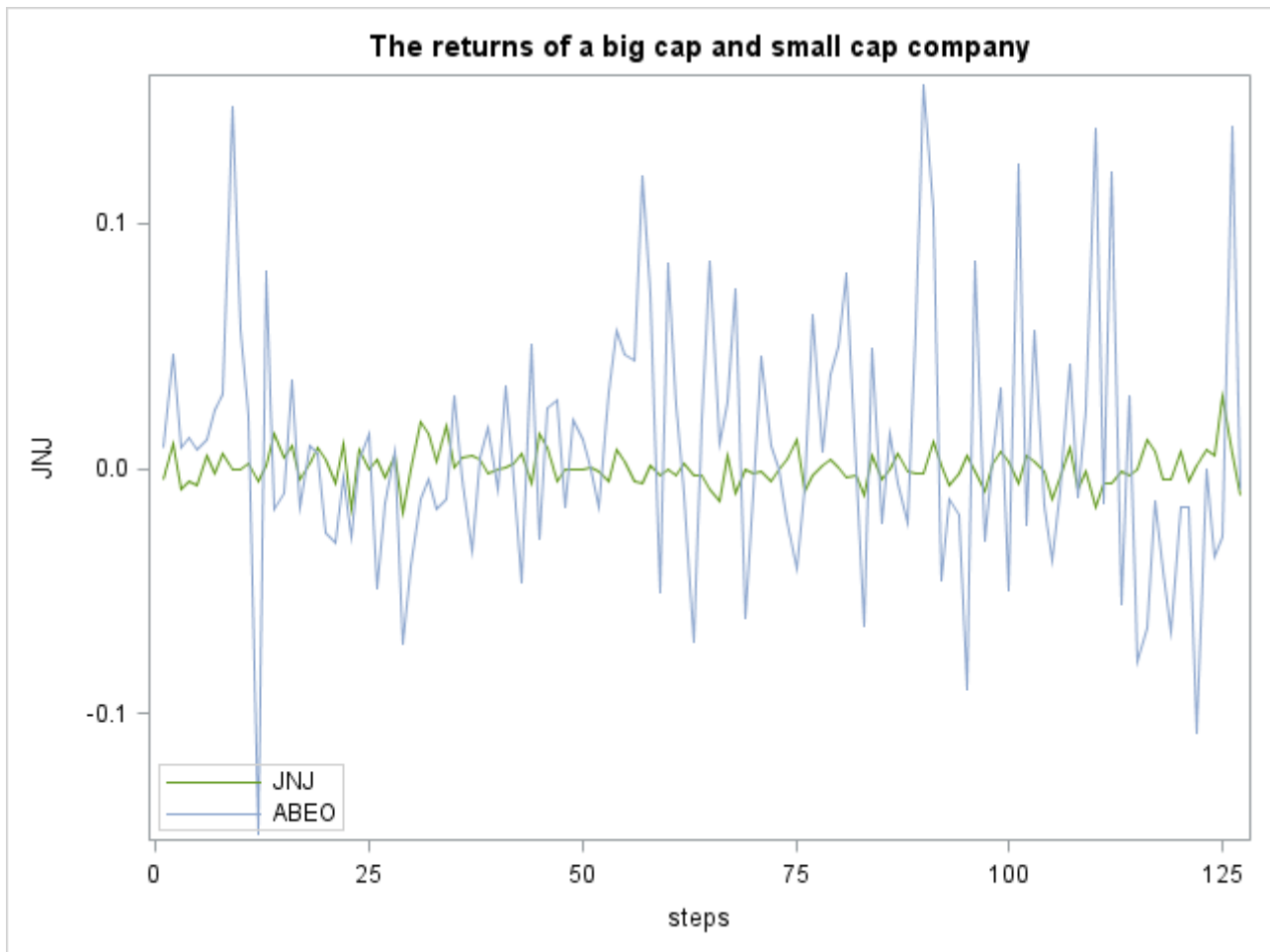
- **Failure of a stock price**

Finally it is necessary to define a failure for such a process. This is done in the following way. For a highly stressed company, that is to say small company, the percentage return is calculated. The loss of this small company is then calculated as  $L_t = -R_t$ . The  $(1 - \alpha)$ th percentile of the loss is denoted and any percentage loss, for any size company, above this now fixed percentile value, is considered a failure. The stock prices are observed at certain intervals (inspection times) and it is then noted how many different companies (different market CAPs) exhibited a loss above this fixed threshold. With large companies having the possibility of never displaying a loss above this level. Only the number of companies experiencing a loss above this level is known, while the exact time until this level was reached, is not known. Also, the company might have experienced this loss well before the inspection time, or might only experience it after the exception time. Leading to observations being left- and right censored.

A visual inspection may assist in deciding if the premise that smaller companies are more stressed than larger companies could be entertained. Figure 3.1 shows the Normal returns for a small CAP and large CAP company. The market CAP of the companies are displayed in Table 3.7.

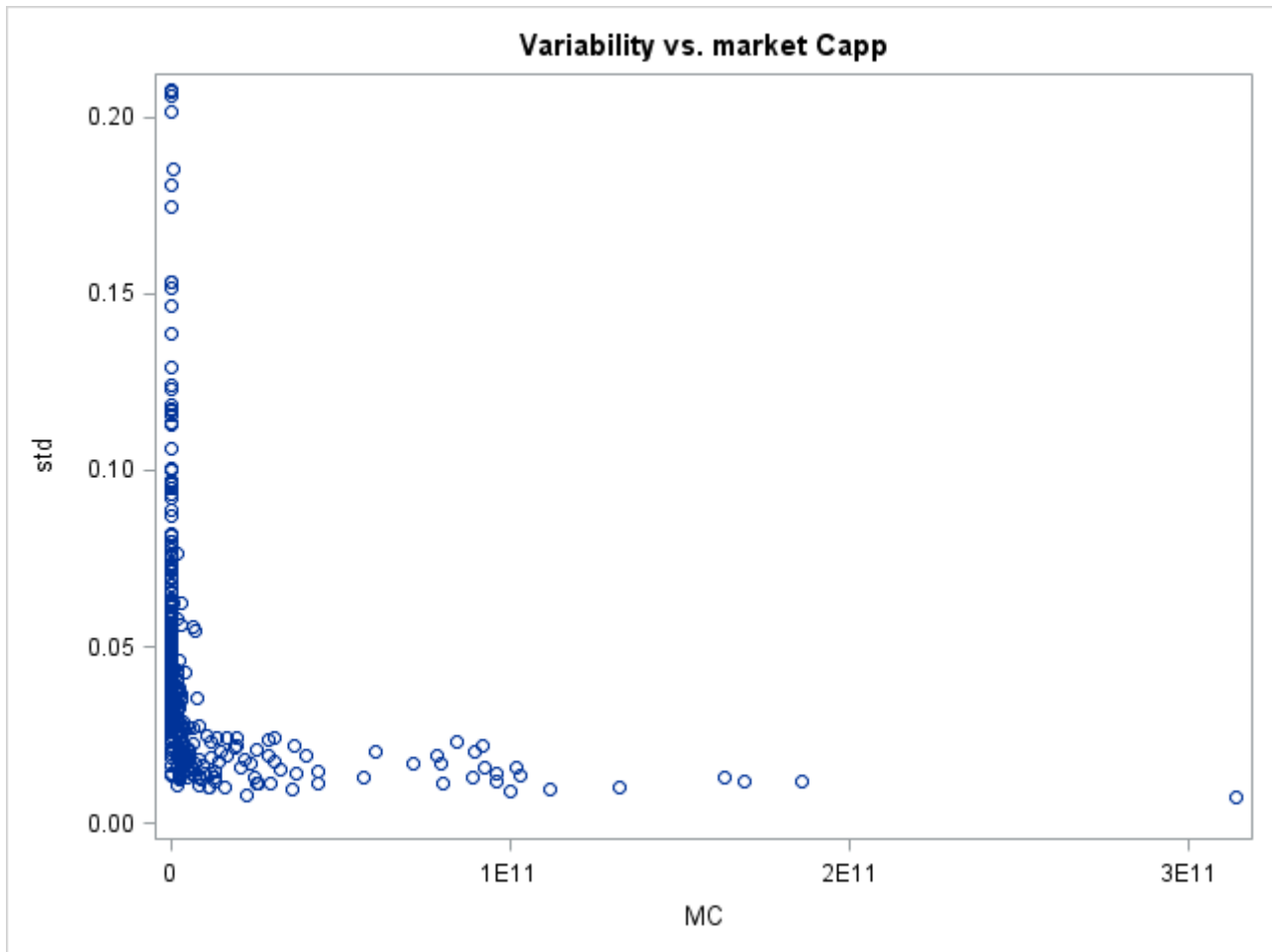
**Table 3.7:** The market CAP of two companies for which the returns are compared.

Obs	JNJ	ABEO
1	314000000000	213015214.1



**Figure 3.1:** The returns of a big CAP company vs. that of small CAP company that are given in the table above.

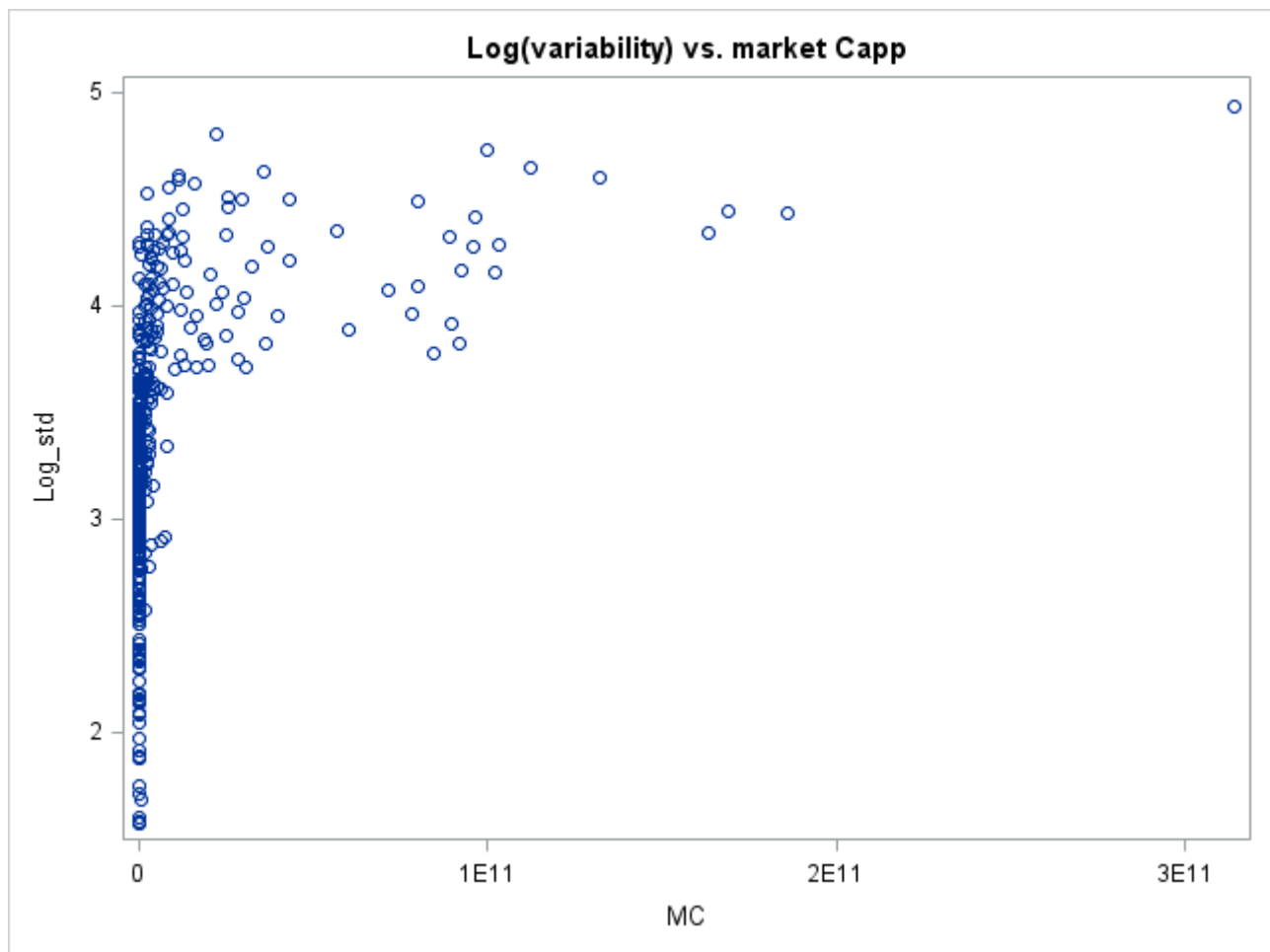
It appears as though companies that have a smaller market CAP are more volatile in relation to that of companies with large market CAP. The phrase in relation is important as it is noted that the volatility is mean reverting towards zero and is seldom less than, or greater than  $-0.15$



**Figure 3.2:** The standard deviation of the returns vs. the market CAP.

and 0.15 respectively. To further investigate what was intuitively and visually inspected above, the standard deviation of the Normal returns,  $\sigma$ , are regressed on the market CAP. First note that the standard deviations are very sensitive to a change in market CAP as can be seen in Figure 3.2 where the standard deviation ranges from 0 to 0.2 for market CAP values ranging from 0 to 3 billion.

The standard deviation of the log-returns explode as the market CAP approach zero. Confirming, once again, that as market CAP decrease the returns of a company become more volatile. It is difficult to simulate data when working with such a sensitive scale. To make the data more tractable a negative log transformation is performed on the standard deviation of the



**Figure 3.3:** The log transformed standard deviation of the returns vs. the market CAP.

returns, see now the updated Figure 3.3

A non-linear model is now used to regress the log of the standard deviation of the returns on market CAP

$$-\log(\sigma) = \alpha + \beta \ln(\text{MarketCAP}) .$$

The parameter estimates are given in Table 3.9

**Table 3.9:** The parameter estimates of the non-linear model.

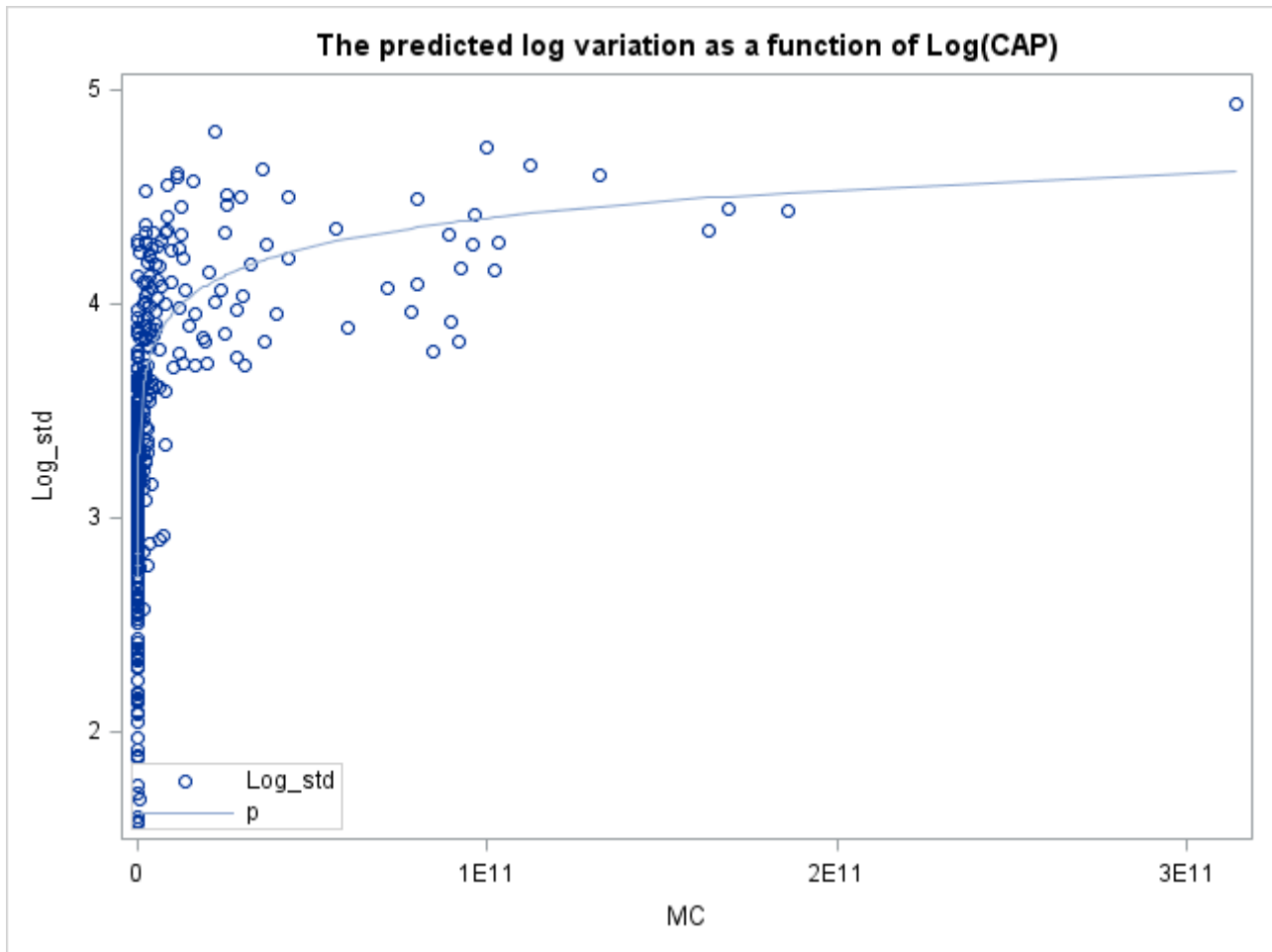
Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
beta	0.1931	0.00827	0.1769	0.2094
alpha	-0.4899	0.1659	-0.8159	-0.1639

so that

$$\begin{aligned}
 -\log(\sigma) &= \alpha + \beta \ln(\text{MarketCAP}) \\
 &= -0.4899 + 0.1931 \ln(\text{MarketCAP}) .
 \end{aligned}
 \tag{3.4}$$

A graphical representation of the fit is given in Figure 3.4





**Figure 3.4:** The negative log transformed standard deviation of the returns vs. the market CAP.

It is now possible to simulate realistic data that compares well to that which is observed in the stock market. In the next section such data is simulated using the assumptions introduced in this section.

## 3.2 Simulation results

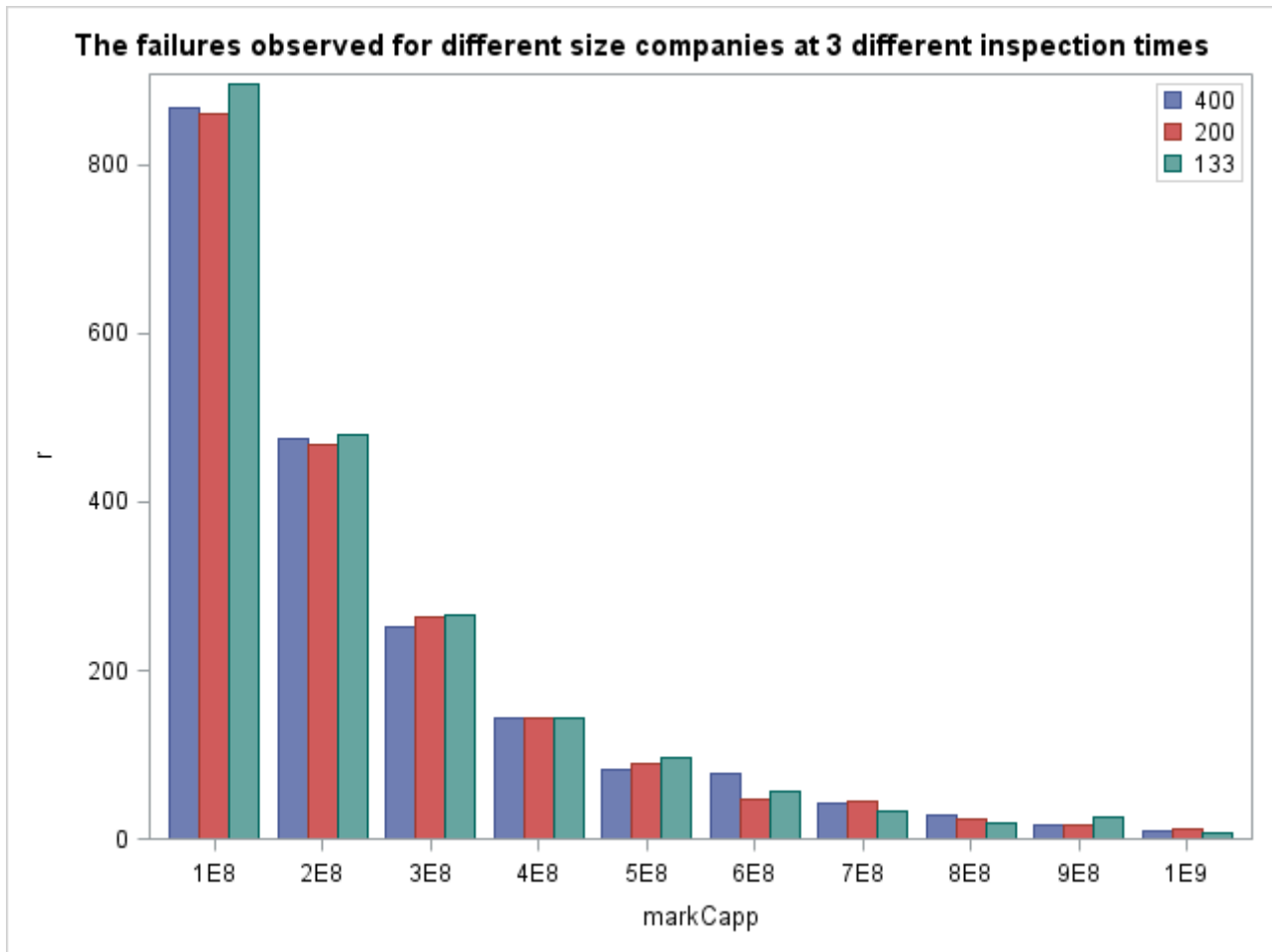
Following the model assumptions set out in Section 3.1 and adopting notation discussed in Section 2.2, the Normal returns of  $M = 1000$  companies are simulated, sigma as in Equation 3.4 and  $\mu$  assumed zero. For a highly stressed company, that is to say small company, the percent-

age return is calculated. The loss of this small company is then calculated as  $L_t = -R_t$ . The  $(1 - 0.05)$ th percentile of the loss is denoted. Any percentage loss, for any size/stressed company, above this now fixed percentile value, is considered a failure. These returns are inspected at 3 different inspection times,  $\mathbf{T} = (400, 200, 133)$ , during a 400 day period. The companies operate under 100 different stress levels i.e. market CAPs,  $\mathbf{x} = (0.1, 0.11, 0.12, \dots, 1)$  in billions. This leads to 300 test conditions. So that  $I = 300$  with  $i = 1, 2, \dots, I$ . The simulation results obtained for 30 of the 300 test conditions are given below.

Obs	markCapp	r	IT	M
1	100000000	891	400	1000
2	100000000	873	200	1000
3	100000000	870	133	1000
4	200000000	504	400	1000
5	200000000	514	200	1000
6	200000000	474	133	1000
7	300000000	265	400	1000
8	300000000	244	200	1000
9	300000000	265	133	1000
10	400000000	153	400	1000
11	400000000	138	200	1000
12	400000000	153	133	1000
13	500000000	83	400	1000
14	500000000	94	200	1000
15	500000000	97	133	1000
16	600000000	53	400	1000
17	600000000	50	200	1000

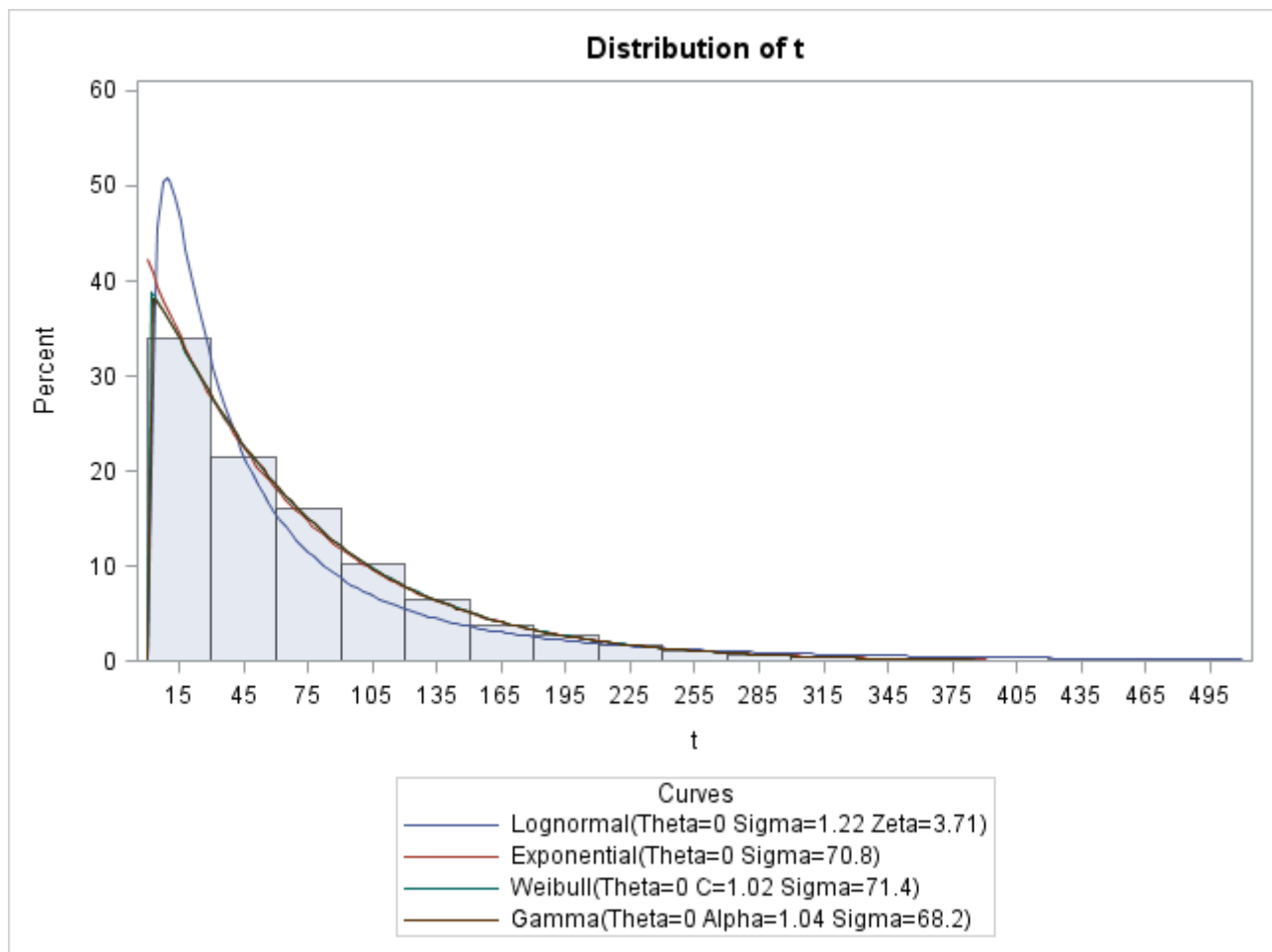
Obs	markCapp	r	IT	M
18	600000000	43	133	1000
19	700000000	35	400	1000
20	700000000	32	200	1000
21	700000000	27	133	1000
22	800000000	25	400	1000
23	800000000	16	200	1000
24	800000000	21	133	1000
25	900000000	15	400	1000
26	900000000	15	200	1000
27	900000000	14	133	1000
28	1000000000	14	400	1000
29	1000000000	15	200	1000
30	1000000000	16	133	1000

The data is graphically represented in [3.5](#) from where it can be seen that the number of failures decrease as companies become less stressed.



**Figure 3.5:** The failure data for companies of different market CAPs evaluated at different inspection times.

The following step towards implementing a one-shot device testing scheme to the data is to investigate the lifetime distribution. This distribution is assumed to have the same PDF across the different levels of stress although the parameters may differ. The stress, market CAP, is thus set and it is observed how long it takes for two thousand companies, operating under such stress, to fail. The failure threshold still defined in terms of the smaller more stressed company. The empirical distribution is given in Figure 3.6



**Figure 3.6:** The empirical lifetime distribution with a Log-normal, Exponential, Weibull and Gamma distribution superimposed.

and it is tested whether the distribution could possible be of form Gamma, Weibull, Exponential or Log-normal. The results of these test are given below.

Goodness-of-Fit Tests for Gamma Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.01577822	Pr > D	>0.250
Cramer-von Mises	W-Sq	0.05313777	Pr > W-Sq	>0.500

Goodness-of-Fit Tests for Gamma Distribution				
Test	Statistic		p Value	
Anderson-Darling	A-Sq	0.53696894	Pr > A-Sq	0.191

Goodness-of-Fit Tests for Weibull Distribution				
Test	Statistic		p Value	
Cramer-von Mises	W-Sq	0.04810609	Pr > W-Sq	>0.250
Anderson-Darling	A-Sq	0.50222404	Pr > A-Sq	0.222

Goodness-of-Fit Tests for Exponential Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.01680341	Pr > D	>0.250
Cramer-von Mises	W-Sq	0.06842699	Pr > W-Sq	>0.500
Anderson-Darling	A-Sq	0.61994318	Pr > A-Sq	>0.250

Goodness-of-Fit Tests for Lognormal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.0731895	Pr > D	<0.010
Cramer-von Mises	W-Sq	3.5885961	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	23.2801268	Pr > A-Sq	<0.005

It is seen that only the Log-Normal distribution is rejected as fit for the data. It is now possible to use all of the above obtained information in the SAS reliability procedure. The reliability procedure is implemented with the Exponential distribution as lifetime distribution, since it was introduced in Chapter 2 and was not rejected as a possible fit for the lifetime

distribution. It is also a simple distribution in the sense that only one parameter can be related to the stress factor. For completeness, the PDF and CDF are given again below.

$$f(x, \eta) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}} \text{ where } x \in (0, \infty); \lambda > 0$$

$$F(x, \lambda) = 1 - \exp -\frac{x}{\lambda}.$$

Further more, a linear link function is implemented. However, it is seen that the model collapses for this link function (Table 3.16).

**Table 3.16:** The estimates obtained by implementing a linear link function.

Exponential Parameter Estimates				
Parameter	Estimate	Standard Error	Asymptotic Normal	
			95% Confidence Limits	
			Lower	Upper
Intercept	4.4913	0.0098	4.4720	4.5105
Mc	0.0000	0.0000	0.0000	0.0000
Shape	1.0000	0.0000	1.0000	1.0000

So a power link function is implemented.

**Table 3.18:** The estimates obtained by implementing a power link function.

Exponential Parameter Estimates				
Parameter	Estimate	Standard Error	Asymptotic Normal	
			95% Confidence Limits	
			Lower	Upper
Intercept	-34.6564	0.1566	-34.9633	-34.3495
Mc	2.1248	0.0081	2.1089	2.1407
Shape	1.0000	0.0000	1.0000	1.0000

The power link function relates the parameter  $\lambda$  of the exponential distribution to the stress factor, market CAP, in the following way

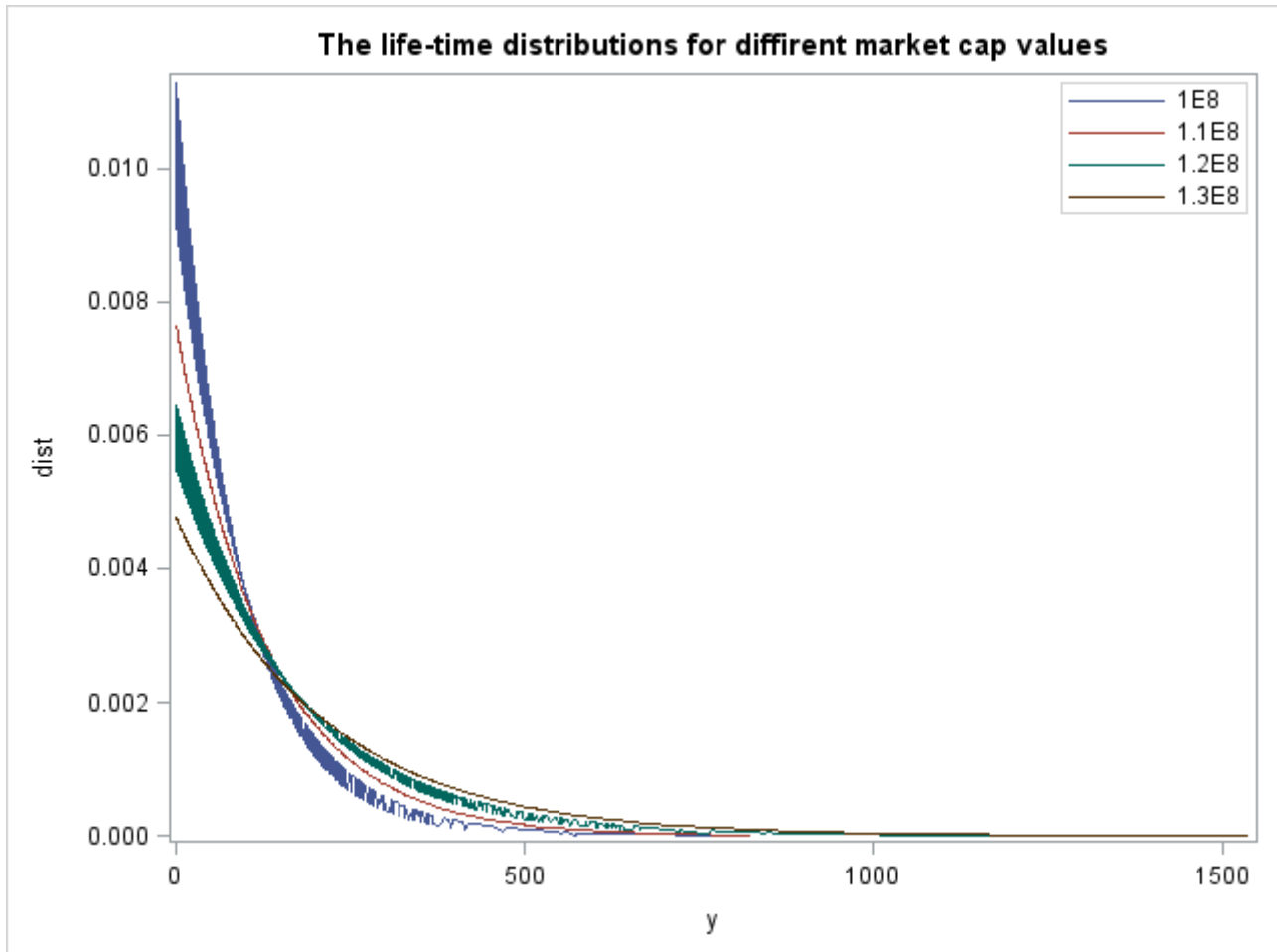
$$\begin{aligned} \lambda &= e^{\beta_0 + \beta_1 \ln(\text{marketCAP})} \\ &= e^{-34.6564 + 2.1248 \ln(\text{marketCAP})} \end{aligned} \quad (3.5)$$

It is now possible to answer questions such as, what is the probability, for a company, having a market CAP of 1 billion, to experience a loss exceeding our cut off point, within the next 100 days? Stated differently, what is the probability of this portfolio failing within the next 100 days?

$$\begin{aligned} F(100) &= \int_0^{100} \frac{1}{\lambda} e^{-\frac{x}{\lambda}} dx \\ &= \int_0^{100} \frac{1}{e^{-34.6564 + 2.1248 \ln(\text{marketCAP})}} \exp\left(-\frac{x}{e^{-34.6564 + 2.1248 \ln(\text{marketCAP})}}\right) dx \\ &= \int_0^{100} \frac{1}{e^{-34.6564 + 2.1248 \ln(\text{marketCAP})}} \exp\left(-\frac{x}{e^{-34.6564 + 2.1248 \ln(\text{marketCAP})}}\right) dx \quad (3.6) \\ &= 1 - \exp\left(-\frac{100}{e^{-34.6564 + 2.1248 \ln(\text{marketCAP})}}\right) \\ &= 0.0084342 \end{aligned}$$



A graph of the exponential PDF with parameter defined as in Equation 3.5 is give in Figure 3.7 for 4 companies of different market CAP.



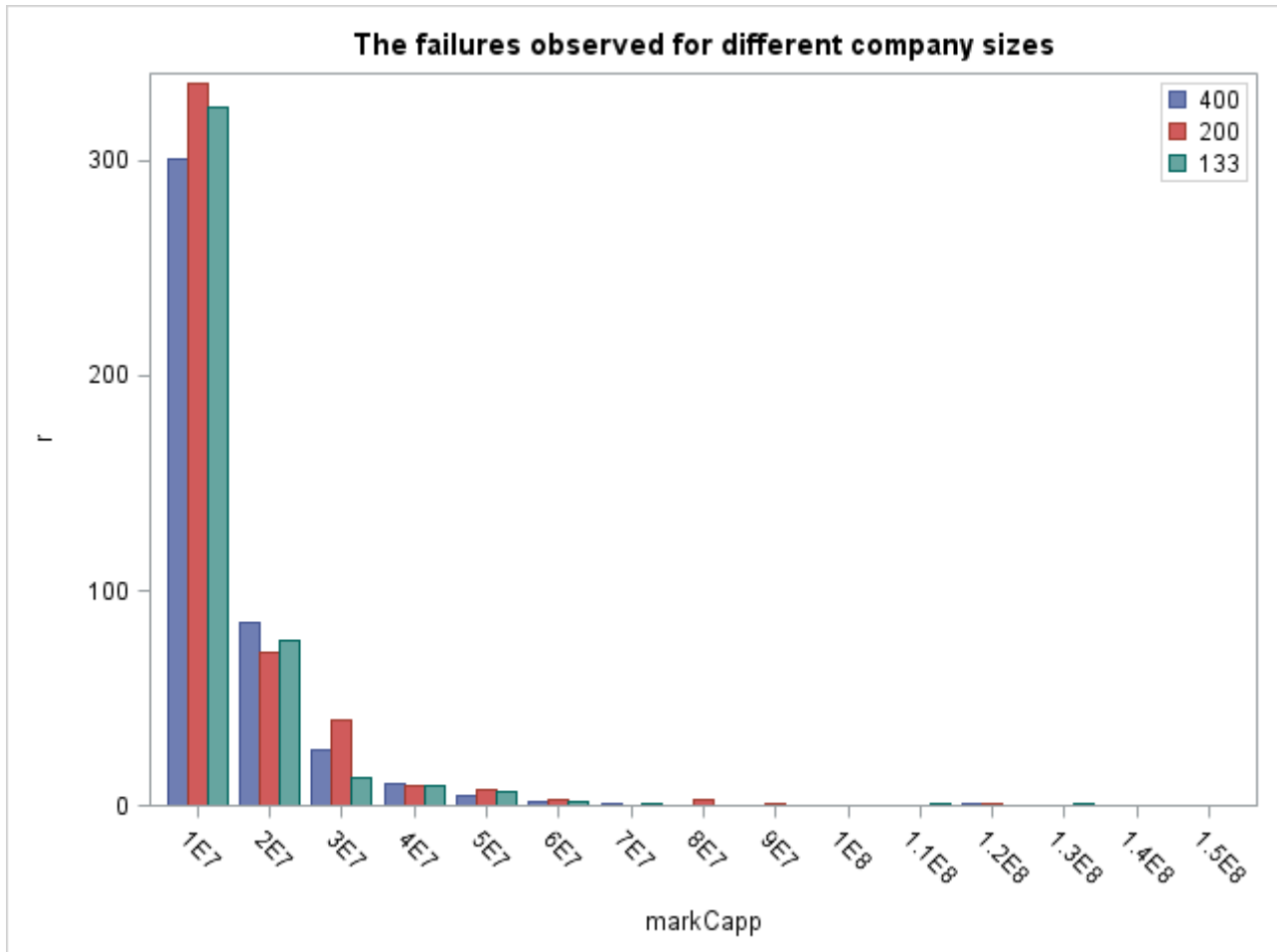
**Figure 3.7:** The lifetime distribution as obtained from the reliability procedure for 4 companies of different market CAP.

From the above graph it can clearly be seen that a linear link would not be appropriate to relate the stress factor, market CAP, to the parameter of the exponential lifetime distribution.

The whole process described up and to this point is now repeated with a failure of a company defined at a higher percentile/threshold. For a highly stressed company, that is to say small company, the percentage return is calculated. The loss of this small company is then calculated as  $L_t = -R_t$ . The  $(1 - 0.01)$ th percentile of the loss is set. Any percentage loss,

for any size/stressed company, above this now fixed percentile value, is considered a failure.

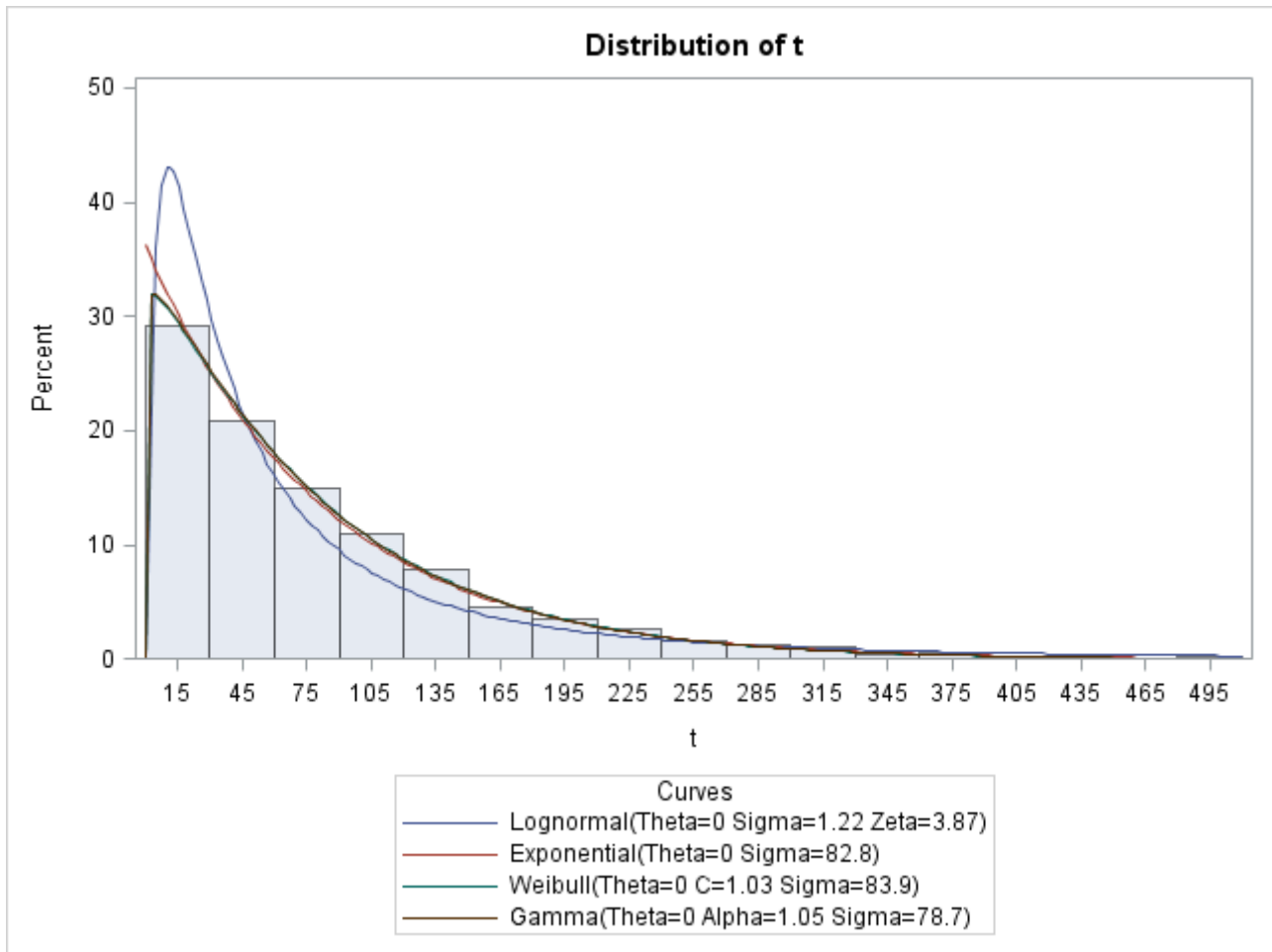
The relevant output follows.



**Figure 3.8:** The failure data for companies of different market CAPs evaluated at different inspection times.

The data becomes censored much sooner with mostly 100% non-failures already occurring at a market CAP value of 0.09 billion and so, to be able to model accurately, it is necessary to simulate companies with intermediate market CAP values as well.

It can again be seen that the exponential distribution is not rejected as possible lifetime distribution;



**Figure 3.9:** The empirical lifetime distribution with a Log-normal, Exponential, Weibull and Gamma distribution superimposed.

Goodness-of-Fit Tests for Exponential Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.01707516	Pr > D	>0.250
Cramer-von Mises	W-Sq	0.09689988	Pr > W-Sq	>0.250
Anderson-Darling	A-Sq	0.71751227	Pr > A-Sq	>0.250

Once again, the linear-link function delivers no results.

**Table 3.21:** The estimates obtained by implementing a linear link function.

Exponential Parameter Estimates				
Parameter	Estimate	Standard Error	Asymptotic Normal	
			95% Confidence Limits	
			Lower	Upper
Intercept	5.3759	0.0561	5.2661	5.4858
Mc	0.0000	0.0000	0.0000	0.0000
Shape	1.0000	0.0000	1.0000	1.0000

The power link function is modelled.

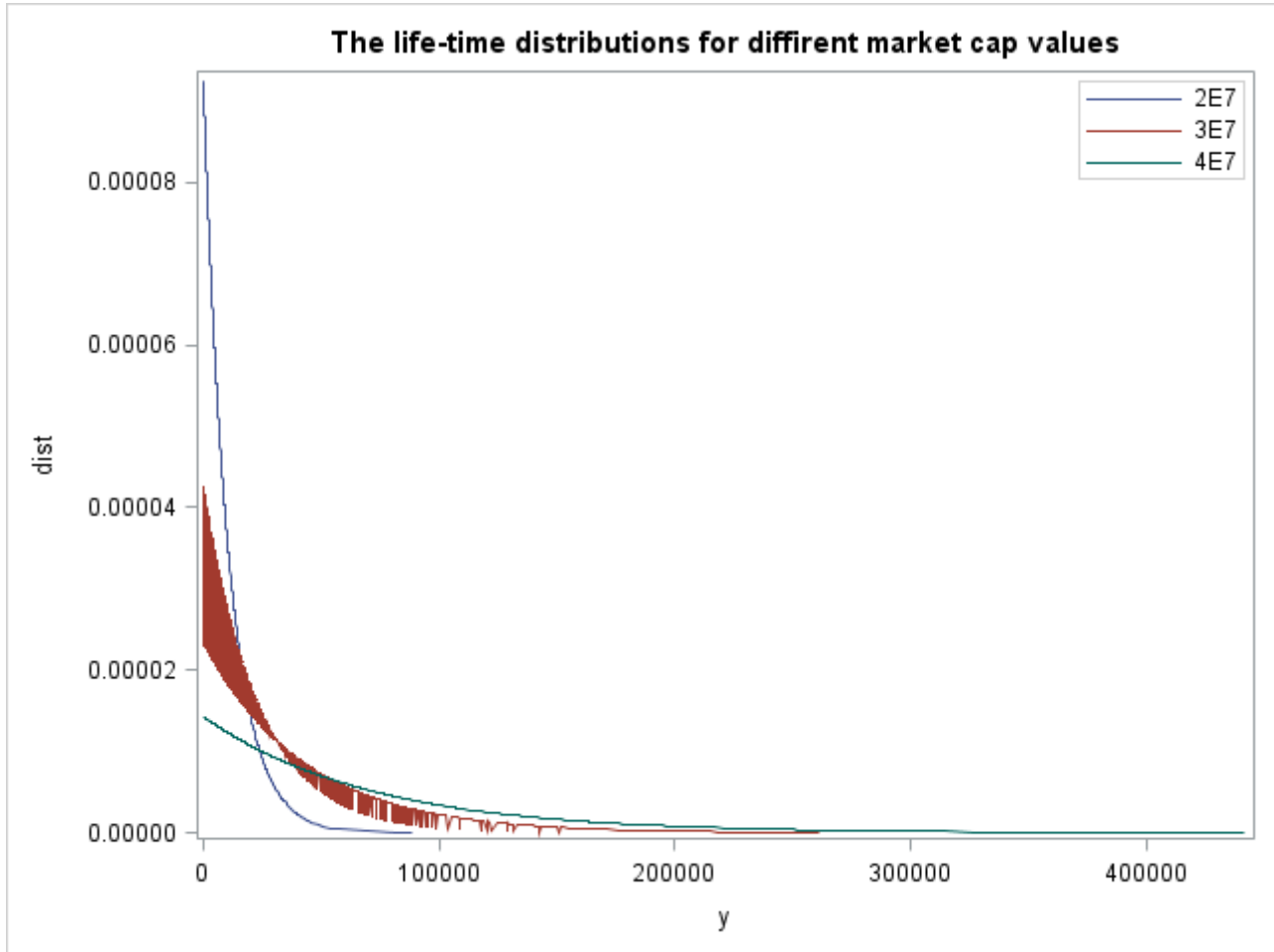
**Table 3.23:** The estimates obtained by implementing a power link function.

Exponential Parameter Estimates				
Parameter	Estimate	Standard Error	Asymptotic Normal	
			95% Confidence Limits	
			Lower	Upper
Intercept	-35.8415	0.8766	-37.5597	-34.1234
Mc	2.6216	0.0535	2.5167	2.7264
Shape	1.0000	0.0000	1.0000	1.0000

The power link function relates the parameter  $\lambda$  of the exponential distribution to the stress factor, market CAP, in the following way

$$\begin{aligned}
 \lambda &= e^{\beta_0 + \beta_1 \ln(\text{marketCAP})} \\
 &= e^{-35.8415 + 2.6216 \ln(\text{marketCAP})} .
 \end{aligned}
 \tag{3.7}$$

A graph of the exponential PDF with parameter defined as in Equation 3.7 is give in Figure 3.7 for 3 companies of different market CAP. From the graph, it is clear why a linear link function was not suitable.



**Figure 3.10:** The lifetime distribution as obtained from the reliability procedure for 4 companies of different market CAP.

It is now possible to answer questions such as, what is the probability, for a company, having a market CAP of 1 billion, to experience a loss exceeding our cut off point of 99.9%, within the next 100 days? Stated differently, what is the probability of this portfolio failing within the next 100 days?



$$\begin{aligned} F(100) &= \int_0^{100} \frac{1}{\lambda} e^{-\frac{x}{\lambda}} dx \\ &= \int_0^{100} \frac{1}{e^{-35.8415+2.6216 \ln(\text{marketCAP})}} \exp\left(-\frac{x}{e^{-35.8415+2.6216 \ln(\text{marketCAP})}}\right) dx \\ &= \int_0^{100} \frac{1}{e^{-35.8415+2.6216 \ln(\text{marketCAP})}} \exp\left(-\frac{x}{e^{-35.8415+2.6216 \ln(\text{marketCAP})}}\right) dx \quad (3.8) \\ &= 1 - \exp\left(-\frac{100}{e^{-35.8415+2.6216 \ln(\text{marketCAP})}}\right) \\ &= 0.0000009 . \end{aligned}$$

The probability in Equation 3.8 is largely reduced from the probability obtained in Equation 3.6. This is expected. As the failure threshold moves into the tail of the distribution, the failure becomes less and less likely.

Other interesting analyses are also possible such as observing how the probability to fail behaves within a fixed time, for companies operating under different levels of stress i.e. market CAPs. This is displayed in Figure 3.11 and 3.12. Figure 3.12 could be considered an enlarged view of Figure 3.11, obtained by focusing on a smaller range of market CAP values.

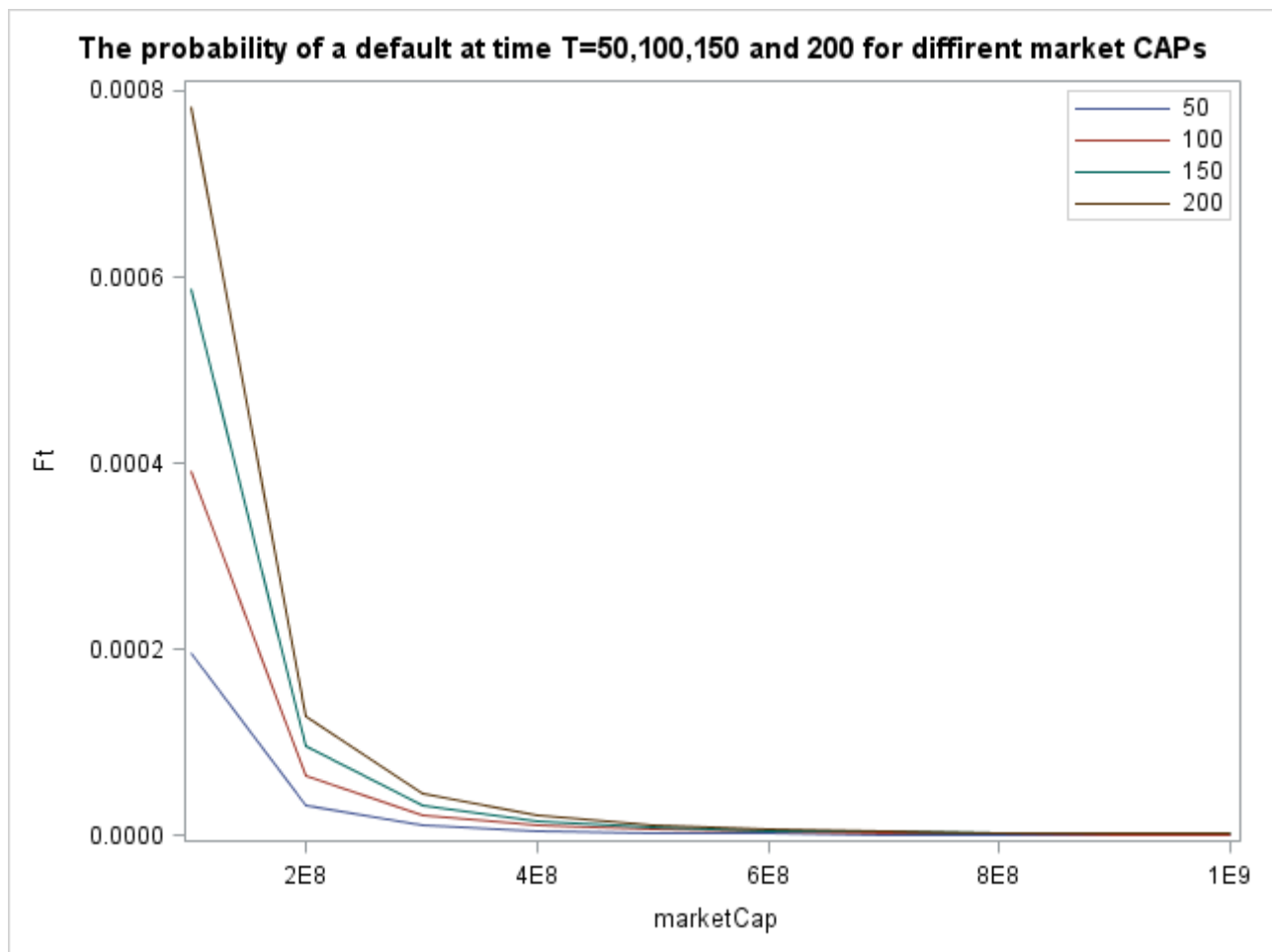
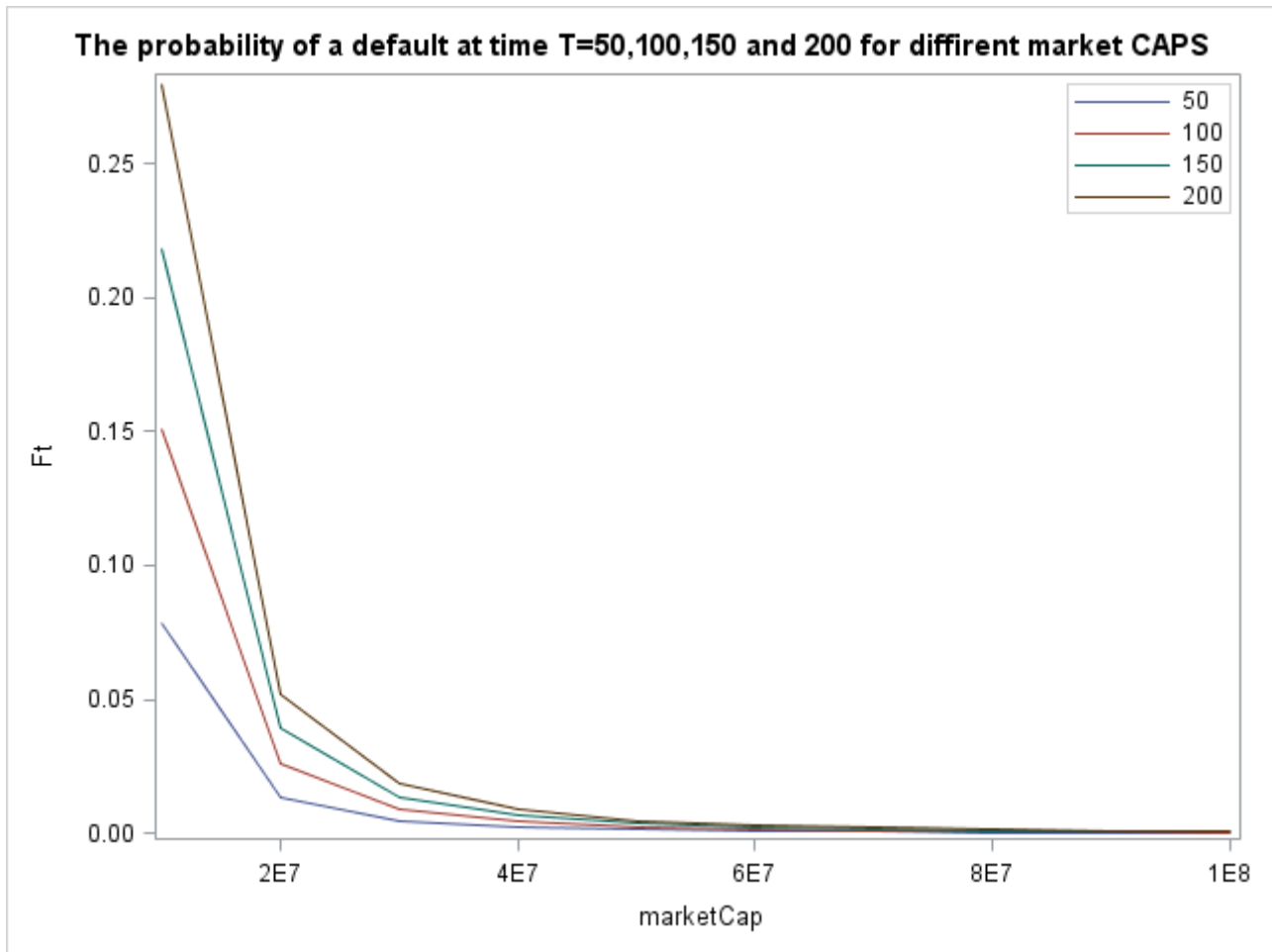


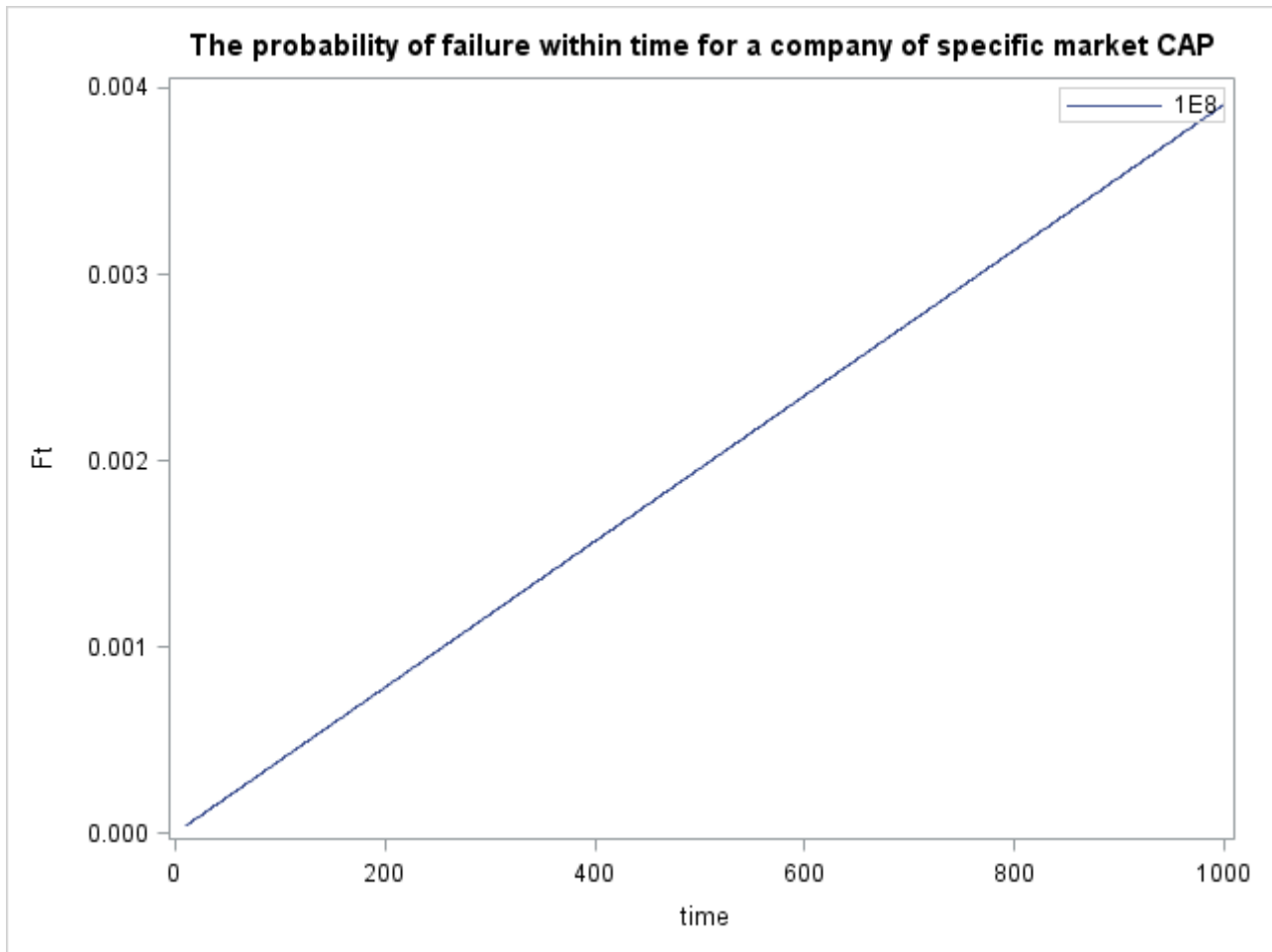
Figure 3.11: The probability of a default at a fixed time for companies of different market CAPs.



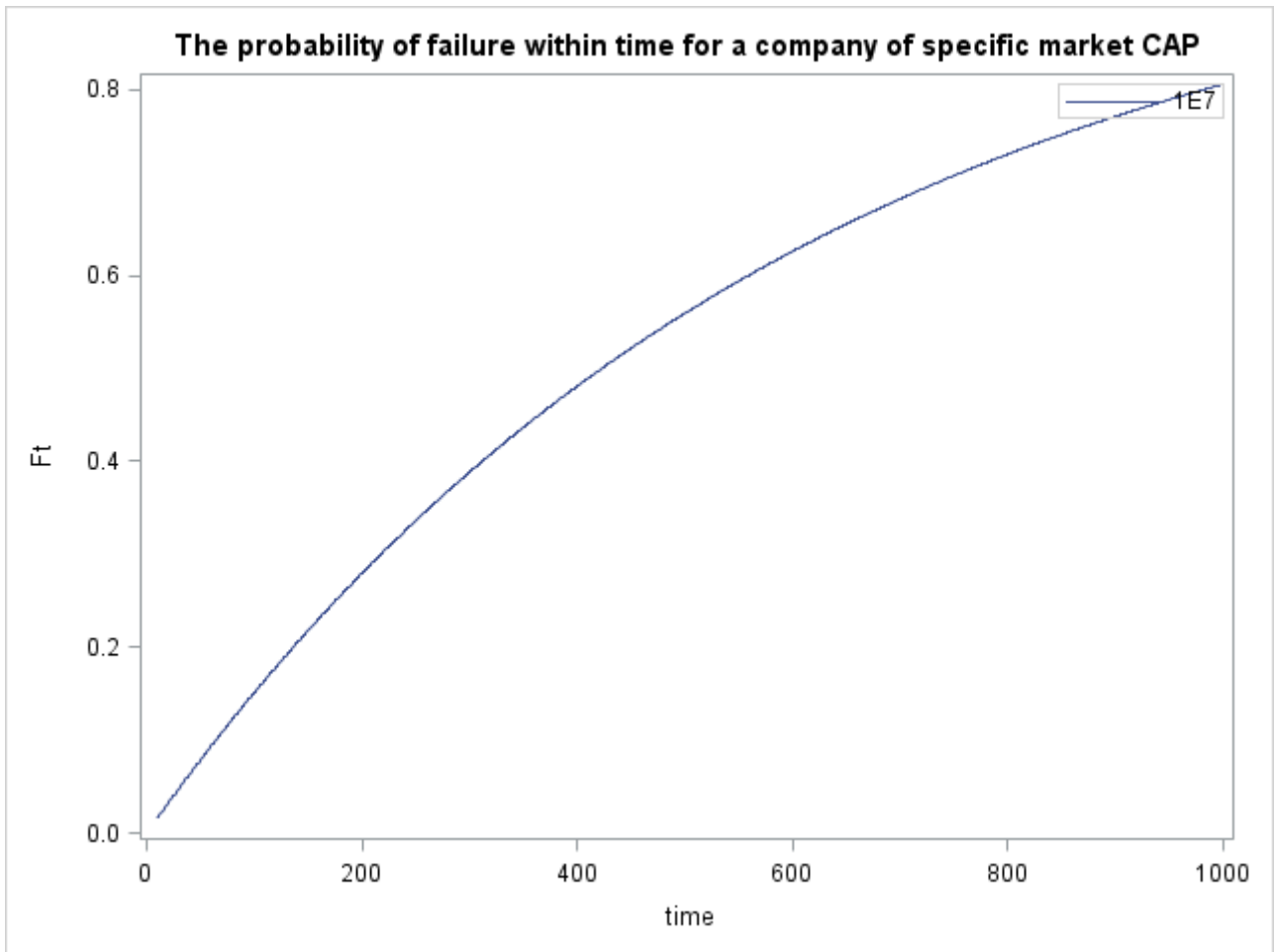
**Figure 3.12:** The probability of a default at a fixed time for companies of different market CAPs from a smaller range.

It is clear that as the company becomes less stressed (greater market CAP) the probability of a failure decreases exponentially. However, the probability of failure for any size company is larger if considered for a longer period of time. This can be seen more clearly in Figure 3.13 and Figure 3.14.





**Figure 3.13:** The probability of a default at different times for companies of different market CAPs.



**Figure 3.14:** The probability of a default at different times for companies of different market CAPs from a smaller range.

## Chapter 4

# Concluding comments

Financial stock return data was considered and one-shot device testing under ATs was applied to the stock market data of 431 companies trading in the health sector. One-shot device testing is a case of interval censored data which required us to discretize the data. In doing so, the default probability for a low-default portfolio (reliability at mission time) could be estimated.

Section 4.1 gives a summary of the study and Section 4.2 gives recommendations on future research.

### 4.1 Summary

The problem addressed is that of very little (or even none) data available for low-default portfolios. Towards this it was seen that different methods exist in dealing with low-default portfolios, such as extreme value theory or deriving the cumulative accuracy profile curve or using proportional hazard models. A possible new method has been proposed in dealing with such portfolios. The proposed method applies the framework of AT and consequently the basic building blocks of ATs needed to be established. Towards understanding estimation techniques used in ATs, the EM algorithm was explored. Understanding the programming component of this method was also investigated and presented for two popular examples in the literature. There is vast literature available on ATs. However, there are none available on how to apply these methods to financial data, or more specifically, to financial time series data. It was seen that the framework used in ATs could be applied to stock market data and deliver basic initial

results. This new proposed model should by no means be viewed as a completed model, but merely serves as a point of departure into exploring the parallel between ATs and the problem of insufficient data for low-default portfolios. These basic initial results made it possible to obtain the probability of a failure for a stock and to observe the relationship between the length of time the stock is traded and the failure probability as well as the size of the company and the failure probability.

## 4.2 Future work

Research can be launched into more accurately defining the lifetime distributions, but also the validity of the assumption that the distribution remains the same across different stress levels. Once confident in the distribution, it can be investigated if multiple parameters could be linked to multiple stress factors. In ATs there are physical models concerning temperature or textile strengths which justify the use of certain link function. As there is not yet such justifications when applying ATs to financial data, a strong possibility for future research also exist here.

If a different AT were to be applied such as ALT, discretizing the data would not be necessary as it was when doing one-shot device testing. Essentially different censoring schemes could be considered. When the data was discretized, inspection times were chosen and data simulated accordingly.  $M$  companies of market capitalization size  $i$  were tested. Sampling techniques and optimal test designs can be further researched in this regard.

Another strong assumption that needs to be addressed is that of independence. Testing the validity of the assumption that failure events for companies trading in the same sector are in fact independent of each other.

# Bibliography

- Abdel Ghaly, A. A., Aly, H. M., and Salah, R. N. (2015). Different estimation methods for constant stress accelerated life test under the family of the exponentiated distributions. *Quality and Reliability Engineering International*, 32:1095–1108.
- Balakrishnan, N. and Ling, M. H. (2012). Multiple-stress model for one-shot device testing data under exponential distribution. *Reliability, IEEE Transactions on*, 61(3):809–821.
- Balakrishnan, N. and Ling, M. H. (2013). Expectation maximization algorithm for one shot device accelerated life testing with weibull lifetimes, and variable parameters over stress. *IEEE Transactions on Reliability*, 62(2):537–551.
- Balakrishnan, N. and Ling, M. H. (2014). Gamma lifetimes and one-shot device testing analysis. *Reliability Engineering & System Safety*, 126:54–64.
- Banasik, J., Crook, J. N., and Thomas, L. C. (1999). Not if but when will borrowers default. *The Journal of the Operational Research Society*, 50(12):1185–1190. Copyright - Copyright (c) 1999 Operational Research Society; Last updated - 2013-10-04; CODEN - OPRQAK.
- Boyes, W. J., Hoffman, D. L., and Low, S. A. (1989). An econometric analysis of the bank credit scoring problem. *Journal of Econometrics*, 40(1):3–14.
- Chen, D. and Lio, Y. (2010). Parameter estimations for generalized exponential distribution under progressive type-i interval censoring. *Computational Statistics & Data Analysis*, 54(6):1581 – 1591.
- Cheng, M.-Y., Lin, J.-Y., Hsiao, T.-Y., and Lin, T. W. (2008). Censoring model for evaluating intellectual capital value drivers. *Journal of Intellectual Capital*, 9(4):639–654.

- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, 39 (1):1–38.
- Embrechts, P., Resnick, S. I., and Samorodnitsky, G. (1999). Extreme value theory as a risk management tool. *North American Actuarial Journal*, 3(2):30–41.
- Escobar, L. A. and Meeker, W. Q. (2006). A review of accelerated test models. *Statistical Science*, 21 (4):552–577.
- Flury, B. and Zoppè, A. (2000). Exercises in em. *The American Statistician*, 54(3):207–209.
- Freels, J. K., Pignatiello, J. J., Warr, R. L., and Hill, R. R. (2015). Bridging the gap between quantitative and qualitative accelerated life tests. *Quality and Reliability Engineering International*, 31(5):789–800.
- Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin.
- Gilli, M. and Këllezi, E. (2006). An application of extreme value theory for measuring financial risk. *Computational Economics*, 27(2-3):207–228.
- Green, P. J. (1990). Bayesian reconstructions from emission tomography data using a modified em algorithm. *IEEE transactions on medical imaging*, 9(1):84–93.
- Hartley, H. O. (1958). Maximum likelihood estimation from incomplete data. *Biometrics*, 14(2):174–194.
- Kay, J. (1997). The em algorithm in medical imaging. *Statistical methods in medical research*, 6(1):55–75.
- Kiefer, N. M. (2009). Default estimation for low-default portfolios. *Journal of Empirical Finance*, 16(1):164–173.
- Kim, C. and Bai, D. (2002). Analyses of accelerated life test data under two failure modes. *International Journal of Reliability, Quality and Safety Engineering*, 9(02):111–125.

- Kitamura, Y., Moriguchi, M., Kaneko, H., Morisaki, H., Morisaki, T., Toyama, K., and Kamatani, N. (2002). Determination of probability distribution of diplotype configuration (diplotype distribution) for each subject from genotypic data using the em algorithm. *Annals of human genetics*, 66(03):183–193.
- Klein, J. P. and Moeschberger, M. L. (2005). *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media.
- Kundu, D. and Dey, A. K. (2009). Estimating the parameters of the marshallÓolkin bivariate weibull distribution by {EM} algorithm. *Computational Statistics & Data Analysis*, 53(4):956 – 965.
- Li, X., Gao, P., and Sun, F. (2015). Acceptance sampling plan of accelerated life testing for lognormal distribution under time-censoring. *Chinese Journal of Aeronautics*, 28(3):814–821.
- Ling, M. H. (2012). *INFERENCE FOR ONE-SHOT DEVICE TESTING DATA*. PhD thesis, McMaster University.
- Ma, X.-b., Lin, F.-c., and Kang, R. (2014). Thermal degradation behavior of silicon dioxide-coated indium–tin–oxide thin films on glass in a storage status under highly accelerated life testing. *IEEE Transactions on Device and Materials Reliability*, 14(1):203–212.
- McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance*, 7(3):271–300.
- Meeker, W. Q. (1984). A comparison of accelerated life test plans for weibull and lognormal distributions and type i censoring. *Technometrics*, 26(2):157–171.
- Morris, M. D. (1987). A sequential experimental design for estimating a scale parameter from quantal life testing data. *Technometrics*, 29(2):173–181.
- Nelson, W. (1975). Graphical analysis of accelerated life test data with a mix of failure modes. *Reliability, IEEE Transactions on*, 24(4):230–237.

- Nelson, W. and Kielpinski, T. J. (1976). Theory for optimum censored accelerated life tests for normal and lognormal life distributions. *Technometrics*, 18(1):105–114.
- Nelson, W. B. (2009). *Accelerated testing: statistical models, test plans, and data analysis*, volume 344. John Wiley & Sons.
- Ott, J. (1977). Counting methods (em algorithm) in human pedigree analysis: linkage and segregation analysis. *Annals of human genetics*, 40(4):443–454.
- Phillips, R. A., Rosenblatt, E., and Vanderhoff, J. H. (1996). The probability of fixed-and adjustable-rate mortgage termination. *The Journal of Real Estate Finance and Economics*, 13(2):95–104.
- Ramirez, J., Gore, W., and Johnston, G. (2001). New methods for modeling reliability using degradation data. *Statistics Data Analysis and Data Mining*, pages 263–226.
- Rauk, A. P., Guo, K., Hu, Y., Cahya, S., and Weiss, W. F. (2014). Arrhenius time-scaled least squares: A simple, robust approach to accelerated stability data analysis for bioproducts. *Journal of pharmaceutical sciences*, 103(8):2278–2286.
- Ren, S. (2004). A comparative study of an accelerated life-test model and a toxicokinetics-based model for the analysis of porcellio scaber survival data. *Environmental toxicology and chemistry*, 23(1):232–237.
- Scheike, T. H. and Sun, Y. (2007). Maximum likelihood estimation for tied survival data under cox regression model via em-algorithm. *Lifetime Data Analysis*, 13(3):399–420. Copyright - Springer Science+Business Media, LLC 2007; Document feature - Equations; Tables; ; Last updated - 2014-08-31.
- Silverman, M. (1998). Summary of halt and hass results at an accelerated reliability test center. In *Reliability and Maintainability Symposium, 1998. Proceedings., Annual*, pages 30–36. IEEE.
- Sinha, D., Patra, K., and Dey, D. K. (2003). Modelling accelerated life test data by using a bayesian approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(2):249–259.



- Stepanova, M. and Thomas, L. (2002). Survival analysis methods for personal loan data. *Operations Research*, 50(2):277–289.
- Van der Burgt, M. J. (2008). Calibrating low-default portfolios, using the cumulative accuracy profile. *Journal of Risk Model Validation*, 1(4):17–33.
- Vernic, R., Teodorescu, S., and Pelican, E. (2009). Two lognormal models for real data. *Annals of Ovidius University, Series Mathematics*, 17(3):263–277.
- Wang, W. and Kececioglu, D. B. (2000). Fitting the weibull log-linear model to accelerated life-test data. *Reliability, IEEE Transactions on*, 49(2):217–223.
- Wang, X., Balakrishnan, N., Guo, B., and Jiang, P. (2015). Residual life estimation based on bivariate non-stationary gamma degradation process. *Journal of Statistical Computation and Simulation*, 85(2):405–421.
- Yildirim, Y. (2008). Estimating default probabilities of cmbs loans with clustering and heavy censoring. *The Journal of Real Estate Finance and Economics*, 37(2):93–111.
- Zhang, J., Liu, C., Chen, X., Cheng, G., and Zhou, A. (2014). Study on constant-step stress accelerated life tests in white organic light-emitting diodes. *Luminescence*, 29(7):933–937.

# Appendix A

## The coding Appendix for Chapter 2

The relevant SAS code for examples presented in Chapter 2 is presented here. The code is available from an online GitHub repository: [Online code](#)

### A.1 SAS code for Example 3 from Chapter 2

```
*For LaTeX output;  
ods _all_ close;  
%let MyPath = C:\Users\Anel\Desktop\D_dissertation\chapters\  
    chapter1;  
ods graphics / outputfmt=png;  
  
proc iml;  
*Set a seed value to make results comparable;  
call randseed(45);  
  
*The 4 class multinomial;  
x={125,18,20,34};  
*The desired level of accuracy;  
acc=0.0001;  
*The initial guess for theta;
```

```

theta=0.1;

*Repeat until desired accuracy is achieved OR if convergence
  does not occur, avoid infinite loop;
do j=1 to 100 until(diff=1);
  k=j-1;*For PRINT matrix;
  v2=x[1,]#(0.25#theta)/(0.5+0.25#theta);*The E-step;
  matrix=matrix/(v2||k||theta);*Update the PRINT matrix
  ;
  theta_old=theta;
  theta=(v2+x[4,])/(v2+x[2,]+x[3,]+x[4,]);*The M-step;
  k=j;*For PRINT matrix;
  matrix=matrix/(v2||k||theta);*Update the PRINT matrix
  ;
  diff=((abs(theta_old-theta))<acc); *Test if accuracy
    has been achieved;
end;

*For LaTeX output;
ods html file="&MyPath.\Output\EX1\GeneticLinkage.htm" gpath="
  &MyPath.\Figures\EX1";
ods tagsets.tablesonlylatex file="&MyPath.\EX1\Genetic.tex"(
  notop nobot)
newfile=table stylesheet="sas.sty"(url="sas") gpath="&MyPath.\
  Figures\EX1";

*PRINT matrix to show each convergence step;
print matrix[label="" colname={"v" "k" "theta"}];

quit;

```

## A.2 SAS code for Example 4 from Chapter 2

```
*For LaTeX output;
ods _all_ close;
%let MyPath = C:\Users\Anel\Desktop\D_dissertation\chapters\
    chapter1;
ods graphics / outputfmt=png;

proc iml;
*Set a seed value to make results comparable;
call randseed(2);

acc=0.001;*The desired level of accuracy;
n=15;*The number of items for which the life times are known;
eta=8; *The real value of eta;
m=20; *The number of items for which we don't know the
    lifetimes;
t=3;*The inspection time;
y=J(n,1,.);*The matrix that will contain the known lifetimes;
v=J(m,1,.);*The matrix that will contain the "unknown"
    lifetimes;

call randgen(y,'exponential',eta);*Generate the known
    lifetimes;
call randgen(v,'exponential',eta);*Generate the "unknown"
    lifetimes;
r=(v<=t)[+,];*The number of items that failed—we don't know
    the exact lifetimes;
pr=r||((y[+,])/nrow(y)); *To be used in the E-step;

*For LaTeX output;
```

```
ods tagsets.tablesonlylatex file="&MyPath.\EX2\r.tex"(notop
  nobot)
newfile=table stylesheet="sas.sty"(url="sas") gpath="&MyPath.\
  Figures\EX2";
print pr[label= "" colname={"The number of failures" "y-bar"}]
  ;

eta=2;*The initial guess for eta;
*Repeat until desired accuracy is achieved OR if convergence
  does not occur, avoid infinite loop;
do j=1 to 100 until(diff=1);
  k=j-1;*For PRINT matrix;
  v2=(m-r)#(t+eta)+r#eta+(-r#t#exp(-t/eta))/(1-exp(-t/
    eta));*The E-step;
  matrix=matrix/(v2||k||eta);*Update the PRINT matrix;
  eta_old=eta;
  eta=(n#(y[+,]/nrow(y))+v2)/(n+m);*The M-step;
  k=j; *For PRINT matrix;
  matrix=matrix/(v2||k||eta); *Update the PRINT matrix;
  diff=((abs(eta_old-eta))<acc); *Test if accuracy has
    been achieved;
end;

*For LaTeX output;
ods html file="&MyPath.\Output\EX2\Exp.htm" gpath="&MyPath.\
  Figures\EX2";
ods tagsets.tablesonlylatex file="&MyPath.\EX2\exp.tex"(notop
  nobot)
newfile=table stylesheet="sas.sty"(url="sas") gpath="&MyPath.\
  Figures\EX2";
```

```
*PRINT matrix to show each convergence step;  
print matrix[label="" colname={"v" "k" "theta"}];  
  
quit;
```

## Appendix B

### The coding Appendix for Chapter 3

The relevant SAS code for examples presented in Chapter 3 is presented here. The code is available from an online GitHub repository: [Online code](#)

#### B.1 SAS code for Section 3.1 from Chapter 3

```
*Import the stock data ;
PROC IMPORT OUT= DISSN.MERGEDDATA
            DATAFILE= "C:\Users\Anel\Desktop\D_dissertation\
            chapters\chapter2\RealData\merged1479114290141.csv"
            DBMS=CSV REPLACE;
            GETNAMES=YES;
            DATAROW=2;
RUN;

proc iml;
*Store the stock data in matrix X with column names equal to
the stock names;
use dissn.mergedData;
read all into x[colname=names];
```

```
close dissN.mergedData;

*Create a column vector from the market capitalization values;
MC=x[1,]';
call sort(MC);
MC=MC[1:250];
create dissN.MCcol from MC[colname={ 'MC' }];
append from MC;
quit;

*Import an excerpt of the stock data to be printed;
PROC IMPORT OUT= DISSN.DataMerged_e
              DATAFILE= "C:\Users\Anel\Desktop\D_dissertation\
              chapters\chapter2\RealData\DataMerged_e.xlsx"
              DBMS=EXCEL REPLACE;
              RANGE="Sheet1$";
              GETNAMES=YES;
              MIXED=NO;
              SCANTEXT=YES;
              USEDATE=YES;
              SCANTIME=YES;
RUN;

*For LATEX;
ods _all_ close;
%let MyPath = C:\Users\Anel\Desktop\D_dissertation\chapters\
              chapter2;
ods graphics / outputfmt=png;
```



```
ods html file = "&MyPath.\Output\RealData\dataE.htm" gpath = "&
  MyPath.\Figures\RealData\dataE";
ods tagsets.tablesonlylatex file = "&MyPath.\RealData\dataE.tex
  "(notop nobot)
newfile=table stylesheet="sas.sty"(url="sas") gpath = "&MyPath.\
  Figures\RealData\dataE";
*Print and expert of stock data to show what data table looks
  like;
proc print data=DISSN.DataMerged_e ;
run;
quit;

proc iml;

*For LATEX;
ods html file = "&MyPath.\Output\RealData\NumberOfCompanies.htm"
  gpath = "&MyPath.\Figures\RealData\NumberOfCompanies";
ods tagsets.tablesonlylatex file = "&MyPath.\RealData\
  NumberOfCompanies.tex"(notop nobot)
newfile=table stylesheet="sas.sty"(url="sas") gpath = "&MyPath.\
  Figures\RealData\NumberOfCompanies";

*Store the stock data in matrix x with column names equal to
  the stock names;
use dissN.mergedData;
read all into x[colname=names];
close dissN.mergedData;

MC=x[1,];
*Do not include the first row: Market capitalization;
x=x[2:nrow(x),];
```

```

mx=ncol(x);*The number of columns in the matrix i.e. the
  number of companies;
print mx[label='' colname={'Number_of_companies'}];
nx=nrow(x);*The number of rows in the matrix i.e. the number
  of prices for each company;
xt_1=x[1:nx-1,];*Create the lagged prices to be used in return
  calculation;
xt=x[2:nx,];

RtX=log(xt/xt_1);*Define a return by the log of the returns;
step=do(1,nx-1,1);*Create the time variable against which each
  return value will be plotted;
step=step';
RtXs=step||RtX;

s2=(((RtX-(Rtx[+,]/(nx-1)))##2)[+,])/(nx-2);*Calculate the
  standard deviation of the returns for each company;
*print s2;
std=(Mc//sqrt(s2))'; *Create a matrix compromising of the
  Market capitalization and standard deviation of the returns
  , for each company;
call sort(std,1);

*Create a data table from the return values for each company;
create dissN.ALL_Rt from RtXS[colname=({'steps'}||names)];
append from RtXs;

*Create a data table from the market capitalization values for
  each company;
create dissN.MC from MC[colname=(names)];

```

```
append from MC;

*Create a data table from the market capitalization values and
  standard deviation of the returns , for each company;
create dissN.std from std[colname={ 'MC' 'std' }];
append from std;
quit;

*For LATEX output;
ods html file="&MyPath.\Output\RealData\Normality\Normality.
  htm" gpath="&MyPath.\Figures\RealData\Normality";
ods tagsets.tablesonlylatex file="&MyPath.\RealData\Normality\
  Normality.tex"(notop nobot)
newfile=table stylesheet="sas.sty"(url="sas") gpath="&MyPath.\
  Figures\RealData\Normality";

*Test if it is reasonable to assume that the logged returns
  are normally distributed;
*Select companies JNJ AGRX MNOV NDRM TLGT log returns;
*Use ods trace to print only specific tables;
ods trace on;
ods select testsfornormality;
proc univariate data=dissN.ALL_Rt normal;
var JNJ AGRX MNOV NDRM TLGT;
output out=dissn.univar probn= JNJ AGRX MNOV NDRM TLGT;
run;
ods trace off;
quit;

*For LATEX output;
```

```
ods html file="&MyPath.\Output\RealData\Normality\Normalityp.
  htm" gpath="&MyPath.\Figures\RealData\Normality";
ods tagsets.tablesonlylatex file="&MyPath.\RealData\Normality\
  Normalityp.tex"(notop nobot)
newfile=table stylesheet="sas.sty"(url="sas") gpath="&MyPath.\
  Figures\RealData\Normality";

*Print only the Shapiro–Wilk normality test p–values;
proc print data=dissN.univar;
run;
quit;

*For LATEX output;
ods graphics on/imagename="smallVSbig";
ods html file="&MyPath.\Output\RealData\smallVSbig.html" gpath
  ="&MyPath.\Figures\RealData";
ods tagsets.tablesonlylatex file="&MyPath.\RealData\smallVSbig
  .tex"(notop nobot)
newfile=table stylesheet="sas.sty"(url="sas") gpath="&MyPath.\
  Figures\RealData";

*****Graphing procedure to graph the log returns of a
  selected small–cap and large–cap company against time;
*****Code exported from ODS Graphics designer;
proc template;
define statgraph sgdesign;
dynamic _STEPS _JNJ _STEPS2 _ABEO;
begingraph;
  entrytitle halign=center 'The_returns_of_a_big_cap_and_
    small_cap_company';
```

```

layout lattice / rowdatarange=data columndatarange=data
rowgutter=10 columngutter=10;
layout overlay;
  seriesplot x=_STEPS y=_JNJ / name='series'
  connectorder=xaxis lineattrs=(color=CX639A21 );
  seriesplot x=_STEPS2 y=_ABEO / name='series2'
  connectorder=xaxis lineattrs=(color=CX8CA6CE );
  discretelegend 'series' 'series2' / opaque=false
  border=true halign=left valign=bottom
  displayclipped=true across=1 order=rowmajor
  location=inside;
endlayout;
endlayout;
endgraph;
end;
run;

proc sgrender data=DISSN.ALL_RT template=sgdesign;
dynamic _STEPS="STEPS" _JNJ="JNJ" _STEPS2="STEPS" _ABEO="ABEO
";
run;
*****End of graphing procedure;

*For LATEX output;
ods graphics on/imagename="MC";
ods html file="&MyPath.\Output\RealData\MC.html" gpath="&
MyPath.\Figures\RealData";
ods tagsets.tablesonlylatex file="&MyPath.\RealData\MC.tex"(
notop nobot)
newfile=table stylesheet="sas.sty"(url="sas") gpath="&MyPath.\
Figures\RealData";

```

```
*Print the Market capitalization values for the companies for
  which the returns were graphed in above procedure;
proc print data=dissn.MC;
var JNJ ABEO;
run;
quit;

*For LATEX output;
ods graphics on/imagenam="std.";
ods html file="&MyPath.\Output\RealData\std.html" gpath="&
  MyPath.\Figures\RealData";
ods tagsets.tablesonlylatex file="&MyPath.\RealData\std.tex"(
  notop nobot)
newfile=table stylesheet="sas.sty"(url="sas") gpath="&MyPath.\
  Figures\RealData";

*****Graphing procedure to graph the standard
  deviation of the log-returns for each company against the
  market capitalization values for each company;
*****Code exported form ODS Graphics designer;
proc template;
define statgraph sgdesign;
dynamic _MC _STD;
begingraph;
  entrytitle halign=center 'Variability_vs._market_Capp';
  layout lattice / rowdatarange=data columndatarange=data
    rowgutter=10 columngutter=10;
  layout overlay;
    scatterplot x=_MC y=_STD / name='scatter';
  endlayout;
endlayout;
```

```
endgraph ;
end ;
run ;

proc sgrender data=DISSN.STD template=sgdesign ;
dynamic _MC="MC" _STD="STD" ;
run ;
*****End of graphing procedure ;

proc iml ;
use dissn.std ;
read all into x ;
*Transform the standard deviations by applying the negative
  log ;
x[,2]= -log(x[,2]) ;

create dissn.std_Log from x[colname={ 'MC' 'Log_std' } ] ;
append from x ;
quit ;

*For LATEX output ;
ods graphics on/imagename="log_std" ;
ods html file="&MyPath.\Output\RealData\std_Log.html" gpath="&
  MyPath.\Figures\RealData\" ;
ods tagsets.tablesonlylatex file="&MyPath.\RealData\std_Log.
  tex"(notop nobot)
newfile=table stylesheet="sas.sty"(url="sas") gpath="&MyPath.\
  Figures\RealData\" ;

*****Graphing procedure to graph the negative log
  standard deviation (transformed standard deviation) of the
```

```

log-returns for each company against the market
capitalization values for each company;
*****Code exported form ODS Graphics designer;
proc template;
define statgraph sgdesign;
dynamic _MC _LOG_STD;
begingraph;
  entrytitle halign=center 'Log( variability )_vs._market_Capp'
  ;
  layout lattice / rowdatarange=data columndatarange=data
  rowgutter=10 columngutter=10;
  layout overlay;
    scatterplot x=_MC y=_LOG_STD / name='scatter';
  endlayout;
endlayout;
endgraph;
end;
run;

proc sgrender data=DISSN.STD_LOG template=sgdesign;
dynamic _MC="MC" _LOG_STD=" 'LOG_STD' n";
run;
*****End of graphing procedure;

*For LATEX output;
ods graphics on/imagename="nlin_model";
ods html file="&MyPath.\Output\RealData\nlin_model.html" gpath
  ="&MyPath.\Figures\RealData\";
ods tagsets.tablesonlylatex file="&MyPath.\RealData\nlin_model
.tex"(notop nobot)

```



```
newfile=table stylesheet="sas.sty"(url="sas") gpath="&MyPath.\
  Figures\RealData\";

*Perform a non-linear regression to understand how the
  transformed standard deviations move with market
  capitalization;
goptions reset=all;
proc nlin data=dissn.std_log ;
  parameters beta=2 alpha=2.6;
  switch=log(mc);
  model log_std=alpha+beta*switch;
  output out=dissn.nlinout_log pred=p lcl=lcl ucl=ucl parms=
    beta_0 beta_1 ;
run;

*For LATEX output;
ods graphics on/imagename="nlin";
ods html file="&MyPath.\Output\RealData\std_Log.html" gpath="&
  MyPath.\Figures\RealData\";
ods tagsets.tablesonlylatex file="&MyPath.\RealData\nlin.tex"(
  notop nobot)
newfile=table stylesheet="sas.sty"(url="sas") gpath="&MyPath.\
  Figures\RealData\";

*****Graphing procedure to graph the non-linear model
  which relates the transformed standard deviations to the
  market capitalization;
*****Code exported form ODS Graphics designer;
proc template;
define statgraph sgdesign;
dynamic _MC _LOG_STD _MC2 _P;
```

```

begingraph;
  entrytitle halign=center 'The predicted log variation as a
    function of Log(CAP)';
  layout lattice / rowdatarange=data columndatarange=data
    rowgutter=10 columngutter=10;
  layout overlay;
    scatterplot x=_MC y=_LOG_STD / name='scatter';
    seriesplot x=_MC2 y=_P / name='series' connectorder=
      xaxis lineattrs=(color=CX8CA6CE );
    discretelegend 'scatter' 'series' / opaque=false
      border=true halign=left valign=bottom
      displayclipped=true across=1 order=rowmajor
      location=inside;
  endlayout;
endlayout;
endgraph;
end;
run;

proc sgrender data=DISSN.NLINOUT_LOG template=sgdesign;
dynamic _MC="MC" _LOG_STD=" 'LOG_STD' n" _MC2="MC" _P="P";
run;

```

## B.2 SAS code for Section 3.2 from Chapter 3

### B.2.1 This code generates Figure 3.5

The code also generates the simulation results given in the table before Figure 3.5

```

*For LATEX output;
ods _all_ close;
%let MyPath = C:\Users\Anel\Desktop\D_dissertation\chapters\
  chapter2;

```

```
ods graphics / outputfmt=png;

*Simulate data to be used in one-shot device test;
proc iml;
call seed(27);
*import the parameters of the non-linear model;
use dissn.nlinout_log ;
read all into std;

std=std[1,ncol(std)-1:ncol(std)];
*print std;
*the percentile value at which a failure will be defined;
alpha=0.05;
N=400;*actually the time variable , let the process run from t
=0 to t=400;
M=1000;*The number of companies each of market capitalization
size S;
mu=0;
*****Select the failure threshold based on the
percentile of a small market capitalization company;
smallCap=0.01*(10**9);
X=log(smallCap)//{1};
factor=std*X;
sigmaf=exp(-factor);*create the standard deviation of the
return based on the output from the non-linear model;
*print sigmaf;
Tcll=(probit(alpha)*sigmaf)+mu;*The percentile which will be
the failure threshold;
*print Tcll;
Tclu=(probit(1-alpha)*sigmaf)+mu;
*print Tclu;
```

```

*****End;

do s=0.1 to 1 by 0.1; *change 0.1 to 0.01 make comparable;
    markCapp=s*(10**9);*The stress factor in Billions;
    X=log (markCapp) // {1};
    factor=std*X;
    sigma=exp(-factor);*create the standard deviation of
        the return based on the output from the non-linear
        model;
    *print sigma;
    call randseed(0);
    Rt=J(N,1,0);
    count=J(1,3,0);
    *****;
    do r=1 to 3; *The number of inspection times;
        do k=1 to M;* The number devices under current
            test condition;
                call randgen(Rt, "Normal", Mu, Sigma);
                *Populate the return matrix;
                ***Let's say we are interested in a
                    loss and not a gain;
                Rt=-Rt;
                *count the number of failed companies
                    under current test condition, a failure is a loss above
                    threshold;
                if (((Rt[1: int(N-1/r),])>=Tclu)[+])>0
                    then count[,r]=count[,r]+1;
                end;
            *Create data matrix;
            DATAM=DataM // (markCapp || count[,r] || int(N/r));
        end;
    end;

```

```
end ;

*Create_the_data_matrix ;
NMat=repeat (M, nrow (DATAM) ) ;
print_Nmat ;
DATAM=DATAM||NMat ;
create_dissN_relia_from_DATAM[colname={ 'markCapp' 'r' 'IT' 'M'
    }];
append_from_DATAM ;

quit ;

*For_LaTeX_output ;
ods_html_file ="&MyPath . \ Output \ SIMM \ data . htm" _gpath ="&MyPath . \
    Figures \ SIMM" ;
ods_tagsets . tablesonlylatex_file ="&MyPath . \ SIMM \ data . tex " (
    notop_nobot )
newfile=table_stylesheet="sas . sty "( url="sas ") _gpath ="&MyPath . \
    Figures \ SIMM" ;

*Print_the_data_matrix ;
proc_print_data=dissN_relia ;
run ;
quit ;

*For_LaTeX_output ;
ods_graphics_on/imagename=" fails " ;
ods_html_file ="&MyPath . \ Output \ SIMM \ fails . htm" _gpath ="&MyPath
    . \ Figures \ SIMM" ;
ods_tagsets . tablesonlylatex_file ="&MyPath . \ SIMM \ fails . tex "(
    notop_nobot )
```

```

newfile=table_styleheet="sas.sty"(url="sas")_gpath="&MyPath.\
  Figures\SIMM";

*****Graphing_procedure_to_graph_the_number_of_failed
  _companies_under_each_test_condition
*****Code_exported_form_ODS_Graphics_designer;
proc_template;
define_statgraph_sgdesign;
dynamic_MARKCAPP_R_IT;
begingraph;
  _entrytitle_halign=center_'The failures observed for
    different size companies at 3 different inspection times';
  _layout_lattice_/_rowdatarange=data_columndatarange=data_
    rowgutter=10_columngutter=10;
  _layout_overlay_/_xaxisopts=(discreteopts=(
    tickvaluefitpolicy=splitrotate));
  _barchart_category=_MARKCAPP_response=_R/_group=_IT_
    name='bar'_stat=mean_groupdisplay=Cluster_clusterwidth
    =0.85;
  _discretelegend_'bar'/_opaque=false_border=true_
    halign=right_valign=top_displayclipped=true_across=1_order=
    rowmajor_location=inside;
  _endlayout;
endlayout;
endgraph;
end;
run;

proc_sgrender_data=DISSN.RELIA_template=sgdesign;
dynamic_MARKCAPP="MARKCAPP" _R="R" _IT="IT";
run;

```

```
*****End;
```

## B.2.2 This code generates Figure 3.6

The code also generates the goodness-of-fit tables for the Log-Normal, Exponential, Gamma and Weibull distribution given after Figure 3.6

```
*For LATEX output;
ods _all_ close;
%let MyPath = C:\Users\Anel\Desktop\D_dissertation\chapters\
  chapter2;
ods graphics / outputfmt=png;

ods html file="&MyPath.\Output\SIMM\Life\life.htm" gpath="&
  MyPath.\Figures\SIMM\Life";
ods tagsets.tablesonlylatex file="&MyPath.\SIMM\Life\life.tex
  "(notop nobot)
newfile=table stylesheet="sas.sty"(url="sas") gpath="&MyPath.\
  Figures\SIMM\Life";

*Determine the life distribution by simulation;
proc iml;
*To make output comparable if needed;
call randseed(45);

*Import the parameters of the non-linear model;
use dissn.nlinout_log ;
read all into std;
std=std[1,ncol(std)-1:ncol(std)];
*print std;
*The percentile value at which a failure will be defined;
alpha=0.05;
```

```
N=500;* Actually the time variable;
mu=0;
*****Select the failure threshold based on the
  percentile of a small market capitalization company;
smallCap=0.01*(10**9);
X=log(smallCap)//{1};
factor=std*X;
sigmaf=exp(-factor);*Create the standard deviation of the
  return based on the output from the non-linear model;
*print sigmaf;
Tcll=(probit(alpha)*sigmaf)+mu;*The percentile which will be
  the failure threshold;
*print Tcll;
Tclu=(probit(1-alpha)*sigmaf)+mu;
*print Tclu;

*The stress level at which the life distribution will be
  assessed. Assuming the distribution remains the same at all
  stress levels;
s=0.045;
markCapp=s*(10**9);*the stress factor in Billions;
X=log(markCapp)//{1};
factor=std*X;
sigma=exp(-factor);*create the standard deviation of the
  return based on the output from the non-linear model;
*print sigma;
*call randseed(0);
Rt=J(N,1,0);

do k=1 to 2000;
```



```
call randgen(Rt, "Normal", Mu, Sigma);*Populate the
return matrix;
if ((Rt>Tclu)[+])>0 then do;
    idx=loc(Rt>=Tclu);
    min=min(idx); *Determine the first time at
    which the returns went above the loss
    threshold;
    t = t//min;
end;
end;

create dissn.life from t[colname={'t'}];
append from t;

quit;

*For LATEX output;
ods graphics on/imagename="lifeh";
ods html file="&MyPath.\Output\SIMM\Life\lifeh.htm" gpath="&
MyPath.\Figures\SIMM\Life";
ods tagsets.tablesonlylatex file="&MyPath.\SIMM\Life\lifeh.tex
(notop nobot)
newfile=table stylesheet="sas.sty"(url="sas") gpath="&MyPath.\
Figures\SIMM\Life";

*Test if the life distribution is Weibull, Lognormal, Gamma or
Exponential;
proc univariate data=dissn.life normal ;
var t;
histogram/ weibull lognormal gamma exponential;
run;
```

```
quit;
```

### B.2.3 This code generates Table 3.16–3.18

```
*For LATEX output;  
ods _all_ close;  
%let MyPath = C:\Users\Anel\Desktop\D_dissertation\chapters\  
    chapter2;  
ods graphics / outputfmt=png;  
  
*****Prepare the data for the  
    reliability procedure;  
data dissn.relia;  
set dissn.relia;  
obs=_n_;  
run;  
quit;  
  
proc iml;  
use dissn.relia;  
read all into x[colname=names];  
x=repeat(x,2);  
call sort(x,ncol(x));  
*print x;  
create dissn.relia1 from x[colname=names];  
append from x;  
quit;  
data dissn.fin_relia;  
set dissn.relia1(drop=obs);  
obs=_n_;
```

```

*If T1 is missing (.), then T2 represents a left-censoring
  time;
if mod(obs,2)=0 then do;
t1=.;
t2=IT;
fail=r;
Mc=markCapp;
end;
*If T2 is missing, T1 represents a right-censoring time;
if mod(obs,2)^=0 then do;
t1=IT;
t2=.;
fail=M-r;
Mc=markCapp;
end;
run;

data dissn.fin_relia;
set dissn.fin_relia(drop= IT r markCapp M);
run;
quit;
*****End data
  preparation;
*For LATEX output;
ods graphics on/imagename="relia";
ods html file="&MyPath.\Output\SIMM\relia\relia.htm" gpath="&
  MyPath.\Figures\SIMM\Relia";
ods tagsets.tablesonlylatex file="&MyPath.\SIMM\Relia\relia.
  tex"(notop nobot)
newfile=table stylesheet="sas.sty"(url="sas") gpath="&MyPath.\
  Figures\SIMM\Relia";

```

```
*Run the reliability procedure with a linear-link function;
ods trace on;
ods select modprmet;
proc reliability data=dissn.fin_relia;
distribution exponential;
freq fail;
*logscale mc;
model (t1 t2)= Mc;
rplot (t1 t2)= Mc/
    pplot
    fit = model
    noconf
    plotdata
    plotfit 10 50 90
;
run;
ods trace off;
quit;
```

```
*Run the reliability procedure with a power-link function;
ods trace on;
ods select modprmet parmest;
proc reliability data=dissn.fin_relia;
distribution exponential;
freq fail;
*logscale mc;
model (t1 t2)= Mc/ relation=pow ;
make 'modprmet' out=dissn.fin_relia_graphMEst;
make 'parmest' out=dissn.fin_relia_graphest;
rplot (t1 t2)= Mc/
```

```
pplot  
fit = model  
    relation=pow  
noconf  
plotdata  
plotfit 10 50 90  
;  
run;  
ods trace off;  
quit;
```

#### **B.2.4 This code generates Figure 3.7**

```
*For LaTeX output;  
ods _all_ close;  
%let MyPath = C:\Users\Anel\Desktop\D_dissertation\chapters\  
    chapter2;  
ods graphics / outputfmt=png;  
  
proc iml;  
*Import the exponential parameter at each market cap level;  
use dissn.Fin_relia_graphest;  
read all into x[colname=names];  
close dissn.Fin_relia_graphest;  
call seed(45);  
  
*Import the link function parameters;  
use dissn.Fin_relia_graphMest;  
read all into xy[colname=names];  
close dissn.Fin_relia_graphMest;  
*The link function parameters;  
B0=xy[1,1];
```

```

B1=xy[2,1];

*Import the market cap values;
use dissn.relia;
read all into mc;
close dissn.relia;
mc=mc[,1];

*****Simulate the exponential distribution for
different market cap values;
M=1000;
v=J(M,1,.);
u=J(M,1,.);
y=J(M,1,.);
eta=x[,2];
do j=1 to 11 by 2;*plot only for 5 companies
    mmc=mmc//mc[j,];
    y[,1]=mc[j,];*The market cap value;
    parm=eta[j+1,];
    call randgen(v,'exponential',parm);*An exponential
        variable;
    *if j<=3 then print v;
    u[,1]=(1/parm)*exp(-v/parm);*The exponential pdf
        evaluated at v, the exponential variables;
    DataM=dataM//(y||v||u);
end;

link=1/exp(b0+b1*log(dataM[,1]));
DataM=DataM||link;

*Store the simulated data;

```

```

create dissn.Fin_relia_graph from DataM[colname={ 'marketCapp' ,
  'y', 'dist' , 'link' }]];
append from DataM;

quit;
*For LaTeX output;
ods graphics on/imagename="densities ";
ods html file="&MyPath.\Output.htm" gpath="&MyPath.\Figures ";
ods tagsets.tablesonlylatex file="&MyPath.densities.tex"(notop
  nobot)
newfile=table stylesheet="sas.sty"(url="sas") gpath="&MyPath.\
  Figures ";

*****Graphing procedure to graph exponential
  distribution for different market capitalization;
*****Code exported form ODS Graphics designer;
proc template;
define statgraph sgdesign;
dynamic _Y _DIST _MARKETCAPP;
begingraph;
  entrytitle halign=center 'The_life-time_distributions_for_
    diffirent_market_cap_values';
  layout lattice / rowdatarange=data columndatarange=data
    rowgutter=10 columngutter=10;
  layout overlay;
    seriesplot x=_Y y=_DIST / group=_MARKETCAPP name='
      series' connectorder=xaxis;
  discretelegend 'series' / opaque=false border=true
    halign=right valign=top displayclipped=true across
    =1 order=rowmajor location=inside;
endlayout;

```

```
endlayout ;
endgraph ;
end ;
run ;

proc sgrender data=DISSN.FIN_RELIA_GRAPH template=sgdesign ;
dynamic _Y="Y" _DIST="DIST" _MARKETCAP="MARKETCAP" ;
run ;
```

### B.2.5 This code generates Figure 3.11– 3.14

```
*For LaTeX output ;
ods _all_ close ;
%let MyPath = C:\Users\Anel\Desktop\D_dissertation\chapters\
chapter2\round2 ;
ods graphics / outputfmt=png ;

proc iml ;
*Import the reliability parameters for the link function ;
use dissn.Fin_relia_99_graphMest ;
read all into xy[colname=names] ;
close dissn.Fin_relia_99_graphMest ;
B0=xy[1,1] ;
B1=xy[2,1] ;
time=50 ;

*Evaluate for a fixed time , how probability of failure changes
with market CAP ;
do r=1 to 4 ;
T=time#r ;
do j=0.1 to 1 by 0.1 ;
k=j*10**9 ;
```



```

    val=1-exp(-T/(exp(B0+B1#log(k)))));
    dataM=dataM//(T||k||val);
    end;
end;

create dissn.InterpGraph from DataM[colname={ 'time ', 'marketCap
', 'Ft' }];
append from DataM;

quit;

*For LaTeX output;
ods graphics on/imagename="interp1";
ods html file="&MyPath.\Output.htm" gpath="&MyPath.\Figures";
ods tagsets.tablesonlylatex file="&MyPath.interp1.tex"(notop
nobot)
newfile=table stylesheet="sas.sty"(url="sas") gpath="&MyPath.\
Figures";

*****Graphing procedure to graph above;
*****Code exported form ODS Graphics designer;
proc template;
define statgraph sgdesign;
dynamic _marketCap _Ft _TIME;
begingraph;
    entrytitle halign=center 'The probability of a default at
time T=50,100,150 and 200 for different market CAPs';
layout lattice / rowdatarange=data columndatarange=data
rowgutter=10 columngutter=10;
layout overlay;

```

```
seriesplot x=_marketCap y=_Ft / group=_TIME name='
series' clusterwidth=0.85 connectorder=xaxis
grouporder=data;
discretelegend 'series' / opaque=false border=true
halign=right valign=top displayclipped=true across
=1 order=rowmajor location=inside;
endlayout;
endlayout;
endgraph;
end;
run;

proc sgrender data=DISSN.INTERPGRAPH template=sgdesign;
dynamic _marketCap="marketCap" _Ft="Ft" _TIME="TIME";
run;
*****End graphing procedure;

proc iml;
*Import the reliability parameters for the link function;
use dissn.Fin_relia_99_graphMest;
read all into xy[colname=names];
close dissn.Fin_relia_99_graphMest;
B0=xy[1,1];
B1=xy[2,1];
time=50;

*Evaluate for a fixed time, how probability of failure changes
with market CAP;
do r=1 to 4;
T=time#r;
*For smaller market CAP values;
```

```

do j=0.01 to 0.1 by 0.01;
k=j*10**9;
val=1-exp(-T/(exp(B0+B1#log(k))));
dataM=dataM//(T||k||val);
end;
end;

create dissn.InterpGraph2 from DataM[ colname={ 'time', '
marketCap', 'Ft' }];
append from DataM;

quit;

*For LaTeX output;
ods graphics on/imagename="interp2";
ods html file="&MyPath.\Output.htm" gpath="&MyPath.\Figures";
ods tagsets.tablesonlylatex file="&MyPath.interp2.tex"(notop
nobot)
newfile=table stylesheet="sas.sty"(url="sas") gpath="&MyPath.\
Figures";

*****Graphing procedure to graph above;
*****Code exported form ODS Graphics designer;
proc template;
define statgraph sgdesign;
dynamic _marketCap _Ft _TIME;
begingraph;
entrytitle halign=center 'The_probability_of_a_default_at_
time_T=50,100,150_and_200_for_diffirent_market_CAPS';
layout lattice / rowdatarange=data columndatarange=data
rowgutter=10 columngutter=10;

```

```

    layout overlay;
      seriesplot x=_marketCap y=_Ft / group=_TIME name='
        series' connectorder=xaxis;
      discretelegend 'series' / opaque=false border=true
        halign=right valign=top displayclipped=true across
        =1 order=rowmajor location=inside;
    endlayout;
  endlayout;
endgraph;
end;
run;

proc sgrender data=DISSN.INTERPGRAPH2 template=sgdesign;
dynamic _marketCap="marketCap" _Ft="Ft" _TIME="TIME";
run;
*****End graphing procedure;

proc iml;
*Import the reliability parameters for the link function;
use dissn.Fin_relia_99_graphMest;
read all into xy[colname=names];
close dissn.Fin_relia_99_graphMest;
B0=xy[1,1];
B1=xy[2,1];
marketC=0.1*10**9;

*Evaluate how probability of failure changes over time for a
  fixed market CAP value;
do j=10 to 1000 by 1;
  val=1-exp(-j/(exp(B0+B1#log(marketC))));
  dataM=dataM/(j||marketC||val);

```

```
end;

create dissn.InterpGraph3 from DataM[colname={ 'time ', '
  marketCap ', 'Ft' }]];
append from DataM;

quit;

*For LaTeX output;
ods graphics on/imagename="interp3 ";
ods html file="&MyPath.\Output.htm" gpath="&MyPath.\Figures ";
ods tagsets.tablesonlylatex file="&MyPath.interp3.tex"(notop
  nobot)
newfile=table stylesheet="sas.sty"(url="sas") gpath="&MyPath.\
  Figures";

*****Graphing procedure to graph above;
*****Code exported form ODS Graphics designer;
proc template;
define statgraph sgdesign;
dynamic _TIME _FT _MARKETCAP;
begingraph;
  entrytitle halign=center 'The_probability_of_failure_within
    _time_for_a_company_of_specific_market_CAP';
  layout lattice / rowdatarange=data columndatarange=data
    rowgutter=10 columngutter=10;
  layout overlay;
    seriesplot x=_TIME y=_FT / group=_MARKETCAP name='
      series' connectorder=xaxis;
  discretelegend 'series' / opaque=false border=true
    halign=right valign=top displayclipped=true across
```

```

        =1 order=rowmajor location=inside;
    endlayout;
endlayout;
endgraph;
end;
run;

proc sgrender data=DISSN.INTERPGRAPH3 template=sgdesign;
dynamic _TIME="TIME" _FT="FT" _MARKETCAP="MARKETCAP";
run;
*****End graphing procedure;

proc iml;
*Import the reliability parameters for the link function;
use dissn.Fin_relia_99_graphMest;
read all into xy[colname=names];
close dissn.Fin_relia_99_graphMest;
B0=xy[1,1];
B1=xy[2,1];
marketC=0.01*10**9;* smaller market CAP value

*Evaluate how probability of failure changes over time for a
fixed market CAP value;
do j=10 to 1000 by 1;
    val=1-exp(-j/(exp(B0+B1#log(marketC))));
    dataM=dataM//(j || marketC || val);
end;

create dissn.InterpGraph4 from DataM[colname={'time',
marketCap', 'Ft' }];
append from DataM;

```

```
quit;

*For LaTeX output;
ods graphics on/imagename="interp4 ";
ods html file="&MyPath.\Output.htm" gpath="&MyPath.\Figures ";
ods tagsets.tablesonlylatex file="&MyPath.interp4.tex"(notop
  nobot)
newfile=table stylesheet="sas.sty"(url="sas") gpath="&MyPath.\
  Figures";

*****Graphing procedure to graph above;
*****Code exported form ODS Graphics designer;
proc template;
define statgraph sgdesign;
dynamic _TIME _FT _MARKETCAP;
begingraph;
  entrytitle halign=center 'The_probability_of_failure_within
    _time_for_a_company_of_specific_market_CAP';
  layout lattice / rowdatarange=data columndatarange=data
    rowgutter=10 columngutter=10;
  layout overlay;
    seriesplot x=_TIME y=_FT / group=_MARKETCAP name='
      series' connectorder=xaxis;
    discretelegend 'series' / opaque=false border=true
      halign=right valign=top displayclipped=true across
      =1 order=rowmajor location=inside;
  endlayout;
endlayout;
endgraph;
end;
```

```
run ;
```

```
proc sgrender data=DISSN.INTERPGRAPH4 template=sgdesign ;
```

```
dynamic _TIME="TIME" _FT="FT" _MARKETCAP="MARKETCAP" ;
```

```
run ;
```

```
*****End graphing procedure ;
```



# Appendix C

## Acronyms

Provide a short introduction to the appendix here. Mention that acronyms are listed alphabetically and typeset in bold, with the meaning of the acronym alongside. Also note that you must include acronyms using `useacronym` for them to show up here. If there are no acronyms defined in your text, this introduction will also not be displayed:

<b>ABT</b>	Accelerated binary test
<b>ADDT</b>	Accelerated destructive degradation test
<b>ALT</b>	Accelerated life test
<b>ARMDT</b>	accelerated repeated measures degradation test
<b>AT</b>	accelerated test
<b>CAP</b>	capitalization
<b>CDF</b>	cumulative density function
<b>EM</b>	Expectation-Maximization
<b>IID</b>	independent and identically distributed
<b>LS</b>	Least Squares
<b>MFFT</b>	Mean time to failure

---

<b>ML</b>	Maximum Likelihood
<b>PDF</b>	probability density function
<b>rv</b>	random variable
<b>VaR</b>	Value at Risk
<b>WLS</b>	Weighted Least Squares

# Appendix D

## Derived publications

The following publication is derived from this dissertation:

- A. Opper, M.T. Loots, F.J.C. Beyers. “Parallel consideration of low probability events in stock markets and one-shot device testing.” *Scandinavian Actuarial Journal (In preparation)*.