Multi-Parameter Regression
Survival Models

Ph.D. Thesis

Kevin Burke

Supervisors:
Prof. Gilbert MacKenzie
Prof. Eamonn Murphy
Acknowledgements

This work was funded by the Irish Research Council (www.research.ie) during the period October 1st 2010 – September 31st 2013 and, thereafter, by the Mathematics and Statistics Department, University of Limerick. Throughout my Ph.D. studies, a number of people have been very helpful to me; I would like to use this opportunity to thank them.

First and foremost, I would like to thank my supervisors, Prof. Gilbert MacKenzie and Prof. Eamonn Murphy, for their constant encouragement and guidance which is very much appreciated and for so generously supplementing my travel funds - this has greatly enhanced my Ph.D. experience. I also thank them both for always being available to talk to and for always having my best interests in mind.

Thank you to my colleagues at the Department of Mathematics and Statistics for providing me with a pleasant and stimulating work environment; in particular, I have had many interesting discussions with Dr. Joseph Lynch and Dr. Ali Sheikhi. I am also very grateful to the current and former heads of department, Dr. Alan Hegarty and Dr. Mark Burke, who have been very kind to me over these last few years, and to Niamh Dooley, the department administrator, whose assistance has always been most efficient.

I would like to thank Prof. Niels Keiding for facilitating my visit to the Department of Biostatistics, University of Copenhagen, where I spent the month of May 2013. I am very grateful to Niels for welcoming me to the department and involving me in the various activities which took place during my stay. I would also like to thank Dr. Christian Bressen Pipper and Dr. Thomas Gerds for taking an interest in my work and Prof. Per Kragh Andersen for kindly lending me a bicycle - this allowed me to explore many parts of Copenhagen and made my stay all the more enjoyable.

Of course, a very special thanks goes to my parents for providing me with the opportunities that have led to where I am today.

And whilst last, furthest from least, Joanna - thank you for always being there to support me.
Summary

In general, parametric regression models can be motivated by allowing the parameters of a probability distribution to depend on covariates. Furthermore, it is standard practice to relate covariates to one parameter of particular interest; we will refer to this approach as single parameter regression (SPR). In these SPR models, the role of the other (covariate independent) parameters is often little more than to provide the model with sufficient generality to adapt to data. A more flexible approach is to also regress these other parameters on covariates; we call this multi-parameter regression (MPR). The primary focus of this thesis is the development of MPR models in the setting of survival analysis (of course, MPR models are not limited to the field of survival analysis).

In Chapter 1 we review some basic concepts of survival analysis - these are standard and may be skipped by the reader familiar with the area. Chapter 2 is largely concerned with developing likelihood theory for survival data which, again, is quite standard and may be skipped. However, in Section 2.3.2 we propose a method - m.l.e. simulation - for calculating the standard error / confidence intervals for functions of parameters. M.l.e. simulation, which competes with the well-known delta method and method of bootstrapping, is based on simulating a sample of \( \hat{\theta} \) vectors, \( \{ \hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(m)} \} \), from \( \hat{\theta}^{(b)} \sim N(\hat{\theta}, \hat{\Sigma}) \) and is used throughout the thesis. In Chapter 3 we discuss a method for simulating survival data and, furthermore, we extend this method to handle models that support a cured proportion (Section 3.5). This is followed by some interesting simulation studies (Section 3.6) where, among other things, we compare the delta method to m.l.e. simulation and investigate how reliably the cured proportion can be estimated (if it exists).

We consider standard regression models for survival data in Chapter 4; in particular, Section 4.18 contains a brief review of some commonly used SPR survival models. Chapter 5 contains our development of MPR survival models: we display the flexibility of MPR (relative to SPR) and discuss the consequences of the approach in terms of interpreting covariate effects (via the hazard ratio), carrying out hypothesis tests (on regression coefficients) and variable selection procedures. Motivated by the need to enhance inter-
pretability of MPR models (and indeed any regression model), in Chapter 6 we propose a least squares approximation to covariate-dependent model quantities, e.g., the hazard function. The proposed method allows straightforward interpretation of covariate effects in terms of the quantity in question but, of course, depends on the adequacy of the approximation. In Chapter 7 we consider frailty modelling - an area of survival analysis concerned with the analysis of unexplained variation (or heterogeneity). In particular, we go through the straightforward algebra of multiplicative gamma frailty which can be used to generalise any parametric model, e.g., Weibull MPR model with multiplicative gamma frailty. Furthermore, using gamma frailty as our starting point, we propose some extensions which combine the ideas of multi-parameter regression and frailty. Finally, we close with a discussion in Chapter 8.
Contents

Acknowledgements iii

Summary v

List of Tables xiii

List of Figures xv

1 Review of Survival Analysis 3
   1.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3
   1.2 Theory and Relationships . . . . . . . . . . . . . . . . . . . . 6
     1.2.1 Standard Probability Functions . . . . . . . . . . . . . 6
     1.2.2 Survival Functions . . . . . . . . . . . . . . . . . . . . 7
     1.2.3 Relationships Between Functions . . . . . . . . . . . . . 8
   1.3 Non-Parametric Methods . . . . . . . . . . . . . . . . . . . . . 12
     1.3.1 Empirical Survivor Function . . . . . . . . . . . . . . . 13
     1.3.2 Kaplan-Meier Estimate of Survivor Curve . . . . . . . 13
     1.3.3 Nelson-Aalen Estimate of Cumulative Hazard Function 17
     1.3.4 Relationship Between Estimators . . . . . . . . . . . . . 18
   1.4 Parametric Survival Distributions . . . . . . . . . . . . . . . 20
     1.4.1 Exponential . . . . . . . . . . . . . . . . . . . . . . . . 21
     1.4.2 Weibull . . . . . . . . . . . . . . . . . . . . . . . . . . 23
     1.4.3 Gompertz . . . . . . . . . . . . . . . . . . . . . . . . . 25
     1.4.4 Log-Logistic . . . . . . . . . . . . . . . . . . . . . . . 27
     1.4.5 Burr Type XII . . . . . . . . . . . . . . . . . . . . . . 29
     1.4.6 Time-Dependent Logistic . . . . . . . . . . . . . . . . . 30

vii
## CONTENTS

1.4.7 Piecewise Exponential .................................. 32

2 Inference .................................................. 35

2.1 Likelihood ................................................. 35

2.1.1 Maximising the Likelihood Function ................. 35

2.1.2 Right-Censored Likelihood .......................... 37

2.2 Comparing Models ....................................... 41

2.2.1 Nested Models ....................................... 42

2.2.2 Non-Nested Models ................................... 42

2.3 Uncertainty .............................................. 45

2.3.1 Uncertainty in the M.L.E. ........................... 45

2.3.2 Uncertainty in Functions of the M.L.E. ............ 47

3 Simulating Survival Data ............................... 55

3.1 Introduction ............................................. 55

3.2 Reasons for Simulating ................................. 56

3.3 Inverse Survivor Function ............................... 58

3.4 Random Censoring ..................................... 59

3.5 Extension to Include Cure Rate Models ............... 61

3.6 Some Simulation Studies ............................... 63

3.6.1 True Model Versus an Incorrect Model .............. 63

3.6.2 Extrapolating Beyond the Data .................... 66

3.6.3 Bias, S.E. and Correlation in Estimates ........... 67

3.6.4 Delta Method Versus M.L.E. Simulation .......... 70

3.6.5 Estimating the Cured Proportion .................... 74

3.6.6 Decreasing Hazards or Cured Proportion? .......... 75

4 Regression Models ....................................... 79

4.1 Introduction ............................................. 79

4.2 The Proportional Hazards Model ....................... 80

4.3 Other Regression Models ............................... 82

4.4 Cox’s Semi-Parametric Model ......................... 83

4.5 Parametric Models ...................................... 87

4.5.1 Weibull ............................................... 89

4.5.2 Gompertz ............................................. 89
CONTENTS

4.5.3 Log-logistic ........................................... 90
4.5.4 Burr ..................................................... 90
4.5.5 Time-Dependent Logistic ............................... 91
4.5.6 Piecewise Exponential ................................. 92

5 Multi-Parameter Regression 97

5.1 Introduction ............................................. 97
5.2 Multi-Parameter Regression Models ..................... 99
  5.2.1 Weibull ............................................. 100
  5.2.2 Gompertz ............................................. 102
  5.2.3 Log-Logistic ....................................... 104
  5.2.4 Burr ............................................... 106
  5.2.5 Time-Dependent Logistic ........................... 109
  5.2.6 Piecewise Exponential .............................. 111
5.3 The Hazard Ratio: Covariate Dependence ............... 120
5.4 Hypothesis Testing and Variable Selection ............ 142
5.5 Simulation Studies .................................. 158
  5.5.1 Correlation Matrix for Estimated Parameters ....... 158
  5.5.2 Bias of Estimated Parameters ....................... 161
  5.5.3 Variable Selection ................................ 163

6 Implied Regression Models 167

6.1 Introduction ............................................. 167
6.2 Approximating Implied Regression Models ............... 169
  6.2.1 Least Squares Solution ............................ 170
  6.2.2 Uncertainty ....................................... 171
6.3 Life Time ............................................... 173
  6.3.1 Restricted Mean Life ............................... 174
  6.3.2 Median Life ...................................... 179
6.4 Hazard Ratio ........................................... 181

7 Frailty Models 189

7.1 Introduction ............................................. 189
7.2 Marginal Functions .................................... 190
7.3 Estimation .............................................. 192
Appendix C  Orthogonal Weibull  285
  C.1  Introduction ........................................... 285
  C.2  Gamma and Related Functions .......................... 285
  C.3  Standard Weibull Expected Information ............ 287
  C.4  Derivation of the Orthogonal Weibull ............... 289
  C.5  Orthogonal Weibull MPR Model ....................... 291

Appendix D  Frailty  295
  D.1  Introduction ........................................... 295
  D.2  Marginal Hazard Function ............................. 295
  D.3  Laplace Transform .................................... 297
      D.3.1  Marginal Likelihood ............................ 298
      D.3.2  Gamma Frailty .................................... 299

Bibliography  300
List of Tables

1.1 Kaplan-Meier Estimates for Lung Cancer Data .............. 16
2.1 Rules of Thumb for $AIC$ Differences ......................... 43
2.2 Model Comparison ........................................ 44
2.3 Time Taken to Produce Confidence Intervals ............. 53
3.1 Simulation Settings .......................................... 63
3.2 Average $AIC$ Differences .................................. 64
3.3 Average Parameter Estimates with Standard Error and Relative Bias ........................................ 68
3.4 Correlation Between Estimates ............................... 69
3.5 Coverage Percentage for Confidence Intervals ............. 72
3.6 Simulation Settings .......................................... 74
3.7 Estimates of Cured Proportion for Gompertz Data .......... 75
3.8 Estimates of Cured Proportion for Non-Cure (Weibull) Data 76
4.1 Linearity of Regression Models ............................... 83
4.2 Cox Model Fit ................................................ 86
4.3 SPR Model Fits .............................................. 93
5.1 MPR Model Fits ............................................. 113
5.2 Burr MPR Model Fit ........................................ 117
5.3 Piecewise Exponential MPR Model Fit ...................... 119
5.4 Log-Logistic MPR Full Covariate Analysis ................. 124
5.5 Log-Logistic MPR Overall Hazard Ratios with C.I.s .......... 132
5.6 Gompertz MPR Full Covariate Analysis ..................... 134
5.7  Gompertz MPR Overall Cure Probabilities with C.I.s . . . . . . . . . . 141
5.8  Log-Logistic MPR: Significance of Scale and Shape Effects . . . . . 148
5.9  Likelihood Ratio Tests and AIC differences . . . . . . . . . . . . . . 151
5.10 Reduced Log-Logistic MPR Model Fit . . . . . . . . . . . . . . . . . 152
5.11 Variable Selection in the Lung Cancer Data . . . . . . . . . . . . . . 156
5.12 Weibull MPR Model: Average Relative Bias . . . . . . . . . . . . . 162
5.13 Orthogonal Weibull MPR Model: Average Relative Bias . . . . . . . 162
5.14 Relative Frequency of Selections . . . . . . . . . . . . . . . . . . . . 164

6.1  Restricted Mean Regression: Log-Logistic MPR Model . . . . . . . . 177
6.2  Median Regression: Log-Logistic MPR Model . . . . . . . . . . . . . 180
6.3  Log-Logistic MPR Hazard Ratios with C.I.s . . . . . . . . . . . . . . 187

7.1  Weibull and Weibull-Gamma SPR Full Covariate Analyses . . . . . . 203
7.2  Summary of Fitted Weibull Models . . . . . . . . . . . . . . . . . . . 205
7.3  A Comparison of Weibull Models . . . . . . . . . . . . . . . . . . . . 207
7.4  Weibull-Gamma Structured Dispersion Full Covariate Analysis . . 216

A.1  Lung Cancer Covariates . . . . . . . . . . . . . . . . . . . . . . . . . 228

B.1  Weibull MPR Simulation Settings . . . . . . . . . . . . . . . . . . . 254
B.2  Orthogonal Weibull MPR Simulation Settings . . . . . . . . . . . . 254
B.3  Log-Logistic MPR Simulation Settings . . . . . . . . . . . . . . . . 255
B.4  Gompertz MPR Simulation Settings . . . . . . . . . . . . . . . . . . 255
B.5  Log-Logistic MPR Model: Average Relative Bias . . . . . . . . . . 257
B.6  Gompertz MPR Model: Average Relative Bias . . . . . . . . . . . . 257

C.1  Standard and Orthogonal Weibull MPR Fits . . . . . . . . . . . . . . 294
List of Figures

1.1 Survival of six individuals .............................................. 5
1.2 Kaplan-Meier curve with confidence intervals .................... 16
1.3 Weibull hazard function ................................................. 24
1.4 Kaplan-Meier and Weibull survivor curves ....................... 25
1.5 Gompertz hazard function ................................................. 26
1.6 Log-logistic hazard function ............................................. 28
1.7 Burr hazard function ...................................................... 30
1.8 TDL hazard function ...................................................... 31
1.9 Piecewise exponential hazard function ................................ 33
2.1 Kaplan-Meier, Weibull and log-logistic survivor curves .......... 44
2.2 Confidence intervals for estimated survivor curves ............. 52
3.1 Average survivor and hazard curves for two scenarios ........... 65
3.2 Extrapolating survivor and hazard curves beyond the data .... 66
3.3 Scatter plots of estimated regression coefficients ................ 69
3.4 Confidence intervals for the first 50 repetitions ................ 73
4.1 Kaplan-Meier and Cox survivor curves ............................. 86
4.2 Estimated hazard function, hazard ratio and survivor function for three SPR models ........................................ 94
4.3 Hazard ratios with confidence intervals for three SPR models .. 95
5.1 Weibull MPR hazard ratio ................................................. 102
5.2 Gompertz MPR hazard ratio .............................................. 104
5.3 Log-logistic MPR hazard ratio .......................................... 106
5.4 Burr MPR hazard ratio ........................................... 108
5.5 Time-dependent logistic MPR hazard ratio ......................... 110
5.6 Estimated hazard function, hazard ratio and survivor function for three MPR models ........................................ 114
5.7 Hazard ratios with confidence intervals for three MPR models 115
5.8 AIC difference versus number of intervals .......................... 118
5.9 Hazard ratios for the piecewise exponential MPR model .......... 120
5.10 Correlation matrix for estimated regression coefficients in the log-logistic MPR model ......................................... 125
5.11 Treatment hazard ratios (log-logistic MPR) ......................... 126
5.12 Age hazard ratios (log-logistic MPR) ............................. 128
5.13 WHO status hazard ratios (log-logistic MPR) ...................... 128
5.14 Sex hazard ratio (log-logistic MPR) ............................... 129
5.15 Smoker hazard ratios (log-logistic MPR) .......................... 129
5.16 Cell type hazard ratios (log-logistic MPR) ......................... 130
5.17 Metastases hazard ratios (log-logistic MPR) ...................... 130
5.18 Sodium hazard ratios (log-logistic MPR) .......................... 131
5.19 Albumen hazard ratios (log-logistic MPR) ......................... 131
5.20 Correlation matrix for estimated regression coefficients in the Gompertz MPR model ........................................... 135
5.21 Treatment hazard ratios (Gompertz MPR) .......................... 135
5.22 Age hazard ratios (Gompertz MPR) ............................... 136
5.23 WHO status hazard ratios (Gompertz MPR) ...................... 136
5.24 Sex hazard ratio (Gompertz MPR) ............................... 137
5.25 Smoker hazard ratios (Gompertz MPR) ......................... 137
5.26 Cell type hazard ratios (Gompertz MPR) .......................... 138
5.27 Metastases hazard ratios (Gompertz MPR) ...................... 138
5.28 Sodium hazard ratios (Gompertz MPR) .......................... 139
5.29 Albumen hazard ratios (Gompertz MPR) .......................... 139
5.30 Joint 95% confidence ellipses and individual 95% confidence intervals ......................................................... 149
5.31 Treatment hazard ratios (log-logistic MPR - reduced) .......... 153
5.32 WHO status hazard ratios (log-logistic MPR - reduced) ....... 153
5.33 Cell type hazard ratios (log-logistic MPR - reduced) .......... 154
LIST OF FIGURES

5.34 Metastases hazard ratios (log-logistic MPR - reduced) . . . . . 154
5.35 Sodium hazard ratios (log-logistic MPR - reduced) . . . . . . 155
5.36 Albumen hazard ratios (log-logistic MPR - reduced) . . . . . 155
5.37 Scatter matrices of estimated regression coefficients for the
Weibull MPR model . . . . . . . . . . . . . . . . . . . . . . . 159
5.38 Scatter matrices of estimated regression coefficients for the
Orthogonal Weibull MPR model . . . . . . . . . . . . . . . . . 160
5.39 Relative frequency of selection of each covariate . . . . . . 165
5.40 Boxplots of estimated regression coefficients for scenarios with
$p = 50\%$ . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 166

6.1 Treatment hazard ratios (log-logistic MPR) . . . . . . . . . . 182
6.2 Age hazard ratios (log-logistic MPR) . . . . . . . . . . . . . . 183
6.3 WHO status hazard ratios (log-logistic MPR) . . . . . . . . . 183
6.4 Sex hazard ratio (log-logistic MPR) . . . . . . . . . . . . . . 184
6.5 Smoker hazard ratios (log-logistic MPR) . . . . . . . . . . . 184
6.6 Cell type hazard ratios (log-logistic MPR) . . . . . . . . . . 185
6.7 Metastases hazard ratios (log-logistic MPR) . . . . . . . . . . 185
6.8 Sodium hazard ratios (log-logistic MPR) . . . . . . . . . . . 186
6.9 Albumen hazard ratios (log-logistic MPR) . . . . . . . . . . 186

B.1 Scatter matrices of estimated regression coefficients for the
Log-logistic MPR model . . . . . . . . . . . . . . . . . . . . . 256
B.2 Scatter matrices of estimated regression coefficients for the
Gompertz MPR model . . . . . . . . . . . . . . . . . . . . . . 256
Chapter 1

Review of Survival Analysis

1.1 Introduction

Survival data typically measure the time elapsed from a well-defined time origin until the occurrence of an event, for example, time from diagnosis of a disease until death from that disease, treatment start date until patient is cured, marriage until divorce, birth of first child until birth of second child, start-up of a business until its closure, unemployment date until re-employment date or commencement of operation until failure of a mechanical component. The survival time is the length of this interval. The use of the word “survival" refers to surviving in a particular state, for example, the state of living, the state of being married or the state of being unemployed. It is clear therefore, that we are not necessarily describing survival in a literal sense with the endpoint being death. This terminology comes from the field of medicine, where death often is the endpoint, but survival-type data arise in many other areas (e.g. engineering, social sciences and econometrics) as is clear from the examples mentioned above.

Survival analysis has arisen as a rich field of statistics due to the special requirements of, and interesting challenges presented by, these particular types of data. One issue is that the data are typically skewed to the right. This is easily dealt with, for example, by using a parametric distribution which can handle skewness or by transforming the data to remove it. However, the most distinctive feature of survival data is a particular type of missingness called
censoring and it is this feature that necessitates special theory and methods of analysis. In particular we consider right-censoring which is the most common form of censoring. An individual’s survival time is right-censored when that particular individual has not yet experienced the event of interest on the last observation date. The event therefore occurs at some point in time which is greater than, i.e. to the right of, the final observation date. We therefore observe some survival times and have a lower bound on others. Censoring occurs due to the way survival times are measured; we must wait for an event to happen to retrieve a measurement. This differs from other data where measurements are typically recorded without such a delay, for example, distance, weight, age, temperature, income etc. A study of survival will take place over some period of time. By the end of the study, not all individuals may have experienced an event and it is usually infeasible to run such studies until all individuals do experience the event on account of, for example, financial or ethical reasons. More practically, we may have gained enough relevant information from our current sample without waiting for further events to occur. In a clinical setting some individuals can become lost to follow-up, that is, they may stop reporting to the clinician for potentially unknown reasons. Again, we only know that the event occurs sometime after the last observation date for this individual.

An artificial example of survival data is depicted in Fig. 1.1. The left panel shows the study in calendar time beginning at zero when the study commenced. There is typically a recruitment period, after which no more individuals are introduced into the sample, and finally the study ends on some date. We see that individuals 3, 4 and 6 experience an event during the study. Individuals 2 and 5 do not experience an event before the study ends and are therefore right-censored. Individual 1 is also right-censored as he/she dropped out during the study. The right panel is realigned so that time zero corresponds to the individual’s specific starting point which may be date of birth, diagnosis of a disease, treatment commencement etc.

Another form of censoring is interval censoring which occurs when the event time is unavailable but is known to lie in an interval. A separate idea is that of left truncation (also called delayed entry). This occurs when some individuals’ current survival times are not at time zero at recruitment,
that is to say, they have survived up until time of recruitment and have yet to experience an event. Thus, analysis must be made conditional on having survived up to recruitment time. We will not consider these issues but various censoring mechanisms are discussed in detail in Kalbfleisch & Prentice (2002, chap. 3) and Lawless (2003, chap. 2).

It should be clear from the above that standard statistical methods will not suffice for survival data due to the presence of censoring. For example, calculation of the mean or the median of the observed times or producing graphical summaries such as a histogram or box-plot are all rendered meaningless, as are standard hypothesis testing procedures and regression analyses.
CHAPTER 1. REVIEW OF SURVIVAL ANALYSIS

1.2 Theory and Relationships

Here we define some probability functions which we split into two groups: standard probability functions and survival functions. We put the probability density function and the cumulative distribution function into the first group, being familiar functions which appear in the usual probability and statistical theory, and the survivor function and hazard function into the second group. All of these are in fact probability functions, and are all equivalent representations of the underlying distribution, but the distinction is made because the latter two are regarded as being more fundamental in survival theory.

1.2.1 Standard Probability Functions

We start by defining \( T \) to be the random variable representing the survival time. Here we only consider the case where \( T \) is continuous although it is possible for \( T \) to be discrete or to have both continuous and discrete components. Thus \( T \in [0, \infty) \) has probability density function, \( f(t) \geq 0 \), such that \( \int_0^\infty f(t)dt = 1 \) and the probability of \( T \) falling in some interval \([t_1, t_2] \subseteq [0, \infty)\) is given by

\[
\Pr(t_1 \leq T \leq t_2) = \int_{t_1}^{t_2} f(t)dt.
\]

Note that

\[
\Pr(t \leq T < t+\Delta t) \approx f(t)\Delta t,
\]

for small \( \Delta t \). The cumulative distribution function, \( F(t) \), is the probability that \( T \) is less than or equal \( t \),

\[
F(t) = \Pr(T \leq t) = \int_0^t f(u)du.
\]

Hence,

\[
\frac{d}{dt}F(t) = f(t).
\]
### 1.2.2 Survival Functions

The *survivor function*, $S(t)$, is given by

$$S(t) = \Pr(T > t) = \int_t^\infty f(u)\,du = 1 - F(t),$$  \hspace{1cm} (1.4)

and represents the probability of surviving past time $t$. It is a monotone decreasing function with $S(0) = 1$ and $S(\infty) = \lim_{t \to \infty} S(t) = 0$. In certain situations it is useful to allow $S(\infty) > 0$ which leads to an improper, or *defective*, distribution. In survival literature such models are often called *cure rate* models.

It is common in survival analysis to report quantities such as “the 5-year survival probability” which is the probability of living longer than 5 years, i.e., $S(5)$ if $T$ represents years. Survivor curves for different groups of individuals are often plotted to compare, for example, a new treatment with an existing one. The survivor function is also essential in likelihood construction to handle right-censoring. This is discussed briefly in Section 1.2.3 and more formally in Chapter 2.

The *hazard function*, $\lambda(t)$, also known as the hazard rate, failure rate or intensity, represents the instantaneous risk (hazard) of an event occurring at time $t$. It is defined as

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t+\Delta t \mid T \geq t)}{\Delta t},$$  \hspace{1cm} (1.5)

and can be any non-negative function. The hazard function specifies entirely the distribution of $T$, as do the functions $f(t)$, $F(t)$ and $S(t)$, but it is the hazard which is regarded as the key quantity in survival analysis and is therefore seen as the starting point.

It is instructive to write

$$\Pr(t \leq T < t+\Delta t \mid T \geq t) \approx \lambda(t)\Delta t,$$  \hspace{1cm} (1.6)

and compare this to (1.1) which represents the unconditional probability of failure in the interval $[t, t+\Delta t)$. Equation (1.6) however represents failure in the interval $[t, t+\Delta t)$ given survival up to that point. Thus, conditioning on $T \geq t$ gives rise to a quantity which focusses on that specific moment in time - hence the importance of the hazard function.
Understanding the shape of the hazard will therefore give insights into the process under study which may be, for example, biological, mechanical or sociological. Some common shapes of the hazard are: constant, monotone increasing/decreasing or hazards which first increase and then decrease (Lawless, 2003, pp. 14–16). Different shapes will fit different situations, for example, an increasing hazard describes a wearout process where longer duration leads to an increased number of failures. The importance of the shape of the hazard is discussed in Aalen & Gjessing (2001). Indeed, the effect of covariates on the shape of the hazard is one of the main focuses of this thesis and is discussed in Chapter 5.

Another important function is the cumulative hazard function, also called integrated hazard function, which is defined as

$$\Lambda(t) = \int_0^t \lambda(u) du,$$

and represents the cumulative risk of an event occurring from time zero to time \( t \). It also has the interpretation of the expected number of events at time \( t \). Defining \( N(t) \) as a process counting the observed number of events at time \( t \), it turns out that

$$M(t) = N(t) - \Lambda(t)$$

defines a type of residual, i.e. observed minus expected. More importantly, it is a martingale. Using this fact, asymptotics of various survival estimators can be derived via martingale theory. The counting process approach to survival is due to Aalen (1975) and provides a unified framework to study more complex survival models (e.g., competing risks and multistate models) and censoring patterns. More information can be found in Andersen et al. (1993) and Fleming & Harrington (2005).

### 1.2.3 Relationships Between Functions

Some important relationships exist between the above survival functions which will now be proved. Applying the law of conditional probability to
(1.5) gives

\[
\lambda(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t \cap T \geq t)}{\Delta t} \frac{1}{\Pr(T \geq t)}
\]

\[
= \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t)}{\Delta t} \frac{1}{\Pr(T \geq t)}
\]

\[
= \lim_{\Delta t \to 0} \frac{\Pr(T \leq t + \Delta t) - \Pr(T \leq t)}{\Delta t} \frac{1}{\Pr(T \geq t)}
\]

\[
= \frac{F'(t)}{S(t)}
\]

\[
= \frac{f(t)}{S(t)}.
\]  

(1.9)

Equivalently

\[
f(t) = \lambda(t)S(t).
\]  

(1.10)

Splitting up the density function in this way is useful in order to incorporate right-censoring into the likelihood function. For a right-censored individual, we only know that the event occurs after time \(t_i\), where \(t_i\) is the observed time for that particular individual. This individual therefore contributes \(S(t_i)\) to the likelihood function. For an individual who experiences an event at time \(t_i\), we know that they survived up to time \(t_i\) but then had an event at that time. This individual contributes \(S(t_i)\lambda(t_i) = f(t_i)\) to the likelihood. Putting this together suggests the form \(\lambda(t_i)\delta_i S(t_i)\) where \(\delta_i \in \{0, 1\}\) is an indicator variable taking the value one for event times and zero for right-censored times. Further discussion on the right-censored likelihood function is deferred until Chapter 2.

It turns out that a relationship between the two key survival functions, the hazard and survivor functions, can be established. From (1.9) and the fact that \(F'(t) = -S'(t)\) it follows that

\[
\lambda(t) = -\frac{S'(t)}{S(t)}
\]

\[
= -\frac{d}{dt} \log S(t),
\]  

(1.11)
a differential equation with solution
\[ S(t) = \exp \left[ - \int_0^t \lambda(u)du \right] = \exp[-\Lambda(t)]. \tag{1.12} \]

We may equivalently write
\[ \Lambda(t) = -\log S(t). \tag{1.13} \]

It is clear from the above that we require \( \lim_{t \to \infty} \Lambda(t) = \infty \), i.e., the integrated hazard must diverge, to define a proper distribution. However, as previously mentioned, defective distributions have a place in survival analysis and so we may allow the integrated hazard to converge to a (positive) constant.

Combining (1.10) and (1.12) leads to
\[ f(t) = \lambda(t) \exp \left[ - \int_0^t \lambda(u)du \right]. \tag{1.14} \]

That the density can be expressed solely in terms of the hazard function further highlights the importance of this function.

The relationship given in (1.12) can be motivated by different means and it is instructive to do so; we follow a similar line of argument to Lancaster (1990, p. 11). First divide the interval \([0, t]\) into \(m\) sub-intervals \([t_0, t_1), [t_1, t_2), \ldots, [t_{m-1}, t_m)\) where \(t_0 = 0\) and \(t_m = t\). We will assume that all intervals are of length \(\Delta t\), although the proceeding results hold without this requirement.

\[
\Pr(T \geq t) = \Pr(T \geq t_0 \cap T \geq t_1 \cap \ldots \cap T \geq t_m) \\
= \Pr(T \geq t_0) \times \Pr(T \geq t_1 | T \geq t_0) \times \cdots \\
\times \Pr(T \geq t_m | T \geq t_{m-1} \cap \ldots \cap T \geq t_0) \\
= 1 \times \Pr(T \geq t_1 | T \geq t_0) \times \cdots \times \Pr(T \geq t_m | T \geq t_{m-1}) \\
= \prod_{i=1}^{m} \Pr(T \geq t_i | T \geq t_{i-1}). \tag{1.15}
\]
We can write

\[ \Pr(T \geq t_i \mid T \geq t_{i-1}) = 1 - \Pr(T < t_i \mid T \geq t_{i-1}) = 1 - \Pr(t_{i-1} \leq T < t_i \mid T \geq t_{i-1}) = 1 - \Pr(t_{i-1} \leq T < t_{i-1} + \Delta t \mid T \geq t_{i-1}) \approx 1 - \lambda(t_{i-1}) \Delta t, \]  

(1.16)

where the last line follows from (1.6) and our assumption that the intervals are all of length \( \Delta t \). Inserting (1.16) into (1.15) gives

\[ \Pr(T \geq t) \approx \prod_{i=1}^{m} [1 - \lambda(t_{i-1}) \Delta t], \]

which has the following intuitive interpretation: survival to time \( t \) is equivalent to not having experienced an event in the sequence of sub-intervals \([t_{i-1}, t_i), i = 1, \ldots, m\). Taking the limit as \( m \to \infty \) leads to

\[ \Pr(T \geq t) = \lim_{m \to \infty} \prod_{i=1}^{m} [1 - \lambda(t_{i-1}) \Delta t] = \mathcal{P}_0^t [1 - \lambda(u)du] = \mathcal{P}_0^t [1 - d\Lambda(u)] \]  

(1.17)

where the limit \( m \to \infty \) ensures that \( \Delta t \to 0 \) and the second line follows from the definition of the product integral\(^1\). Here we use \( \mathcal{P} \) as the symbol for product integration. From the properties of product integration, and the fact that \( T \) is continuous, (1.17) becomes \( \exp[-\Lambda(t)] \) (Gill & Johansen, 1990).

Thus, we arrive at (1.12) via probability arguments which highlight the deep connection between the hazard and survivor functions. In fact (1.17) is more general than (1.12) and includes the cases where \( T \) is discrete or a mixture of both discrete and continuous components. In these cases (1.12) does not hold but (1.17) does. The product integral representation is an important one which, among other things, underpins counting processes and likelihood construction (Kalbfleisch & Prentice, 2002, sec. 1.2.3).

\(^1\)Note that product integrals are to products, as standard (Riemann) integrals are to sums, i.e. \( \mathcal{P} \) is to \( \prod \) as \( \int \) is to \( \sum \).
1.3 Non-Parametric Methods

Although the focus of this thesis is mainly parametric approaches, we will first discuss non-parametric estimators as they can provide us with useful preliminary summaries of data and sometimes help in suggesting a parametric model. Standard non-parametric procedures, e.g., the empirical distribution function or histogram, cannot simply be applied to survival data due to censoring and, in addition to this, we would like some impression of the hazard function (given its importance in survival analysis) which is not a quantity estimated in other areas of statistics.

The two most important non-parametric survival estimators are the Kaplan-Meier estimate of the survivor function, denoted $\hat{S}_{KM}(t)$, and the Nelson-Aalen estimate of the cumulative hazard function, denoted $\hat{\Lambda}_{NA}(t)$. Both estimators can be derived via maximum likelihood arguments, without making any distributional assumptions, thereby giving the closest estimates of the true functions, $S(t)$ and $\Lambda(t)$, on the space of all possible functions. For this reason they are often treated as the truth and used for graphical comparison with a corresponding parametric estimator in order to establish the fit of the parametric model. It is also possible to estimate $f(t)$ and $\lambda(t)$, based on $\hat{S}_{KM}(t)$ and $\hat{\Lambda}_{NA}(t)$ respectively, but we will not consider this as it is more difficult, owing to kernel and bandwidth selection, and can be subject to high variability.

In what follows, let’s assume we have a sample of $n$ individuals with observed times denoted $t_i, i = 1, \ldots, n$. Some of these will be event times and some of these will be censoring times. Let $t_{(1)} < t_{(2)} < \cdots < t_{(r)}$ represent the $r$ distinct ordered event times from our sample which, in the presence of tied event times, will correspond to more than $r$ individuals’ times, i.e., $r \leq n$. Also define $t_{(0)} = 0$ and $t_{(r+1)} = \infty$, respectively, and construct $r + 1$ intervals $I_j = [t_{(j)}, t_{(j+1)})$, $j = 0, \ldots, r$. Each of these intervals will contain $d_j$ events and $m_j$ censored observations. Note that $d_0 = 0$. Any censoring times in $I_j$ recorded as being equal to the event time $t_{(j)}$ are conventionally assumed to be infinitesimally larger than $t_{(j)}$ so that the interval $I_j$ begins with the event time. We also define $n_j = (d_j + m_j) + (d_{j+1} + m_{j+1}) + \cdots + (d_r + m_r)$ to be the number of individuals at risk just prior to $t_{(j)}$. Thus, $n_0 = n$. 

1.3. NON-PARAMETRIC METHODS

1.3.1 Empirical Survivor Function

First we introduce the *empirical survivor function* which is given by

\[ \hat{S}_n(t) = \frac{\text{Number of event times} > t}{n}, \]  
(1.18)

and is the most simple estimate of the survivor function but cannot be used when there are censored times. The empirical survivor function equals one for \( t < t^{(1)} \), zero for \( t \geq t^{(r)} \) and is assumed to be constant between adjacent event times. Therefore \( \hat{S}_n(t) \) is a step-function which decreases by \( d_j/n \) just after each observed event time. We mention this estimator due to its simplicity and relationship with the familiar empirical distribution function, \( \hat{F}_n(t) \), i.e., \( \hat{S}_n(t) = 1 - \hat{F}_n(t) \), but it is of little practical use for survival data due to censoring.

1.3.2 Kaplan-Meier Estimate of Survivor Curve

The *Kaplan-Meier* estimator, also known as the product limit estimator, generalises the empirical survivor function to allow for right-censored data. It was first proposed by Böhmer (1912) but was not pursued further until Kaplan & Meier (1958) reintroduced it, providing formal treatment including its derivation as a non-parametric maximum likelihood estimator. Assuming that censoring occurs independently of event times\(^2\), it can be shown that the relevant log-likelihood function is

\[ \ell(\theta) = \sum_{j=0}^{r} d_j \log \lambda_j + (n_j - d_j) \log(1 - \lambda_j) \]  
(1.19)

where \( \theta = (\lambda_0, \lambda_1, \ldots, \lambda_r)^T \) and \( \lambda_j = \Pr(T = t^{(j)} | T \geq t^{(j)}) \) which is a discrete version of (1.5). Maximising (1.19), by solving \( d\ell/d\lambda_j = 0 \) for \( j = 0, \ldots, r \), leads to \( \hat{\lambda}_j = d_j/n_j \) which is the empirical hazard. The Kaplan-

\(^2\)See Section 2.1.2 for information on likelihood construction for right-censored data.
Meier estimator is then given by

\[
\hat{S}_{KM}(t) = \prod_{j \mid t \geq t(j)} (1 - \hat{\lambda}_j)
\]

\[
= \prod_{j \mid t \geq t(j)} \frac{n_j - d_j}{n_j},
\]

(1.20)

and is a step-function, equal to one for \( t < t_{(1)} \), dropping by the factor \( (n_j - d_j)/n_j \) just after each event time. When the largest observed time corresponds to an event, i.e., \( t_{\text{max}} = t_{(r)} \), then \( \hat{S}_{KM}(t) = 0 \) for \( t > t_{(r)} \), whereas, if it is a censored observation then \( \hat{S}_{KM}(t) \) does not drop to zero and is undefined for \( t > t_{\text{max}} \). Note also that the Kaplan-Meier estimator reduces to the empirical survivor function when there are no censored observations.

We now discuss briefly the calculation of the standard error of \( \hat{S}_{KM}(t) \) at a given time, \( t \). Using standard likelihood theory (reviewed in Chapter 2) we find that

\[
\text{var}(\hat{\lambda}_j) = \frac{d_j(n_j - d_j)}{n_j^3}.
\]

Writing

\[
\log \hat{S}_{KM}(t) = \sum_{j \mid t \geq t(j)} \log(1 - \hat{\lambda}_j),
\]

an application of the delta method (Section 2.3.2) gives

\[
\text{var}\left[ \log \hat{S}_{KM}(t) \right] = \sum_{j \mid t \geq t(j)} \frac{1}{(1 - \hat{\lambda}_j)^2} \text{var}(\hat{\lambda}_j)
\]

\[
= \sum_{j \mid t \geq t(j)} \frac{d_j}{n_j(n_j - d_j)}.
\]

and a further application of the delta method gives

\[
\hat{\sigma}_{KM}^2(t) = \text{var}\left[ \hat{S}_{KM}(t) \right] = \left[ \hat{S}_{KM}(t) \right]^2 \text{var}\left[ \log \hat{S}_{KM}(t) \right].
\]

Thus

\[
\hat{\sigma}_{KM}(t) = \hat{S}_{KM}(t) \left[ \sum_{j \mid t \geq t(j)} \frac{d_j}{n_j(n_j - d_j)} \right]^{1/2},
\]

(1.21)
which is known as Greenwood’s formula and is due to Greenwood (1926) who
derived it as the standard error of the actuarial life-table estimator. One may
therefore construct a 95% confidence interval for $\hat{S}_{KM}(t)$ via

$$\hat{S}_{KM}(t) \pm 1.96 \hat{\sigma}_{KM}(t). \tag{1.22}$$

However, due to the fact that $\hat{S}_{KM}(t) \in [0, 1]$, it is well known that (1.22)
can produce unsatisfactory results, i.e., values that are larger than one or
negative. The log\([-\log(\cdot)]\) transformation maps $\hat{S}_{KM}(t)$ onto \((-\infty, \infty)\). It is
easily found, by applying the delta method again, that

$$\text{var}\left\{\log \left[ -\log \hat{S}_{KM}(t) \right] \right\} = \frac{1}{\left[ \log \hat{S}_{KM}(t) \right]^2} \text{var} \left[ \log \hat{S}_{KM}(t) \right]$$

$$= \frac{\hat{\sigma}_{KM}^2(t)}{\left[ \hat{S}_{KM}(t) \log \hat{S}_{KM}(t) \right]^2},$$

and therefore, through back-transformation,

$$\left[ \hat{S}_{KM}(t) \right]^{\exp\left\{ \pm 1.96 \hat{\sigma}_{KM}(t) / [\hat{S}_{KM}(t) \log \hat{S}_{KM}(t)] \right\}} \tag{1.23}$$
gives a 95% confidence interval with better properties than (1.22).

Note that in all of the above, the likelihood function (1.19) has been
treated as though $\theta$ is a vector of parameters. This is not true as $\theta$
is data-dependent and $\text{dim}(\theta)$ increases as $n$ increases leading to an infinite-
dimensional parameter space as $n \to \infty$. More formal results can be found

**Example 1.1. Lung Cancer in Northern Ireland**

We now calculate the Kaplan-Meier survivor curve for a subset of 19 indi-
viduals from the lung cancer dataset described in Appendix A; the sub-
set in question is the group of males under the age of 50. The estimated sur-
vival probabilities along with standard errors (calculated using (1.21)) and
confidence intervals (obtained using (1.22) and (1.23), respectively) are sum-
marised in Table 1.1. The corresponding graph can be seen in Fig. 1.2.
### Table 1.1. Kaplan-Meier Estimates for Lung Cancer Data

<table>
<thead>
<tr>
<th>$t_{(j)}$</th>
<th>$d_j$</th>
<th>$n_j$</th>
<th>$\hat{S}_{KM}(t)$ No transform.</th>
<th>$\log(-\log(\cdot))$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.132</td>
<td>1</td>
<td>19</td>
<td>0.947 (0.051) [0.847, 1.048]</td>
<td>[0.681, 0.992]</td>
</tr>
<tr>
<td>0.230</td>
<td>1</td>
<td>18</td>
<td>0.895 (0.070) [0.757, 1.033]</td>
<td>[0.641, 0.973]</td>
</tr>
<tr>
<td>0.460</td>
<td>1</td>
<td>17</td>
<td>0.842 (0.084) [0.678, 1.006]</td>
<td>[0.587, 0.946]</td>
</tr>
<tr>
<td>0.493</td>
<td>1</td>
<td>16</td>
<td>0.789 (0.094) [0.606, 0.973]</td>
<td>[0.532, 0.915]</td>
</tr>
<tr>
<td>0.986</td>
<td>1</td>
<td>15</td>
<td>0.737 (0.101) [0.539, 0.935]</td>
<td>[0.479, 0.881]</td>
</tr>
<tr>
<td>1.710</td>
<td>1</td>
<td>14</td>
<td>0.684 (0.107) [0.475, 0.893]</td>
<td>[0.428, 0.844]</td>
</tr>
<tr>
<td>1.841</td>
<td>1</td>
<td>13</td>
<td>0.632 (0.111) [0.415, 0.848]</td>
<td>[0.379, 0.804]</td>
</tr>
<tr>
<td>3.945</td>
<td>1</td>
<td>12</td>
<td>0.579 (0.113) [0.357, 0.801]</td>
<td>[0.332, 0.763]</td>
</tr>
<tr>
<td>4.603</td>
<td>1</td>
<td>11</td>
<td>0.526 (0.115) [0.302, 0.751]</td>
<td>[0.287, 0.719]</td>
</tr>
<tr>
<td>6.016</td>
<td>1</td>
<td>10</td>
<td>0.474 (0.115) [0.249, 0.698]</td>
<td>[0.244, 0.673]</td>
</tr>
<tr>
<td>8.942</td>
<td>1</td>
<td>9</td>
<td>0.421 (0.113) [0.199, 0.643]</td>
<td>[0.204, 0.625]</td>
</tr>
<tr>
<td>9.008</td>
<td>1</td>
<td>8</td>
<td>0.368 (0.111) [0.152, 0.585]</td>
<td>[0.165, 0.575]</td>
</tr>
<tr>
<td>12.756</td>
<td>1</td>
<td>4</td>
<td>0.276 (0.115) [0.051, 0.502]</td>
<td>[0.088, 0.506]</td>
</tr>
</tbody>
</table>

**Note:** $\hat{S}_{KM}(t)$ is defined up to $t = 19.627$ as this is the largest observed time and it is a censored observation.

---

**Figure 1.2.** Kaplan-Meier curve with C.I.s based on the untransformed estimator (dash) and the $\log(-\log(\cdot))$ transformation (dot) respectively.
One may wish to estimate the $p$th percentile using the formula

$$\hat{t}_{p/100} = \left[ t_{(j)} \mid \hat{S}_{KM}(t_{(j)}) = 1 - p/100 \right].$$

However, due to the fact that the Kaplan-Meier estimator is a step-function, it is usually the case that there is no $t_{(j)}$ for which Kaplan-Meier is exactly equal to $1 - p/100$. For example, in Table 1.1, since $\hat{S}_{KM}(4.603) = 0.526$ and $\hat{S}_{KM}(6.016) = 0.474$, we may choose to estimate the median as $\hat{t}_{0.5} = (4.603 + 6.016)/2 = 5.31$ months. One may prefer an interval estimate rather than a point estimate, and this has been considered by many authors (Efron, 1981; Reid, 1981; Brookmeyer & Crowley, 1982; Emerson, 1982), but we will not pursue such approaches here.

### 1.3.3 Nelson-Aalen Estimate of Cumulative Hazard Function

The *Nelson-Aalen* estimator of the cumulative hazard function was first introduced by Nelson (1969, 1972) and was independently derived by Altshuler (1970). The estimator was later studied by Aalen (1975, 1978) in detail using the counting process formulation (briefly mentioned in Section 1.2.2) who, among other things, extended its use to Markov chains and more general event history models.

The Nelson-Aalen estimator is given by

$$\hat{\Lambda}_{NA}(t) = \sum_{j \mid t \geq t_{(j)}} \hat{\lambda}_j = \sum_{j \mid t \geq t_{(j)}} \frac{d_j}{n_j}.$$

It is noteworthy that this estimator is formed using the empirical hazard estimates which maximise the non-parametric log-likelihood (1.19). Thus, the estimators (1.20) and (1.24) come from the same estimation procedure which highlights the fact that they are the non-parametric maximum likelihood estimators of the survivor and cumulative hazard functions respectively. In fact, the two estimators are functionally related and we will show this in Section 1.3.4.
Using a similar approach to that shown in the previous section, one may derive the standard error of (1.24) as

$$\hat{\sigma}_{\text{NA}}(t) = \left[ \sum_{j \mid t \geq t(j)} \frac{d_j(n_j - d_j)}{n_j^2} \right]^{1/2},$$

and the corresponding 95% confidence interval is

$$\hat{\Lambda}_{\text{NA}}(t) \pm 1.96 \hat{\sigma}_{\text{NA}}(t).$$

Taking a log transformation maps $\hat{\Lambda}_{\text{NA}}(t)$ onto $(-\infty, \infty)$. Thus,

$$\text{var} \left[ \log \hat{\Lambda}_{\text{NA}}(t) \right] = \frac{\hat{\sigma}^2_{\text{NA}}(t)}{\left[ \hat{\Lambda}_{\text{NA}}(t) \right]^2},$$

and therefore, through back-transformation,

$$\left[ \hat{\Lambda}_{\text{NA}}(t) \right] \exp \left[ \pm 1.96 \hat{\sigma}_{\text{NA}}(t)/\hat{\Lambda}_{\text{NA}}(t) \right]$$

gives a 95% confidence interval which has better properties than (1.26).

### 1.3.4 Relationship Between Estimators

The fundamental relationship given in (1.17) shows how the survivor function is closely related to the corresponding cumulative hazard function for a general distribution, i.e., continuous, discrete or mixed. It is easy to verify that $\hat{S}_{\text{KM}}(t)$ and $\hat{\Lambda}_{\text{NA}}(t)$ are related to each other via (1.17):

$$\hat{S}_{\text{KM}}(t) = \prod_{j \mid t \geq t(j)} (1 - \hat{\lambda}_j)
= \mathcal{P}_0 \left[ 1 - d\hat{\Lambda}_{\text{NA}}(u) \right],$$

which follows from the discrete nature of $\hat{\Lambda}_{\text{NA}}(u)$, i.e., it is a step function (see Gill & Johansen (1990)). This tells us that by maximising the log-likelihood in (1.19), we obtain a (discrete) non-parametric estimator for the distribution
of $T$ with survivor function and cumulative hazard function given by $\hat{S}_{\text{KM}}(t)$ and $\hat{\Lambda}_{\text{NA}}(t)$ respectively.

Textbooks based on the counting process approach typically regard estimation of the cumulative hazard function as being most important and derive the Nelson-Aalen estimator first, albeit, using martingale theory (Aalen et al., 2008; Andersen et al., 1993). The Kaplan-Meier estimator is then motivated via (1.28) and, consequently, its asymptotic properties are derived (Andersen et al., 1993, sec. IV.3). It is worth noting that the asymptotic variances obtained using martingales differ slightly from those given in (1.21) and (1.25). Discussion of these different variance estimates can be found in Klein (1991).

In addition to the Kaplan-Meier and Nelson-Aalen estimators, one may motivate the following alternative estimators

$$\tilde{\Lambda}(t) = -\log \hat{S}_{\text{KM}}(t), \quad (1.29)$$
$$\tilde{S}(t) = \exp \left[-\hat{\Lambda}_{\text{NA}}(t)\right], \quad (1.30)$$

using the relationships given in (1.13) and (1.12) respectively. For example, (1.30) is sometimes called the Breslow estimator (Breslow, 1972); note that Fleming & Harrington (1984) compare this to the Kaplan-Meier estimator. Clearly (1.29) and (1.30) are not non-parametric maximum likelihood estimators as $\tilde{\Lambda}(t) \neq \hat{\Lambda}_{\text{NA}}(t)$ and $\tilde{S}(t) \neq \hat{S}_{\text{KM}}(t)$. However, these will be approximately equal when the $\hat{\lambda}_j$'s are small:

$$\tilde{\Lambda}(t) = -\log \hat{S}_{\text{KM}}(t)$$
$$= -\log \prod_{j \mid t \geq t_{(j)}} (1 - \hat{\lambda}_j)$$
$$= -\sum_{j \mid t \geq t_{(j)}} \log(1 - \hat{\lambda}_j)$$
$$\approx \sum_{j \mid t \geq t_{(j)}} \hat{\lambda}_j$$
$$= \hat{\Lambda}_{\text{NA}}(t),$$

where we have used the fact that $\log(1-x) \approx -x$ when $x$ is small. Thus, when $d_j$ is small relative to $n_j$ ($j = 0, \ldots, r$), $\hat{\Lambda}_{\text{KM}}(t)$ is close to the Nelson-Aalen
estimator. One may similarly show that $\hat{S}_{NA}(t) \approx \hat{S}_{KM}(t)$ when the $\hat{\lambda}_j$'s are small.

1.4 Parametric Survival Distributions

The advantage of the non-parametric methods mentioned above is that no assumptions about the data are made and thus we “let the data speak for themselves”. However, parametric models have many advantages. Often a parametric model can adequately fit the data and in this case it is much preferable to describe, for example, the survivor function with a compact formula, such as $S(t) = \exp(-t^{1.8})$, than a table such as Table 1.1 (which grows with the sample size). Having simple mathematical formulae makes estimation of quantities, such as the mean or median survival time, and their corresponding standard errors quite straightforward. In addition to this, estimates can have higher efficiency (i.e., lower variance) than the corresponding non-parametric estimates (see Miller (1983) and Cox & Oakes (1984, sec. 8.5)). We may also simulate from the model to estimate more complex quantities or extrapolate beyond the data.\(^3\) Parametric survival functions typically vary smoothly with $t$, unlike the non-parametric estimators, and this may be preferable as it is usually assumed that the true functions are also smooth (when $t$ is continuous).

Parametric modelling allows the use standard likelihood theory (see Chapter 2) for inference purposes whereas non-parametric methods require more specialised treatment (as mentioned in Section 1.3). Furthermore, the use of parametric regression models has been advocated by Efron (1988), Royston & Parmar (2002) and Nardi & Schenper (2003) among others and, indeed, the incorporation of covariates is straightforward in a parametric framework owing again to the use standard likelihood theory (see Chapters 4 and 5).

For all of these reasons, parametric analyses can be very fruitful in helping us gain understanding of the phenomenon at hand. Furthermore, as noted in Aalen et al. (2008, p. 207):

\(^3\)Extrapolation far beyond the range of the data is of course ill-advised (see Sections 3.6.2, 3.6.5 and 3.6.6).
It has become a tradition to use non- and semi-parametric methods ... to analyse censored survival data ... (Practitioners) would gain from the use of a wider range of statistical methods, including parametric methods, than is the current practice.

This is a point of view with which we agree. Therefore, our approach to survival will mainly be parametric due to the typical over-emphasis of non-parametric methods in the survival literature.

In the proceeding sections we will describe some useful parametric survival distributions (Johnson et al., 1994, 1995; Lawless, 2003; Marshall & Olkin, 2007). For each distribution we will start with the hazard function and derive the cumulative hazard and survivor function, respectively, using (1.7) and (1.12). We will also derive the inverse survivor function, $S^{-1}(u)$, which is useful for simulation (see Chapter 3) and calculation of percentiles. We mainly consider two-parameter models ($\lambda =$ scale parameter and $\gamma =$ shape parameter) where, in our setup, the scale parameter controls the magnitude of the hazard while the shape parameter controls its time evolution. The categorisation of various types of parameters is treated much more formally in Marshall & Olkin (2007, chap. 7); our setup does not necessarily adhere to these definitions. Note also that we only consider parametric models where $S(t)$ has a closed form, leading a closed form likelihood function. Thus we do not consider, for example, the gamma, log-normal or inverse-Gaussian distributions (though the latter two are similar in shape to the log-logistic considered below). Information on these models can be found in the above references.

1.4.1 Exponential

The exponential distribution is the most simple parametric survival model. The hazard function is given by

$$\lambda(t) = \lambda,$$  \hspace{1cm} (1.31)

where $\lambda > 0$ is a scale parameter. Regardless of the point in time, $t$, the hazard is constant, thereby assuming that events can occur randomly at any time.
CHAPTER 1. REVIEW OF SURVIVAL ANALYSIS

The cumulative hazard is given by

\[ \Lambda(t) = \lambda t \]  \hspace{1cm} (1.32)

and the survivor function is

\[ S(t) = \exp(-\lambda t). \]  \hspace{1cm} (1.33)

The inverse survivor function is given by

\[ t = S^{-1}(u) = \frac{1}{\lambda} \log u, \]  \hspace{1cm} (1.34)

where \( u \in (0, 1) \), and is useful for simulation and calculating percentiles. The \( p \)th percentile is given by setting \( u = 1 - p/100 \). Thus

\[ t_{p/100} = -\frac{1}{\lambda} \log(1 - p/100), \]  \hspace{1cm} (1.35)

for example, \( t_{0.5} = \frac{(\log 2)}{\lambda} \).

The assumption of constant hazard is usually too strict for the model to be of much practical use and more flexible distributions are discussed in the proceeding sections. However, we introduce this model for two reasons. The first is historical. The exponential distribution was one of the first widely used survival models due to its simplicity. For early examples of its use see Davis (1952) and Epstein & Sobel (1953). Bartholomew (1957) showed that the maximum likelihood estimator of \( \lambda \) is given by the closed form expression

\[ \hat{\lambda} = \frac{\text{The total number of events}}{\text{The sum of all observed event times and censoring times}}, \]

and, therefore, in a time when computing was less powerful, one could easily calculate \( \hat{\lambda} \) on a pocket calculator for example.

The second reason we introduce the exponential distribution is that all survival distributions are related to it via

\[ T \sim \text{“any distribution”} \]

\[ \Rightarrow \Lambda(t) \sim \text{Exp}(\lambda = 1). \]
Here $\Lambda(t)$ is the cumulative hazard function corresponding to $T$ and $\text{Exp}(\lambda = 1)$ is the unit exponential distribution, i.e., an exponential distribution with $\lambda = 1$. This is easy to show. First, define the transformation $X = \Lambda(t)$, where the random variable $T$ can have any distribution. Thus, we have that

$$\Pr(T > t) = \Pr[\Lambda(T) > \Lambda(t)] = \Pr(X > x).$$

Furthermore, from (1.12), we also know that

$$\Pr(T > t) = \exp[-\Lambda(t)] = \exp(-x).$$

Hence, combining the above expressions gives

$$\Pr(X > x) = \exp(-x),$$

which is the survivor function for a unit exponential distribution. Therefore, the cumulative hazard function can be viewed as a function which transforms $T$ into a unit exponential variable. This property is discussed in Singpurwalla (2006) who, among other things, uses it to generate multivariate survival distributions.

### 1.4.2 Weibull

The *Weibull distribution* is a very popular distribution - not only in survival analysis - and this is clear from the wide variety of applications mentioned in Johnson et al. (1994, pp. 684–685). Its popularity is both due to its flexibility, as originally demonstrated by Weibull (1951), and its theoretical derivation as one of the three possible extremal distributions (Fisher & Tippett, 1928). The hazard function is given by

$$\lambda(t) = \lambda \gamma t^{\gamma-1},$$

(1.36)

where $\lambda > 0$ and $\gamma > 0$ are the scale and shape parameters respectively. The hazard function can increase ($\gamma > 1$), decrease ($\gamma < 1$) or remain constant ($\gamma = 1$) over time (see Fig. 1.3). The Weibull distribution therefore generalises the exponential distribution.
The cumulative hazard is given by

\[ \Lambda(t) = \lambda t^\gamma, \]  

(1.37)

and the survivor function is

\[ S(t) = \exp(-\lambda t^\gamma). \]  

(1.38)

The inverse survivor function is given by

\[ t = S^{-1}(u) = \left(-\frac{1}{\lambda} \log u\right)^{1/\gamma}, \]  

(1.39)

where \( u \in (0, 1) \), and the \( p \)th percentile is

\[ t_{p/100} = \left[-\frac{1}{\lambda} \log(1 - p/100)\right]^{1/\gamma}. \]  

(1.40)

**Example 1.2. Lung Cancer Data in Northern Ireland**

Upon fitting the Weibull model to the data described in Example 1.1 we
obtain the maximum likelihood estimates $\hat{\lambda} = 0.221$ and $\hat{\gamma} = 0.625$ respectively. Figure 1.4 shows that the estimated Weibull survivor function, $\hat{S}(t) = \exp(-0.221t^{0.625})$, closely agrees with the Kaplan-Meier estimate. The estimate of the median based on the Weibull model is $\hat{t}_{0.5} = 6.22$ months.

Figure 1.4. Kaplan-Meier and Weibull survivor curves overlayed.

1.4.3 Gompertz

The Gompertz distribution was derived by Gompertz (1825), a practicing actuary, as a model for human mortality and has received much attention in actuarial literature (Olshansky & Carnes, 1997). The model is based on the assumption that the change in the hazard, at time $t$, is proportional to its current value. This can be represented via the following differential equation

$$\frac{d\lambda(t)}{dt} = A\lambda(t),$$

which has solution

$$\lambda(t) = \exp(At + B).$$
In our notation, the hazard is given by

$$\lambda(t) = \lambda \exp(\gamma t),$$  \hspace{1cm} (1.41)

where $\lambda > 0$ and $\gamma \in (-\infty, \infty)$ are the scale and shape parameters respectively. The hazard function can increase ($\gamma > 0$), decrease ($\gamma < 0$) or remain constant ($\gamma = 0$) over time (see Fig. 1.5). The Gompertz hazard is similar to the Weibull but, due to the exponentiation of time, has a faster rate of growth/decay than the Weibull which grows/decays according to a power of time.

![Figure 1.5. Gompertz hazard function with median time fixed at $t_{0.5} = 5$ and $\gamma = -0.5, 0.0, 0.5$ and 1.5 ⇒ $\lambda \approx 0.38, 0.14, 0.03$ and 0.0006.](image)

The cumulative hazard is given by

$$\Lambda(t) = \frac{\lambda}{\gamma} [\exp(\gamma t) - 1],$$  \hspace{1cm} (1.42)

and the survivor function is

$$S(t) = \exp \left\{ -\frac{\lambda}{\gamma} [\exp(\gamma t) - 1] \right\}. \hspace{1cm} (1.43)$$
Clearly, when \( \gamma < 0 \), \( S(\infty) = \exp(\lambda/\gamma) \Rightarrow \) proportion of non-susceptible individuals or cured fraction. It is worth noting how this differs from the Weibull model. Decreasing Gompertz hazards imply that not all individuals experience the event whereas decreasing Weibull hazards still imply that the event is inevitable.

The Gompertz inverse survivor function is given by

\[
t = S^{-1}(u) = \frac{1}{\gamma} \log \left( -\frac{\gamma}{\lambda} \log u + 1 \right),
\]

where \( u \in (0,1) \) if \( \gamma \geq 0 \) and \( u \in (0, \exp(\lambda/\gamma)] \) if \( \gamma < 0 \). The \( p \)th percentile is given by

\[
t_{p/100} = \frac{1}{\gamma} \log \left( -\frac{\gamma}{\lambda} \log(1 - p/100) + 1 \right),
\]

and, if \( \gamma < 0 \), then \( t_{p/100} = \infty \) for \( p > 100[1 - \exp(\lambda/\gamma)] \).

### 1.4.4 Log-Logistic

The log-logistic distribution is defined via \( T = \exp(Y) \) where \( Y \sim \) logistic distribution (see Johnson et al. (1995, chap. 23) for details on the logistic distribution), analogous to the relationship between the log-normal and the normal distribution. In fact, the log-logistic distribution is very similar to the log-normal in shape but has the advantage of closed form survivor and hazard functions. This makes it more useful in survival analysis. Bennett (1983) provides an early application of the log-logistic distribution to lung cancer data. In econometrics literature it is often called the “Fisk” distribution after Fisk (1961).

The Weibull and Gompertz hazards are monotonic increasing/decreasing which may not be appropriate in certain situations. Hazards are commonly found to rise initially to a peak followed thereafter by a decline, for example, survival after heart transplantation - the patient faces an increased risk of death in the initial days after the transplant while the body is adapting. The hazard function for the log-logistic distribution allows for this possibility and is given by

\[
\lambda(t) = \frac{\lambda \gamma t^{\gamma-1}}{1 + \lambda t^\gamma},
\]

(1.46)
where $\lambda > 0$ and $\gamma > 0$ are the scale and shape parameters respectively. For $\gamma \leq 1$, the hazard decreases over time while for $\gamma > 1$ the hazard increases to a single mode and then decreases; the mode occurs at $t_{\text{mode}} = \left(\frac{\gamma - 1}{\lambda}\right)^{1/\gamma}$, i.e., the highest-risk point. A variety of log-logistic hazard shapes can be seen in Fig. 1.6.

The cumulative hazard is given by

$$\Lambda(t) = \log(1 + \lambda t^\gamma),$$

and the survivor function is

$$S(t) = (1 + \lambda t^\gamma)^{-1}.$$  \hspace{1cm} (1.47)  \hspace{1cm} (1.48)

The inverse survivor function is given by

$$t = S^{-1}(u) = \left(\frac{u^{-1} - 1}{\lambda}\right)^{1/\gamma},$$

where $u \in (0, 1)$, and the $p$th percentile is

$$t_{p/100} = \left[\frac{(1 - p/100)^{-1} - 1}{\lambda}\right]^{1/\gamma}.$$  \hspace{1cm} (1.49)  \hspace{1cm} (1.50)
1.4.5 Burr Type XII

The *Burr Type XII distribution* contains both the Weibull and log-logistic distributions as special cases and can be derived via frailty arguments (see Section 7.5). Burr (1942) introduced twelve probability distributions and paid particular attention to the twelfth (hence “type XII”) which has become the most prominent in the literature (Rodriguez, 1977; Tadikamalla, 1980).

The hazard function is

\[
\lambda(t) = \frac{\lambda \gamma t^{\gamma - 1}}{1 + \lambda \rho t^\gamma},
\]

where \( \lambda > 0 \) is a scale parameter and \( \gamma > 0 \) and \( \rho > 0 \) are the shape parameters. One easily sees that the hazard reduces to the log-logistic hazard when \( \rho = 1 \) and approaches the Weibull hazard as \( \rho \to 0 \). The Burr hazard decreases for \( \gamma \leq 1 \) and increases to a mode followed by a decline for \( \gamma > 1 \). The mode occurs at \( t_{\text{mode}} = \left( \frac{\gamma - 1}{\lambda \rho} \right)^{1/\gamma} \). This behaviour is very similar to the log-logistic hazard, i.e., the hazard cannot theoretically increase without peaking and declining. However, since \( t_{\text{mode}} \to \infty \) as \( \rho \to 0 \), the hazard can be treated as increasing-only over a relevant time-range (when \( \rho \approx 0 \)). Accordingly, the Burr hazard is very flexible (see Fig. 1.7).

The cumulative hazard is given by

\[
\Lambda(t) = \frac{1}{\rho} \log(1 + \lambda \rho t^\gamma),
\]

and the survivor function is

\[
S(t) = (1 + \lambda \rho t^\gamma)^{-1/\rho}.
\]

The inverse survivor function is given by

\[
t = S^{-1}(u) = \left( \frac{u^{-\rho} - 1}{\lambda \rho} \right)^{1/\gamma},
\]

where \( u \in (0, 1) \), and the \( p \)th percentile is

\[
t_{p/100} = \left[ \frac{(1 - p/100)^{-\rho} - 1}{\lambda \rho} \right]^{1/\gamma}.
\]
1.4.6 Time-Dependent Logistic

The time-dependent logistic distribution, TDL, was introduced by MacKenzie (1996). It was motivated by using a logistic function, \( \exp(z)/(1 + \exp(z)) \), as a model for the hazard function. Thus,

\[
\lambda(t) = \frac{\exp(\gamma t + \lambda)}{1 + \exp(\gamma t + \lambda)},
\]

where \( \lambda, \gamma \in (-\infty, \infty) \) are the scale and shape parameters respectively\(^4\). When the parameter \( \lambda \) is modelled as a linear function of covariates (1.56) generalises logistic regression to time-dependence (see Chapter 4 for more information on regression models). The hazard function can increase \( (\gamma > 0) \), decrease \( (\gamma < 0) \) or remain constant \( (\gamma = 0) \) over time (see Fig. 1.8). The TDL distribution therefore generalises the exponential distribution.

\(^4\)That presented in MacKenzie (1996) is a more general three-parameter hazard but we only consider a two-parameter sub-model due to identifiability issues concerned with the third parameter.
The cumulative hazard is given by

\[
\Lambda(t) = \frac{1}{\gamma} \log \left( \frac{1 + \exp(\gamma t + \lambda)}{1 + \exp(\lambda)} \right),
\]

(1.57)

and the survivor function is

\[
S(t) = \left[ \frac{1 + \exp(\gamma t + \lambda)}{1 + \exp(\lambda)} \right]^{-1/\gamma}.
\]

(1.58)

Note that, like the Gompertz model, when \( \gamma < 0 \) the TDL is a cure rate model with \( S(\infty) = [1 + \exp(\lambda)]^{1/\gamma} \). The TDL inverse survivor function is given by

\[
t = S^{-1}(u) = \frac{1}{\gamma} \left( \log \left\{ u^{-\gamma} [1 + \exp(\lambda)] - 1 \right\} - \lambda \right),
\]

(1.59)

where \( u \in (0, 1) \) if \( \gamma \geq 0 \) and \( u \in (0, [1 + \exp(\lambda)]^{1/\gamma}) \) if \( \gamma < 0 \). The \( p \)th percentile is given by

\[
t_{p/100} = \frac{1}{\gamma} \left( \log \left\{ (1 - p/100)^{-\gamma} [1 + \exp(\lambda)] - 1 \right\} - \lambda \right),
\]

(1.60)

and, if \( \gamma < 0 \), then \( t_{p/100} = \infty \) for \( p > 100 \{1 - [1 + \exp(\lambda)]^{1/\gamma}\} \).
1.4.7 Piecewise Exponential

The piecewise exponential model, also known as the piecewise constant model, is a flexible parametric model which lies between the parametric and non-parametric approaches mentioned above. It has been considered by Holford (1976), Breslow (1974) and Friedman (1982). The hazard function is assumed to be constant between specified intervals $I_j = [t_{(j-1)}, t_{(j)})$, $j = 1, \ldots, m$, where $t_{(0)} = 0$ and $t_{(m)} = \infty$. Note that, unlike the non-parametric methods, there will not be an interval for every distinct event time in the dataset. Thus,

$$\lambda(t) = \lambda_j \quad t \in I_j$$

where $\lambda = (\lambda_1, \ldots, \lambda_m)^T \in \mathbb{R}_+^m$ is the vector of positive parameters and $a(t)$ is an $m$-dimensional vector function of $t$ whose $j$th component is given by

$$a_j(t) = \begin{cases} 1 & \text{if } t \in I_j, \\ 0 & \text{otherwise}. \end{cases}$$

Thus $a(t)$ is a vector indicating which interval $t$ lies in, e.g., if $m = 4$ and $t \in I_3$ then $a(t) = (0, 0, 1, 0)^T$. The piecewise exponential hazard is clearly very flexible as it can approximate any arbitrary hazard shape (Fig. 1.9), essentially letting the data speak for themselves whilst keeping the dimension low relative to non-parametric methods.

The cumulative hazard is given by

$$\Lambda(t) = \int_0^t a(u)^T \lambda du$$

$$= \int_0^{t_{(1)}} \lambda_1 du + \cdots + \int_{t_{(j-1)}}^t \lambda_j du + 0 + \cdots + 0 \quad t \in I_j$$

$$= \sum_{k=1}^{j-1} (t_k - t_{(k-1)}) \lambda_k + (t - t_{(j-1)}) \lambda_j \quad t \in I_j$$

$$= d(t)^T \lambda,$$
where \( d(t) \) is an \( m \)-dimensional vector function of \( t \) whose \( j \)th component is such that

\[
d_j(t) = \begin{cases} 
t(j) - t_{(j-1)} & \text{if } t > t_{(j)}, \\
t - t_{(j)} & \text{if } t_{(j-1)} \leq t \leq t_{(j)}, \\
0 & \text{if } t < t_{(j-1)}.
\end{cases}
\] (1.65)

Thus \( d(t) \) represents the time spent in each interval for a particular value of \( t \), e.g., if \( m = 4 \) and \( t \in I_3 \) then \( d(t) = (t_{(1)} - t_{(0)}, t_{(2)} - t_{(1)}, t - t_{(2)}, 0)^T \). The survivor function is therefore given by

\[
S(t) = \exp \left[ -d(t)^T \lambda \right].
\] (1.66)

Using the fact that \( d(t)^T \lambda = d(t_{(j-1)})^T \lambda + (t - t_{(j-1)}) \lambda_j \) for \( t \in I_j \), the inverse survivor function is given by

\[
t = S^{-1}(u) = t_{(j-1)} - \frac{\log u + d(t_{(j-1)})^T \lambda}{\lambda_j},
\] (1.67)

for \( u \in (S(t_{(j)}), S(t_{(j-1)})\] and the \( p \)th percentile is given by

\[
t_{p/100} = t_{(j-1)} - \frac{\log(1 - p/100) + d(t_{(j-1)})^T \lambda}{\lambda_j},
\] (1.68)
for \( 1 - p/100 \in (S(t_{ij}), S(t_{i(j-1)})]. \)

The above functions are discontinuous at the cut points which is a disadvantage of the piecewise exponential model. Alternatively one could model the hazard function using splines (polynomials which join at the cut points). Continuity is then gained at the expense of some added computational effort relative to the piecewise constant model. Hazard modelling using splines has been considered by Kooperberg et al. (1995), Rosenberg (1995) and Royston & Parmar (2002) but we will not consider these approaches.
Chapter 2

Inference

2.1 Likelihood

As in many other areas of statistics, the primary vehicle for parameter estimation and inference in survival analysis is the so-called likelihood function. Fisher (1922, 1925, 1934) originally developed the theory of maximum likelihood, setting forth many of the fundamental concepts of modern statistics\(^1\). As noted by Efron (1998):

Statistics went from an ad hoc collection of ingenious techniques to a coherent discipline (due to the work of Fisher).

Because maximum likelihood methods are now well-known and covered in many texts (such as Silvey (1975), Pawitan (2001) and Cox (2006)), we will only summarise the main results here. These standard likelihood methods must be adapted for survival-type data (to handle right-censoring) which we discuss in Section 2.1.2.

2.1.1 Maximising the Likelihood Function

First we consider the case of full information, i.e., no censoring. In this case the data consist of \(n\) observed times, \(t_1, \ldots, t_n\), all of which correspond to

\(^1\)We will not consider Bayesian methods here. For details see Gelman et al. (2003) and Robert & Casella (2010). A Bayesian treatment of survival analysis can be found in Ibrahim et al. (2005).
event times. Given that these times arise according to a probability distribution with p.d.f. \( f(t \mid \theta) \), where \( \theta \) is a \( k \)-dimensional vector of unknown parameters, the likelihood function is

\[
L(\theta) = \prod_{i=1}^{n} f(t_i \mid \theta), \quad (2.1)
\]

where we can multiply densities by assuming that observations are independent. This function tells us how likely a particular \( \theta \) vector is given the observed data. Thus, we wish to find the parameter vector which is most likely. This vector, denoted \( \hat{\theta} \), maximises the likelihood function and is known as the maximum likelihood estimator (m.l.e.) of \( \theta \). We seldom work with the likelihood function itself. Rather we use the log-likelihood function which is given by

\[
\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(t_i \mid \theta). \quad (2.2)
\]

This is due to the fact that it is easier to maximise a sum than a product and both functions attain their maximum at \( \hat{\theta} \), i.e.,

\[
\hat{\theta} = \arg\max_{\theta} L(\theta) = \arg\max_{\theta} [\log L(\theta)]. \quad (2.3)
\]

We maximise \( \ell(\theta) \), using standard calculus methodology, by setting the first derivatives equal to zero, i.e.,

\[
\frac{\partial \ell(\theta)}{\partial \theta_j} = 0, \quad j = 1, \ldots, k. \quad (2.4)
\]

These equations are referred to as the score equations and are solved simultaneously to give the m.l.e. \( \hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)^T \). This solution is typically not analytic necessitating the use of numerical methods, such as the Newton-Raphson method. Note that a parametrisation of the model where \( \theta \in \mathbb{R}^k \) is usually preferable. This prevents optimisation algorithms from moving into regions outside of a constrained parameter space, which may cause the algorithm to fail. For example, if \( \theta \in \mathbb{R}_+^k \) in a particular parametrisation, then one should work with \( \theta^* = \log \theta \in \mathbb{R}^k \).
It can be shown that the m.l.e. is a consistent estimator of the true parameter vector, i.e.,
\[ \hat{\theta} \to \theta \quad \text{as} \quad n \to \infty, \]
where \( \theta \) denotes the true parameter vector (Pawitan, 2001, sec. 9.3). Furthermore, \( \hat{\theta} \) has an asymptotic multivariate normal distribution centered at \( \theta \) (see Section 2.3.1). Similarly, any function of \( \theta \) is also estimated consistently, i.e.,
\[ g(\hat{\theta}) \to g(\theta) \quad \text{as} \quad n \to \infty, \]
for example, the median survival time \( \hat{t}_{0.5} = S^{-1}(0.5 \mid \hat{\theta}) \to S^{-1}(0.5 \mid \theta) \). Indeed, like \( \theta \), \( g(\hat{\theta}) \) also has an asymptotic multivariate normal distribution (see Section 2.3.2).

Although the notation above suggests a basic (no covariate) model as in Section 1.4, the methodology applies equally for regression models (Chapters 4 and 5). In this case we use the notation \( f(t \mid x, \theta) \) where we now make the assumption that observations are conditionally independent, given covariates, to motivate the likelihood function (2.1). For regression models, the vector \( \theta \) will contain regression coefficients in addition to distributional parameters.

### 2.1.2 Right-Censored Likelihood

#### Random Censoring

In the previous section we considered the case of full information. However, in survival data, we will typically have right-censored times\(^2\). In this case the response variable is not an event time but rather a pair \((t_i, \delta_i)\), \(i = 1, \ldots, n\), where \(t_i\) is the observed survival time and \(\delta_i\) is an indicator variable such that
\[ \delta_i = \begin{cases} 1 & \text{if } t_i \text{ is an event time,} \\ 0 & \text{if } t_i \text{ is a censoring time.} \end{cases} \tag{2.5} \]

We will also have explanatory variables but, as in the previous section, we will suppress \(x\)-dependence for notational convenience.

\(^2\)We may also have interval-censored or left-truncated times but these will not be considered here (see Kalbfleisch & Prentice (2002, chap. 3) and Lawless (2003, chap. 2)).
We need to form a joint distribution for the pair of random variables $(T, \delta)$ from which we can derive the likelihood function using (2.1). First we will discuss the important case of random censoring, where both event and censoring times are random variables, which includes type I censoring, where censoring times are fixed by design, as a special case. Let $\tilde{T}$ and $C$ represent the true event time and the censoring time respectively. Moreover, these variables are assumed to be independent (given covariates) with density functions given by $f(t \mid \theta)$ and $g(t \mid \phi)$, and survivor functions given by $S(t \mid \theta)$ and $G(t \mid \phi)$ respectively. Note the assumption that the distributions for $\tilde{T}$ and $C$ depend on different parameters; this is called non-informative censoring. The observed survival time is therefore $T = \min(\tilde{T}, C)$ and the censoring indicator is $\delta = 1(\tilde{T} \leq C) = 1(T = \tilde{T})$. Thus

$$\Pr(T = t, \delta = 1) = \Pr(\tilde{T} = t, C > t) = f(t \mid \theta)G(t \mid \phi),$$

(by independence of $\tilde{T}$ and $C$)

$$\Pr(T = t, \delta = 0) = \Pr(\tilde{T} > t, C = t) = S(t \mid \theta)g(t \mid \phi).$$

(by independence of $\tilde{T}$ and $C$)

The above can be combined to give the joint density

$$\Pr(T = t, \delta) = [f(t)G(t)]^\delta[S(t)g(t)]^{1-\delta},$$

and, for a given dataset comprising $(t_i, \delta_i)$, $i = 1, \ldots, n$, the likelihood function (for $\theta$ and $\phi$) is

$$L(\theta, \phi) = \prod_{i=1}^{n} [f(t_i)G(t_i)]^{\delta_i}[S(t_i)g(t_i)]^{1-\delta_i}.$$

As with (2.1), we require that individual observations are independent (or conditionally independent given $x$) to multiply densities in the above. Typically we are only interested in modelling the survival distribution and, hence, we can drop the terms involving the censoring distribution. Thus,

$$L(\theta) = \prod_{i=1}^{n} [f(t_i)]^{\delta_i}[S(t_i)]^{1-\delta_i},$$
and, using (1.10),

\[ L(\theta) = \prod_{i=1}^{n} [\lambda(t_i)S(t_i)]^\delta_i [S(t_i)]^{1-\delta_i} \]

\[ = \prod_{i=1}^{n} [\lambda(t_i)]^\delta_i S(t_i). \quad (2.6) \]

Hence, the log-likelihood is

\[ \ell(\theta) = \sum_{i=1}^{n} \delta_i \log \lambda(t_i) + \log S(t_i), \quad (2.7) \]

which holds for any possible independent censoring distribution. We only assume a (parametric) model for survival; the censoring distribution is arbitrary. To fit such model we insert the relevant hazard and survival functions into (2.7) and solve the score equations (2.4).

From (1.12), the above log-likelihood can be rewritten as

\[ \ell(\theta) = \sum_{i=1}^{n} \delta_i \log \lambda(t_i) - \Lambda(t_i) \]

\[ = \sum_{i=1}^{n} \delta_i \log \lambda(t_i) - \int_{0}^{t_i} \lambda(t) \, dt, \quad (2.8) \]

which shows the importance of the hazard function, i.e., the likelihood contribution for each individual is fully encompassed by the hazard. Another important re-expression of the log-likelihood function (similar to that in Aalen et al. (2008, sec. 5.1)) is given by

\[ \ell(\theta \mid t^*) = \sum_{i=1}^{n} \int_{0}^{t_i} \log \lambda(t) \, dN_i(t) - \int_{0}^{t^*} Y_i(t) \lambda(t) \, dt, \quad (2.9) \]

which is written in counting process notation\(^3\). Here \(N_i(t) = 1(t_i \leq t, \delta_i = 1)\) is the observed counting process and \(Y_i(t) = 1(t_i \geq t)\) is the at-risk process for the \(i\)th individual, respectively. Setting \(t^* = \infty\) ensures equality with (2.8);

\(^3\)See Andersen et al. (1993, chap. 2), Kalbfleisch & Prentice (2002, chap. 5) and Aalen et al. (2008, chap. 1) for details on counting processes.
in fact we require that \( t^* \geq t_{\text{max}} = \max(t_1, \ldots, t_n) \), i.e., \( \ell(\theta) \mid t^* \geq t_{\text{max}} = \ell(\theta) \). More generally, when viewed as a function of \( t^* \in [0, \infty) \), (2.9) describes the likelihood contribution from the whole dataset as it unfolds over time; the total contribution is attained at \( t^* = t_{\text{max}} \). Viewing the survival experience of the whole dataset in this way underpins the study of more general censoring schemes which we now discuss briefly (see also Kalbfleisch & Prentice (2002, secs. 3.1-3.4) and Aalen et al. (2008, sec. 5.1)).

**Independent Censoring**

We have seen that the log-likelihood function given by (2.7) holds for any independent censoring distribution. It turns out that (2.7) is even more general than this. We have only considered the case of random censoring, where both the survival and censoring times follow probability distributions. This does not cover censoring processes such as type II censoring, where the study is terminated after \( n_e \) events have occurred (Lawless, 2003, pp. 55–56). In this case, censoring times depend on the course of the study rather than coming from a distribution. However, it can be shown using martingale techniques, that (2.9) (which is equivalent to (2.7)) is the likelihood function for right-censored data arising from any independent censoring process (see Kalbfleisch & Prentice (2002, sec. 6.2) and Aalen et al. (2008, sec. 5.1)). Such a process must be independent of the survival process at time \( t \), given the history of the study up to that time (and covariates), but is otherwise arbitrary. Hence the likelihood function above is quite general.

It is worth noting that censoring was assumed to be non-informative (i.e., does not depend on \( \theta \)). Even if censoring is informative, the use of (2.7) can still be justified as a partial likelihood (as defined by Cox (1975)) with some loss of efficiency in estimating \( \theta \). However, Kalbfleisch & Prentice (2002, p. 196) note that it is difficult to construct a realistic example where censoring is informative yet independent of survival. Therefore, independence is the key assumption. The censoring mechanism would not be independent if individuals were censored (i.e., withdrawn from a study) whenever they had a particularly high/low risk of an event. We would be unable to make valid conclusions based on such data unless the nature of the dependence could be modelled appropriately.
2.2 Comparing Models

In Section 2.1 it has been implicitly assumed that the model we are fitting to the data, \( f(t) \), is the true model and that the m.l.e. converges to the true parameter vector. In practice however, this is rarely the case. Indeed any model we assume is a simplification of reality and is therefore not the true model. The true model may be sufficiently complex that it cannot be expressed as a simple mathematical formula. Nonetheless, statistical models (which approximate reality) can provide estimates of key features, such as the median survival time or the hazard function, and give insights into the data generating process. Hence the famous phrase: “all models are wrong but some are useful” (Box & Draper, 1987).

For the purposes of this section we introduce the notation

\[
\text{Assumed Model: } f(t | \theta), \tag{2.10}
\]

\[
\text{True Model: } f^*(t | \theta^*).
\]

We can then define the Kullback-Leibler distance (Kullback & Leibler, 1951) as

\[
KL(f^*, f) = E \log \left[ \frac{f^*(T)}{f(T)} \right], \tag{2.11}
\]

which measures the average information lost when \( f^* \) is approximated by \( f \) (Burnham & Anderson, 2002, sec. 2.1). It can be shown that \( KL(f^*, f) \geq 0 \) with equality when \( f(t) = f^*(t) \), i.e., the best fitting model is the true model (Pawitan, 2001, sec. 13.2). Thus, \( KL(f^*, f) \) is typically greater than zero, but there is some \( \theta = \hat{\theta} \) which minimises its value. In other words \( f(t | \hat{\theta}) \) is the closest we can get to \( f^*(t | \theta^*) \) among all models in the family \( f(t | \theta) \).

In light of the above, it may seem reasonable to use \( KL(f^*, f) \) as an objective function (to be minimised) for model estimation. This is not possible however as \( KL(f^*, f) \) depends on the unknown distribution \( f^* \). It is easy to show that maximising the likelihood function is equivalent to minimising \( KL(f^*, f) \) (Pawitan, 2001, sec. 13.3) and, thus, the m.l.e. \( \hat{\theta} \) is a consistent estimator of \( \theta \) - the parameter vector for which \( f(t) \) is closest to \( f^*(t) \). In practice we may fit a variety of different models to the data. The model with the highest likelihood is closest to the truth but we must also take into account its complexity. If a more complex models only offers a modest increase
in the value of the likelihood, the simpler model is typically preferable. This is a philosophy often referred to as *Occam’s razor* which states that “entities must not be multiplied beyond necessity”.

### 2.2.1 Nested Models

In the case where the models are *nested*, we can compare them by means of a *likelihood ratio test*. Let’s assume we have two models: \( f = f(t \mid \theta \in \Theta) \) and \( f_0 = f(t \mid \theta_0 \in \Theta_0 \subset \Theta) \), where \( \dim(\Theta) = k \) and \( \dim(\Theta_0) = k - r \). In other words \( f \) reduces to \( f_0 \) by setting \( r \) of its parameters equal to specified constants, e.g., \( \text{Exp}(\lambda) = \text{Weibull}(\lambda, \gamma = 1) \) (see Section 1.4). Wilks (1938) showed that

\[
W = 2 \log \frac{L(\hat{\theta})}{L_0(\hat{\theta}_0)} = 2[\ell(\hat{\theta}) - \ell_0(\hat{\theta}_0)] \sim \chi^2_r,
\]

as \( n \to \infty \) (see also Silvey (1975, chap. 7)). Thus, if \( W > \chi^2_{r,1-\alpha} \), we conclude that the fit provided by \( f \) is significantly better than \( f_0 \) (at the \( \alpha \) level of significance).

### 2.2.2 Non-Nested Models

More generally, models can be *non-nested*, i.e., there is no configuration of parameters which makes the two models equal. Therefore, we cannot use (2.12). The most widely applied method for comparing models is the so-called *Akaike information criterion* (Akaike, 1973, 1974) which is given by

\[
AIC = -2\ell(\hat{\theta}) + 2k,
\]

for a model \( f(t \mid \theta) \) where \( \dim(\theta) = k \). The AIC is a measure of fit where smaller values indicate superior fit. AIC can be motivated in a number of different ways (see Pawitan (2001, secs. 3.5 and 13.6) and Burnham & Anderson (2002, sec. 7.2)) and can be used as a general model selection
tool for both nested and non-nested models. Burnham & Anderson (2002) advocate its general use (see also Lindsey (1999)).

It is important to note that $AIC$ is on an interval scale due the omission of an unknown constant in its derivation (equal to $E\log f^*(T)$ from (2.11)). Thus, in a set of competing models, $\{f_1, \ldots, f_m\}$, Burnham & Anderson (2002, chap. 2) define the $AIC$ difference as

$$\Delta_j = AIC_j - AIC_{\text{min}}, \quad j = 1, \ldots, m,$$

(2.14)

where $AIC_{\text{min}}$ is the minimum $AIC$ value in the set. Burnham & Anderson also provide some rules of thumb shown in Table. 2.1 below. The use of $\Delta_j$ makes presentation/comparison clearer and also avoids the issue of mistakenly interpreting $AIC$ values on a ratio scale, for example, thinking no significant difference exists between $AIC_1 = 100,000$ and $AIC_2 = 100,020$ due to their magnitude when, in truth, $\Delta_2 = 20$ means that there is little support for model 2. The $AIC$ differences can be used further (in a pseudo-Bayesian manner) to define individual model probabilities/odds and for the purpose of model averaging (Burnham & Anderson, 2002, chap. 2-4).

Another method for model comparison, based on Bayesian arguments, is the Bayesian information criterion, also known as the Schwarz criterion (Schwarz, 1978), which is given by

$$BIC = -2\ell(\hat{\theta}) + \log(n)k.$$

(2.15)

$BIC$ is used in the same way as $AIC$ (we aim to minimise its value) but has a larger penalty for complexity when $n \geq 8$ (which it typically is). For a detailed account of model selection based on information criteria see Burnham & Anderson (2002). In particular see Burnham & Anderson (2002, chap. 6) for a comparison of selection methods, e.g., $AIC$ versus $BIC$.
Example 2.1. *Lung Cancer Data in Northern Ireland*

We fitted the Weibull and log-logistic models to the lung cancer dataset using maximum likelihood and have summarised the results in Table 2.2. The log-logistic model has a lower $AIC$ value and is therefore closer to the truth. We can also check the fit graphically by comparing the estimated Weibull and log-logistic survivor curves to the Kaplan-Meier curve (recall that the Kaplan-Meier curve is the closest estimate of the true survivor function). In Fig. 2.1 we can see that the log-logistic curve matches the Kaplan-Meier curve more closely than the Weibull.

<table>
<thead>
<tr>
<th>Model</th>
<th>Shape $\hat{\gamma}$</th>
<th>Scale $\hat{\lambda}$</th>
<th>$AIC$</th>
<th>$\Delta_{AIC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weibull</td>
<td>0.858</td>
<td>0.170</td>
<td>4126.8</td>
<td>12.3</td>
</tr>
<tr>
<td>Log-Logistic</td>
<td>1.151</td>
<td>0.173</td>
<td>4114.5</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Figure 2.1. Kaplan-Meier curve (black) with predicted Weibull (red) and log-logistic (green) curves overlayed.
2.3. **Uncertainty**

2.3.1 Uncertainty in the M.L.E.

It is generally unsatisfactory to produce a point estimate without conveying some sense of uncertainty associated with this estimate. The m.l.e., $\hat{\theta}$, is found by solving the score equations, (2.4), i.e., the first derivatives of the log-likelihood function, $\ell(\theta)$. The precision of the estimate however, comes from the *second* derivatives - these measure *curvature*. High curvature at $\hat{\theta}$ implies a tight peak and, hence, less uncertainty about the true parameter vector $\theta$. The *Hessian* matrix (named after German mathematician Ludwig Otto Hesse) is the $k \times k$ symmetric matrix of all second-order partial derivatives, i.e.,

$$H(\theta) = \begin{pmatrix}
\frac{\partial^2 \ell}{\partial \theta_1^2} & \frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_k} \\
\frac{\partial^2 \ell}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \ell}{\partial \theta_2^2} & \cdots & \frac{\partial^2 \ell}{\partial \theta_2 \partial \theta_k} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 \ell}{\partial \theta_k \partial \theta_1} & \cdots & \cdots & \frac{\partial^2 \ell}{\partial \theta_k \partial \theta_k}
\end{pmatrix}$$  \hspace{1cm} (2.16)

$$= \nabla_\theta \nabla_\theta^T \ell(\theta),$$  \hspace{1cm} (2.17)

where $\nabla_\theta = (\partial/\partial \theta_1, \ldots, \partial/\partial \theta_k)^T$ denotes the operation of partial differentiation with respect to $\theta$.

In likelihood theory the Hessian is not the key quantity, but rather its negation, the so-called *observed information matrix*,

$$I_o(\theta) = -H(\theta).$$  \hspace{1cm} (2.18)

The observed information quantifies the information carried in the dataset at hand. Moreover, the *expected information matrix*, or Fisher information matrix, is defined as

$$\mathcal{I}(\theta) = E[I_o(\theta)],$$  \hspace{1cm} (2.19)

where expectation is taken with respect to the data generating process which depends on both the survival and censoring processes. Note that, without censoring, expectation is with respect to the probability distribution of $T$. 
While the observed information varies from sample to sample, the expected information is an average quantity (over all possible datasets) that does not depend on any particular dataset. Therefore, $\mathcal{I}(\theta)$ tells us how difficult it is to estimate $\theta$ on average; parameters with higher expected information attain a greater level of precision for a given sample size. In particular, if we have some idea \textit{a priori} as to the value of $\theta$, then $\mathcal{I}(\theta)$ can be used for sample size determination.

Related to the above discussion is the fact that, as $n \to \infty$,

$$
\hat{\theta} \sim N[\theta, \Sigma = \mathcal{I}(\theta)^{-1}], \quad (2.20)
$$

where $\theta$ is the true parameter vector. This is a well-known result of likelihood theory (see Silvey (1975, pp. 74-78) and Pawitan (2001, secs. 9.4 and 9.9)) which can be shown to extend to right-censored survival data using the martingale central limit theorem (see Borgan (1984), Andersen et al. (1993, sec. 6.1) and Kalbfleisch & Prentice (2002, sec. 5.8)).

We noted in Section 2.1.1 that, for the purpose of optimisation, one should work with a model parametrisation where $\theta \in \mathbb{R}$. In addition to this, an unconstrained parametrisation means that (2.20) cannot place density on unallowable regions. Therefore, hypothesis tests and confidence intervals will have better properties if $\theta \in \mathbb{R}$ and, for this reason, we typically choose to parametrise our models in this way. With this in mind, (2.20) can be used to test the hypothesis $H_0 : \theta = \theta_0$ or to form joint confidence regions for the parameters. It is more common however to test hypotheses pertaining to individual parameters. As a consequence of (2.20), we have that

$$
\hat{\theta}_j \sim N[\theta_j, \sigma_{\theta_j}^2], \quad j = 1, \dots, k, \quad (2.21)
$$

where $\sigma_{\theta_j}^2$, the variance of $\hat{\theta}_j$, is the $j$th diagonal element of the covariance matrix $\Sigma$ in (2.20). Thus, the hypothesis $H_0 : \theta_j = \theta_{j0}$ can be tested using the Wald statistic, or Z-score,

$$
Z = \frac{\hat{\theta}_j - \theta_{j0}}{\sigma_{\theta_j}} \sim N(0, 1). \quad (2.22)
$$

Similarly, a $(1 - \alpha)100\%$ confidence interval for $\theta_j$ is given by

$$
\hat{\theta}_j \pm z_{1-\alpha/2} \sigma_{\theta_j}, \quad (2.23)
$$
where \( z_{1-\alpha/2} \) is the \((1 - \alpha/2)100\)th percentile of the standard normal distribution.

In the above construction of Z-scores and confidence intervals, an estimate of \( \Sigma \) is required. The two possibilities are \( \hat{\Sigma} = \mathcal{I}(\hat{\theta})^{-1} \) or \( \hat{\Sigma} = \mathcal{I}_o(\hat{\theta})^{-1} \), i.e., the expected or observed information matrix evaluated at the m.l.e. It turns out that the use of the observed information is preferable for a number of reasons:

**Computation:** Calculation of \( \mathcal{I} \) requires integration which is usually intractable whereas \( \mathcal{I}_o \) is always readily available. Even if \( \mathcal{I} \) is tractable, one must then programme its functional form into statistical software which can be laborious. Standard maximisation software (e.g. \texttt{nlm} or \texttt{optim} in R) provide the Hessian matrix, \( H(\hat{\theta}) \), as part of the output from which we can calculate \( \mathcal{I}_o(\hat{\theta}) = -H(\hat{\theta}) \). Thus, we only need to programme the log-likelihood function to get an estimate of \( \Sigma \).

**Censoring:** In order to calculate \( \mathcal{I} \), one must integrate over the data generating process which includes both the survival and the censoring processes. The dependence on censoring means it is impossible to calculate \( \mathcal{I} \) unless we assume a model for the censoring process - something which we prefer to keep arbitrary (see Section 2.1.2).

**Relevant Information:** The superiority of \( \mathcal{I}_o \) over \( \mathcal{I} \) was investigated by Efron & Hinkley (1978) who argued that inference based on \( \mathcal{I}_o(\hat{\theta}) \) is approximately conditional on relevant information from the dataset (called ancillary information). This leads to a better estimate of \( \Sigma \) and, hence, better behaved Z-scores and confidence intervals. The expected information matrix is averaged over all possible datasets and, therefore, does not have access to dataset-specific information. For further discussion see Pawitan (2001, secs. 8.3, 9.6 and 9.7) and Cox (2006, sec. 6.6).

### 2.3.2 Uncertainty in Functions of the M.L.E.

In the previous section we showed how to account for uncertainty in the estimation of parameters via the asymptotic distribution of \( \hat{\theta} \) given in (2.20).
However, often we are not interested in $\hat{\theta}$ itself but rather a function of $\hat{\theta}$, for example, the survivor function or median. Thus, methods for producing confidence intervals or hypothesis tests for such estimated functions are required.

**Delta Method**

The *delta method* is a well known classical method for deriving the asymptotic distribution of statistical estimates. Assuming that $g(\hat{\theta})$ is some function of the m.l.e. then, as $n \to \infty$,  

$$g(\hat{\theta}) \sim N[g(\theta), \nabla_\theta^T g(\theta) \Sigma \nabla_\theta g(\theta)], \quad (2.24)$$

where $\nabla_\theta = (\partial/\partial \theta_1, \ldots, \partial/\partial \theta_k)^T$ denotes the operation of partial differentiation with respect to $\theta$ and $\Sigma = \mathcal{I}(\theta)^{-1}$ (see Pawitan (2001, sec. 4.7) and Bishop et al. (2007, sec. 14.6)). Thus, we can estimate the variance of $g(\hat{\theta})$ via  

$$\hat{\sigma}^2_{g(\theta)} = \nabla_\theta^T g(\hat{\theta}) \ I_\theta(\hat{\theta})^{-1} \nabla_\theta g(\hat{\theta}), \quad (2.25)$$

which can be used in the usual way to test hypotheses or compute confidence intervals.

An important special case of (2.24) arises when $g(\cdot)$ is a function of just one of the estimated parameters, the $j$th say. In this case  

$$g(\hat{\theta}_j) \sim N[g(\theta_j), g'(\theta_j)^2 \sigma^2_{\theta_j}], \quad (2.26)$$

where $\sigma^2_{\theta_j}$ is the variance of $\hat{\theta}_j$, i.e., the $j$th diagonal element of the covariance matrix $\Sigma$. This is due to the fact that $\nabla_\theta g(\theta_j) = (0, \ldots, 0, g'(\theta_j), 0, \ldots, 0)^T$.

We previously mentioned the advantages of using a model parametrisation where $\theta$ is unconstrained; this is also relevant to $g(\theta)$. Often the range of the function $g(\theta)$ is constrained, i.e., $g(\theta) \in \mathcal{C} \subset \mathbb{R}$. Confidence intervals based on (2.24) may have bad properties if the true value of $g(\theta)$ lies near the upper/lower limit of $\mathcal{C}$. It is typically better to transform the function such that $g^*[g(\theta)] \in \mathbb{R}$. Using (2.26) we get  

$$g^*[g(\hat{\theta})] \sim N\{g^*[g(\theta)], g''^*[g(\theta)]^2 \sigma^2_{g(\theta)}\}, \quad (2.27)$$
2.3. UNCERTAINTY

from which we can compute a confidence interval for $g^*[g(\hat{\theta})]$. Applying $g^*^{-1}(\cdot)$, i.e., back-transforming, gives a confidence interval for $g(\hat{\theta})$ which remains within the constrained space $C$. This was discussed in Section 1.3 in relation to the Kaplan-Meier and Nelson-Aalen estimators. In particular see Example 1.1.

M.L.E. Simulation

The method of *m.l.e. simulation* is something suggested by the author that does not seem to appear elsewhere in the literature. This is an easy to implement method which avoids calculating/programming analytic derivatives: a tedious process where we are left open to making errors.

We know that $\hat{\theta}$ has a multivariate normal distribution (asymptotically) given by (2.20). We can then simulate $\hat{\theta}^{(b)}$ vectors, $b = 1, \ldots, m$, directly from

$$\hat{\theta}^{(b)} \sim N[\hat{\theta}, I_o(\hat{\theta})^{-1}],$$

(2.28)

where $\hat{\theta}$ is the m.l.e. and $I_o(\hat{\theta})$ is the observed information matrix in a particular application. The empirical distribution of our simulated sample of “m.l.e.” vectors $\{\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(m)}\}$ will then approximate the asymptotic distribution in (2.20) for $m$ sufficiently large. In other words, the simulated sample $\{\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(m)}\}$ is equivalent to a sample of m.l.e. vectors arising from $m$ independent datasets. All we need is a routine for simulating multivariate normal vectors; the function *mvtnorm* in the R package *MASS* does just this. As *mvtnorm* can simulate a sample of 10,000 100-dimensional multivariate normal vectors in a fraction of a second, the computational burden of setting $m$ large is very minimal. It is important however that $\theta$ is unconstrained for m.l.e. simulation, i.e., $\theta \in \mathbb{R}$. Otherwise we may simulate values which are not allowed, leading to nonsensical results.

For some function $g(\cdot)$, we can calculate $\hat{g}^{(b)} = g(\hat{\theta}^{(b)})$, $b = 1, \ldots, m$, to produce a univariate sample $\{\hat{g}^{(1)}, \ldots, \hat{g}^{(m)}\}$ whose empirical distribution will approximate the sampling distribution of $g(\hat{\theta})$. The standard deviation of this sample gives $se[g(\hat{\theta})]$ if required. However, in order to construct a $(1 - \alpha)100\%$ confidence interval for $g(\hat{\theta})$ we suggest using

$$\left[\hat{g}_{\alpha/2}, \hat{g}_{1-\alpha/2}\right],$$
where \( \hat{g}_\alpha \) denotes the \( \alpha \)-percentile of the simulated sample \( \{\hat{g}^{(1)}, \ldots, \hat{g}^{(m)}\} \).

Note that if \( g(\cdot) \in C \subset \mathbb{R} \) then \( \hat{g}^{(b)} \in C \forall b \), i.e., confidence intervals based on m.l.e. simulation will always respect \( C \) without the need to transform \( g(\theta) \) as in (2.27).

It is clear that this method can be implemented easily without tedious calculations and with much less room for error (compared with the delta method). It may appear similar to computational Bayesian approaches which simulate from a posterior distribution. However, we are not treating the true parameter vector, \( \theta \), as random here; nor is (2.28) viewed as a posterior distribution. Rather, simulating from (2.28) is used as a computational tool to approximate the asymptotic distribution of the m.l.e., \( \hat{\theta} \). Although philosophically different from the Bayesian view, a simulated sample of \( \hat{\theta}^{(b)} \) vectors from (2.28) may be very similar to that of a corresponding Bayesian analysis in cases where \( n \) is large and reasonably flat priors are assumed.

**Bootstrapping**

The method of *bootstrapping* is a technique based on resampling the observed data in order to determine the sampling distribution of statistical estimates. This simple, widely applicable, method was first introduced by Efron (1979) and has since become a standard computational tool in statistics (Efron & Tibshirani, 1994; Davison & Hinkley, 1997).

The observed data \( \{y_1, \ldots, y_n\} \) have an empirical distribution function which associates a probability mass of \( 1/n \) to each observation; this empirical distribution approximates the true distribution of the data. Thus we can draw samples (of size \( n \)) \( \{y_1^{(b)}, \ldots, y_n^{(b)}\}, b = 1, \ldots, m, \) from the empirical distribution which are equivalent to independent samples of data. This is done by sampling with replacement from the original set \( \{y_1, \ldots, y_n\} \). Note that for survival data \( y_i = (t_i, \delta_i, x_i^T) \) (Efron, 1981). Upon fitting our assumed model to each of the \( m \) datasets, using maximum likelihood, we attain a sample of m.l.e. vectors, \( \{\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(m)}\} \). We can then calculate \( \hat{g}^{(b)} = g(\hat{\theta}^{(b)}), b = 1, \ldots, m, \) to produce a sample \( \{\hat{g}^{(1)}, \ldots, \hat{g}^{(m)}\} \). As for the method of m.l.e. simulation, we can form confidence intervals for \( g(\theta) \) based on \( \{\hat{g}^{(1)}, \ldots, \hat{g}^{(m)}\} \).

Bootstrapping is a very general procedure that can be used in any situa-
2.3. Uncertainty

where replicate datasets are desired, e.g., cross validation, estimation of prediction error or calculation of standard error for non-parametric estimators. However, if we only require an estimate of the sampling distribution of \( g(\hat{\theta}) \), as is the goal of this section, m.l.e. simulation will perform well (if \( n \) is large enough). Moreover, m.l.e. simulation is much less computationally expensive as it does not require refitting of the model to multiple datasets.

Example 2.2. Confidence Intervals for Weibull Survivor Function

Here we parametrise the Weibull model in terms of \( \theta = (\alpha, \beta)^T \in \mathbb{R}^2 \) where \( \alpha = \log \gamma \) and \( \beta = \log \lambda \) respectively. Thus, the survivor function is

\[
g(\theta) = S(t | \theta) = \exp[-\exp(\beta) t^{\exp(\alpha)}]. \tag{2.29}
\]

The delta method gives

\[
\hat{\sigma}^2_g = \nabla_{\theta} g(\hat{\theta}) \hat{\Sigma} \nabla_{\theta} g(\hat{\theta}),
\]

where \( \nabla_{\theta} g(\theta) = [-\exp(\beta) \exp(\alpha) t^{\exp(\alpha)} S(t | \theta) \log t, -\exp(\beta) t^{\exp(\alpha)} S(t | \theta)]^T \), and \( \hat{\Sigma} = \hat{I}_{\theta}^{-1}(\hat{\theta}) \). Therefore

\[
g(\hat{\theta}) \pm 1.96\hat{\sigma}_g \tag{2.30}
\]

is the corresponding 95% confidence interval (which depends on \( t \)). Given that \( g(\theta) \in [0, 1] \) we can apply the transformation \( g^*(\cdot) = \log[-\log(\cdot)] \) which gives \( g^*[g(\theta)] \in (-\infty, \infty) \). From (2.27), and the fact that \( g^*(x) = 1/(x \log x) \),

\[
\hat{\sigma}^2_{g^*og} = \frac{\hat{\sigma}^2_g}{[g(\hat{\theta}) \log g(\hat{\theta})]^2}.
\]

A 95% confidence interval for \( g^*[g(\theta)] \) is then given by \( g^*[g(\hat{\theta})] \pm 1.96\hat{\sigma}_{g^*og} \). Back-transforming, i.e., applying \( g^{-1}(\cdot) = \exp[-\exp(\cdot)] \), gives

\[
g(\hat{\theta})^{\exp(\pm \hat{\sigma}_{g^*og})} = g(\hat{\theta})^{\exp(\pm \hat{\sigma}_g/[g(\hat{\theta}) \log g(\hat{\theta})])}, \tag{2.31}
\]

which is a 95% confidence interval for \( g(\theta) \) constrained to lie in \([0, 1]\).

We fitted the Weibull model to three subsets of the lung cancer dataset, namely: (i) individuals aged 90 and older, (ii) individuals aged 86 - 90 and (iii) individuals aged 82 - 86. We also fitted the model to (iv) the full lung cancer dataset. The sample sizes are 6, 19, 58 and 855 respectively. For
each fitted model we evaluated the survivor function (2.29) at 100 values of $t$ ranging from $t = 0$ to $t = 20$. The corresponding 95% confidence intervals (at each $t$ value) were calculated analytically using the delta method, (2.30), and the transformed delta method, (2.31). Additionally, 95% confidence intervals were calculated using m.l.e. simulation and bootstrapping, both with $m = 1000$. The results are shown in Fig.2.2.

We can see that, for the 90+ group ($n = 6$) the delta method produces confidence intervals which breach the $[0, 1]$ bounds; this is rectified using the transformed delta method. This breach of allowable bounds is much less...
2.3. UNCERTAINTY

prominent for the 86 - 90 group \((n = 19)\) and, moreover, there is little difference between the confidence intervals produced by any of the methods. For the 82 - 86 group \((n = 58)\), and also for the full dataset, the confidence intervals are virtually indistinguishable. Thus we can see that, even for relatively small datasets, the methods produce similar results. Furthermore, in Section 3.6.4, we carry out a simulation study to evaluate confidence intervals constructed using m.l.e. simulation and the delta method; again we find close agreement.

As the methods produce similar results, we must be aware of the ease of implementation and the computational expense required. We have already established that both m.l.e. simulation and bootstrapping require very little effort to implement as calculation/programming of derivatives is not required. In order to determine computational expense we recorded the time taken to produce the confidence intervals shown in Fig. 2.2. These times are displayed in Table 2.3 and exclude the time taken to initially fit the model, as this is something that must be carried out in any case.

Table 2.3. Time Taken (in seconds) to Produce Confidence Intervals in Fig. 2.2

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Delta</th>
<th>Delta (g^*)</th>
<th>M.l.e. Sim.</th>
<th>Boot.</th>
<th>Boot. 0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age (\geq 90)</td>
<td>6</td>
<td>0.007</td>
<td>0.007</td>
<td>0.109</td>
<td>12.877</td>
<td>17.357</td>
</tr>
<tr>
<td>86 (\leq) age (&lt; 90)</td>
<td>19</td>
<td>0.006</td>
<td>0.007</td>
<td>0.110</td>
<td>13.302</td>
<td>18.915</td>
</tr>
<tr>
<td>82 (\leq) age (&lt; 86)</td>
<td>58</td>
<td>0.005</td>
<td>0.009</td>
<td>0.110</td>
<td>13.155</td>
<td>20.504</td>
</tr>
<tr>
<td>All ages</td>
<td>855</td>
<td>0.006</td>
<td>0.008</td>
<td>0.110</td>
<td>27.815</td>
<td>44.340</td>
</tr>
<tr>
<td>Relative to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age (\geq 90)</td>
<td>6</td>
<td>1.00</td>
<td>1.00</td>
<td>15.57</td>
<td>1839.57</td>
<td>2479.57</td>
</tr>
<tr>
<td>86 (\leq) age (&lt; 90)</td>
<td>19</td>
<td>1.00</td>
<td>1.17</td>
<td>18.33</td>
<td>2217.00</td>
<td>3152.50</td>
</tr>
<tr>
<td>82 (\leq) age (&lt; 86)</td>
<td>58</td>
<td>1.00</td>
<td>1.80</td>
<td>22.00</td>
<td>2631.00</td>
<td>4100.80</td>
</tr>
<tr>
<td>All ages</td>
<td>855</td>
<td>1.00</td>
<td>1.33</td>
<td>18.33</td>
<td>4635.83</td>
<td>7390.00</td>
</tr>
<tr>
<td>Delta</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age (\geq 90)</td>
<td>6</td>
<td>0.06</td>
<td>0.06</td>
<td>1.00</td>
<td>118.14</td>
<td>159.24</td>
</tr>
<tr>
<td>86 (\leq) age (&lt; 90)</td>
<td>19</td>
<td>0.05</td>
<td>0.06</td>
<td>1.00</td>
<td>120.93</td>
<td>171.95</td>
</tr>
<tr>
<td>82 (\leq) age (&lt; 86)</td>
<td>58</td>
<td>0.05</td>
<td>0.08</td>
<td>1.00</td>
<td>119.59</td>
<td>186.40</td>
</tr>
<tr>
<td>All ages</td>
<td>855</td>
<td>0.05</td>
<td>0.07</td>
<td>1.00</td>
<td>252.86</td>
<td>403.09</td>
</tr>
</tbody>
</table>

Note: Times shown above are in fact averages over 50 repetitions (carried out on an Intel® Core™ i7 620M 2.67 Ghz). This was done to absorb any random fluctuations in computation time.
The table shows two versions of the bootstrap, namely: “Boot.” and “Boot. 0.01”. The former uses the m.l.e., \( \hat{\theta} = (\hat{\alpha}, \hat{\beta})^T \), from the initial model fit, as a starting value for the Newton-Raphson algorithm in each of the 1000 bootstrapped fits. The latter uses generic starting values: \((0.01, 0.01)^T\). Since each bootstrapped dataset will not be vastly different from the original dataset, we expect that \( \hat{\theta}^{(b)} \approx \hat{\theta} \), \( b = 1, \ldots, 1000 \). Thus, using \( \hat{\theta} \) as a starting value should help the Newton-Raphson algorithm to converge to \( \hat{\theta}^{(b)} \) more quickly; this indeed turned out to be the case as we can see from Table 2.3. Even with this speed-up however, the bootstrap takes much longer than the delta method or m.l.e. simulation. Obviously, for larger sample sizes (e.g., when \( n = 855 \)) the bootstrap time increases as we must refit the model 1000 times to large replicate datasets. Neither the delta method nor m.l.e. simulation depend on sample size; the former is instantaneous whilst the latter takes a fraction of a second. Thus, we favour m.l.e. simulation in this case as it is both simple to implement and computationally inexpensive.
Chapter 3

Simulating Survival Data

3.1 Introduction

In a simulation study we generate data from a known statistical model. This gives us access to something we rarely have in reality - the truth. Accordingly, we can investigate a variety of model properties, in a controlled environment, affording us useful knowledge which is transferable to our applied work. First we must choose a model from which we will simulate, \( f(t \mid \theta) \), and values for the elements of the simulation scenario vector - sample size, censored proportion and parameter values - denoted by \((n, p, \theta) \in \mathcal{N} \times \mathcal{P} \times \Theta = \mathcal{S}\). The scenario space, \( \mathcal{S} \), is the set of all scenarios to be considered in our simulation study and is of finite dimension, i.e., \( \text{dim}(\mathcal{S}) = r \ll \infty \). For a particular scenario \( s_k = (n_k, p_k, \theta_k) \), \( k = 1, \ldots, r \), we proceed as follows:

1. Generate data from \( f(t) \) for scenario \( s_k \).
2. Fit a model to these data.
4. Repeat the steps 1 - 3 \( m \) times to produce simulation replicates.

We carry out the above procedure for all \( s_k \in \mathcal{S} \) in order to explore the behaviour of the responses across the scenario space. Thus, we could employ the use of standard experimental design techniques to design our simulation
CHAPTER 3. SIMULATING SURVIVAL DATA

study efficiently; this fact is touched on in Ripley (1987, sec. 5.5) and is the subject of Kleijnen (2008). However, we will opt for a full factorial design in our simulation work, i.e., \( \text{dim}(S) = \text{dim}(N) \cdot \text{dim}(P) \cdot \text{dim}(\Theta) \).

Uncertainty can be reduced to any desired level by increasing the number of replicates (which only depends on our patience in waiting for the results of our simulation). As noted by Ripley (1987): “randomness was introduced by the experimenter and hence is under his or her complete control”. Thus, sufficient reduction in uncertainty enables “proof” by simulation.

3.2 Reasons for Simulating

The objective of analysing simulation experiments is to answer questions on properties of statistical models. As we ultimately apply these models to real data, understanding their properties reinforces such applications. Some items of potential interest, which can be investigated by simulation, are outlined below (for further information on the uses of simulation see Ripley (1987)).

- We may wish to discover how well features of \( f(t \mid \theta) \) (e.g., the survivor function, hazard function, median survival time etc.) can be recovered using \( f(t \mid \hat{\theta}) \). By monitoring these estimates over the scenario space, we can determine how badly inference may be compromised in less than ideal situations, e.g., small sample size with high censoring.

- Standard likelihood theory (Chapter 2) says that \( \hat{\theta} \) is an unbiased estimator of \( \theta \) and follows a multivariate normal distribution asymptotically. By examining the sets of m.l.e. vectors \( \{\hat{\theta}_{1}^{(k)}, \ldots, \hat{\theta}_{m}^{(k)}\} \), for \( k = 1, \ldots, r \), we can determine when such asymptotic theory is reasonable. Furthermore, we can investigate the properties of confidence intervals for parameters, or functions of parameters, and check that they have the correct coverage, i.e., that 95% confidence intervals include the true parameter 95% of the time.

- We can generate data from \( f_{1}(t) \) and fit an alternative model, \( f_{2}(t) \), to these data. In this way we can understand how well a misspecified model can perform (cf. Ha & MacKenzie (2010)).
3.2. REASONS FOR SIMULATING

- Orthogonality of parameters is often a desirable property which ensures asymptotic independence of parameter estimates (Cox & Reid, 1987). It is possible to determine orthogonality (or lack thereof) analytically in terms of the expected information matrix but, as mentioned in Section 2.3.1, the expected information matrix can be difficult to work out (particularly when censoring is present). However, it is easy to calculate the simulation-based correlation matrix for estimated parameters (in a given scenario) using the fact that we have $m$ replicates of $\hat{\theta}$. Moreover, we can examine the pattern of correlation across the scenario space.

- In a regression setting it is useful to know how often important covariates can be found using a particular variable selection procedure. We can check this by simulation and compare different procedures.

With everything under our “complete control” we may investigate any model property of interest. In certain cases deriving analytic results is difficult/impossible, whereas a simulation study can easily be carried out to uncover the necessary results. Furthermore, advances in modern computing power make large-scale simulation studies feasible which, therefore, make complicated modelling questions accessible.

As is clear from the above, simulation is very powerful. However, it does have its limitations. Firstly, a statistical model is just that: a model, or idealisation, of reality (see Section 2.2). Real life data-generating mechanisms are complex and may potentially be so complex that they cannot be expressed as a mathematical formula. However, we simulate data from models which typically have a relatively simple mathematical form. A second issue is that there is nothing to prevent us from simulating scenarios which lead to unrealistic data. We may choose a very implausible point on the parameter space leading to simulated data unlike anything found in practice.

Of course, the above criticisms do not invalidate any simulation study from a mathematical perspective; all results are still relevant within the framework of the simulation, i.e., for the given model and scenario space. However, these findings may not be exactly reproducible in real life applications where, for example, there are many more sources of variation than our theoretical model can allow for. If desired, we can generate data which more
closely resemble real data through *data directed simulation*. This involves using experience fitting models to real data to inform us as to what parameter values typically arise in reality. However, it is worth noting that the ability to simulate “unrealistic” data can also be seen as a strength; we can explore, for example, limiting cases on the boundary of the scenario space. Indeed, just because we have not seen such data, does not mean we will not.

### 3.3 Inverse Survivor Function

We simulate using the *method of inversion* which is based on the well-known fact that, for any distribution, the corresponding cumulative distribution function, \( F(t) \), is uniformly distributed on \([0, 1]\) (see Devroye (1986, sec. 2.2)). This implies that the survivor function, \( S(t) \), is also uniformly distributed on \([0, 1]\). Thus, given \( U \sim \text{Uniform}(0, 1) \),

\[
S(T) \sim U \Rightarrow T \sim S^{-1}(U),
\]  

(3.1)

where we use \( S(t) \) rather than \( F(t) \) due its central role in survival analysis. Therefore, upon generating a sample of uniform variables, \( \{u_1, \ldots, u_n\} \), we can produce a sample of survival times using (3.1): \( \{t_1 = S^{-1}(u_1), \ldots, t_n = S^{-1}(u_n)\} \). Note that we have given the functional form of \( S^{-1}(\cdot) \) for a variety of commonly used survival distributions in Section 1.4.

We may also wish to simulate data which depend on covariates. In this case we must first generate a vector of covariates for each individual \( x_i = (x_{1i}, \ldots, x_{pi})^T, i = 1, \ldots, n \). We can then allow the model parameters to depend on these covariates in some way, i.e., \( \theta_i = g(x_i) \) for some function \( g(\cdot) \). Our sample of survival times, which depend on covariates, is then \( \{t_1 = S^{-1}(u_1 | x_1, \theta_1), \ldots, t_n = S^{-1}(u_n | x_n, \theta_n)\} \). For more information on regression models see Chapters 4 and 5.

The above clearly depends on our ability to generate \( \{u_1, \ldots, u_n\} \). Computer algorithms used to produce such a sample aim to mimic randomness but are, in fact, deterministic; the \( j \)th number is constructed using the sequence of numbers preceding it. Thus, such a sample is referred to as *pseudo-random*. Generating a sample of pseudo-random variables, and assessing their properties, is non-trivial. This specialist topic is covered in detail in Gentle (2003,
3.4 Random Censoring

In the previous section we have shown how to simulate uncensored survival data. As we typically have censored data in practice, we would like to incorporate this feature in our simulated data. Here we use random censoring as a mechanism for censoring simulated observations. Using the notation of Section 2.1.2, we let \( \tilde{T} \) and \( C \) represent the event time and the censoring time respectively. These random variables are independent with density functions given by \( f(t|\theta) \) and \( g(t|\phi) \), and survivor functions given by \( S(t|\theta) \) and \( G(t|\phi) \) respectively. Using (3.1), we generate

\[
\tilde{T} = S^{-1}(U_1) \quad \text{and} \quad C = G^{-1}(U_2),
\]

where \( U_1 \perp U_2 \Rightarrow \tilde{T} \perp C \). The observed survival time is then given by \( T = \min(\tilde{T}, C) \) and the censoring indicator is \( \delta = 1(\tilde{T} \leq C) = 1(T = \tilde{T}) \).

Thus, from a simulated sample \( \{(u_{11}, u_{21}), \ldots, (u_{1n}, u_{2n})\} \), a censored sample of survival data, \( \{(t_1, \delta_1), \ldots, (t_n, \delta_n)\} \), is produced.

We now show how to control the censored proportion in a simulated sample by following the work of MacKenzie (1994). The probability that \( T \) is a censoring time is given by

\[
p = \Pr(\tilde{T} > C) = \int_0^\infty \Pr(\tilde{T} > t \cap C = t) \, dt
\]

\[
= \int_0^\infty \Pr(\tilde{T} > t) \Pr(C = t) \, dt \quad (\tilde{T} \perp C)
\]

\[
= \int_0^\infty S(t|\theta)g(t|\phi) \, dt.
\]

We would like to be able to generate data with a prespecified censored proportion, \( p_{\text{fix}} \), so that we can determine the effect of varying its value. In
(3.3) above, $p$ is implied by the chosen values of $\theta$ and $\phi$. However, we are generally only interested in the value of $\theta$ and prefer $\phi$ to be implied by the chosen values of $\theta$ and $p_{\text{fix}}$. Therefore, setting $\theta$ in advance, we consider $p$ above to be a function of $\phi$, i.e.,

$$p(\phi) = \int_0^\infty S(t \mid \theta)g(t \mid \phi) \, dt.$$  

Furthermore, in our simulations, we assume that $C \sim \text{Exp}(\phi)$. Thus,

$$p(\phi) = \int_0^\infty S(t \mid \theta) \phi \exp(-\phi t) \, dt,$$  

(3.4)

and $\phi$ is chosen as the solution of $p(\phi) = p_{\text{fix}}$, i.e., $\phi = p^{-1}(p_{\text{fix}})$ where $p^{-1}(\cdot)$ is the inverse of (3.4). This $\phi$ value ensures that the proportion censored in a simulated sample is approximately equal to $p_{\text{fix}}$, i.e., $1 - \sum_{i=1}^n \delta_i/n \approx p_{\text{fix}}$. If we wish, we can iteratively generate data (using a while loop) until we arrive at a dataset where the censored proportion is exactly equal to $p_{\text{fix}}$.

If our simulated data depends on covariates then, as mentioned in Section 3.3, we have $\theta_i = \theta(x_i)$ for $i = 1, \ldots, n$. In this case, (3.4) becomes

$$p_i(\phi) = \int_0^\infty S(t \mid \theta_i) \phi \exp(-\phi t) \, dt.$$

(3.5)

If there are $n$ distinct covariate profiles then there are $n$ distinct $\theta_i$ values, each with a corresponding $\phi$ value: $\phi_i = p^{-1}_i(p_{\text{fix}})$ for $i = 1, \ldots, n$. However, if there are $m < n$ covariate profiles then we only need to invert (3.5) $m$ times, e.g., if our simulated data depend on a single binary covariate then there are only two $\theta$ values and, hence, two $\phi$ values.

The functional form of (3.4) (or (3.5)) is typically not analytic and must be approximated using numerical integration. Therefore, $p^{-1}(\cdot)$ is itself non-analytic and our desired $\phi$ value does not have a closed-form solution. However, MacKenzie’s $J$-function, given by

$$J(\phi) = [p_{\text{fix}} - p(\phi)]^2,$$

(3.6)

can be minimised using numerical optimisation to yield the $\phi$ value for which $p(\phi)$ is closest to $p_{\text{fix}}$ (MacKenzie, 1994). It is worth mentioning a simple case
where (3.4) is analytic: when $\tilde{T} \sim \text{Exp}(\lambda)$,
\[ p(\phi) = \int_0^\infty \phi \exp[-(\lambda + \phi)t] \, dt = \frac{\phi}{\lambda + \phi}. \]
\[ \Rightarrow p^{-1}(p_{\text{fix}}) = \frac{p_{\text{fix}}}{1 - p_{\text{fix}}} \lambda. \]

### 3.5 Extension to Include Cure Rate Models

Cure rate models (also known as defective or tail deficit models) are those supporting survivor curves which do not fall to zero but to some positive limit, i.e., $\lim_{t \to \infty} S(t) \neq 0$. The Gompertz and time-dependent logistic models (Sections 1.4.3 and 1.4.6) are both examples of cure rate models. In particular, these two models imply a cured proportion when their shape parameter, $\gamma$, is negative. Thus, we generalise the method of MacKenzie (1994) described in the previous section in order to facilitate the simulation of cure rate data.

First we define $p_{\text{cure}} = \lim_{t \to \infty} S(t)$, the cured proportion. Following Section 3.4, we simulate data based on the method of inversion and assume random censoring. However, in the case of a cure rate model

\[ \tilde{T} = \begin{cases} S^{-1}(U_1) & \text{if } U_1 < 1 - p_{\text{cure}} \\ \infty & \text{if } U_1 \geq 1 - p_{\text{cure}}, \end{cases} \]

i.e., the model generates cured individuals with probability $p_{\text{cure}}$ and these individuals have a survival time of infinity. This intuitively means that such individuals never experience the event. As before, censoring times are generated via $C = G^{-1}(U_2)$.

In this setting the censored proportion is

\[ p = \Pr(\tilde{T} > C) \]
\[ = \Pr(\tilde{T} > C \cap \tilde{T} < \infty) + \Pr(\tilde{T} > C \cap \tilde{T} = \infty) \]
\[ = \Pr(\tilde{T} < \infty) \Pr(\tilde{T} > C \mid \tilde{T} < \infty) + \Pr(\tilde{T} = \infty) \Pr(\tilde{T} > C \mid \tilde{T} = \infty) \]
\[ = (1 - p_{\text{cure}}) p^* + p_{\text{cure}}, \]
where the last line comes from fact that all cured individuals are censored by definition, i.e., \( \Pr(\tilde{T} > C | \tilde{T} = \infty) = 1 \). Note also that we have defined \( p^* = \Pr(\tilde{T} > C | \tilde{T} < \infty) \) which is the censored proportion in the non-cured individuals. The value of \( p_{\text{cure}} \) is implied by the parameter vector \( \theta \) and therefore, in order to control \( p \), we must control \( p^* \). It is clear from (3.8) that we now have a lower bound on \( p \), i.e., \( p \geq p_{\text{cure}} \).

As in Section 3.4, we assume that \( C \sim \text{Exp}(\phi) \). However, in the case of cure rate data we choose a value of \( \phi \) such that \( p^*(\phi) = p^*_{\text{fix}} \) where \( p^*_{\text{fix}} \) is a prespecified censored proportion \textit{for the non-cured individuals}. Considering \( p^* \) (and hence \( p \)) to be a function of \( \phi \) and rearranging (3.8) gives

\[
p^*(\phi) = \frac{p(\phi) - p_{\text{cure}}}{1 - p_{\text{cure}}}
\]

where \( p(\phi) \) is still defined as in (3.4). Thus, \( \phi \) is chosen as the solution of \( p^*(\phi) = p^*_{\text{fix}} \). To this end, we define the \( J^- \)-function (generalising MacKenzie’s \( J \)-function):

\[
J^-(\phi) = [p^*_{\text{fix}} - p^*(\phi)]^2,
\]

which is minimised using numerical methods. The solution of the \( J^- \)-function produces a \( \phi \) value such that \( 1 - \sum_{i=1}^{n} \delta_i(\tilde{t}_i < \infty)/\sum_{i=1}^{n} 1(\tilde{t}_i < \infty) \approx p^*_{\text{fix}} \) in any simulated sample. Furthermore, we may wish to iteratively generate data (using a while loop) until we arrive at a dataset where this proportion is exactly equal to \( p^*_{\text{fix}} \).

In the case where parameters depend on covariates we have \( \theta_i = \theta(x_i) \), for \( i = 1, \ldots, n \), and, hence, \( p_{\text{cure},i} = p_{\text{cure}}(x_i) \). Thus, (3.9) becomes

\[
p^*_{i}(\phi) = \frac{p_i(\phi) - p_{\text{cure},i}}{1 - p_{\text{cure},i}},
\]

where \( p_i(\phi) \) is defined in (3.5) and \( J^i(\phi) = [p^*_{\text{fix}} - p^i(\phi)]^2 \) must be solved for \( i = 1, \ldots, n \). In this case the overall censored proportion is

\[
p = (1 - \bar{p}_{\text{cure}}) p_{\text{fix}} + \bar{p}_{\text{cure}},
\]

where \( \bar{p}_{\text{cure}} = \sum_{i=1}^{n} p_{\text{cure},i}/n \) and is the lower bound of \( p \).

Clearly the above can be used as a general procedure for simulating survival data: both cure rate and standard survival models. In the latter case, \( p_{\text{cure}} = 0 \) and hence \( p^* = p \), \( p^*_{\text{fix}} = p_{\text{fix}} \) and \( J^* = J \).
3.6 Some Simulation Studies

We simulated data from a Weibull model with parameter vector $\theta = (\alpha = \log \gamma, \beta = \log \lambda)^T$. In total there were $2^4 = 16$ simulation scenarios arising from all combinations of values shown in Table 3.1. Each of the 16 scenarios was repeated 1000 times. At each repetition we fitted a Weibull model (i.e., the true model) to the simulated data. All of the studies below (except that of Section 3.6.5) are based on this simulation setup and are intended to show some of the uses of simulation (as mentioned in Section 3.2).

### Table 3.1. Simulation Settings

<table>
<thead>
<tr>
<th>$n$</th>
<th>$p$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values</td>
<td>100</td>
<td>20%</td>
<td>-0.8</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>80%</td>
<td>0.8</td>
</tr>
</tbody>
</table>

3.6.1 True Model Versus an Incorrect Model

In addition to fitting the true (Weibull) model to the simulated data, we also fitted a log-logistic model. At each repetition of a particular scenario we calculated the difference between $AIC$ values, \[ \Delta_{AIC,j} = AIC_{\text{log-logistic},j} - AIC_{\text{Weibull},j}, \]
for $j = 1, \ldots, 1000$. Hence the average $AIC$ difference for that particular scenario is \[ \bar{\Delta}_{AIC} = \frac{1}{1000} \sum_{j=1}^{1000} \Delta_{AIC,j}. \]

The average $AIC$ differences from all scenarios are shown in Table 3.2.

All $\bar{\Delta}_{AIC}$ are positive which means that the true model provides a better fit on average. It is clear that $\bar{\Delta}_{AIC}$ increases as information increases (i.e., $n \uparrow$ or $p \downarrow$) which we would expect; it is well-known that it is easier to choose among competing models/hypotheses when the level of information is high. In addition to this, the $\bar{\Delta}_{AIC}$ value is lower for $\alpha = -0.8$ compared with $\alpha = 0.8$. This is also to be expected given the shapes of hazard supported
CHAPTER 3. SIMULATING SURVIVAL DATA

Table 3.2. Average AIC Differences

<table>
<thead>
<tr>
<th>p</th>
<th>α</th>
<th>β</th>
<th>$\Delta_{AIC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.8</td>
<td>4.7</td>
<td>50.8</td>
</tr>
<tr>
<td>20%</td>
<td>0.8</td>
<td>4.8</td>
<td>51.1</td>
</tr>
<tr>
<td></td>
<td>-0.8</td>
<td>8.9</td>
<td>91.9</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>9.1</td>
<td>91.1</td>
</tr>
<tr>
<td></td>
<td>-0.8</td>
<td>0.0</td>
<td>0.4</td>
</tr>
<tr>
<td>80%</td>
<td>0.8</td>
<td>0.0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>-0.8</td>
<td>0.8</td>
<td>10.2</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>1.0</td>
<td>10.1</td>
</tr>
</tbody>
</table>

by the two models (see Section 1.4). When $\alpha > 0$ the Weibull hazard is increasing monotonically whereas the closest shape supported by the log-logistic model is a hazard that increases to a peak followed by monotonic decrease. On the other hand, when $\alpha < 0$ the Weibull hazard is decreasing monotonically which is a hazard shape also supported by the log-logistic model.

We denote by (a) and (b) the scenarios with $(n, p, \alpha, \beta) = (100, 80\%, -0.8, -0.8)$ and $(n, p, \alpha, \beta) = (1000, 20\%, 0.8, -0.8)$ respectively. These represent cases where the fitted models are both difficult and easy to distinguish from each other (as $\Delta_{AIC}^a = 0.0$ and $\Delta_{AIC}^b = 91.9$ respectively). In these scenarios we now judge the fit provided by both models in terms of their estimated survivor and hazard functions. The average estimated survivor curve (arising from either fitted model), evaluated at time $t$, is given by

$$\bar{S}(t) = \frac{1}{1000} \sum_{j=1}^{1000} \hat{S}_j(t),$$

where $\hat{S}_j(t) = S(t | \hat{\alpha}_j, \hat{\beta}_j)$ is the $j$th ($j = 1, \ldots, 1000$) estimated survivor curve evaluated at time $t$. The corresponding 95% confidence intervals are given by percentile 2.5 and 97.5 in the sample $\{\hat{S}_1(t), \ldots, \hat{S}_{1000}(t)\}$. We can
also calculate the average estimated hazard function, and corresponding 95% confidence interval, in the same way.

Figure 3.1 compares the average estimated survivor and hazard curves (using both fitted models) with the true curves in scenario (a) and (b). The results are in agreement with the AIC values: the two models provide a similar fit to the data arising from scenario (a) whereas the failings of the log-logistic model are clear for scenario (b) where, as previously noted, this model cannot support the monotonic increasing hazard.

Figure 3.1. Average estimated survivor and hazard curves (solid) with 95% confidence intervals (dash) for two scenarios: (a) \((n, p, \alpha, \beta) = (100, 80\%, -0.8, -0.8)\) and (b) \((n, p, \alpha, \beta) = (1000, 20\%, 0.8, -0.8)\).
3.6.2 Extrapolating Beyond the Data

In Fig. 3.1 we have not evaluated the survivor curves in scenario (a) to the point where they are near zero. This is because the estimated curves are evaluated at time-points within the range of the data. In this particular scenario no observed time was greater than $t = 2.5$, due to the high probability of being censored. Thus, we can say that the two models fit the data adequately given the information at hand. Figure 3.2 shows what happens if we continue our predictions beyond the data (up to $t = 25$).

![Figure 3.2. Extrapolating estimated survivor and hazard curves beyond the range of the data for scenario (a) where $(n,p,\alpha,\beta) = (100,80\%,-0.8,-0.8)$. Average curves (solid) with C.I. (dash) shown.](image)

We can see that the log-logistic hazard decreases faster than the true hazard which results in overestimated survival probabilities. The Weibull estimate remains close to the truth as it is the correct model. Thus, either model is suitable for describing the data within the observed time range ($t = 0$ to $t = 2.5$) but only the correct model maintains suitability beyond this. In reality we typically do not know which model is the correct model (or if any fitted model is “correct”). It is therefore very risky to extrapolate any estimated quantity too far beyond the data. This advice is, of course, well-known but has been confirmed by our simulation (See Sections 3.6.5...
and 3.6.6 also where we investigate the estimation of a cured proportion, i.e., extrapolating the survivor curve to \( t = \infty \).

### 3.6.3 Bias, S.E. and Correlation in Estimates

Here we investigate the bias in estimating parameters for the fitted Weibull model (as this is the data generating model). We also look at the standard error in these estimates and the correlation between them. In each scenario we calculated the average estimated parameters,

\[
\bar{\alpha} = \frac{1}{1000} \sum_{j=1}^{1000} \hat{\alpha}_j, \quad \bar{\beta} = \frac{1}{1000} \sum_{j=1}^{1000} \hat{\beta}_j, \tag{3.13}
\]

where \( \hat{\alpha}_j \) and \( \hat{\beta}_j \) are the estimated parameters from the \( j \)th repetition (\( j = 1, \ldots, 1000 \)) of that scenario. Similarly, we calculated the average relative bias of the estimates,

\[
\text{rbias}(\hat{\alpha}) = \frac{\bar{\alpha} - \alpha}{|\alpha|}, \quad \text{rbias}(\hat{\beta}) = \frac{\bar{\beta} - \beta}{|\beta|}, \tag{3.14}
\]

and the standard errors (i.e., standard deviation over simulation replicates),

\[
\text{sd}(\hat{\alpha}) = \sqrt{\frac{\sum_{j=1}^{1000} (\hat{\alpha}_j - \bar{\alpha})^2}{1000 - 1}}, \quad \text{sd}(\hat{\beta}) = \sqrt{\frac{\sum_{j=1}^{1000} (\hat{\beta}_j - \bar{\beta})^2}{1000 - 1}}. \tag{3.15}
\]

The results are summarised in Table 3.3.

Clearly bias decreases as information increases (i.e., \( n \uparrow \) or \( p \downarrow \)). Bias is larger when less information is available, as we may expect, but is not alarmingly high even in the worst case scenario with \( n = 100 \) and \( p = 80\% \). This explains why we could extrapolate beyond the data reliably using the Weibull model in Fig. 3.2 where \( (n, p, \alpha, \beta) = (100, 80\%, -0.8, -0.8) \); even though information is low, the bias in \( \hat{\theta} = (\hat{\alpha}, \hat{\beta})^T \) remains relatively low (on average) and we can recover the true survivor and hazard functions (on average). Turning our attention to the standard errors, we can see that they decrease as information increases (i.e., \( n \uparrow \) or \( p \downarrow \)).
Table 3.3. *Average Parameter Estimates with Standard Error and Relative Bias*

<table>
<thead>
<tr>
<th>n</th>
<th>p</th>
<th>α</th>
<th>β</th>
<th>ŭα S.E.</th>
<th>R.Bias</th>
<th>ŭβ S.E.</th>
<th>R.Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>20%</td>
<td>-0.8</td>
<td>-0.8</td>
<td>-0.791 (0.089)</td>
<td>0.011</td>
<td>-0.808 (0.122)</td>
<td>-0.010</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.8</td>
<td>-0.789 (0.095)</td>
<td>0.014</td>
<td>0.810 (0.092)</td>
<td>0.013</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>-0.8</td>
<td>0.807 (0.085)</td>
<td>0.008</td>
<td>-0.797 (0.148)</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.8</td>
<td>0.808 (0.086)</td>
<td>0.009</td>
<td>0.814 (0.118)</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>-0.8</td>
<td>0.8</td>
<td>-0.779 (0.209)</td>
<td>0.026</td>
<td>-0.775 (0.139)</td>
<td>0.032</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.8</td>
<td>-0.778 (0.204)</td>
<td>0.028</td>
<td>0.897 (0.488)</td>
<td>0.122</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>-0.8</td>
<td>0.839 (0.162)</td>
<td>0.048</td>
<td>-0.792 (0.179)</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.8</td>
<td>0.819 (0.154)</td>
<td>0.024</td>
<td>0.863 (0.301)</td>
<td>0.079</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>-0.8</td>
<td>0.8</td>
<td>-0.799 (0.028)</td>
<td>0.001</td>
<td>-0.800 (0.038)</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.8</td>
<td>-0.801 (0.028)</td>
<td>-0.001</td>
<td>0.801 (0.028)</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>-0.8</td>
<td>0.802 (0.028)</td>
<td>0.002</td>
<td>-0.802 (0.047)</td>
<td>-0.003</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.8</td>
<td>0.801 (0.026)</td>
<td>0.002</td>
<td>0.802 (0.038)</td>
<td>0.002</td>
<td></td>
</tr>
</tbody>
</table>

Plotting the estimated parameters against each other (Fig. 3.3) reveals that the estimates tend to be quite correlated with each other in general (see Table 3.4 for corresponding correlation values). Moreover, the correlation increases when \( p \) increases except in the case with \((α, β) = (0.8, -0.8)\) where correlation becomes lower. We can see that all scatter plots are centred at the true values in all cases which, in light of Table 3.3, is to be expected. Similarly, the precision of the estimates, and how it depends on \( n \) and \( p \), can be clearly visualised from these scatter plots.
3.6. SOME SIMULATION STUDIES

Figure 3.3. Scatter plots of estimated regression coefficients for each scenario with least squares line (red) overlayed. Reference lines (grey) at the true parameter values are also shown.

Table 3.4. Correlation Between Estimates

<table>
<thead>
<tr>
<th>p</th>
<th>n</th>
<th>α</th>
<th>β</th>
<th>corr(\hat{\alpha}, \hat{\beta})</th>
</tr>
</thead>
<tbody>
<tr>
<td>20%</td>
<td>100</td>
<td>-0.8</td>
<td>-0.8</td>
<td>-0.78 -0.77</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.8</td>
<td>0.8</td>
<td>0.58 0.55</td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>0.8</td>
<td>-0.8</td>
<td>-0.62 -0.67</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.8</td>
<td>0.8</td>
<td>0.39 0.41</td>
</tr>
<tr>
<td>80%</td>
<td>100</td>
<td>-0.8</td>
<td>-0.8</td>
<td>0.91 0.92</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.8</td>
<td>0.8</td>
<td>0.98 0.99</td>
</tr>
<tr>
<td></td>
<td>80%</td>
<td>0.8</td>
<td>-0.8</td>
<td>-0.10 -0.03</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.8</td>
<td>0.8</td>
<td>0.81 0.80</td>
</tr>
</tbody>
</table>
When parameter estimates are highly correlated, the stability of the estimation procedure, and reliability of such estimates, may be of concern (Lee & Whitmore, 2006). However, in our simulation work, the models fitted the simulated data without issue (using the Newton-Raphson method) despite the fact that the parameter estimates are correlated. Furthermore, the estimates are unbiased (as seen in Table 3.3) and we can recover quantities of interest such as the survivor or hazard function (Fig. 3.1). Thus, we conclude that, even in cases where estimated parameters are correlated or information is low, the estimation procedure remains stable.

3.6.4 Delta Method Versus M.L.E. Simulation

Here we compare our proposed method of m.l.e. simulation with the classical delta method (both described in Section 2.3.2). In Example 2.2 we showed a case where the resulting confidence intervals are numerically very close. From that we may conjecture that the coverage provided by either method should also be quite similar\(^1\). We investigate this now, more formally, by simulation. We will produce confidence intervals for two quantities, namely: (i) the median, \(t_{0.5}\), and (ii) \(S(t_{0.99})\), i.e., the survivor function evaluated at the 99th percentile. Note that the value of the median depends on the \((\alpha, \beta)\) combination whereas \(S(t_{0.99}) \equiv 0.01\).

The confidence intervals are based on fitting the Weibull model (the true model) to the simulated data. In the case of the median (i.e., \(g(\hat{\theta}) = t_{0.5} = [\exp(-\beta) \log 2]^{\exp(-\alpha)}\)), the estimated variance is given by \(\hat{\sigma}^2_{t_{0.5}} = \nabla_{\theta}t_{0.5}^T \hat{\Sigma} \nabla_{\theta}t_{0.5}\) where \(\nabla_{\theta}t_{0.5} = [-t_{0.5} \log t_{0.5}, -\exp(-\alpha)t_{0.5}]^T\), \(\hat{t}_{0.5} = g(\hat{\theta})\) and \(\hat{\Sigma} = I^{-1}_o(\hat{\theta})\). Thus, the delta method gives the 95% confidence interval

\[
\hat{t}_{0.5} \pm 1.96\hat{\sigma}_{t_{0.5}}.
\]

Given that the median is constrained to be positive, we also applied the transformation \(g^*(\cdot) = \log(\cdot) \Rightarrow \log t_{0.5} \in (-\infty, \infty)\). Using the fact that \((\log x)' = 1/x\), we get \(\hat{\sigma}^2_{\log t_{0.5}} = \hat{\sigma}^2_{t_{0.5}}/(\hat{t}_{0.5})^2\) and the 95% confidence interval for \(\log t_{0.5}\) is then given by \(\log \hat{t}_{0.5} \pm 1.96\hat{\sigma}_{\log t_{0.5}}\). Back-transforming gives the

\(^1\)Of course this does not guarantee that either method attains the specified level of confidence.
3.6. SOME SIMULATION STUDIES

transformed delta method 95% confidence interval

\[ \hat{t}_{0.5} \exp\left( \pm \hat{\sigma}_{t_{0.5}} / \hat{t}_{0.5} \right), \]  

which is constrained to lie in \([0, \infty)\). In the case of \(S(t_{0.99})\), confidence intervals, produced using the delta method and transformed delta method, can be calculated using the algebra in Example 2.2.

The method of m.l.e. simulation (with 1000 replicates) provides us with samples \(\{\hat{t}_{0.5}^{(1)}, \ldots, \hat{t}_{0.5}^{(1000)}\}\) and \(\{\hat{S}(t_{0.99})^{(1)}, \ldots, \hat{S}(t_{0.99})^{(1000)}\}\). Confidence intervals for \(t_{0.5}\) and \(S(t_{0.99})\) are then formed by selecting percentiles 2.5 and 97.5 from each sample; these confidence intervals will always lie within the allowable range of each quantity.

For each of the 16 simulation scenarios, confidence intervals for \(t_{0.5}\) and \(S(t_{0.99})\) were constructed using the delta method, transformed delta method and m.l.e. simulation as described above. The coverage percentage associated with each method was then calculated as

\[
\left[ \frac{1}{1000} \sum_{j=1}^{1000} 1(l_j \leq g(\theta) \leq u_j) \right] \times 100,
\]

where \(g(\theta)\) is one of the quantities of interest (either \(t_{0.5}\) or \(S(t_{0.99})\)) and \([l_j, u_j]\) is the 95% confidence interval for \(g(\theta)\) in the \(j\)th repetition of the particular simulation scenario. The results are given in Table 3.5.

With respect to the confidence intervals for \(t_{0.5}\), all methods have very similar coverage levels. The only exceptions to this are the cases where \((n, p, \alpha, \beta) = (100, 80\%, -0.8, -0.8)\) and \((n, p, \alpha, \beta) = (100, 80\%, -0.8, 0.8)\). It is noteworthy that, for all methods, coverage is almost uniformly higher than the stated 95%. While one may wish to address this issue, it is not our intention to do so. We only aim to show that the method of m.l.e. simulation produces results similar to more conventional methods without the burden of calculating/programming analytic derivatives. Turning to the confidence intervals for \(S(t_{0.99})\) we can see that those constructed using the untransformed delta method have very bad properties when the level of information is lower, i.e., coverage much lower than 95%, whereas the transformed delta method and m.l.e. simulation have better coverage.
#### Table 3.5. Coverage Percentage for Confidence Intervals Calculated Using the Delta Method, Transformed Delta Method and M.L.E. Simulation

<table>
<thead>
<tr>
<th>n</th>
<th>100</th>
<th>1000</th>
<th>20%</th>
<th>80%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Delta</td>
<td>Delta g*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Coverage of 95% C.I.s for $t_{0.5}$</td>
<td>Coverage of 95% C.I.s for $S(t_{0.99})$</td>
</tr>
<tr>
<td>$p$</td>
<td>$\alpha$</td>
<td>$\beta$</td>
<td>Delta</td>
<td>Delta g*</td>
</tr>
<tr>
<td>-0.8</td>
<td>-0.8</td>
<td>97.3</td>
<td>98.4</td>
<td>98.9</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>98.1</td>
<td>98.4</td>
<td>98.5</td>
</tr>
<tr>
<td>0.8</td>
<td>-0.8</td>
<td>93.6</td>
<td>94.0</td>
<td>93.5</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>95.3</td>
<td>95.0</td>
<td>95.7</td>
</tr>
<tr>
<td>-0.8</td>
<td>-0.8</td>
<td>93.0</td>
<td>97.4</td>
<td>99.0</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>93.4</td>
<td>98.0</td>
<td>99.5</td>
</tr>
<tr>
<td>0.8</td>
<td>-0.8</td>
<td>96.4</td>
<td>97.1</td>
<td>98.0</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>97.3</td>
<td>97.4</td>
<td>98.2</td>
</tr>
</tbody>
</table>

We now look at some of the individual confidence intervals produced using the three methods for the two quantities of interest. In particular, Fig. 3.4 shows confidence intervals for the first 50 repetitions of the first scenario, $(n, p, \alpha, \beta) = (100, 20\%, -0.8, -0.8)$, where $t_{0.5} \approx 2.6$ and, of course, $S(t_{0.99}) = 0.01$. We can see that all confidence intervals for $t_{0.5}$ are numerically quite close. The defect in the confidence intervals for $S(t_{0.99})$ produced
using the untransformed delta method is clear: they extend below zero and are lower than those from the other two methods. This explains the results in Table 3.5.

Figure 3.4. Confidence intervals for the first 50 repetitions of the first simulation scenario, \((n, p, \alpha, \beta) = (100, 20\%, -0.8, -0.8)\). Horizontal reference lines are shown at the true values.

This study shows that m.l.e. simulation produces results which are comparable to standard methods. In particular, the confidence intervals are numerically closer to the transformed delta method; both methods respect the allowable range of \(g(\theta)\) which is advantageous. Thus, we conclude that m.l.e. simulation is a useful and easily applied method for constructing confidence intervals for functions of the parameters.
3.6.5 Estimating the Cured Proportion

In this simulation study we investigate the properties of estimating the proportion of cured individuals in a dataset. Thus, we simulated data from a Gompertz model with parameter vector $\theta = (\alpha = \gamma, \beta = \log \lambda)^T$. However, rather than setting $\alpha$ and $\beta$ values, we set $p_{\text{cure}}$ and $\beta$. The implied $\alpha$ value is then given by

$$
\alpha = \frac{\exp(\beta)}{\log p_{\text{cure}}},
$$

as $p_{\text{cure}} = S(\infty) = \exp \left[ \exp(\beta)/\alpha \right]$ for the Gompertz model.

In total there were $2^4 = 16$ simulation scenarios arising from all combinations of values shown in Table 3.6. Each of the 16 scenarios was repeated 1000 times. At each repetition we fitted a Gompertz model (i.e., the true model) to the simulated data from which we estimated the cured proportion via $\hat{p}_{\text{cure}} = \exp[\exp(\hat{\beta})/\hat{\alpha}]$. Furthermore, the average estimate, relative bias and standard error were calculated for each scenario (analogous to (3.13), (3.14) and (3.15) respectively).

The results are shown in Table 3.7 and are very clear. When the censored proportion of non-cured individuals is low ($p^* = 20\%$), $\hat{p}_{\text{cure}}$ is unbiased with a small standard error. However, when $p^*$ is higher ($= 80\%$), the estimation of $p_{\text{cure}}$ becomes more unstable, i.e., we observe bias (which is worse when $p_{\text{cure}}$ is higher) and large standard errors. Furthermore, the bias and large standard errors persist even at larger sample sizes. In other words, it is hard to recover the true value of $p_{\text{cure}}$ when the censored proportion (of the non-cured individuals) is high. Note that $p_{\text{cure}}$ is a tail probability and it is well known that the presence of (high) censoring obscures our ability to estimate the tail reliably as it is not observed in the data. Clearly, we are

<table>
<thead>
<tr>
<th>Table 3.6. Simulation Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
</tr>
<tr>
<td>Values</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>1000</td>
</tr>
</tbody>
</table>

Note: Recall that, for cure rate data, the censored proportion applies to the non-cured individuals which we denote by $p^*$ (see Section 3.5).
extrapolating beyond the data (i.e., to time infinity) when estimating the cured proportion and, of course, extrapolation is always advised against (as discussed in Section 3.6.2).

Table 3.7. Estimates of Cured Proportion for Gompertz Data

<table>
<thead>
<tr>
<th>$p^*$</th>
<th>$p_{cure}$</th>
<th>$\beta$</th>
<th>$\hat{p}_{cure}$</th>
<th>S.E.</th>
<th>R.Bias</th>
<th>$\hat{p}_{cure}$</th>
<th>S.E.</th>
<th>R.Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>20%</td>
<td>-0.8</td>
<td>0.299</td>
<td>(0.061)</td>
<td>-0.005</td>
<td>0.300</td>
<td>(0.019)</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>60%</td>
<td>-0.8</td>
<td>0.596</td>
<td>(0.062)</td>
<td>-0.006</td>
<td>0.599</td>
<td>(0.018)</td>
<td>-0.001</td>
<td></td>
</tr>
<tr>
<td>80%</td>
<td>-0.8</td>
<td>0.600</td>
<td>(0.058)</td>
<td>0.000</td>
<td>0.600</td>
<td>(0.019)</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

| 20%   | 0.8       | 0.298   | (0.061)          | -0.005 | 0.300 | (0.019)          | 0.000 |        |
| 60%   | 0.8       | 0.596   | (0.062)          | -0.006 | 0.599 | (0.018)          | -0.001 |        |
| 80%   | 0.8       | 0.596   | (0.062)          | -0.006 | 0.599 | (0.018)          | -0.001 |        |

| 30%   | -0.8      | 0.322   | (0.303)          | 0.074  | 0.280 | (0.180)          | -0.067 |        |
| 80%   | 0.8       | 0.309   | (0.297)          | 0.030  | 0.289 | (0.183)          | -0.037 |        |
| 60%   | -0.8      | 0.467   | (0.371)          | -0.221 | 0.519 | (0.237)          | -0.134 |        |
| 80%   | 0.8       | 0.504   | (0.364)          | -0.159 | 0.522 | (0.229)          | -0.130 |        |

The numerical findings of this simulation study are in line with previous anecdotal evidence in the literature. Farewell (1982) found that “for some data sets the likelihood function ... may be quite flat (over a range of $p_{cure}$ values).” Based on our simulation work, it seems very likely that these data sets contained a large proportion of censored individuals. Indeed, Lawless (2003, p. 183) noted that “in settings where censoring is heavy ... an imprecise estimate of ($p_{cure}$ is observed).” Lambert et al. (2007) also discuss the instability of cure proportion estimates.

### 3.6.6 Decreasing Hazards or Cured Proportion?

We continue with our investigation of cure rate modelling. In particular, for the Gompertz model (and also the time-dependent logistic model), a negative value of the shape parameter implies both a decreasing hazard and a cured proportion. Of course, this is plausible in itself: a proportion of individuals never experience an event and, hence, the hazard decreases over time. How-
ever, in other models (e.g., the Weibull) a decreasing hazard does not imply a cured proportion. In these models events occur less frequently over time but eventually every individual experiences an event. This suggests that, if the latter case is true, a Gompertz model will imply a cured proportion when, in fact, it does not exist.

In order to investigate the above conjecture, we fitted the Gompertz model to the simulated Weibull data (described in Table 3.1) and estimated the cured proportion. For each scenario, we calculated the average estimate and the standard error (over 1000 replications). The results are shown in Table 3.8.

<table>
<thead>
<tr>
<th></th>
<th>$p$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\hat{p}_{cure}$</th>
<th>S.E.</th>
<th>$\hat{p}_{cure}$</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>20%</td>
<td>-0.8</td>
<td>-0.8</td>
<td>0.8</td>
<td>0.077 (0.022)</td>
<td></td>
<td>0.075 (0.007)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td></td>
<td></td>
<td>0.075 (0.023)</td>
<td></td>
<td>0.075 (0.007)</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>-0.8</td>
<td>0.8</td>
<td></td>
<td>0.000 (0.000)</td>
<td></td>
<td>0.000 (0.000)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td></td>
<td></td>
<td>0.000 (0.000)</td>
<td></td>
<td>0.000 (0.000)</td>
<td></td>
</tr>
<tr>
<td>80%</td>
<td>-0.8</td>
<td>-0.8</td>
<td>0.8</td>
<td>0.653 (0.101)</td>
<td></td>
<td>0.667 (0.023)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td></td>
<td></td>
<td>0.654 (0.098)</td>
<td></td>
<td>0.667 (0.023)</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>-0.8</td>
<td>0.8</td>
<td></td>
<td>0.000 (0.000)</td>
<td></td>
<td>0.000 (0.000)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td></td>
<td></td>
<td>0.000 (0.000)</td>
<td></td>
<td>0.000 (0.000)</td>
<td></td>
</tr>
</tbody>
</table>

The findings are as expected. In order for the Gompertz model to fit the situation of decreasing hazards, it must also imply a cure rate. Thus, whenever $\alpha = -0.8$ for the Weibull model, the fitted Gompertz model estimates a non-zero cure probability (the value of which is very high when censoring is high). Note that the standard errors are small (especially when $n = 1000$) meaning that model consistently (but incorrectly) assigns a cured proportion to the data.

The Gompertz model is constrained in the sense that the cured proportion and the shape of the hazard are controlled by the same parameter (this
also applies to the time-dependent logistic model). Thus, the classical cure mixture model (Boag, 1949; Berkson & Gage, 1952; Farewell, 1977, 1982, 1986; Maller & Zhou, 1995) which is defined by

\[ S(t) = (1 - p_{\text{cure}})S_0(t) + p_{\text{cure}}, \]  

(3.19)

may seem preferable as \( p_{\text{cure}} \) is a free parameter. Here \( S_0(t) \) is the survivor curve for the non-cured individuals, for example, if \( S_0(t) \) is a Weibull survivor curve then the model has three parameters and can, in theory, distinguish between the two types of decreasing hazard mentioned above (i.e., decreasing due to cure or simply decreasing). However, Farewell (1986) noted that “there is a degree of non-identifiability” between \( p_{\text{cure}} \) and the shape parameters of \( S_0(t) \) in this mixture model for some data sets (i.e., when censoring rate is high). Thus, even though the Weibull mixture model has three parameters, it can behave as if it has two when the information is low. This is not surprising in light of the difficulties we found in estimating \( p_{\text{cure}} \) reliably (when it does exist) in low information cases (Section 3.6.5).

It is clear from the above, and Section 3.6.5, that one should exercise some caution when reporting estimates of the cured proportion. Moreover, Farewell (1986) advised caution on the grounds of scientific plausibility also, stating that estimates of cured proportions may be indefensible in certain cases. However, this does not imply that cure rate models cannot be used in these cases. The issues that we have discussed arise solely due to extrapolation beyond the data. Just because the survivor curve has an asymptote as \( t \to \infty \), does not mean we have to report this value. For example, Lambert et al. (2007) suggested that “even when it is not reasonable to assume a cure fraction . . . the models may still fit the data well for the observed follow-up period”. Indeed, as is clear from Table 3.8, the Gompertz model recognises when the true (Weibull) hazard is decreasing and when it is increasing. Thus, the fitted Gompertz model may explain the data quite well, and produce reasonable estimates, within the observed time range. Furthermore, Perperoglou et al. (2007) found that a cure rate analysis resulted in conclusions similar to that of other (non-cure) analyses (in terms of estimated survival probabilities within the observed time range).
Chapter 4

Regression Models

4.1 Introduction

In the methods considered thus far, we have not been concerned with determining covariate effects (although we have alluded to covariate analyses in Chapter 2 and simulation from regression models in Chapter 3). While the methods considered up to now are useful for determining quantities such as the population survivor curve, hazard function or percentiles, data generally do not arise from a homogeneous population. Individual survival times typically depend on a variety of features which cause heterogeneity in the data. Indeed, interest usually lies in explaining this heterogeneity through the incorporation of covariates, or explanatory variables, in our analysis. Hence the need to generalise the methods of the previous sections to include covariates. As is standard in statistical modelling, regression approaches are used to facilitate this requirement. The advantage of regression models is that they allow us to determine if a particular covariate influences the (survival) process and to measure, via the corresponding regression coefficient, the magnitude of this influence while adjusting for the other covariates present in the model.

Given the central role played by the hazard function in survival analysis (as described in Sections 1.2.2 and 1.2.3), it is natural to consider the effect of covariates via \( \lambda(t \mid x) \), where \( \lambda(\cdot) \) is the hazard function and \( x = (x_1, \ldots, x_p)^T \) is a vector of measured covariates. Thus, we must specify how \( \lambda(t \mid x) \) depends on \( x \).
4.2 The Proportional Hazards Model

The most popular regression model in survival analysis is the so-called proportional hazards, PH, model. This model assumes that covariate effects act multiplicatively on a baseline hazard function, i.e.,

$$\lambda(t \mid x) = \phi(x)\lambda_0(t), \quad \text{(4.1)}$$

where $\phi(x)$ is a proportionality factor which depends on the vector of covariates, $x = (1, x_1, \ldots, x_p)^T$, and the corresponding regression coefficients $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T$. The function $\lambda_0(t)$ is the baseline hazard function common to all individuals. Thus, the effect of covariates is to increase or decrease the overall magnitude of the hazard according to $\phi(x)$ being greater than or less than one respectively. At a given time, $t$, the ratio of the hazards for two individuals, with covariate vectors $x_i$ and $x_j$ respectively, is given by

$$\psi(x_i, x_j) = \frac{\lambda(t \mid x_i)}{\lambda(t \mid x_j)} = \frac{\phi(x_i)\lambda_0(t)}{\phi(x_j)\lambda_0(t)}$$

$$= \frac{\phi(x_i)}{\phi(x_j)}, \quad \text{(4.2)}$$

which is a constant - in the sense that it does not depend on $t$ - and hence the hazards are proportional.

Typically, covariates enter $\phi$ through a log-link to ensure positivity of hazards:

$$\log \phi = \beta_0 + x_1\beta_1 + \ldots + x_p\beta_p$$

$$= x^T \beta,$$

$$\Rightarrow \phi = \exp(x^T \beta).$$

Other functional forms for $\phi$ are possible (see Fiegel & Zelen (1965) and Hudec & Platz (1983)) but will not be considered here. Thus, the hazard function is

$$\lambda(t \mid x) = \exp(x^T \beta)\lambda_0(t), \quad \text{(4.3)}$$
4.2. THE PROPORTIONAL HAZARDS MODEL

and the hazard ratio is

$$
\psi(x_i, x_j) = \frac{\exp(x_i^T \beta)}{\exp(x_j^T \beta)} = \exp[(x_i - x_j)^T \beta]. \tag{4.4}
$$

Assuming, without the loss of generality, that all covariates are equal apart from the first, $x_1$, gives

$$
\psi(s) = \exp[(x_1 - x_1)\beta_1] = \exp(s\beta_1),
\Rightarrow \psi(1) = \exp(\beta_1). \tag{4.5}
$$

Thus, exponentiating a $\beta$ coefficient gives the factor by which the hazard changes for a one unit increase in the corresponding $x$, assuming all other covariates are equal, e.g., “the hazard (or risk) in group A is $\exp(\beta)$ times that of group B”. This simple relative risk interpretation is quite appealing which explains the popularity of the proportional hazards model.

Given that the hazard function for the PH model is (4.3), we can derive the other survival functions using the theory outlined in Section 1.2.3. The cumulative hazard function is

$$
\Lambda(t \mid x) = \int_0^t \lambda(u \mid x)du = \exp(x^T \beta) \int_0^t \lambda_0(u)du = \exp(x^T \beta) \Lambda_0(t), \tag{4.6}
$$

and the survivor function is

$$
S(t \mid x) = \exp[-\Lambda(t \mid x)] = \exp[-\exp(x^T \beta)\Lambda_0(t)] = S_0(t)^{-\exp(x^T \beta)}. \tag{4.7}
$$

The baseline hazard function, $\lambda_0(t)$, may be given a parametric specification, leading to a fully parametric PH model, or may be a non-parametric
function. In the latter case the model is *semi-parametric* as it comprises a non-parametric component, \( \lambda_0(t) \), and a parametric component, \( \exp(x^T \beta) \).

Cox (1972) introduced the semi-parametric PH model, typically referred to as “the Cox model”, which we will describe briefly in Section 4.4. First we will highlight some other regression specifications.

### 4.3 Other Regression Models

The proportional hazards model, although the most popular, is but one way of incorporating covariates in survival analysis. Some other possibilities are

\[
\begin{align*}
\text{Accelerated Failure Time:} & \quad \lambda(t \mid x) = \exp(x^T \beta) \lambda_0[\exp(x^T \beta)t], \\
& \quad S(t \mid x) = S_0[\exp(x^T \beta)t], \quad (4.8) \\
\text{Additive Hazards:} & \quad \lambda(t \mid x) = \lambda_0(t) + x^T \beta, \quad (4.9) \\
\text{Proportional Odds:} & \quad \frac{S(t \mid x)}{1 - S(t \mid x)} = \exp(x^T \beta) \frac{S_0(t)}{1 - S_0(t)}, \quad (4.10)
\end{align*}
\]

where the baseline function (indicated by a subscript 0) may be parametric or non-parametric.

The accelerated failure time model is commonly used in an engineering setting where \( \lambda_0(t) \) is typically a parametric function (Lawless, 2003, chap. 6). Semi-parametric accelerated failure time models have also been developed but are less popular (Buckley & James, 1979; Ying, 1993; Jin et al., 2003). Details on the additive hazard model can be found in Aalen (1989), McKeague & Sasieni (1994) and Aalen et al. (2008, chap. 4). The proportional odds model links covariates to the baseline survivor function, rather than the hazard function, and has been considered by Bennett (1983) and Pettitt (1984). It is worth noting that all of the above regression models, (4.3) and (4.8) – (4.10), are linear on some scale (summarised in Table 4.1).
4.4 Cox’s Semi-Parametric Proportional Hazards Model

Although our approach to the analysis of survival data will mainly be parametric, Cox’s model warrants specific mention due to its popularity. Indeed this model has become the standard method for analysing survival data, is widely available in statistical software and routinely used. Cox (1972) introduced an estimation procedure which allows the $\beta$ coefficients to be estimated while leaving $\lambda_0(t)$ completely unspecified using the likelihood function

$$L(\beta) = \prod_{j=1}^{r} \frac{\exp(x_{(j)}^T\beta)}{\sum_{k \in R(t_{(j)})} \exp(x_{k}^T\beta)},$$

(4.11)

where the subscript $(j)$ denotes the $j$th of $r$ ordered event times in a sample (see Section 1.3 for notation) and $R(t_{(j)})$ is the set of individuals at risk just prior to $t_{(j)}$, i.e., those currently uncensored and yet to experience an event just before $t_{(j)}$. This likelihood function does not arise in the usual way (as a product of probability densities), rather it is a partial likelihood: an object formally defined in Cox (1975) which ignores some information from the full likelihood. Nonetheless, for the purposes of model fitting and inference, standard likelihood theory (described in Chapter 2) can be applied to this partial likelihood (Andersen & Gill, 1982).

It is clear from (4.11) that only individuals who experience events contribute to the numerator. Moreover, only one individual can contribute to this numerator at each event time via the term $\exp(x_{(j)}^T\beta)$. Therefore, (4.11) cannot be used if any events occur at the same time, known as tied events\(^1\).

\(^1\)Censoring times being tied with each other or with event times present no difficulties. Their contributions appear in the denominator of (4.11) irrespective of this.
CHAPTER 4. REGRESSION MODELS

One way to deal with ties is to artificially break them. Let's assume that two event times occur at time $t_{(j)}$ which we denote by $t_{(j_1)}$ and $t_{(j_2)}$. We can define two new variables, $t'_{(j_1)} = t_{(j_1)} + \epsilon_1$ and $t'_{(j_2)} + \epsilon_2$, where $\epsilon_1$ and $\epsilon_2$ are small random disturbances, e.g., $\epsilon_1, \epsilon_2 \sim N(0, \sigma^2)$ with $\sigma^2$ small. We can then proceed with (4.11) as the times are now distinct. However, the most popular approach is to treat the times as distinct without perturbation (Breslow, 1974). Thus the product of contributions at time $t_{(j)}$ is

$$\frac{\exp(x_{(j_1)}^T \beta)}{\sum_{k \in R(t_{(j_1)})} \exp(x_k^T \beta)} \cdot \frac{\exp(x_{(j_2)}^T \beta)}{\sum_{k \in R(t_{(j_2)})} \exp(x_k^T \beta)} = \frac{\exp[(x_{(j_1)} + x_{(j_2)})^T \beta]}{\left[\sum_{k \in R(t_{(j)})} \exp(x_k^T \beta)\right]^2}.$$ 

More generally, Breslow’s likelihood function is

$$L(\beta) = \prod_{j=1}^r \frac{\exp(s_{(j)}^T \beta)}{\left[\sum_{k \in R(t_{(j)})} \exp(x_k^T \beta)\right]^{d_j}}, \quad (4.12)$$

where $d_j$ is the number of events at time $t_{(j)}$ and $s_{(j)}$ is the sum of covariate vectors from these $d_j$ individuals, i.e., $s_{(j)} = \sum_{l=1}^{d_j} x_{(j_l)}$ where $x_{(j_l)}$ is the covariate vector for the $l$th individual at the $j$th event time.\(^2\) Other methods for handling ties are available and these are discussed in Therneau & Grambsch (2000, sec. 3.3) and Kalbfleisch & Prentice (2002, sec. 4.2.3).

After obtaining $\hat{\beta}$ values by maximising (4.12), one can then estimate the baseline cumulative hazard function, $\Lambda_0(t)$, using the Breslow estimator (Breslow, 1972, 1974)\(^3\). This is given by

$$\hat{\Lambda}_0(t) = \sum_{j \mid t \geq t_{(j)}} \frac{d_j}{\sum_{k \in R(t_{(j)})} \exp(x_k^T \beta)}, \quad (4.13)$$

and can be motivated either as an approximation to a maximum likelihood estimator proposed by Kalbfleisch & Prentice (1973) or using martingale theory (Andersen & Gill, 1982). Given the discrete nature of (4.13), one may formulate an estimator for $S_0(t)$ based on product integration (see Sections

\(^2\)Note that (4.12) reduces to (4.11) in the absence of tied event times.

\(^3\)It is also known as the Nelson-Aalen estimator, as it reduces to (1.24) when $\hat{\beta} = 0$.\)
1.2.3 and 1.3.4). However, in this context, most authors suggest the use of

\[ \hat{S}_0(t) = \exp[-\hat{\Lambda}_0(t)] \]

\[ = \prod_{j | t \geq t(j)} \exp \left[ -\frac{d_j}{\sum_{k \in R(t(j))} \exp(x_k^T \beta)} \right] \]  \hspace{1cm} (4.14)

as an estimator for \( S_0(t) \). Using (4.6) and (4.7) we have

\[ \hat{\Lambda}(t \mid x) = \exp(x^T \beta) \hat{\Lambda}_0(t), \]  \hspace{1cm} (4.15)

and

\[ \hat{S}(t \mid x) = \hat{S}_0(t)^{\exp(x^T \beta)}. \]  \hspace{1cm} (4.16)

Note that one could also estimate \( \lambda_0(t) \), and hence \( \lambda_0(t \mid x) \), by differencing \( \hat{\Lambda}_0(t) \) but this estimator is inconsistent (Burr, 1994); consistency can be achieved via the use of splines (Whittemore & Keller, 1986; Rosenberg, 1995) or kernel smoothers (Ramlau-Hansen, 1983; Tanner & Wong, 1984).

The advantage of the Cox model is that we can retrieve estimates of the covariate effects without specifying the underlying baseline distribution. On the surface this seems a very flexible approach as we avoid parametric assumptions about the distribution. However, we are making the assumption that covariates enter the hazard through a parametric multiplicative component (4.3). This assumption that hazards are proportional is a strong one in itself. Also, because Cox’s framework allows us, and in fact encourages us, to estimate the covariate effects (using (4.11)) without worrying about the underlying distributional functions, many analyses are carried out without thinking about these functions and the insights they may provide.

**Example 4.1. Cox Analysis of Lung Cancer Data**

We fitted the Cox model to the lung cancer dataset using only the treatment covariate which has five levels, namely: palliative care (the reference category), surgery, chemotherapy, radiotherapy and chemotherapy + radiotherapy combined. The results are summarised in Table 4.2 where the estimated coefficients come from maximising (4.12) and the standard errors are obtained using standard likelihood theory (see Chapter 2). The signs of the coefficients are all negative implying that all treatments reduce the hazard
relative to palliative care; this reduction is statistically significant in all cases ($|Z| > 1.96$). As outlined in Section 4.2, we can interpret the coefficients in terms of hazard ratios through exponentiation. Thus, an individual receiving surgery has $\exp(-2.18) = 0.11$ times the risk of death compared with an individual receiving palliative care. Similarly, the hazard ratios for the other three treatments (relative to palliative care) are 0.68, 0.57 and 0.43 respectively.

We note that while the proportional hazards model provides us with ease

| Treatment | $\hat{\beta}$ | S.E. | $|Z|$ |
|-----------|-------------|------|------|
| Palliative | 0.00 | —— | —— |
| Surgery   | -2.18 (0.23) | 9.68 |
| Chemo     | -0.38 (0.17) | 2.23 |
| Radio     | -0.55 (0.09) | 6.34 |
| C+R       | -0.85 (0.20) | 4.18 |

Figure 4.1. Kaplan-Meier (solid) curves for each treatment group with Cox model (dash) fitted curves overlayed.
of interpretation, appropriate use of the model depends on the validity of the proportional hazards assumption in (4.3). The estimated Cox survivor curves for each treatment group are shown in Figure 4.1. The corresponding Kaplan-Meier curves, which are the closest to the data by definition, are also shown for comparison. The Cox model’s simplifying assumption of proportionality approximates reality but may be questionable in this case.

4.5 Parametric Models

As mentioned in Section 1.4, parametric distributions have the advantage of offering mathematical formulae to describe the survivor/hazard functions or quantities such as mean/median survival time. When we generalise to parametric regression models, we can investigate the effects of covariates on any of these functions/quantities. Of course we must understand the capabilities of each parametric distribution and the interpretation of parameters but, when equipped with this knowledge, we can gain many insights into the underlying survival process. In addition to this, parametric regression models are fitted to data using the classical (full) likelihood function which is well understood and where the presence of tied event times presents no difficulty. The theory in Chapter 2 holds for parametric regression models but the likelihood function, where individuals are now assumed conditionally independent given covariates, generalises (2.7) and is given by

$$\ell(\theta) = \sum_{i=1}^{n} \delta_i \log \lambda(t_i \mid x_i) + \log S(t_i \mid x_i), \quad (4.17)$$

where $x_i = (x_{i1}, \ldots, x_{ip})^T$ is the vector of covariates for the $i$th individual and $\theta$ now contains regression coefficients in addition to distributional parameters.

In order to generate such a parametric regression model, one could first start with the regression specification (from Sections 4.2 and 4.3) and then select a parametric baseline function, e.g., one may choose a proportional hazards specification with $\lambda_0(t) = \lambda \gamma t^{\gamma-1}$ giving a Weibull PH model. The approach we will take however is to generalise the parametric distributions
of Section 1.4 by allowing their distributional parameters to depend on covariates. The hazard regression is then implied by the chosen model. Our parameter naming scheme in Section 1.4 is such that, for each distribution, the parameter $\lambda$ corresponds to the standard “interest” parameter in survival literature. This is the scale parameter which controls the overall magnitude of the hazard. Thus, one arrives at a parametric regression model by letting

$$g(\lambda) = x^T \beta,$$  \hspace{1cm} (4.18)

where $g(\cdot)$ is an appropriate link function, $x = (1, x_1, \ldots, x_p)^T$ is the vector of covariates and $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T$ is the corresponding vector of regression coefficients. For most of the distributions in Section 1.4 the scale parameter is constrained to be positive and so, in these cases, $g(\cdot) = \log(\cdot)$. Note that one could also allow other distributional parameters to depend on covariates (simultaneously with $\lambda$). We will refer to such models, where multiple parameters depend on covariates, as multi-parameter regression (MPR) models. MPR models are not traditionally considered in survival analysis but are the main focus of this thesis. At this stage however, we will only discuss the standard single-parameter regression (SPR) models defined by (4.18).

Given the popularity of the proportional hazards model in survival, practitioners have become accustomed to hazard ratios as a means of summarising covariate effects. For this reason we give the functional form of the hazard ratio, which we denote by $\psi$, for each of the regression models below. Our discussion is brief as these SPR models are special cases of the MPR models which are developed in detail in Chapter 5. First we will introduce some notation. Without the loss of generality, let’s assume that we are interested in determining the effect of the (binary) covariate $x_1$ via its hazard ratio. In what follows, it is helpful to decompose the linear predictor $x^T \beta$ into two parts: the term involving $x_1$ and the remaining terms, i.e.,

$$x^T \beta = \beta_0 + x_1 \beta_1 + \ldots + x_p \beta_p$$

$$= x_1 \beta_1 + \beta_0 + \ldots + x_p \beta_p$$

$$= x_1 \beta_1 + \tilde{x}^T \beta,$$

where $\tilde{x} = (1, 0, x_2, \ldots, x_p)^T$. As we will see, the term $\tilde{x}^T \beta$ appears in the hazard ratio for some of the SPR models below. We defer our discussion of such terms until Section 5.3.
4.5. **PARAMETRIC MODELS**

4.5.1 **Weibull**

For the Weibull SPR model, \( \log(\lambda) = x^T \beta \). Thus, the hazard and survivor functions are given by

\[
\lambda(t \mid x) = \exp(x^T \beta) \gamma t^{\gamma-1},
\]

and

\[
S(t \mid x) = \exp[- \exp(x^T \beta)t^\gamma],
\]

where \( \beta \in \mathbb{R}^{p+1} \) and \( \gamma > 0 \).

At time \( t \), the hazard ratio for \( x_1 \) is given by

\[
\psi(t) = \frac{\lambda(t \mid x_1 = 1)}{\lambda(t \mid x_1 = 0)} = \exp(\beta_1),
\]

which is time constant and, therefore, we see that the Weibull SPR model is a PH model, i.e., it has the *proportional hazards property*.

4.5.2 **Gompertz**

For the Gompertz SPR model, \( \log(\lambda) = x^T \beta \). Thus, the hazard and survivor functions are given by

\[
\lambda(t \mid x) = \exp(x^T \beta) \exp(\gamma t)
\]

\[
= \exp(x^T \beta + \gamma t),
\]

and

\[
S(t \mid x) = \exp \left\{ - \frac{\exp(x^T \beta)}{\gamma} [\exp(\gamma t) - 1] \right\},
\]

where \( \beta \in \mathbb{R}^{p+1} \) and \( \gamma \in \mathbb{R} \).

At time \( t \), the hazard ratio for \( x_1 \) is given by

\[
\psi(t) = \exp(\beta_1).
\]

and we see that the Gompertz SPR model also has the proportional hazards property.
4.5.3 Log-logistic

The log-logistic SPR model also requires a log-link for the scale parameter. Thus, the hazard and survivor functions are given by

$$\lambda(t \mid x) = \frac{\exp(x^T \beta) t^{\gamma - 1}}{1 + \exp(x^T \beta) t^\gamma},$$  \hspace{1cm} (4.25)

and

$$S(t \mid x) = \frac{[1 + \exp(x^T \beta) t^\gamma]^{-1}}{\rho},$$ \hspace{1cm} (4.26)

where $\beta \in \mathbb{R}^{p+1}$ and $\gamma > 0$.

At time $t$, the hazard ratio for $x_1$ is given by

$$\psi(t) = \exp(\beta_1) \frac{1 + \exp(\tilde{x}^T \beta) t^\gamma}{1 + \exp(\beta_1 + \tilde{x}^T \beta) t^\gamma},$$ \hspace{1cm} (4.27)

which is time-dependent. Therefore, the log-logistic model may be a useful alternative to a PH model.

It is worth noting that $\lim_{t \to \infty} \psi(t) = 1$. Therefore, the log-logistic model naturally handles convergent hazards. The rate of convergence is controlled by the shape parameter $\gamma$. In practice this could arise, for example, where an initial treatment effect wears off over time.

4.5.4 Burr

The Burr SPR model again requires a log-link for the scale parameter. Thus, the hazard and survivor functions are given by

$$\lambda(t \mid x) = \frac{\exp(x^T \beta) t^{\gamma - 1}}{1 + \exp(x^T \beta) t^\gamma},$$ \hspace{1cm} (4.28)

and

$$S(t \mid x) = \frac{[1 + \exp(x^T \beta) t^\gamma]^{-1/\rho}}{\rho},$$ \hspace{1cm} (4.29)

where $\beta \in \mathbb{R}^{p+1}$ and $\gamma, \rho > 0$.

The hazard ratio, at time $t$, for $x_1$ is given by

$$\psi(t) = \exp(\beta_1) \frac{1 + \exp(\tilde{x}^T \beta) t^\gamma}{1 + \exp(\beta_1 + \tilde{x}^T \beta) t^\gamma}. $$ \hspace{1cm} (4.30)
This has a similar form to the log-logistic hazard ratio and, as in the log-logistic case, converges to one with time. However, the Burr hazard ratio has the additional property that \( \lim_{\rho \to 0} \psi(t) = \exp(\beta_1) \). This is a consequence of the fact that when \( \rho \to 0 \), the Burr model reduces to the Weibull model (see Section 1.4.5). Thus, the Burr SPR model unifies proportional hazards regression and non-proportional hazards regression making it a flexible parametric model which supersedes both the Weibull and log-logistic models. Note however that, because \( \rho \) is a constant, all covariate effects are either proportional on the hazard scale or non-proportional. This is something that we can relax by letting \( \rho \) depend on covariates, i.e., a multi-parameter regression Burr model. Such a model will allow some effects to be proportional and some non-proportional. This Burr MPR model will be very flexible indeed (see Section 5.2.4).

### 4.5.5 Time-Dependent Logistic

The scale parameter in the TDL distribution is unconstrained and therefore \( \lambda = x^T \beta \). The hazard and survivor functions are given by

\[
\lambda(t \mid x) = \frac{\exp(\gamma t + x^T \beta)}{1 + \exp(\gamma t + x^T \beta)}, \quad (4.31)
\]

and

\[
S(t \mid x) = \left[ \frac{1 + \exp(\gamma t + x^T \beta)}{1 + \exp(x^T \beta)} \right]^{-1/\gamma}, \quad (4.32)
\]

where \( \beta \in \mathbb{R}^{p+1} \) and \( \gamma \in \mathbb{R} \). It is worth noting that

\[
\text{logit} \lambda(t \mid x) = \log \left( \frac{\lambda(t \mid x)}{1 - \lambda(t \mid x)} \right) = \gamma t + x^T \beta,
\]

which is a generalisation of the standard logistic regression model to time-dependence. Hence the name “time-dependent logistic”.

The hazard ratio, at time \( t \), for \( x_1 \) is given by

\[
\psi(t) = \exp(\beta_1) \frac{1 + \exp(\gamma t + x^T \beta)}{1 + \exp(\gamma t + \beta_1 + x^T \beta)}. \quad (4.33)
\]
4.5.6 Piecewise Exponential

Covariates are introduced into the piecewise exponential model by assuming a proportional hazards specification, \( \lambda(t \mid x) = \exp(x^T \beta)\lambda_0(t) \). Recall that for the piecewise exponential distribution (Section 1.4.7) the time axis is split (not necessarily at event times) into intervals \( I_j = [t_{(j-1)}, t_{(j)}) \), \( j = 1 \ldots, m \), where \( t_{(0)} = 0 \) and \( t_{(m)} = \infty \). Thus,

\[
\lambda(t \mid x) = \exp(x^T \beta)a(t)^T \lambda, \quad (4.34)
\]

where \( x = (x_1, \ldots, x_p)^T \), \( \beta = (\beta_1, \ldots, \beta_p)^T \), \( \lambda = (\lambda_1, \ldots, \lambda_m)^T \) and \( a(t) \) is an \( m \)-dimensional vector indicating which interval \( t \) lies in, e.g., if \( m = 4 \) and \( t \in I_3 \) then \( a(t) = (0, 0, 1, 0)^T \). The survivor function is given by

\[
S(t) = \exp \left[ -\exp(x^T \beta)d(t)^T \lambda \right], \quad (4.35)
\]

where \( d(t) \) represents the time spent in each interval for a particular value of \( t \), e.g., if \( m = 4 \) and \( t \in I_3 \) then \( d(t) = (t_2 - t_0, t_3 - t_2, t - t_3, 0)^T \).

At time \( t \), the hazard ratio for \( x_1 \) is given by

\[
\psi(t) = \exp(\beta_1), \quad (4.36)
\]
as this is a PH model.

This regression model may also be viewed as letting the \( j \)th distributional parameter of the piecewise exponential model depend on covariates via

\[
\log \lambda_j = \beta_{0j} + x^T \beta, \quad (4.37)
\]
for \( j = 1, \ldots, m \). In other words, the intercept term varies in each time interval. It is also worth noting that if we split the time axis at each distinct event time (creating \( r + 1 \) intervals where \( r \) is the total number of events) we get the Cox model (Section 4.4). Thus, assuming \( m \ll r+1 \), the PH piecewise exponential model is a flexible parametric PH model of lower dimension than the semi-parametric model.

Example 4.2. Single Parameter Regression Analysis of Lung Cancer Data
We continue the analysis of the lung cancer data presented in Example 4.1. Here we fit the Weibull, log-logistic and TDL single-parameter regression
models to the data using the likelihood function given in (4.17). The estimated parameters for each model are summarised in Table 4.3 along with the maximised likelihood and $AIC$ values. The results are similar across models. The $\beta$ values are all negative indicating that all treatments reduce the hazard compared with palliative care. The models are in agreement as to the relative merit of each treatment; surgery offers the largest hazard reduction followed by the combined treatment (C+R), radiotherapy and finally chemotherapy.

<table>
<thead>
<tr>
<th>Weibull</th>
<th>Log-logistic</th>
<th>TDL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est. S.E.</td>
<td>Z</td>
</tr>
<tr>
<td>Shape $\gamma$</td>
<td>0.93 (0.03)</td>
<td>——</td>
</tr>
<tr>
<td>Intercept $\beta_0$</td>
<td>-1.48 (0.08)</td>
<td>——</td>
</tr>
<tr>
<td>Palliative —</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Surgery $\beta_1$</td>
<td>-2.22 (0.23)</td>
<td>9.89</td>
</tr>
<tr>
<td>Chemo $\beta_2$</td>
<td>-0.40 (0.17)</td>
<td>2.30</td>
</tr>
<tr>
<td>Radio $\beta_3$</td>
<td>-0.57 (0.09)</td>
<td>6.50</td>
</tr>
<tr>
<td>C+R $\beta_4$</td>
<td>-0.87 (0.20)</td>
<td>4.30</td>
</tr>
<tr>
<td>$\ell(\hat{\theta})$</td>
<td>-1960.75</td>
<td>——</td>
</tr>
<tr>
<td>$AIC$</td>
<td>3933.51</td>
<td>——</td>
</tr>
<tr>
<td>$\Delta AIC$</td>
<td>34.19</td>
<td>——</td>
</tr>
</tbody>
</table>

The estimated hazard, hazard ratio and survivor functions for each model are shown in Figure 4.2. Confidence intervals for the hazard ratios are given in Fig. 4.3 and have been calculated using m.l.e. simulation (Section 2.3.2). We see that the Weibull and TDL analyses are roughly in agreement, that is to say, in this instance, the TDL model produces similar results to the PH analysis even though it is a non-PH model. On the other hand, the log-logistic model does produce non-proportional hazards which imply that the effectiveness of each treatment diminishes with time.

The log-logistic model provides the best fit to the data based on a much lower $AIC$ value and the fact that the estimated survivor curves fit the Kaplan-Meier curves more closely. However, clearly none of the models provide an adequate fit to the C+R group. This is a result of the fact that they
lack the flexibility to model crossing hazards, due to the constant shape parameter. We can overcome this in the more general setting of multi-parameter regression models in Chapter 5.

Figure 4.2. Estimated hazard function, hazard ratio and survivor function for the Weibull, log-logistic and TDL SPR models. The Kaplan-Meier curves (step function) appear in bottom panels for comparison.
Figure 4.3. Hazard ratios with confidence intervals for the Weibull, log-logistic and TDL SPR models.
Chapter 5

Multi-Parameter Regression

5.1 Introduction

The classical approach to (parametric) regression analysis is to have a clear parameter of interest which is allowed to depend on covariates, for example, in generalized linear models (McCullagh & Nelder, 1989) where traditionally the location parameter is regressed on a set covariates. Other parameters remain constant and are typically only present to allow the model to adapt to data. We will refer to these models as single parameter regression (SPR) models as the regression structure appears in one distributional parameter (such as the models considered in Section 4.5). When more than one parameter is available, the question is why should one take precedence over the others in terms of covariate analysis? A more flexible approach is to regress multiple parameters simultaneously on covariates. In the context of normal linear regression, Park (1966) and Harvey (1976) modelled the dispersion parameter as a function of covariates to address heteroscedasticity, while Smyth (1989) modelled dispersion in the more general case of generalized linear models. More recently, structured dispersion has been studied by Lee & Nelder (2001, 2006). Other modern frameworks with more than one regression component are generalized additive models for location, scale and shape (Rigby & Stasinopoulos, 2005) and joint mean-covariance modelling in longitudinal data analysis (Pan & MacKenzie, 2003, 2006, 2007). Hereafter, we refer to this flexible approach as multi-parameter regression, MPR.
In our parametric survival setup we have a scale parameter, $\lambda$, which controls the overall magnitude of the hazard, and a shape parameter $\gamma$, which controls the time evolution of the hazard (Section 1.4). As we saw in Section 4.5, the scale parameter is the parameter of interest in standard SPR survival models. We now extend the SPR model, defined by (4.18), to a multi-parameter regression model by setting

$$g(\lambda) = x^T \beta, \quad h(\gamma) = z^T \alpha,$$

(5.1)

where $x = (1, x_1, \ldots, x_p)^T$ and $z = (1, z_1, \ldots, z_q)^T$ are covariate vectors which may or may not contain covariates in common, $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T$ and $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_q)^T$ are the corresponding regression coefficients and $g(\cdot)$ and $h(\cdot)$ are appropriate link functions (often the log link in cases we consider). Thus, the addition of $h(\gamma) = z^T \alpha$ extends the typical SPR models to allow shape to depend on covariates; the resulting MPR models will be able to handle a wider variety of survival data more readily, e.g., crossing, converging or diverging hazards. Moreover, the shape parameter has an important role to play in describing the underlying process (Aalen & Gjessing, 2001) and, therefore, gaining an understanding of how this depends on covariates is of key interest. This is not something available to us within the context of the Cox model, which is mainly focussed on the estimation of relative risk (see Section 4.4), though it is worth noting that the Cox model can be extended in a variety of different ways (Therneau & Grambsch, 2000) to gain further flexibility. While these extensions may require specialised theory, the MPR model, being fully parametric, is estimated straightforwardly via the likelihood function

$$\ell(\theta) = \sum_{i=1}^{n} \delta_i \log \lambda(t_i \mid x_i, z_i) + \log S(t_i \mid x_i, z_i),$$

(5.2)

which only differs from the SPR case, given in (4.17), notationally by the presence of $z_i$. The associated likelihood theory remains unchanged where we are now estimating $p + q + 2$ regression coefficients; there are no extra (constant) distributional parameters in this case as these are all structured to depend on covariates.

Multi-parameter regression is not wholly novel in a survival setting; early references include Taulbee (1979) and Gaynor (1987). More recently, the
approach has been applied to the inverse Gaussian model (Lee & Whitmore, 2006, 2010; Aalen et al., 2008). The reason for this specific interest in the inverse Gaussian model as a candidate for multi-parameter regression is its derivation as the distribution of time until a Wiener process reaches an absorbing barrier/threshold. Thus, its parameters can be interpreted as drift and distance from the threshold where, for example, the underlying Wiener process may represent physical deterioration leading to failure upon reaching some critical level (the threshold)\(^1\). Viewing survival in terms of this particular underlying mechanism is interesting, but of course multi-parameter regression is a broader concept which is not limited to any particular family of models. Furthermore, as mentioned in Section 1.4, we are not considering models, such as the inverse Gaussian model, where the survivor function does not have a closed form; the log-logistic model, which we do consider, has a very similar shape to the inverse Gaussian model.

It is clear from the above that multi-parameter regression in survival has not previously received much attention as a general procedure and is not in mainstream use. This may be, in part, due to the fact that even standard parametric methods (e.g., SPR models) receive less attention in survival analysis compared with non- and semi-parametric counterparts (as we have noted in Section 1.4). We therefore explore and discuss the consequences of the MPR approach and show its flexibility both analytically and via real data analyses.

5.2 Multi-Parameter Regression Models

Parallelling Section 4.5, we will interpret the MPR models below in terms of the (time-dependent) hazard ratio which we denote by \(\psi(t)\). Let \(c\) be a binary covariate whose effect we wish to determine (via \(\psi(t)\)). As we now have scale and shape regressions, we will assume that \(c\) is common to both and, furthermore, that it is the first covariate in the vectors \(x\) and \(z\), i.e.,

\(^1\)It is noteworthy that ascribing a mechanistic interpretation to model parameters is speculative since the data may not have been generated by this mechanism. A similar problem was noted by MacKenzie (1986) in a different context.
$x_1 = z_1 = c$. Thus we have
\[ x^T \beta = \beta_0 + c \beta_1 + \ldots + x_p \beta_p \]
\[ z^T \alpha = \alpha_0 + c \alpha_1 + \ldots + z_q \alpha_q \]
\[ = c \beta_1 + \beta_0 + \ldots + x_p \beta_p \]
\[ = c \alpha_1 + \alpha_0 + \ldots + z_q \alpha_q \]
\[ = c \beta_1 + \hat{x}^T \beta, \quad \text{and} \]
\[ = c \alpha_1 + \hat{z}^T \alpha, \quad (5.3) \]
where \( \hat{x} = (1, 0, x_2, \ldots, x_p)^T \) and \( \hat{z} = (1, 0, z_2, \ldots, z_q)^T \).

The hazard ratio, at time \( t \), for the covariate \( c = x_1 = z_1 \) is therefore given by
\[ \psi(t) = \frac{\lambda(t \mid c = 1)}{\lambda(t \mid c = 0)}, \quad (5.4) \]
which we study for each model by examining its first derivative (w.r.t. \( t \)) and the beginning/end-points,
\[ \psi(0) = \lim_{t \to 0} \psi(t), \quad \psi(\infty) = \lim_{t \to \infty} \psi(t). \]
In the proceeding sections we merely show results; derivations of these results (e.g., differentiation and limit calculation) can be found in Appendix B.3. Moreover, we will see that \( \psi(t) \) is conditional on \( \hat{x} \) or \( \hat{z} \) (i.e., we could write \( \psi(t \mid \hat{x}, \hat{z}) \)) for some models; this is discussed in Section 5.3.

5.2.1 Weibull

For the Weibull MPR model, \( \log(\lambda) = x^T \beta \) and \( \log(\gamma) = z^T \alpha \) where \( (\beta, \alpha)^T \in \mathbb{R}^{p+q+2} \). Thus, the hazard and survivor functions are given by
\[ \lambda(t \mid x, z) = \exp(x^T \beta) \exp(z^T \alpha) t^{\exp(z^T \alpha) - 1}, \quad (5.5) \]
and
\[ S(t \mid x, z) = \exp[- \exp(x^T \beta) t^{\exp(z^T \alpha)}]. \quad (5.6) \]
The hazard ratio, at time \( t \), for \( c = x_1 = z_1 \) is given by
\[ \psi(t) = \frac{\lambda(t \mid c = 1)}{\lambda(t \mid c = 0)} = \exp(\beta_1) \exp(\alpha_1) t^{\exp(z^T \alpha)[\exp(\alpha_1) - 1]}, \quad (5.7) \]
which generalises the Weibull SPR hazard ratio, (4.21), to time-dependence. Thus the Weibull MPR model is a non-PH model. The first derivative is
\[ \frac{d\psi(t)}{dt} = \exp(\hat{z}^T \alpha)[\exp(\alpha_1) - 1] \frac{\psi(t)}{t}, \]
and, therefore, it is the sign of $\alpha_1$ which solely determines the path of the hazard ratio. This coefficient is of key importance, i.e., $\psi(t) = \begin{cases} \text{Increasing} & \text{if } \alpha_1 > 0, \\ \text{Constant} & \text{if } \alpha_1 = 0, \\ \text{Decreasing} & \text{if } \alpha_1 < 0. \end{cases}$

In particular, as $\psi(t)$ is constant ($= \exp(\beta_1)$) if $\alpha_1 = 0$, the Weibull MPR model directly extends the proportional hazards model, thus providing a test of proportionality adjusted for other scale and shape covariates. Furthermore, it is clear that the Weibull MPR model, being more general, can provide a better fit to data than the Weibull PH model (i.e., the SPR model).

Given that $\psi(0) = \begin{cases} 0 & \text{if } \alpha_1 > 0, \\ \exp(\beta_1) & \text{if } \alpha_1 = 0, \\ \infty & \text{if } \alpha_1 < 0, \end{cases}$ we know $\psi(t^*) = 1$ for some $t^*$ if $\alpha_1 \neq 0$. This represents the time point where the hazards cross, i.e., the Weibull MPR model implies that the hazards must cross when $\alpha_1$ is non-zero. Indeed we may be interested in the value of $t^*$ which we easily find, for $\alpha_1 \neq 0$, as

$$t^* = \psi^{-1}(1) = \left[\frac{1}{\exp(\beta_1) \exp(\alpha_1)}\right]^{1/\{\exp(\beta_1/\exp(\alpha_1))-1\}}.$$

While the Weibull MPR model implies a theoretical crossing of hazards, the value of $t^*$ may be such that we can ignore this crossing for practical purposes. For example, if $t^* \approx 0$ then we may consider $\psi(t)$ for values of $t > t^* + \epsilon$ (and $\epsilon$ small) which represents diverging hazards. In this case $\psi(t)$ is either increasing from $\psi(t^* + \epsilon) > 1$ or decreasing from $\psi(t^* + \epsilon) < 1$ (as in Fig. 5.1 when $(\beta_1, \alpha_1) = (-0.5, -0.3)$). Conversely, if $t^* \gg t_{\max}$ and $\psi(t_{\max}) \approx 1$, where $t_{\max}$ is the largest time in the dataset, then we may treat the hazards as converging (as in Fig. 5.1 when $(\beta_1, \alpha_1) = (-1, 0.25)$). Another point worth noting is that, when $\alpha_1 \neq 0$, $\psi(0)$ is either infinity or zero. A hazard ratio of infinity/zero means that having $c = 1$ infinitely increases/decreases the hazard relative to $c = 0$. This is again a theoretical
implication of the model which may not always be useful in practice. Thus, we may prefer to view \( \psi(t) \) for values of \( t > \varepsilon \) (for some small \( \varepsilon \)) such that \( \psi(\varepsilon) \) is a more reasonable (finite, non-zero) value, i.e., \( \psi(\varepsilon) \) is sufficiently smaller than infinity or sufficiently larger than zero. These comments apply equally to the other MPR models considered in Sections 5.2.2 - 5.2.6.

### 5.2.2 Gompertz

For the Gompertz MPR model, \( \log(\lambda) = x^T \beta \) and \( \gamma = z^T \alpha \) where \( (\beta, \alpha)^T \in \mathbb{R}^{p+q+2} \). Thus, the hazard and survivor functions are given by

\[
\lambda(t \mid x, z) = \exp(x^T \beta) \exp(z^T \alpha t), \tag{5.8}
\]

and

\[
S(t \mid x, z) = \exp \left\{ -\frac{\exp(x^T \beta)}{z^T \alpha} [\exp(z^T \alpha t) - 1] \right\}. \tag{5.9}
\]

The hazard ratio, at time \( t \), for \( c = x_1 = z_1 \) is given by

\[
\psi(t) = \exp(\beta_1) \exp(\alpha_1 t), \tag{5.10}
\]
which, like the Weibull MPR model, generalises its SPR counterpart, (4.24), to time-dependence, i.e., non-proportional hazards. The Gompertz MPR hazard ratio has a very simple and appealing form; the effect of \( c \) is fully described by the corresponding scale and shape regression coefficients \( \beta_1 \) and \( \alpha_1 \) respectively. Interestingly,

\[
\frac{d\psi(t)}{dt} = \alpha_1 \psi(t),
\]

which is the same differential representation as that for the Gompertz hazard function in Section 1.4.3. Clearly, the nature of the time-dependence is controlled by the sign of \( \alpha_1 \) and, as in the Weibull MPR case, \( \psi(t) \) can increase (\( \alpha_1 > 0 \)), decrease (\( \alpha_1 < 0 \)) or remain constant (\( \alpha_1 = 0 \)). Thus, like the Weibull MPR model, the Gompertz MPR model provides us with a test of proportionality of hazards (by testing \( H_0 : \alpha_1 = 0 \)).

Note that

\[
\psi(0) = \exp(\beta_1) \quad \forall \alpha_1,
\]

\[
\psi(\infty) = \begin{cases} 
\infty & \text{if } \alpha_1 > 0, \\
\exp(\beta_1) & \text{if } \alpha_1 = 0, \\
0 & \text{if } \alpha_1 < 0,
\end{cases}
\]

and, therefore, unlike the Weibull case, \( \psi(t) \) here is not guaranteed to be equal to one for some \( t = t^* \) when \( \alpha_1 \neq 0 \). However, it is clear that if \( \beta_1 < 0 \) and \( \alpha_1 > 0 \) then \( \psi(t) \) will increase from \( \exp(\beta_1) < 1 \) to \( \infty \), i.e., \( \psi(t^*) = 1 \) for some \( t^* \). Similarly, the hazards cross if \( \beta_1 > 0 \) and \( \alpha_1 < 0 \). Thus, \( \psi(t^*) = 1 \) if \( \text{sgn}(\beta_1) \neq \text{sgn}(\alpha_1) \neq 0 \), and, furthermore,

\[
t^* = \psi^{-1}(1) = \frac{-\beta_1}{\alpha_1}.
\]

It is clear from the above that the Gompertz MPR model has some appealing properties. The shape of \( \psi(t) \) is similar to the Weibull MPR model (see Fig. 5.2) although it does not depend on \( \tilde{z} \) and increases/decreases at a faster rate. Nonetheless, analyses using both models may produce similar results. Moreover, both models offer a straightforward extension into non-proportional hazards modelling.
5.2.3 Log-Logistic

The log-logistic MPR model requires a log-link for the parameters, i.e.,
\[ \log(\lambda) = x^T \beta \quad \text{and} \quad \log(\gamma) = z^T \alpha \] where \((\beta, \alpha)^T \in \mathbb{R}^{p+q+2}\). Thus, the hazard and survivor functions are given by

\[
\lambda(t \mid x, z) = \frac{\exp(x^T \beta) \exp(z^T \alpha) t \exp(z^T \alpha) - 1}{1 + \exp(x^T \beta) t \exp(z^T \alpha)},
\]

and

\[
S(t \mid x, z) = [1 + \exp(x^T \beta) t \exp(z^T \alpha)]^{-1}.
\]

The hazard ratio, at time \(t\), for \(c = x_1 = z_1\) is given by

\[
\psi(t) = \frac{\exp(\beta_1 + \alpha_1) t \exp(\tilde{x}^T \alpha) \exp(\alpha_1) - 1}{1 + \exp(\tilde{x}^T \beta) t \exp(\tilde{z}^T \alpha)} \frac{1 + \exp(\tilde{x}^T \beta) t \exp(\tilde{z}^T \alpha)}{1 + \exp(\beta_1 + \tilde{x}^T \beta) t \exp(\alpha_1 + \tilde{z}^T \alpha)},
\]

which is time-dependent. Of course, as noted in Section 4.5.3, the SPR hazard ratio was already time-dependent. However, the functional form of (5.13) above is more general than (4.27) and will therefore support a wider variety of shapes.
5.2. MULTI-PARAMETER REGRESSION MODELS

It is easy to show that

\[ \psi(0) = \begin{cases} 
0 & \text{if } \alpha_1 > 0, \\
\exp(\beta_1) & \text{if } \alpha_1 = 0, \\
\infty & \text{if } \alpha_1 < 0,
\end{cases} \quad \psi(\infty) = \exp(\alpha_1) \quad \forall \alpha_1. \]

Thus, \( \psi(t) \) increases when \( \alpha_1 > 0 \) (from 0 to \( \exp(\alpha_1) \)), decreases when \( \alpha_1 < 0 \) (from \( \infty \) to \( \exp(\alpha_1) \)) and, when \( \alpha_1 = 0 \), \( \psi(t) \) can increase, decrease or remain equal to one depending on the value of \( \beta_1 \). Furthermore, it is clear that the hazards cross if \( \alpha_1 \neq 0 \). However, unlike the Weibull and Gompertz models, \( t^* \), the crossing point, cannot be expressed analytically but, nonetheless, it can easily be calculated numerically/graphically in practice.

It is clear from the above that \( \alpha_1 \) is key in defining how \( \psi(t) \) evolves over time. Furthermore,

\[ \frac{d\psi(t)}{dt} = \left\{ \exp(z^T \alpha)[\exp(\alpha_1) - 1]\frac{1}{t} + \lambda(t \mid c = 0)[1 - \psi(t)] \right\}\psi(t), \]

and while this derivative is more complicated than the Weibull or Gompertz cases, we can still use it to tell us about the nature of \( \psi(t) \) over time (see Appendix B.3.3). We find that

\[ \psi(t) = \begin{cases} 
\text{Increasing} & \text{if } \alpha_1 > 0, \\
\text{Increasing (monotonic)} & \text{if } \alpha_1 = 0 \text{ and } \beta_1 < 0, \\
\text{Decreasing (monotonic)} & \text{if } \alpha_1 = 0 \text{ and } \beta_1 > 0, \\
\text{Decreasing} & \text{if } \alpha_1 < 0.
\end{cases} \]

Figure 5.3 shows some possible shapes that \( \psi(t) \) can take.

Clearly the log-logistic MPR model extends its SPR counterpart. The SPR model forces \( \psi(t) \) to converge to one whereas the MPR model allows \( \psi(t) \) to converge to \( \exp(\alpha_1) \) and can handle crossing hazards and non-monotonic hazard ratios. Furthermore, whereas the Weibull and Gompertz MPR models assume \( \psi(t) \) continues to increase/decrease over time (when \( \alpha_1 \neq 0 \)), the fact that \( \psi(t) \) levels off over time for the log-logistic MPR model may be more realistic in some situations.

Although we are interpreting the models in terms of the hazard ratio, it is worth noting that the log-logistic model is often interpreted in terms of
the odds of survival (Bennett, 1983). One easily finds that, at time \(t\), the odds ratio for the \(c = x_1 = z_1\) is given by

\[
\frac{S(t \mid c = 1)}{1 - S(t \mid c = 1)} \frac{1 - S(t \mid c = 0)}{S(t \mid c = 0)} = \exp(-\beta_1)\exp(\tilde{z}^T \alpha)[1 - \exp(\alpha_1)],
\]

and thus we see that the SPR log-logistic model (\(\alpha_1 = 0, \tilde{z}^T \alpha = \alpha_0\)) is a proportional odds model (see Section 4.3) whereas the MPR model generalises this to allow the odds ratio to depend on time. When viewed in this way, the \(\alpha_1\) coefficient can be used to test proportionality of odds; this is not something we will pursue.

## 5.2.4 Burr

The Burr distribution (Section 1.4.5) has three distributional parameters and thus the Burr MPR model has three regression components: \(\log(\lambda) = x^T \beta\), \(\log(\gamma) = z^T \alpha\) and \(\log(\rho) = w^T \tau\) where, in the third component, \(w = (1, w_1, \ldots, w_r)^T\) and \(\tau = (\tau_0, \tau_1, \ldots, \tau_r)^T\) respectively. Thus \((\beta, \alpha, \tau)^T \in \mathbb{R}^{p + q + r + 3}\). As the Burr MPR model contains the log-logistic MPR model
as a special case \((w^T \tau = 0)\) and the Weibull MPR model as a limiting case \((w^T \tau \to -\infty)\), it will be capable of modelling a range of phenomena.

The hazard and survivor functions are given by

\[
\lambda(t \mid x, z, w) = \frac{\exp(x^T \beta) \exp(z^T \alpha) t^{\exp(z^T \alpha) - 1}}{1 + \exp(x^T \beta) \exp(w^T \tau) t^{\exp(z^T \alpha)}},
\]

and

\[
S(t \mid x, z, w) = \left[1 + \exp(x^T \beta) \exp(w^T \tau) t^{\exp(z^T \alpha)}\right]^{-1}/\exp(w^T \tau).
\]

The hazard ratio, at time \(t\), for \(c = x_1 = z_1 = w_1\) is given by

\[
\psi(t) = \exp(\beta_1 + \alpha_1) t^{\exp(\mu^T \alpha) \exp(\alpha_1) - 1} \frac{1 + \exp(\tilde{x}^T \beta + \tilde{w}^T \tau) t^{\exp(\tilde{x}^T \alpha)}}{1 + \exp(\beta_1 + \tau_1 + \tilde{x}^T \beta + \tilde{w}^T \tau) t^{\exp(\alpha_1 + \tilde{x}^T \alpha)}},
\]

and its derivative is given by

\[
\frac{d\psi(t)}{dt} = \left\{ \exp(\tilde{x}^T \alpha) \exp(\alpha_1) - 1 \frac{1}{t} + \exp(\tilde{w}^T \tau) \lambda(t \mid c = 0)[1 - \exp(\tau_1)\psi(t)] \right\} \psi(t).
\]

Clearly these are similar to those of the log-logistic MPR model. Furthermore, we find that

\[
\psi(0) = \begin{cases} 
0 & \text{if } \alpha_1 > 0, \\
\exp(\beta_1) & \text{if } \alpha_1 = 0, \\
\infty & \text{if } \alpha_1 < 0,
\end{cases}
\]

\[
\psi(\infty) = \exp(\alpha_1 - \tau_1) \forall \alpha_1.
\]

Remembering that for the log-logistic MPR model \(\psi(\infty) = \exp(\alpha_1)\), we can see the extended flexibility of the Burr model where \(\psi(\infty) = \exp(\alpha_1 - \tau_1)\), i.e., while \(\alpha_1\) describes both the time-evolution and the end point of \(\psi(t)\) in the log-logistic model, the Burr model is not restricted by this constraint. In any case, as with the other models considered so far, \(\alpha_1\) is the key parameter here in characterising the time evolution where

\[
\psi(t) = \begin{cases}
\text{Increasing} & \text{if } \alpha_1 > 0, \\
\text{Increasing (monotonic)} & \text{if } \alpha_1 = 0 \text{ and } \beta_1 + \tau_1 < 0, \\
\text{Constant} & \text{if } \alpha_1 = 0 \text{ and } \beta_1 + \tau_1 = 0, \\
\text{Decreasing (monotonic)} & \text{if } \alpha_1 = 0 \text{ and } \beta_1 + \tau_1 > 0, \\
\text{Decreasing} & \text{if } \alpha_1 < 0.
\end{cases}
\]
Thus, \( \psi(t) \) increases if \( \alpha_1 > 0 \) and decreases if \( \alpha_1 < 0 \) (both possibly non-monotonically). The case where \( \alpha_1 = 0 \) is interesting as \( \psi(t) \) can increase or decrease (monotonically) from \( \exp(\beta_1) \) to \( \exp(-\tau_1) \); the start and end points are finite and non-zero which may be desirable in practice. Of course, the model also handles (approximately) infinite/zero limits depending on the values of \( \beta_1 \) and \( \tau_1 \). Note that when \( \alpha_1 = 0 \) and \( \beta_1 + \tau_1 = 0 \), \( \psi(t) = \exp(\beta_1) \) which is of particular interest given the status of PH regression in survival analysis. Thus, the Burr MPR model provides a test of proportionality which supports a larger variety in alternative hypotheses than the Weibull or Gompertz MPR models. Figure 5.4 shows some possible shapes that \( \psi(t) \) can take.

The Burr MPR model is clearly more flexible than its SPR counterpart (Section 4.5.4), the latter having \( \psi(t) \) constrained to increase/decrease monotonically from \( \exp(\beta_1) \) to one (as \( \alpha_1 = \tau_1 = 0 \)). Furthermore, the Burr SPR model supports PH effects but is limited in the sense that it assumes such effects are consistent across all covariates, i.e., either all covariates have time-constant effects or not. The MPR model supports PH effects on a per
It is clear from the above that the Burr MPR model is capable of modelling a variety of situations. It is a flexible model which relaxes the limitations of other models considered up to now, unifying and extending the (also quite flexible) Weibull MPR and log-logistic MPR models. Furthermore, it contains standard simpler SPR models (Weibull, log-logistic and Burr) as special cases. It can therefore provide a better fit to data than these models and will reduce to them in cases where its full flexibility is not required. Hence, this model provides robust parametric regression.

5.2.5 Time-Dependent Logistic

For the time-dependent logistic MPR model, \( \lambda = x^T \beta \) and \( \gamma = z^T \alpha \) where \((\beta, \alpha)^T \in \mathbb{R}^{p+q+2}\). Thus, the hazard and survivor functions are given by

\[
\lambda(t \mid x, z) = \frac{\exp(z^T \alpha t + x^T \beta)}{1 + \exp(z^T \alpha t + x^T \beta)},
\]

and

\[
S(t \mid x, z) = \left[ \frac{1 + \exp(z^T \alpha t + x^T \beta)}{1 + \exp(x^T \beta)} \right]^{-1/z^T \alpha}.
\]

The hazard ratio, at time \( t \), for \( c = x_1 = z_1 \) is given by

\[
\psi(t) = \exp(\alpha_1 t + \beta_1) \frac{1 + \exp(\tilde{z}^T \alpha t + \tilde{x}^T \beta)}{1 + \exp[(\alpha_1 + \tilde{z}^T \alpha)t + \beta_1 + \tilde{x}^T \beta]},
\]

which generalises the functional form of the hazard ratio for the time-dependent logistic SPR model (Section 4.33). We find that

\[
\frac{d\psi(t)}{dt} = [\alpha_1 + \tilde{z}^T \alpha \lambda(t \mid c = 0) - (\alpha_1 + \tilde{z}^T \alpha)\lambda(t \mid c = 1)] \psi(t),
\]

and thus we see that both \( \alpha_1 \) and \( \tilde{z}^T \alpha \) are important in characterising the time evolution of \( \psi(t) \). This can also be seen in the limits of \( \psi(t) \):

\[
\psi(0) = \exp(\beta_1) \frac{1 + \exp(\tilde{x}^T \beta)}{1 + \exp(\tilde{z}^T \alpha + \tilde{x}^T \beta)}, \quad \forall \alpha_1, \tilde{z}^T \alpha,
\]

\footnote{In our practical work so far, the flexibility of the Burr MPR model has not led to significant gains over the log-logistic MPR model. Moreover, we have found that convergence issues are common when attempting to fit the Burr MPR model.}
and, after some algebra (Appendix B.3.5),

\[
\psi(\infty) = \begin{cases} 
1 & \text{if } \alpha_1 \geq 0 \text{ and } \tilde{z}^T \alpha > 0, \\
\frac{1+\exp(\tilde{x}^T \beta)}{\exp(\tilde{x}^T \beta)} & \text{if } \alpha_1 > 0 \text{ and } \tilde{z}^T \alpha = 0, \\
\infty & \text{if } \alpha_1 > 0 \text{ and } \tilde{z}^T \alpha < 0, \\
\exp(\tilde{\beta}_1) \frac{1+\exp(\tilde{x}^T \beta)}{1+\exp(\tilde{x}^T \beta)} & \text{if } \alpha_1 = 0 \text{ and } \tilde{z}^T \alpha = 0, \\
\exp(\beta_1) & \text{if } \alpha_1 = 0 \text{ and } \tilde{z}^T \alpha < 0, \\
\exp(\tilde{\beta}_1+\tilde{x}^T \beta) \frac{1+\exp(\tilde{\beta}_1+\tilde{x}^T \beta)}{1+\exp(\tilde{\beta}_1+\tilde{x}^T \beta)} & \text{if } \alpha_1 < 0 \text{ and } \tilde{z}^T \alpha = -\alpha_1, \\
0 & \text{if } \alpha_1 < 0 \text{ and } \tilde{z}^T \alpha < -\alpha_1,
\end{cases}
\]  

where cases with \( \tilde{z}^T \alpha \) exactly equal to zero or to \(-\alpha_1\) (not likely in practice) are highlighted in grey. We can see that, unlike the models considered up to now, the \( \alpha_1 \) coefficient alone is not enough to tell us about the time evolution of \( \psi(t) \). Some possible shapes are given in Fig. 5.5. The time-dependent

![Figure 5.5. Time-dependent logistic MPR hazard ratio shapes (solid). In all cases \( \tilde{x}^T \beta = 0 \). Reference line (dot) represents equality of hazards.](image)
logistic SPR model (Section 4.31), with $\alpha_1 = 0$ and $\tilde{z}^T \alpha = \alpha_0$, is limited to $\psi(t)$ increasing/decreasing either to one ($\alpha_0 > 0$) or to $\exp(\beta_1)$ ($\alpha_0 < 0$). A third possibility is that $\psi(t)$ remains constant at $\psi(0)$ ($\alpha_0 = 0$); we discuss this in the proceeding paragraph.

As we’ve mentioned, $\tilde{z}^T \alpha = 0$ is unlikely in practice. However, one way in which this can arise is the case where $\alpha_j = 0$ for $j = 0, 1, \ldots, q$. This leads to $\psi(\infty) = \psi(0)$ and, moreover, $\psi'(t) = 0$, i.e., the hazard ratios are time-independent for all covariates. This special case is a type of proportional hazards model, although not in the classical sense (defined in (4.3)) as the hazard ratios depend on $\tilde{x}$.

### 5.2.6 Piecewise Exponential

The piecewise exponential model is, of course, different to the other parametric models above in that its dimension depends on how many intervals we split the time axis into (see Sections 1.4.7 and 4.5.6). There are $m$ intervals in total, $I_j = [t_{(j-1)}, t_{(j)})$ for $j = 1, \ldots, m$, and hence $m$ distributional parameters: $\lambda_1, \ldots, \lambda_m$. Thus the multi-parameter regression model is defined by modelling each of these parameters via

$$\log \lambda_j = x^T \beta_j, \quad (5.22)$$

where $x = (1, x_1, \ldots, x_p)^T$, $\beta_j = (\beta_{0j}, \beta_{1j}, \ldots, \beta_{pj})^T$ and $j = 1, \ldots, m$ (compare this with (4.37)). The hazard function is given by

$$\lambda(t \mid x) = a(t)^T (\lambda_1, \ldots, \lambda_m)^T, \quad (5.23)$$

where $\lambda_j = \exp(x^T \beta_j)$ and $a(t)$ is an $m$-dimensional vector indicating which interval $t$ lies in, e.g., if $m = 4$ and $t \in I_3$ then $a(t) = (0, 0, 1, 0)^T$. The survivor function is given by

$$S(t) = \exp \left[ -d(t)^T (\lambda_1, \ldots, \lambda_m)^T \right], \quad (5.24)$$

where $d(t)$ represents the time spent in each interval for a particular value of $t$, e.g., if $m = 4$ and $t \in I_3$ then $d(t) = (t_{(1)} - t_{(0)}, t_{(2)} - t_{(1)}, t - t_{(2)}, 0)^T$. 

At time $t \in I_j$, the hazard ratio for $x_1$ is given by

$$
\psi(t_{(j-1)} \leq t < t_{(j)}) = \frac{\exp(\beta_{1j} + \hat{x}^T \beta_j)}{\exp(\hat{x}^T \beta_j)} = \exp(\beta_{1j}),
$$

(5.25)

where $\beta_{1j}$ is the coefficient of $x_1$ in the $j$th time interval. Thus, the hazards are proportional \textit{within} an interval of time; the hazard ratio is a step function when viewed over time. Furthermore, we may write

$$
\psi(t) = \exp[\beta_1(t)],
$$

(5.26)

where $\beta_1(t)$ is a step function such that $\beta_1(t \in I_j) = \beta_{1j}$. The form of the hazard ratio in this model is appealing and easily interpreted by exponentiating the $m$ regression coefficients corresponding to the covariate in question.

Clearly the piecewise exponential MPR model can approximate any covariate effect but the dimension can become high as the number of parameters is $m \times (p + 1)$. The PH piecewise exponential model (Section 4.5.6) is a special case of this model which arises by setting all regression coefficients (apart from the intercept) to constants, i.e., $\beta_k(t) = \beta_k$ for $k = 1, \ldots, p$. We may, of course, specify a model in which some covariate effects are time-dependent and some are time-constant.

\textbf{Example 5.1. Multi-Parameter Regression Analysis of Lung Cancer Data}

In Example 4.2 we analysed the (unadjusted) effect of treatment in the lung cancer data using three SPR models (Weibull, log-logistic and time-dependent logistic). We continue now with this analysis within the more general framework of multi-parameter regression. Hence, we fitted the Weibull, log-logistic and time-dependent logistic MPR models to the data for direct comparison with our previous SPR analysis. The estimated scale and shape regression coefficients for the three models are given in Table 5.1.

In this MPR setting, quantifying covariate effects directly through the individual regression coefficients is more difficult than it is for SPR models. As the effect of a covariate now depends both on its scale effect ($\beta$ coefficient) and its shape effect ($\alpha$ coefficient), it seems reasonable to view the overall effect in terms of a quantity which takes both of these into account. Thus, we use the hazard ratio to measure the overall effect.
5.2. MULTI-PARAMETER REGRESSION MODELS

Table 5.1. MPR Model Fits

<table>
<thead>
<tr>
<th></th>
<th>Weibull</th>
<th>Log-logistic</th>
<th>TDL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{\beta} )</td>
<td>S.E.</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-1.28 (0.08)</td>
<td>——</td>
<td></td>
</tr>
<tr>
<td>Palliative</td>
<td>0.00</td>
<td>——</td>
<td></td>
</tr>
<tr>
<td>Surgery</td>
<td>-3.91 (0.83)</td>
<td>4.69</td>
<td></td>
</tr>
<tr>
<td>Chemo</td>
<td>-0.50 (0.32)</td>
<td>1.56</td>
<td></td>
</tr>
<tr>
<td>Radio</td>
<td>-1.26 (0.19)</td>
<td>6.72</td>
<td></td>
</tr>
<tr>
<td>C+R</td>
<td>-4.06 (0.88)</td>
<td>4.60</td>
<td></td>
</tr>
</tbody>
</table>

|        | \( \hat{\alpha} \) | S.E. | | \( \hat{\alpha} \) | S.E. | | \( \hat{\alpha} \) | S.E. |
|--------|---------|------| | \( \hat{\alpha} \) | S.E. | | \( \hat{\alpha} \) | S.E. |
| Intercept | -0.19 (0.04) | —— | | 0.18 (0.04) | —— | | -0.11 (0.02) | —— |
| Palliative | 0.00 | —— | | 0.00 | —— | | 0.00 | —— |
| Surgery | 0.59 (0.20) | 2.90 | | 0.31 (0.20) | 1.53 | | 0.18 (0.05) | 3.63 |
| Chemo | 0.07 (0.14) | 0.51 | | -0.02 (0.14) | 0.15 | | 0.09 (0.05) | 1.78 |
| Radio | 0.34 (0.07) | 4.66 | | 0.25 (0.07) | 3.39 | | 0.14 (0.03) | 5.25 |
| C+R | 0.97 (0.16) | 5.90 | | 0.93 (0.17) | 5.50 | | 0.30 (0.06) | 5.00 |

Before examining hazard ratios, we inspect the estimates and standard errors in Table 5.1 from which we can draw some preliminary conclusions. The \( \beta \) coefficients are negative which means that all treatments tend to reduce the hazard relative to palliative care. However, the \( \alpha \) coefficients are positive (except chemotherapy in the log-logistic model) and, therefore, the hazards are increasing relative to palliative care. In other words, the effectiveness of each treatment reduces over time. The Z-scores show that chemotherapy is not significant in either the scale or the shape (for the Weibull and log-logistic models), i.e., the overall effect of chemotherapy is not significant.

Recall that the AIC values for the three models in SPR form (Table 4.3) were 3933.51, 3899.32 and 3932.32 respectively. Thus, we can see that the fit has improved in all cases by generalising to MPR form (with AIC reductions of 37.27, 23.85 and 43.37 respectively). The improvement in fit is also evident.
from Fig. 5.6 (bottom panel) where all estimated survivor curves match the corresponding Kaplan-Meier curves more closely than in the SPR case (see Fig. 4.2), particularly for the C+R group. The log-logistic MPR model fits
5.2. MULTI-PARAMETER REGRESSION MODELS

the data most closely (notable in the palliative care and radiotherapy groups) as is reflected in its AIC value.

The top two panels of Fig. 5.6 show the covariate effects in terms of the hazard and these are in agreement with our earlier conclusions based on

![Figure 5.7](image)

*Figure 5.7. Hazard ratios with confidence intervals for the Weibull, log-logistic and TDL MPR models.*
inspecting the $\beta$ and $\alpha$ coefficients in Table 5.1, i.e., the effect of treatment (relative to palliative care) is to reduce the hazard but this is not maintained over the full range of time. Figure 5.6 shows each hazard ratio separately with the corresponding confidence interval (calculated using m.l.e. simulation - see Section 2.3.2). The three models are roughly in agreement.

Surgery significantly reduces the hazard but its effect may wear off towards the end of the time period (according to the log-logistic and time-dependent logistic models). Nonetheless, it is clear that this treatment is superior to the other three. Chemotherapy may reduce the hazard somewhat but this is not statistically significant (apart from a brief period early in time). Both radiotherapy and the combined treatment (chemo + radio) reduce the hazard significantly but their effects wear off after about 5 - 7 months. Of these two treatments, the combined treatment offers a greater reduction in the hazard up to this point. However, it appears to be more hazardous later in time.

It is clear from the above that the MPR approach provides a more flexible framework than conventional SPR methods. With this we can discover more complicated relationships between covariates and survival which simpler methods may fail to capture.

**Example 5.2. MPR Burr Analysis of Lung Cancer Data**

In Section 5.2.4 we showed that the Burr MPR model can support a variety of hazard effects and generalises other models we have considered. Thus, we applied this model to the lung cancer data (again considering the unadjusted effect of treatment only). The resulting estimates are given in the first column of Table 5.2.

In this full Burr MPR model, with three regression components, none of the $\tau$ coefficients are statistically significant. Thus, a reduced Burr MPR model with $\tau = \tau_0$ is shown in the second column of Table 5.2. In this model $\tau_0$ is still not statistically different from zero and, as discussed in Section 5.2.4, $\text{Burr}(\beta, \alpha, \tau = 0) = \text{log-logistic}(\beta, \alpha)$. Indeed the estimated $\beta$ and $\alpha$ coefficients for this reduced Burr MPR model are numerically very close to those from the log-logistic MPR analysis (Table 5.1), as are the hazard ratios (not shown for the Burr model). Therefore, the extra complexity of the Burr model is not needed in this case so it reduces to a log-logistic model.
Table 5.2. *Burr MPR Model Fit*

<table>
<thead>
<tr>
<th>Scale</th>
<th>Burr(β, α, τ)</th>
<th>Burr(β, α, τ&lt;sub&gt;0&lt;/sub&gt;)</th>
<th>Burr(β, α&lt;sub&gt;0&lt;/sub&gt;, τ&lt;sub&gt;0&lt;/sub&gt;)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β</td>
<td>S.E.</td>
<td>Z</td>
</tr>
<tr>
<td>Intercept</td>
<td>-1.16 (0.10)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Palliative</td>
<td>0.00</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Surgery</td>
<td>-4.54 (1.23)</td>
<td>3.68</td>
<td>-4.16 (0.87) 4.81</td>
</tr>
<tr>
<td>Chemo</td>
<td>-0.62 (0.33)</td>
<td>1.91</td>
<td>-0.54 (0.35) 1.56</td>
</tr>
<tr>
<td>Radio</td>
<td>-1.47 (0.22)</td>
<td>6.67</td>
<td>-1.54 (0.22) 7.00</td>
</tr>
<tr>
<td>C+R</td>
<td>-5.57 (1.69)</td>
<td>3.29</td>
<td>-5.28 (1.12) 4.74</td>
</tr>
<tr>
<td>Shape 1</td>
<td>α</td>
<td>S.E.</td>
<td>Z</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.15 (0.08)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Palliative</td>
<td>0.00</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Surgery</td>
<td>0.46 (0.38)</td>
<td>1.22</td>
<td>0.36 (0.21) 1.76</td>
</tr>
<tr>
<td>Chemo</td>
<td>-0.27 (0.16)</td>
<td>1.68</td>
<td>-0.01 (0.14) 0.07</td>
</tr>
<tr>
<td>Radio</td>
<td>0.10 (0.14)</td>
<td>0.73</td>
<td>0.26 (0.07) 3.57</td>
</tr>
<tr>
<td>C+R</td>
<td>0.95 (0.32)</td>
<td>2.97</td>
<td>0.94 (0.17) 5.57</td>
</tr>
<tr>
<td>Shape 2</td>
<td>τ</td>
<td>S.E.</td>
<td>Z</td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.11 (0.27)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Palliative</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Surgery</td>
<td>1.03 (1.71)</td>
<td>0.60</td>
<td>—</td>
</tr>
<tr>
<td>Chemo</td>
<td>-10.52 (12.24)</td>
<td>0.86</td>
<td>—</td>
</tr>
<tr>
<td>Radio</td>
<td>-0.96 (1.01)</td>
<td>0.95</td>
<td>—</td>
</tr>
<tr>
<td>C+R</td>
<td>0.06 (1.11)</td>
<td>0.05</td>
<td>—</td>
</tr>
<tr>
<td>ℓ(θ)</td>
<td>-1925.77</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>AIC</td>
<td>3881.54</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Δ&lt;sub&gt;AIC&lt;/sub&gt;</td>
<td>5.68</td>
<td>0.00</td>
<td>—</td>
</tr>
</tbody>
</table>

Note also that the only statistically significant α coefficient in the full Burr MPR model is that of the C+R group. Thus, we may have tried removing treatment from both shape components (α and τ) leading to the Burr SPR model (shown in the third column) which has increased AIC; hence, we would not choose this model. This is a very simple example of variable selection...
within an MPR model. Of course variable selection in standard SPR models is more straightforward as there is only one regression component in which covariates can appear. We consider MPR variable selection more generally in Section 5.4.

Another noteworthy feature of MPR models, evident in Table 5.2, is that estimated regression coefficients (and standard errors) of a covariate in one component can vary somewhat upon removal of the covariate from another regression component. A good example of this behaviour is noticeable in the $\beta$ coefficients of both surgery and the combined treatment across the three models. This is due to correlation across the distributional parameters (see Section 3.6.3) - a matter which is discussed further, along with its potential consequences for hypothesis testing, in Section 5.4. Moreover, the correlation structure of estimates is investigated by simulation in Section 5.5.1.

**Example 5.3. MPR Piecewise Exponential Analysis of Lung Cancer Data**

Continuing with the single factor (treatment covariate) analysis of lung cancer data, we now look at the piecewise exponential MPR model. The number of time intervals was varied from one (giving an SPR exponential model) to
fifteen. Figure 5.8 shows the $AIC$ differences for the fifteen models considered. Based on this, we will consider the model with four time intervals (a two-interval model is also plausible). Note that the intervals were constructed so that the number of events in each one is approximately equal. In particular, the four-interval model has $I_1 = [0.0, 1.3)$, $I_2 = [1.3, 3.2)$, $I_3 = [3.2, 6.8)$ and $I_4 = [6.8, 21.0)$, respectively, with approximately 168 events contained in each.

Table 5.3. Piecewise Exponential MPR Model Fit

|                | $\hat{\beta}(t)$ | S.E. | $|Z|$ | $\hat{\beta}(t)$ | S.E. | $|Z|$ |
|----------------|-------------------|------|------|-------------------|------|------|
| $t \in [0.0, 1.3)$ |                   |      |      |                   |      |      |
| Intercept      | -1.24 (0.09)      | ——   | ——   | -1.45 (0.10)      | ——   | ——   |
| Palliative     | 0.00              | ——   | ——   | 0.00              | ——   | ——   |
| Surgery        | $-\infty$         | ——   | ——   | -2.45 (0.59)      | 4.19  |      |
| Chemo          | -0.21 (0.30)      | 0.69 |      | -0.78 (0.42)      | 1.86  |      |
| Radio          | -1.51 (0.24)      | 6.32 |      | -0.63 (0.17)      | 3.71  |      |
| C+R            | $-\infty$         | ——   | ——   | -2.71 (1.00)      | 2.70  |      |
| $t \in [1.3, 3.2)$ |                   |      |      |                   |      |      |
| Intercept      | -1.84 (0.11)      | ——   | ——   | -2.11 (0.13)      | ——   | ——   |
| Palliative     | 0.00              | ——   | ——   | 0.00              | ——   | ——   |
| Surgery        | -2.15 (0.46)      | 4.67 |      | -1.55 (0.31)      | 5.03  |      |
| Chemo          | -0.34 (0.35)      | 0.97 |      | -0.13 (0.34)      | 0.38  |      |
| Radio          | -0.37 (0.17)      | 2.19 |      | 0.07 (0.18)       | 0.40  |      |
| C+R            | -0.62 (0.35)      | 1.76 |      | 0.20 (0.28)       | 0.72  |      |
| $t \in [3.2, 6.8)$ |                   |      |      |                   |      |      |
| Intercept      | -1.84 (0.11)      | ——   | ——   | -2.11 (0.13)      | ——   | ——   |
| Palliative     | 0.00              | ——   | ——   | 0.00              | ——   | ——   |
| Surgery        | -2.15 (0.46)      | 4.67 |      | -1.55 (0.31)      | 5.03  |      |
| Chemo          | -0.34 (0.35)      | 0.97 |      | -0.13 (0.34)      | 0.38  |      |
| Radio          | -0.37 (0.17)      | 2.19 |      | 0.07 (0.18)       | 0.40  |      |
| C+R            | -0.62 (0.35)      | 1.76 |      | 0.20 (0.28)       | 0.72  |      |
| $t \in [6.8, 21)$ |                   |      |      |                   |      |      |

The estimated regression coefficients for the model are shown in Table 5.3. The results are consistent with the other models fitted: all treatments reduce the hazard, surgery is significant over the full time range, chemotherapy is not significant and both radiotherapy and the combined treatment become non-significant later in time. Note that there were no events in the surgery
or combined treatment groups in the first time interval. This led to infinite (negative) coefficients as discussed in Appendix B.2.6.

The hazard ratio plots are given in Fig. 5.9. The resemblance with those previously seen in Fig. 5.7 is clear, increasing credence in our earlier findings since the piecewise exponential model does not impose any structure on the data. Essentially, this model lets the data speak for themselves (within the framework of a parametric model) but it does have disadvantages: firstly, it is non-continuous in time and, secondly, its parameters can become inestimable if the number of time-intervals is large (see Appendix B.2.6).

![Hazard ratios with confidence intervals for the piecewise exponential MPR model.](image)

**Figure 5.9.** Hazard ratios with confidence intervals for the piecewise exponential MPR model.

### 5.3 The Hazard Ratio: Covariate Dependence

In Sections 5.2.1 - 5.2.6 we saw that MPR hazard ratios typically depend on terms such as $\tilde{x}^T \beta$ (defined in (5.3)). In other words, MPR hazard ratios are conditional on a specific covariate profile (except in the case of the Gompertz or piecewise exponential models). Up to now (Examples 5.1 - 5.3) we have considered single-factor analyses where, of course, the only other terms in the regression components are the intercepts, i.e., $\tilde{x}^T \beta = \beta_0$. Thus, the issue
of covariate-dependent hazard ratios did not arise. These simple examples serve as a good starting point in exhibiting the increased utility of multi-parameter regression over single-parameter regression analyses. Multi-factor analyses, however, are typically of greater scientific interest; covariate effects are adjusted in the presence of the other covariates in the model\(^3\). In this multi-factor case, covariate-dependent hazard ratios do arise (i.e., \(\tilde{x}^T \beta \neq \beta_0\)).

In the notation of Section 5.2, we first assume an MPR model with scale and shape regression components \(g(\lambda) = x^T \beta = c\beta_1 + \tilde{x}^T \beta\) and \(h(\gamma) = z^T \alpha = c\alpha_1 + \tilde{z}^T \alpha\), respectively, where \(c\) is a binary covariate. The hazard ratio, at time \(t\), for the covariate \(c\) is therefore

\[\psi(t \mid \tilde{x}, \tilde{z}) = \frac{\lambda(t \mid c = 1, \tilde{x}, \tilde{z})}{\lambda(t \mid c = 0, \tilde{x}, \tilde{z})},\]  

(5.27)

which differs from (5.4) notationally in the sense that we now explicitly show the dependence on other covariates via \(\tilde{x}\) and \(\tilde{z}\) respectively, i.e., the effect of \(c\) is conditional on the other covariates in the model via \((\tilde{x}, \tilde{z})\). Thus, one can evaluate (5.27) for different values of \((\tilde{x}, \tilde{z})\), e.g., setting \((\tilde{x}, \tilde{z})\) to values corresponding to high/low risk groups may be of interest. Clearly, examining the hazard ratio for different covariate values may be quite informative. However, Karrison (1987) noted that this approach gives rise to a multiple comparisons problem (in the context of a different covariate-dependent quantity). Essentially, there may be a very large, potentially infinite, number of possible \((\tilde{x}, \tilde{z})\) values which is not helpful in making clear statements about the individual effect of \(c\), the covariate of interest\(^4\).

In light of the above, an overall adjusted measure of the effect of \(c\) is required. Hence, we define the overall hazard ratio

\[\bar{\psi}(t) = \frac{1}{n} \sum_{i=1}^{n} \psi(t \mid \tilde{x}_i, \tilde{z}_i),\]  

(5.28)

\(^3\)Of course, this does not imply that single-factor MPR analyses are not of interest in their own right. In Example 5.1 we found that the MPR approach can enable us to match the Kaplan-Meier curves very closely (much more so than SPR) which is noteworthy as the KM curves are closest to the data by definition (Section 1.3). Thus, the MPR models are very close to the data in this case whilst providing more insight than KM curves.

\(^4\)However, if the analysis is based on a small number of categorical covariates (say, two or three), evaluating \(\psi(t \mid \tilde{x}, \tilde{z})\) at all values of \((\tilde{x}, \tilde{z})\) seems appropriate and useful.
i.e., evaluate the hazard ratio for \( c \) at every value of \((\tilde{x}, \tilde{z})\) observed in the dataset and then compute the average. Quantities of this type have been considered previously by many authors, e.g., Nelder & Lane (1982) in generalized linear models, Shen & Fleming (1997) who estimated mean survival, Karrison (1987), Zucker (1998) and Chen & Tsiatis (2001) in the study of restricted mean life (a quantity we consider in Section 6.3.1) and Martinussen & Pipper (2013) in relation to a concordance measure. Alternatively we may wish to evaluate the average-covariates hazard ratio

\[
\psi(t \mid \tilde{x}, \tilde{z}),
\]

where \((\tilde{x}, \tilde{z})\) is an “average of the covariates” value. In this case it seems reasonable to set each categorical covariate to its modal class and each continuous covariate to its arithmetic mean.

Of course \( \bar{\psi}(t) \neq \psi(t \mid \tilde{x}, \tilde{z}) \) in general and, furthermore, these two measures have different interpretations: \( \psi(t \mid \tilde{x}, \tilde{z}) \) is the effect of \( c \) for an average individual whereas \( \bar{\psi}(t) \) is the average effect of \( c \) over all individuals in the dataset. However, often the two measures are close numerically (although \( \bar{\psi}(t) \) has wider confidence intervals - see Figs. 5.11 - 5.19). A third possibility is to approximate \( \log \lambda(t \mid x, z) \) using a linear model, thus separating covariate effects (at time \( t \)) in a more familiar manner, i.e., through regression coefficients which, upon exponentiation, provide estimates of the hazard ratios. In Chapter 6 we discuss a (least squares) solution to approximating implied regression models; in particular Section 6.4 deals with the hazard ratio.

It is worth noting that the hazard ratio for the Gompertz MPR model, given in (5.10), depends only on the scale and shape coefficients of the covariate of interest, \( c \). Thus, the hazard ratio is fully summarised by these two values and does not depend on other covariates. Similarly, the piecewise exponential MPR hazard ratio, (5.26), is summarised by the \( m \) regression coefficients (one for each time interval) corresponding to the covariate \( c \). This is noteworthy as we do not need to calculate \( \bar{\psi}(t) \), i.e., \( \bar{\psi}(t) = \psi(t) \) for either of these MPR models which simplifies data analysis.

**Example 5.4. Full Covariate Analysis of Lung Cancer Data: Log-Logistic MPR Model**

We continue with the analysis of the lung cancer data. In our previous
analyses, we investigated the unadjusted effect of treatment. We now carry out a full covariate analysis of the data using the log-logistic MPR model.\(^5\)

Unsurprisingly, the \textit{AIC} value for this full model (\(= 3683.91\)) is much lower than any of the single factor models considered in previous examples, i.e., the variation in the data is better explained by considering the effects of all covariates. However, it is clear from the Z-scores in Table 5.4 that a more parsimonious model (and hence lower \textit{AIC}) is achievable. Variable selection, when applied to MPR models, is complicated by the fact that there exists correlation across regression components. Figure 5.10 shows the correlation matrix for the estimated regression coefficients of the fitted model.\(^6\) The pattern of correlation is clear: \((\hat{\beta}, \hat{\alpha})\) pairs corresponding to the same covariate are highly (negatively) correlated whereas other estimates are relatively uncorrelated. Thus, variable selection procedures should be adapted to account for this correlation (see Section 5.4). At this point however, we consider the full model (i.e., saturating the scale and shape) without attempting to reduce its dimension.

Inspecting the individual regression coefficients in Table 5.4 may give us some idea of the effect of particular covariates on the hazard (e.g., as was done in Example 5.1), however it is more instructive to look at the hazard ratios which quantify the total effect. Of course, in this full covariate analysis (i.e., multi-factor model) hazard ratios depend on the specific covariate profile via the presence of \(\tilde{x}\) and \(\tilde{z}\). As discussed previously, we eliminate this covariate-dependence by calculating overall hazard ratios, defined in (5.28). Figures 5.11 - 5.19 show the overall hazard ratios for the log-logistic MPR analysis of the lung cancer data with confidence intervals calculated using the method of m.l.e. simulation (Section 2.3.2). The average-covariates hazard

\(^5\)We also considered Weibull and time-dependent logistic MPR models whose \textit{AIC} values are 3700.88 and 3718.42 respectively. Thus, the log-logistic model has a considerably lower \textit{AIC} value (\(= 3683.91\)). It is noteworthy that this was also the case in Examples 4.2 and 5.1. Moreover, the Burr MPR model has \textit{AIC} = 3682.8 providing only a marginal improvement over the log-logistic model.

\(^6\)Recall that the variance-covariance matrix for the estimated parameters is obtained by inverting the observed information matrix (Section 2.3). Thus, from the variance-covariance matrix, we can work out the corresponding correlation matrix for the estimated parameters.
## Table 5.4. Log-Logistic MPR Full Covariate Analysis

<table>
<thead>
<tr>
<th>Scale</th>
<th>Shape</th>
<th>β</th>
<th>S.E.</th>
<th>Z</th>
<th>α</th>
<th>S.E.</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td></td>
<td>-3.70</td>
<td>(0.80)</td>
<td>——</td>
<td>0.10</td>
<td>(0.22)</td>
<td>——</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td>0.00</td>
<td>——</td>
<td>——</td>
<td>0.00</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>Palliative</td>
<td></td>
<td>0.00</td>
<td>——</td>
<td>——</td>
<td>0.00</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>Surgery</td>
<td></td>
<td>-1.67</td>
<td>(0.90)</td>
<td>1.86</td>
<td>-0.03</td>
<td>(0.20)</td>
<td>0.13</td>
</tr>
<tr>
<td>Chemo</td>
<td></td>
<td>-0.33</td>
<td>(0.43)</td>
<td>0.77</td>
<td>-0.16</td>
<td>(0.15)</td>
<td>1.08</td>
</tr>
<tr>
<td>Radio</td>
<td></td>
<td>-1.16</td>
<td>(0.25)</td>
<td>4.72</td>
<td>0.21</td>
<td>(0.07)</td>
<td>2.85</td>
</tr>
<tr>
<td>C+R</td>
<td></td>
<td>-5.07</td>
<td>(1.22)</td>
<td>4.17</td>
<td>0.64</td>
<td>(0.19)</td>
<td>3.34</td>
</tr>
<tr>
<td>Age Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 50</td>
<td></td>
<td>0.00</td>
<td>——</td>
<td>——</td>
<td>0.00</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>50 - 60</td>
<td></td>
<td>-1.07</td>
<td>(0.52)</td>
<td>2.04</td>
<td>0.40</td>
<td>(0.16)</td>
<td>2.51</td>
</tr>
<tr>
<td>60 - 70</td>
<td></td>
<td>-1.16</td>
<td>(0.46)</td>
<td>2.49</td>
<td>0.44</td>
<td>(0.15)</td>
<td>3.02</td>
</tr>
<tr>
<td>70 - 80</td>
<td></td>
<td>-0.91</td>
<td>(0.47)</td>
<td>1.92</td>
<td>0.32</td>
<td>(0.15)</td>
<td>2.16</td>
</tr>
<tr>
<td>80 +</td>
<td></td>
<td>-0.79</td>
<td>(0.50)</td>
<td>1.58</td>
<td>0.35</td>
<td>(0.17)</td>
<td>2.10</td>
</tr>
<tr>
<td>WHO Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td></td>
<td>0.00</td>
<td>——</td>
<td>——</td>
<td>0.00</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>Light Work</td>
<td></td>
<td>-0.04</td>
<td>(0.53)</td>
<td>0.07</td>
<td>-0.01</td>
<td>(0.12)</td>
<td>0.04</td>
</tr>
<tr>
<td>No Work</td>
<td></td>
<td>1.09</td>
<td>(0.51)</td>
<td>2.11</td>
<td>-0.08</td>
<td>(0.13)</td>
<td>0.66</td>
</tr>
<tr>
<td>&gt; 50% Bed</td>
<td></td>
<td>1.76</td>
<td>(0.53)</td>
<td>3.34</td>
<td>-0.06</td>
<td>(0.14)</td>
<td>0.45</td>
</tr>
<tr>
<td>Bedbound</td>
<td></td>
<td>2.38</td>
<td>(0.64)</td>
<td>3.71</td>
<td>-0.02</td>
<td>(0.21)</td>
<td>0.11</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td>0.00</td>
<td>——</td>
<td>——</td>
<td>0.00</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td>0.15</td>
<td>(0.18)</td>
<td>0.80</td>
<td>-0.04</td>
<td>(0.06)</td>
<td>0.78</td>
</tr>
<tr>
<td>Smoker</td>
<td></td>
<td>0.00</td>
<td>——</td>
<td>——</td>
<td>0.00</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td>-0.01</td>
<td>(0.29)</td>
<td>0.05</td>
<td>0.15</td>
<td>(0.09)</td>
<td>1.62</td>
</tr>
<tr>
<td>Ex-smoker</td>
<td></td>
<td>-0.02</td>
<td>(0.30)</td>
<td>0.07</td>
<td>0.13</td>
<td>(0.09)</td>
<td>1.35</td>
</tr>
<tr>
<td>Missing</td>
<td></td>
<td>0.21</td>
<td>(0.52)</td>
<td>0.40</td>
<td>-0.03</td>
<td>(0.20)</td>
<td>0.15</td>
</tr>
<tr>
<td>Cell Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Squamous</td>
<td></td>
<td>0.00</td>
<td>——</td>
<td>——</td>
<td>0.00</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>Small</td>
<td></td>
<td>1.06</td>
<td>(0.33)</td>
<td>3.20</td>
<td>0.07</td>
<td>(0.11)</td>
<td>0.64</td>
</tr>
<tr>
<td>Adeno.</td>
<td></td>
<td>0.04</td>
<td>(0.36)</td>
<td>0.12</td>
<td>0.09</td>
<td>(0.10)</td>
<td>0.93</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td>0.38</td>
<td>(0.25)</td>
<td>1.54</td>
<td>0.00</td>
<td>(0.07)</td>
<td>0.56</td>
</tr>
<tr>
<td>Metastases</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
<td>0.00</td>
<td>——</td>
<td>——</td>
<td>0.00</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td>1.58</td>
<td>(0.34)</td>
<td>4.57</td>
<td>-0.15</td>
<td>(0.08)</td>
<td>1.87</td>
</tr>
<tr>
<td>Missing</td>
<td></td>
<td>0.86</td>
<td>(0.37)</td>
<td>2.33</td>
<td>-0.12</td>
<td>(0.09)</td>
<td>1.31</td>
</tr>
<tr>
<td>Sodium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 136 mmol/l</td>
<td></td>
<td>0.00</td>
<td>——</td>
<td>——</td>
<td>0.00</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>&lt; 136 mmol/l</td>
<td></td>
<td>0.04</td>
<td>(0.18)</td>
<td>2.42</td>
<td>0.003</td>
<td>(0.05)</td>
<td>0.06</td>
</tr>
<tr>
<td>Missing</td>
<td></td>
<td>-1.06</td>
<td>(0.55)</td>
<td>1.94</td>
<td>0.36</td>
<td>(0.15)</td>
<td>2.41</td>
</tr>
<tr>
<td>Albumen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 35 g/l</td>
<td></td>
<td>0.00</td>
<td>——</td>
<td>——</td>
<td>0.00</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>&lt; 35 g/l</td>
<td></td>
<td>0.73</td>
<td>(0.20)</td>
<td>3.68</td>
<td>-0.01</td>
<td>(0.06)</td>
<td>0.13</td>
</tr>
<tr>
<td>Missing</td>
<td></td>
<td>0.59</td>
<td>(0.35)</td>
<td>1.70</td>
<td>-0.03</td>
<td>(0.11)</td>
<td>0.27</td>
</tr>
</tbody>
</table>

\[ \ell(\hat{\theta}) = -1789.95 \quad AIC = 3683.91 \]
5.3. THE HAZARD RATIO: COVARIATE DEPENDENCE

The hazard ratios, (5.29), are also shown for comparison.

Naturally, the effect of treatment is of special interest from a medical point of view and, hence, we discuss Fig. 5.11 first in detail. Both the overall and average-covariates hazard ratios are numerically very similar for all treatments. Thus, in this particular case, it does not matter whether we wish to look at the effects over all individuals or for a common individual. Comparing these adjusted hazard ratios to the unadjusted ones in Fig. 4.3 (middle column = log-logistic model), we can see that the trajectories are similar in both cases. However, the effects of surgery, radiotherapy and the combined

---

7 All covariates are categorical in the lung cancer dataset and, hence, each is set to its modal class in the calculation of average-covariates hazard ratios. The modal classes are as follows: treatment = “palliative”, age group = “60 - 70”, WHO status = “no work”, sex = “male”, smoker = “yes”, cell type = “other”, metastases = “yes”, sodium = “≥ 136 mmol/l” and albumen = “≥ 35 g/l”.

---

Figure 5.10. Correlation matrix for estimated regression coefficients in the log-logistic MPR model.
treatment have all reduced slightly (closer to one). Conversely, the effect of chemotherapy has increased slightly. Furthermore, the confidence intervals for the adjusted hazard ratios are much wider than for the unadjusted case.

Clearly all treatments are beneficial (adjusted hazard ratios below one) for some period of time. Surgery offers a large reduction in the hazard (relative to palliative care) and its effect is statistically significant until about 12 - 13 months; it has the longest lasting effect. Chemotherapy has a modest beneficial effect for a brief period ending after about 4 - 5 months. Radiotherapy has a larger effect than chemotherapy but wears off after about 3 - 4 months. The combined treatment is very effective early in time (possibly superior to surgery in the first 2 - 3 months) but wears off after about 5 months.

We will now briefly summarise the effects of the other covariates (Figs. 5.12 - 5.19) via the overall adjusted hazard ratios. Note that the average-covariates lead to somewhat similar interpretations but differ mainly in that they have narrower confidence intervals. This is not surprising as these hazard ratios apply to a specific “average” individual whereas the overall hazard ratios take into account the variation across all individuals.
5.3. **THE HAZARD RATIO: COVARIATE DEPENDENCE**

- **Age group**: Older age groups have a greater risk of death but this is only statistically significant for groups 50 - 60 and 60 - 70 and only after 5 - 7 months. In other words, the risk is higher for older individuals who have lung cancer for longer periods of time.

- **WHO status**: The worse one’s health status is, the greater the risk of death (not statistically significantly for the light work group). As can be seen from the y-axis, the hazard ratios are very large in magnitude.

- **Sex**: Males are not significantly different from females.

- **Smoker**: Smokers / ex-smokers have a higher risk of death. Note that the confidence intervals include the value of one, but only just.

- **Cell type**: All cell types appear to be more hazardous than squamous cell but this is only statistically significant for the small cell type (for almost the entire time range).

- **Metastases**: Where metastases are present, the risk of death is much higher. This is significant for about 7 - 8 months. As can be seen from the y-axis, this hazard ratio is large in magnitude.

- **Sodium**: Lower sodium levels leads to a slightly higher risk of death in the first 6 months.

- **Albumen**: Lower albumen levels increase the risk of death (statistically significant for about 9 months).

Clearly, the hazard ratio plots contain useful information. However, we may be interested in their numeric values at some key time points. To this end, Table 5.5 contains the values of the overall adjusted hazard ratios at 1.75, 4.75 and 12 months respectively. These are the 25th, 50th and 75th percentiles as estimated using the Kaplan-Meier curve for the whole sample.
Figure 5.12. Age (reference: < 50) overall (solid) and average-covariates (dash) hazard ratios with confidence intervals for the log-logistic MPR model.

Figure 5.13. WHO status (reference: normal) overall (solid) and average-covariates (dash) hazard ratios with confidence intervals for the log-logistic MPR model.
5.3. THE HAZARD RATIO: COVARIATE DEPENDENCE

Figure 5.14. *Sex (reference: female) overall (solid) and average-covariates (dash) hazard ratios with confidence intervals for the log-logistic MPR model.*

Figure 5.15. *Smoker (reference: no) overall (solid) and average-covariates (dash) hazard ratios with confidence intervals for the log-logistic MPR model.*
Figure 5.16. Cell type (reference: squamous) overall (solid) and average-covariates (dash) hazard ratios with confidence intervals for the log-logistic MPR model.

Figure 5.17. Metastases (reference: no) overall (solid) and average-covariates (dash) hazard ratios with confidence intervals for the log-logistic MPR model.
5.3. THE HAZARD RATIO: COVARIATE DEPENDENCE

Figure 5.18. Sodium (reference: ≥ 136 mmol/l) overall (solid) and average-covariates (dash) hazard ratios with confidence intervals for the log-logistic MPR model.

Figure 5.19. Albumen (reference: ≥ 35 g/l) overall (solid) and average-covariates (dash) hazard ratios with confidence intervals for the log-logistic MPR model.
Table 5.5. Log-Logistic MPR Overall Hazard Ratios with C.I.s

<table>
<thead>
<tr>
<th>Treatment</th>
<th>$t = 1.75$</th>
<th>$t = 4.75$</th>
<th>$t = 12$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Palliative</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Surgery</td>
<td>0.26 (0.10, 0.60)*</td>
<td>0.38 (0.26, 0.58)*</td>
<td>0.58 (0.40, 0.95)*</td>
</tr>
<tr>
<td>Chemo</td>
<td>0.61 (0.38, 0.97)*</td>
<td>0.62 (0.40, 0.99)*</td>
<td>0.69 (0.45, 1.08)</td>
</tr>
<tr>
<td>Radio</td>
<td>0.61 (0.48, 0.77)*</td>
<td>0.95 (0.77, 1.19)</td>
<td>1.18 (0.96, 1.46)</td>
</tr>
<tr>
<td>C+R</td>
<td>0.04 (0.01, 0.21)*</td>
<td>0.31 (0.15, 0.76)*</td>
<td>1.20 (0.63, 2.37)</td>
</tr>
<tr>
<td>Age Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 50</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>50 - 60</td>
<td>0.90 (0.59, 1.49)</td>
<td>1.44 (0.92, 2.37)</td>
<td>1.74 (1.08, 2.98)*</td>
</tr>
<tr>
<td>60 - 70</td>
<td>0.91 (0.65, 1.48)</td>
<td>1.54 (1.01, 2.48)*</td>
<td>1.87 (1.19, 3.13)*</td>
</tr>
<tr>
<td>70 - 80</td>
<td>0.89 (0.62, 1.40)</td>
<td>1.28 (0.85, 2.01)</td>
<td>1.51 (0.97, 2.48)</td>
</tr>
<tr>
<td>80 +</td>
<td>1.00 (0.64, 1.67)</td>
<td>1.45 (0.86, 2.66)</td>
<td>1.66 (0.96, 2.95)</td>
</tr>
<tr>
<td>WHO Status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Light Work</td>
<td>0.96 (0.58, 1.67)</td>
<td>0.96 (0.69, 1.36)</td>
<td>0.97 (0.71, 1.38)</td>
</tr>
<tr>
<td>No Work</td>
<td>2.12 (1.25, 3.68)*</td>
<td>1.51 (1.04, 2.19)*</td>
<td>1.13 (0.81, 1.62)</td>
</tr>
<tr>
<td>&gt; 50% Bed</td>
<td>3.61 (2.06, 6.41)*</td>
<td>2.16 (1.40, 3.46)*</td>
<td>1.34 (0.92, 2.03)</td>
</tr>
<tr>
<td>Bedbound</td>
<td>5.74 (2.49, 13.91)*</td>
<td>2.94 (1.34, 6.63)*</td>
<td>1.56 (0.81, 2.88)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Male</td>
<td>1.04 (0.86, 1.27)</td>
<td>0.97 (0.79, 1.19)</td>
<td>0.95 (0.79, 1.13)</td>
</tr>
<tr>
<td>Smoker</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Yes</td>
<td>1.27 (0.93, 1.82)</td>
<td>1.42 (1.01, 2.03)*</td>
<td>1.38 (0.99, 2.00)</td>
</tr>
<tr>
<td>Ex-smoker</td>
<td>1.22 (0.90, 1.79)</td>
<td>1.35 (0.98, 1.91)</td>
<td>1.32 (0.95, 1.89)</td>
</tr>
<tr>
<td>Missing</td>
<td>1.11 (0.60, 2.19)</td>
<td>1.04 (0.53, 2.43)</td>
<td>1.00 (0.52, 2.16)</td>
</tr>
<tr>
<td>Cell Type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Squamous</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Small</td>
<td>2.49 (1.65, 3.79)*</td>
<td>2.01 (1.33, 3.22)*</td>
<td>1.50 (1.06, 2.24)*</td>
</tr>
<tr>
<td>Adeno.</td>
<td>1.23 (0.88, 1.75)</td>
<td>1.31 (0.93, 1.83)</td>
<td>1.26 (0.90, 1.71)</td>
</tr>
<tr>
<td>Other</td>
<td>1.26 (0.97, 1.65)</td>
<td>1.11 (0.87, 1.41)</td>
<td>1.02 (0.81, 1.29)</td>
</tr>
<tr>
<td>Metastases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Yes</td>
<td>2.75 (1.90, 4.04)*</td>
<td>1.67 (1.30, 2.12)*</td>
<td>1.12 (0.88, 1.41)</td>
</tr>
<tr>
<td>Missing</td>
<td>1.66 (1.13, 2.44)*</td>
<td>1.22 (0.91, 1.65)</td>
<td>0.99 (0.76, 1.28)</td>
</tr>
<tr>
<td>Sodium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 136 mmol/l</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>&lt; 136 mmol/l</td>
<td>1.40 (1.13, 1.70)*</td>
<td>1.26 (1.04, 1.55)*</td>
<td>1.12 (0.95, 1.37)</td>
</tr>
<tr>
<td>Missing</td>
<td>0.86 (0.47, 1.54)</td>
<td>1.61 (0.90, 3.32)</td>
<td>1.74 (1.08, 2.80)*</td>
</tr>
<tr>
<td>Albumen</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 35 g/l</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>&lt; 35 g/l</td>
<td>1.73 (1.37, 2.19)*</td>
<td>1.43 (1.14, 1.81)*</td>
<td>1.18 (0.97, 1.44)</td>
</tr>
<tr>
<td>Missing</td>
<td>1.51 (0.99, 2.16)</td>
<td>1.27 (0.85, 1.95)</td>
<td>1.09 (0.77, 1.60)</td>
</tr>
</tbody>
</table>

Note: Symbol “*” indicates that the C.I. does not contain the value one.
Example 5.5. Full Covariate Analysis of Lung Cancer Data: Gompertz MPR

We have mentioned that the Gompertz MPR model has the advantage of its hazard ratios being covariate-independent, obviating the need to average over the covariate distribution. Hence, we now show the results of the full covariate Gompertz analysis of the lung cancer data. This analysis mirrors the log-logistic analysis of Example 5.4 and is mainly included for the purpose of comparison with the earlier analysis.

Firstly, comparing Tables 5.6 and 5.4, we can see similarities in the signs, relative magnitudes and Z-scores of the $\beta$ and $\alpha$ coefficients. Thus, even though the models are structurally different, it seems they are roughly in agreement as to the size and significance of the various scale and shape effects. Note however that the $AIC$ value for the Gompertz model is quite a bit higher than the log-logistic model (3709.74 versus 3683.91).

We can see from Fig. 5.20 that the estimated regression coefficients are correlated in a manner which is virtually indistinguishable from the log-logistic case (Fig. 5.10). Thus, the correlation pattern that we observed earlier is not just a feature of the log-logistic model, but rather, it applies more generally. Indeed we have found in practice that this pattern emerges for all of the MPR models we consider. Furthermore, we also show by simulation (Section 5.5.1 and Appendix B.4) that the correlation pattern arises across a variety of different scenarios and models.

The Gompertz hazard ratio plots are given in Figs. 5.21 - 5.29. However, we will not go into detail interpreting these hazard ratios. Instead we briefly compare them to those from the log-logistic analysis (Figs. 5.11 - 5.19). We can see that the exact results from the two analyses are slightly different. For example, the Gompertz hazard ratios have wider confidence intervals later in time. Nonetheless, when the hazard ratios are compared side by side, it is apparent that the basic conclusions from both analyses are very similar. We reiterate the fact that the log-logistic model has a much lower $AIC$ value. Thus, while the Gompertz MPR hazard ratio has an appealingly simple functional form, this, of course, does not ensure that the model provides a superior fit.
### Table 5.6. Gompertz MPR Full Covariate Analysis

<table>
<thead>
<tr>
<th></th>
<th>Scale</th>
<th>Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}$</td>
<td>S.E.</td>
</tr>
<tr>
<td>Intercept</td>
<td>-2.92</td>
<td>(0.53)</td>
</tr>
<tr>
<td><strong>Treatment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palliative</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Surgery</td>
<td>-1.47</td>
<td>(0.48)</td>
</tr>
<tr>
<td>Chemo</td>
<td>-0.54</td>
<td>(0.32)</td>
</tr>
<tr>
<td>Radio</td>
<td>-0.65</td>
<td>(0.16)</td>
</tr>
<tr>
<td>C+R</td>
<td>-2.24</td>
<td>(0.50)</td>
</tr>
<tr>
<td><strong>Age Group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 50</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>50 - 60</td>
<td>-0.70</td>
<td>(0.40)</td>
</tr>
<tr>
<td>60 - 70</td>
<td>-0.59</td>
<td>(0.36)</td>
</tr>
<tr>
<td>70 - 80</td>
<td>-0.52</td>
<td>(0.37)</td>
</tr>
<tr>
<td>80 +</td>
<td>-0.48</td>
<td>(0.39)</td>
</tr>
<tr>
<td><strong>WHO Status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Light Work</td>
<td>-0.14</td>
<td>(0.32)</td>
</tr>
<tr>
<td>No Work</td>
<td>0.67</td>
<td>(0.31)</td>
</tr>
<tr>
<td>&gt; 50% Bed</td>
<td>1.11</td>
<td>(0.33)</td>
</tr>
<tr>
<td>Bedbound</td>
<td>1.53</td>
<td>(0.43)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Male</td>
<td>-0.02</td>
<td>(0.12)</td>
</tr>
<tr>
<td><strong>Smoker</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Yes</td>
<td>0.13</td>
<td>(0.20)</td>
</tr>
<tr>
<td>Ex-smoker</td>
<td>0.03</td>
<td>(0.21)</td>
</tr>
<tr>
<td>Missing</td>
<td>0.20</td>
<td>(0.38)</td>
</tr>
<tr>
<td><strong>Cell Type</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Squamous</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Small</td>
<td>0.76</td>
<td>(0.23)</td>
</tr>
<tr>
<td>Adeno.</td>
<td>0.21</td>
<td>(0.22)</td>
</tr>
<tr>
<td>Other</td>
<td>0.25</td>
<td>(0.16)</td>
</tr>
<tr>
<td><strong>Metastases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Yes</td>
<td>1.00</td>
<td>(0.20)</td>
</tr>
<tr>
<td>Missing</td>
<td>0.49</td>
<td>(0.22)</td>
</tr>
<tr>
<td><strong>Sodium</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 136 mmol/l</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>&lt; 136 mmol/l</td>
<td>0.32</td>
<td>(0.12)</td>
</tr>
<tr>
<td>Missing</td>
<td>-0.32</td>
<td>(0.33)</td>
</tr>
<tr>
<td><strong>Albumen</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 35 g/l</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>&lt; 35 g/l</td>
<td>0.60</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Missing</td>
<td>0.40</td>
<td>(0.24)</td>
</tr>
</tbody>
</table>

$\ell(\hat{\theta}) = -1802.87 \quad \text{AIC} = 3709.74$
Figure 5.20. Correlation matrix for estimated regression coefficients in the Gompertz MPR model.

Figure 5.21. Treatment (reference: palliative care) hazard ratios with confidence intervals for the Gompertz MPR model.
Figure 5.22. Age (reference: < 50) hazard ratios with confidence intervals for the Gompertz MPR model.

Figure 5.23. WHO status (reference: normal) hazard ratios with confidence intervals for the Gompertz MPR model.
5.3. THE HAZARD RATIO: COVARIATE DEPENDENCE

Figure 5.24. Sex (reference: female) hazard ratio with confidence intervals for the Gompertz MPR model.

Figure 5.25. Smoker (reference: no) hazard ratios with confidence intervals for the Gompertz MPR model.
Figure 5.26. Cell type (reference: squamous) hazard ratios with confidence intervals for the Gompertz MPR model.

Figure 5.27. Metastases (reference: no) hazard ratios with confidence intervals for the Gompertz MPR model.
5.3. THE HAZARD RATIO: COVARIATE DEPENDENCE

Figure 5.28. Sodium (reference: $\geq 136$ mmol/l) hazard ratios with confidence intervals for the Gompertz MPR model.

Figure 5.29. Albumen (reference: $\geq 35$ g/l) hazard ratios with confidence intervals for the Gompertz MPR model.
Example 5.6. Lung Cancer Cure Probabilities via the Gompertz MPR Model

A novel feature of the Gompertz model (mentioned in Section 1.4.3) is the fact that, when its shape parameter is negative, it has a survivor curve which does not fall to zero but rather to some positive limit\(^8\). This positive limit is the cure probability. Thus, the Gompertz MPR model implies a regression model for the cure probability,

\[
p_{\text{cure}}(x, z) = S(\infty \mid x, z) = \lim_{t \to \infty} \exp \left\{ -\frac{\exp(x^T \beta)}{z^T \alpha} \left[ \exp(z^T \alpha t) - 1 \right] \right\} = \begin{cases} 0 & \text{if } z^T \alpha > 0, \\ \exp \left[ \frac{\exp(x^T \beta)}{z^T \alpha} \right] & \text{if } z^T \alpha < 0. \end{cases} \tag{5.30} \]

From this we can compute the overall cure probability for a given group which, analogous to the overall hazard ratio defined in (5.28), is given by

\[
\overline{p}_{\text{cure}}(c = 1) = \frac{1}{n} \sum_{i=1}^{n} p_{\text{cure}}(c = 1, \tilde{x}_i, \tilde{z}_i), \tag{5.31}
\]

where \(c\) is a binary variable indicating membership of the group of interest and \(\tilde{x}\) and \(\tilde{z}\) are as defined in (5.3). In words, we fix \(c\) to the value of one and average over the full distribution of covariates in the dataset.

Continuing with the Gompertz analysis in Example 5.5, we now investigate the cure probabilities implied by this fitted model. These are shown in Table 5.7 with confidence intervals calculated using m.l.e. simulation (Section 2.3.2). As a reference, it is worth estimating the population cure probability by \(\sum_{i=1}^{n} p_{\text{cure}}(x_i, z_i)/n\) which, for this dataset, is 0.016; clearly there is very little chance of being cured of lung cancer. Furthermore, most of the cure probabilities in Table 5.7 are not statistically different from this reference probability. The only exceptions to this are the youngest individuals (i.e., < 50) with \(\overline{p}_{\text{cure}} = 0.144 (0.024, 0.279)\) and non-smokers with \(\overline{p}_{\text{cure}} = 0.07 (0.024, 0.206)\). It is also noteworthy that in the surgery group \(\overline{p}_{\text{cure}} = 0.1 (0.015, 0.315)\) which only just includes the reference probability.

One caveat to the above results is that calculation of cure probabilities represents extrapolation (i.e., to time infinity) and, therefore, the degree of

\(^8\)This is also the case for the time-dependent logistic model (Section 1.4.6).
### 5.3. The Hazard Ratio: Covariate Dependence

Table 5.7. Gompertz MPR Overall Cure Probabilities with C.I.s

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment</strong></td>
<td></td>
</tr>
<tr>
<td>Palliative</td>
<td>$0.017 (0.010, 0.059)$</td>
</tr>
<tr>
<td>Surgery</td>
<td>$0.100 (0.015, 0.315)$</td>
</tr>
<tr>
<td>Chemo</td>
<td>$0.021 (0.002, 0.179)$</td>
</tr>
<tr>
<td>Radio</td>
<td>$0.006 (0.002, 0.033)$</td>
</tr>
<tr>
<td>C+R</td>
<td>$0.000 (0.000, 0.024)$</td>
</tr>
<tr>
<td><strong>Age Group</strong></td>
<td></td>
</tr>
<tr>
<td>$&lt; 50$</td>
<td>$0.144 (0.024, 0.279)$</td>
</tr>
<tr>
<td>$50 - 60$</td>
<td>$0.002 (0.000, 0.035)$</td>
</tr>
<tr>
<td>$60 - 70$</td>
<td>$0.004 (0.002, 0.039)$</td>
</tr>
<tr>
<td>$70 - 80$</td>
<td>$0.015 (0.007, 0.068)$</td>
</tr>
<tr>
<td>$80 +$</td>
<td>$0.014 (0.004, 0.092)$</td>
</tr>
<tr>
<td><strong>WHO Status</strong></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>$0.034 (0.010, 0.153)$</td>
</tr>
<tr>
<td>Light Work</td>
<td>$0.012 (0.005, 0.058)$</td>
</tr>
<tr>
<td>No Work</td>
<td>$0.021 (0.011, 0.078)$</td>
</tr>
<tr>
<td>$&gt; 50%$ Bed</td>
<td>$0.016 (0.004, 0.076)$</td>
</tr>
<tr>
<td>Bedbound</td>
<td>$0.001 (0.000, 0.124)$</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>$0.016 (0.008, 0.060)$</td>
</tr>
<tr>
<td>Male</td>
<td>$0.016 (0.010, 0.052)$</td>
</tr>
<tr>
<td><strong>Smoker</strong></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>$0.070 (0.024, 0.206)$</td>
</tr>
<tr>
<td>Yes</td>
<td>$0.008 (0.004, 0.044)$</td>
</tr>
<tr>
<td>Ex-smoker</td>
<td>$0.009 (0.005, 0.047)$</td>
</tr>
<tr>
<td>Missing</td>
<td>$0.033 (0.000, 0.280)$</td>
</tr>
<tr>
<td><strong>Cell Type</strong></td>
<td></td>
</tr>
<tr>
<td>Squamous</td>
<td>$0.023 (0.011, 0.074)$</td>
</tr>
<tr>
<td>Small</td>
<td>$0.013 (0.002, 0.084)$</td>
</tr>
<tr>
<td>Adeno.</td>
<td>$0.006 (0.002, 0.043)$</td>
</tr>
<tr>
<td>Other</td>
<td>$0.018 (0.010, 0.061)$</td>
</tr>
<tr>
<td><strong>Metastases</strong></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>$0.023 (0.011, 0.080)$</td>
</tr>
<tr>
<td>Yes</td>
<td>$0.016 (0.008, 0.054)$</td>
</tr>
<tr>
<td>Missing</td>
<td>$0.018 (0.009, 0.064)$</td>
</tr>
<tr>
<td><strong>Sodium</strong></td>
<td></td>
</tr>
<tr>
<td>$\geq 136$ mmol/l</td>
<td>$0.018 (0.012, 0.057)$</td>
</tr>
<tr>
<td>$&lt; 136$ mmol/l</td>
<td>$0.013 (0.006, 0.054)$</td>
</tr>
<tr>
<td>Missing</td>
<td>$0.002 (0.000, 0.079)$</td>
</tr>
<tr>
<td><strong>Albumen</strong></td>
<td></td>
</tr>
<tr>
<td>$\geq 35$ g/l</td>
<td>$0.016 (0.009, 0.059)$</td>
</tr>
<tr>
<td>$&lt; 35$ g/l</td>
<td>$0.023 (0.011, 0.074)$</td>
</tr>
<tr>
<td>Missing</td>
<td>$0.008 (0.001, 0.074)$</td>
</tr>
</tbody>
</table>
belief in such probabilities depends on one’s belief in the model. It is well known that estimated tail probabilities (for survival data) have bad properties due to censoring and, of course, cure probabilities are extreme tail probabilities. Indeed, we have found this to be the case in Sections 3.6.5 and 3.6.6 and generally caution is advised in their use (extrapolation was also discussed in Section 3.6.2). Nonetheless, the results of this particular analysis do not appear to be unreasonable.

### 5.4 Hypothesis Testing and Variable Selection

It is natural to enquire if certain covariates affect the scale, $\lambda$, or the shape, $\gamma$, of the distribution by performing hypothesis tests on regression coefficients (or through the use of variable selection procedures). However, as we observe correlation across the regression components (mentioned in Examples 5.2, 5.4 and 5.5), these components cannot be considered separately. Thus, we must account for this correlation when carrying out any such testing.

Firstly we note that an orthogonal parametrisation is obviously attractive as the regression components are independent (i.e, $\lambda \perp \gamma \Rightarrow \beta \perp \alpha$) and, therefore, can be considered separately. However, most parametrisations are, of course, non-orthogonal. Furthermore, even if an orthogonal parametrisation does exist for a particular model, the parameters of a corresponding non-orthogonal version may have a more appealing interpretation. For example, the orthogonal Weibull model (Appendix C), when generalised to multi-parameter regression form, does not provide us with a test of proportional hazards (via the testing of $\alpha$-coefficients) as in the standard Weibull (see Appendix C.5) and so the standard form may be preferable. In relation to this matter, Cox (2006, sec. 6.4.4) said that “so far as parameters of interest are concerned, subject-matter interpretability has primacy.” In addition to this, orthogonality is defined in terms of the expected information matrix, $\mathcal{I} = E(I_\alpha)$, see Cox & Reid (1987), which is difficult to construct for
5.4. HYPOTHESIS TESTING AND VARIABLE SELECTION

survival data due to censoring. Indeed Cox & Reid (1987) did not consider censored data and derived the orthogonal Weibull model under the condition of full information (see Appendix C.4). Thus “orthogonal” parametrisations, defined in this way, do not retain orthogonality in the presence of censoring which we confirm by simulation in Section 5.5.1.

Clearly, we are much more likely to encounter the use of non-orthogonal models in practice (orthogonal models may be considered a special, often unattainable, case). Therefore we will only consider hypothesis testing within non-orthogonal models. First we will assume, as before, that the scale and shape regression components are \(g(\lambda) = x^T \beta = c \beta_1 + \tilde{x}^T \beta\) and \(h(\gamma) = z^T \alpha = c \alpha_1 + \tilde{z}^T \alpha\), respectively, where \(c\) is a binary covariate.

Note that in single-parameter regression models (SPR), covariates only appear in the scale regression. In this case, the importance of the covariate \(c\) is determined by testing if its estimated effect, \(\hat{\beta}_1\), differs statistically from zero. This effect is adjusted for the other covariates which, being an SPR model, appear in the same regression component as \(c\). However, this situation is generalised in the multi-parameter regression (MPR) case. In the two component (scale and shape) MPR model we are considering, there are two hypotheses of interest, namely: (i) \(H_0 : \beta_1 = 0\) and (ii) \(H_0 : \alpha_1 = 0\). These hypotheses cannot be tested independently due to the correlation that exists across the scale-shape space. The consequence of this correlation is as follows: the effect of \(c\) in one regression is adjusted for the other covariates in that regression and all covariates in the other regression. It is important to note that \(c\) itself appears in the other regression and \(\hat{\beta}_1\) and \(\hat{\alpha}_1\) will be highly correlated. Thus, the scale effect of \(c\) is significantly adjusted for its shape effect (and vice versa). It is clear then that it is not appropriate to judge the importance of \(c\) in one component without it being present in the other component; we can imagine scenarios where a covariate only becomes significant when present in both components. Of course, the individual hypothesis tests are carried out in the usual way, i.e., Wald tests based on the fact that

9Recall from Section 2.3.1 that \(I\) is not usable; its functional form is unknown due to the presence of censoring. Even in the uncensored case \(I\), while known, can be intractable.

10We have found in practice that scale and shape coefficients of the same covariate are highly correlated (see Examples 5.4 and 5.5). Furthermore, we investigate this by simulation in Section 5.5.1 and Appendix B.4.
(asymptotically) \( \hat{\beta}_1 \sim N(\beta_1, \text{var}(\hat{\beta}_1)) \) and \( \hat{\alpha}_1 \sim N(\alpha_1, \text{var}(\hat{\alpha}_1)) \), respectively, where the variances are obtained from the diagonal of the covariance matrix, \( \Sigma(\hat{\theta}) = I_o^{-1}(\hat{\theta}) \) (Section 2.3).

It is also of interest to determine the total (or joint) effect of \( c \), i.e., by testing (iii) \( H_0 : \beta_1 = \alpha_1 = 0 \). Using the fact that (asymptotically) \( (\hat{\beta}_1, \hat{\alpha}_1)^T \sim N([\beta_1, \alpha_1]^T, \Sigma_{\hat{\beta}_1, \hat{\alpha}_1}] \), where \( \Sigma_{\hat{\beta}_1, \hat{\alpha}_1} \) is the relevant \( 2 \times 2 \) covariance sub-matrix of \( \Sigma(\hat{\theta}) \), we have that

\[
[(\hat{\beta}_1, \hat{\alpha}_1)^T - (\beta_1, \alpha_1)^T] \Sigma_{\hat{\beta}_1, \hat{\alpha}_1}^{-1} [(\hat{\beta}_1, \hat{\alpha}_1)^T - (\beta_1, \alpha_1)^T] \sim \chi^2_2. \tag{5.32}
\]

Thus a p-value can be obtained by setting \( (\beta_1, \alpha_1)^T = (0, 0)^T \) in (5.32) and comparing this statistic to the \( \chi^2_2 \) distribution. Furthermore, the corresponding \( (1 - \alpha)100\% \) confidence ellipse for \( (\beta_1, \alpha_1)^T \) is given by the set of \( (\beta_1, \alpha_1)^T \) points defining the contour line such that (5.32) is equal to \( \chi^2_{2,1-\alpha} \) (see Friendly et al. (2013) for detailed account of confidence ellipses).

We now define \( M(x, z) \) to be a model with scale and shape covariate vectors \( x \) and \( z \), respectively. Hence, let \( M_0 = M(\bar{x}, \bar{z}) \) be a model where \( c \) does not appear (i.e., \( c \not\in \bar{x}, \bar{z} \)) and the three models which include \( c \) are: \( M_{\beta_1} = M(\bar{x} \cup c, \bar{z}) \), \( M_{\alpha_1} = M(\bar{x}, \bar{z} \cup c) \) and \( M_{\beta_1, \alpha_1} = M(\bar{x} \cup c, \bar{z} \cup c) \), respectively. Therefore, we can test hypotheses (i), (ii) and (iii) by comparing \( M_{\beta_1, \alpha_1} \) with \( M_{\alpha_1} \), \( M_{\beta_1} \) with \( M_{\beta_1, \alpha_1} \) with \( M_0 \), respectively. This can be done by means of likelihood ratio tests (as the models are nested) or using information criteria (see Section 2.2 and Burnham & Anderson (2002)); such approaches are more in line with variable selection procedures (Miller, 2002). Of course inferential problems associated with variable selection methods are well documented (Miller, 1984; Hurvich & Tsai, 1990; Zhang, 1992) and automatic selection procedures are much criticised. Nonetheless, such procedures may be useful when the number of covariates is large, so that methods for MPR models are required (just as they are for SPR models). To this end we discuss MPR forward selection in the proceeding paragraph.

Algorithm 1 below shows the pseudocode for MPR forward selection (the corresponding R code is given in Appendix B.5.3) which we also describe here in words. The algorithm begins by fitting the null model \( M(x = \emptyset, z = \emptyset) \). Thereafter, in each iteration of the main \textbf{while} loop, the algorithm seeks a model which is better than the current model; “better” in this context...
Algorithm 1: Multi-Parameter Regression Forward Selection

\[ x \leftarrow z \leftarrow \emptyset; \]

\[ \text{Fit model } M_{\text{new}} \leftarrow M(x, z); \]

\[ \text{finish } \leftarrow \text{false}; \]

\[ \textbf{while } \text{finish } = \text{false} \textbf{ do} \]

\[ M_{\text{old}} \leftarrow M_{\text{new}}; \]

\[ \textbf{foreach } c \notin x \textbf{ do} \]

\[ \text{Fit model } M^* \leftarrow M(x \cup c, z); \]

\[ \text{if } M^* \text{ is "better than" } M_{\text{new}} \textbf{ then} \]

\[ M_{\text{new}} \leftarrow M^*; \]

\[ \textbf{end} \]

\[ \textbf{end} \]

\[ \textbf{foreach } c \notin z \textbf{ do} \]

\[ \text{Fit model } M^* \leftarrow M(x, z \cup c); \]

\[ \text{if } M^* \text{ is "better than" } M_{\text{new}} \textbf{ then} \]

\[ M_{\text{new}} \leftarrow M^*; \]

\[ \textbf{end} \]

\[ \textbf{end} \]

\[ \textbf{foreach } c \notin x, z \textbf{ do} \]

\[ \text{Fit model } M^* \leftarrow M(x \cup c, z \cup c); \]

\[ \text{if } M^* \text{ is "better than" } M_{\text{new}} \textbf{ then} \]

\[ M_{\text{new}} \leftarrow M^*; \]

\[ \textbf{end} \]

\[ \textbf{end} \]

\[ \text{if } M_{\text{old}} \neq M_{\text{new}} \textbf{ then} \]

\[ \text{Update } x \text{ and } z; \]

\[ \textbf{else} \]

\[ \text{finish } \leftarrow \text{true} \]

\[ \textbf{end} \]

\[ \textbf{end} \]

\[ \textbf{Result: } M_{\text{new}} \text{ (the "best" model)} \]
may be based on likelihood ratio tests, $AIC$, $BIC$, etc. Of course, at each such iteration, there are candidate covariates to be considered for inclusion in the scale $\{c \mid c \notin x\}$, the shape $\{c \mid c \notin z\}$ and for simultaneous inclusion $\{c \mid c \notin x,z\}$. Hence, this main loop is composed of three for loops where all of these new models are fitted and compared to the current model. If, at the end of the while loop, the algorithm finds a better model, then this becomes the current model and another iteration of the loop commences. Otherwise, the loop ends as the best model has been found. Extending the algorithm to include backward steps as well as forward steps is straightforward (Appendix B.5.3 shows the output of such a procedure applied to the lung cancer data), and, furthermore, the starting model need not be the null model.

If one were not aware of the issues discussed in this section, then other strategies for determining important covariates may seem reasonable. We now discuss some possibilities that such a user may implement. These are all special cases of Algorithm 1 and are, therefore, sub-optimal.

- **Standard ("good") advice** suggests that one should start with a simpler model and then increase complexity. To this end a user may choose to start by considering the scale only (i.e., the standard SPR model). Once the best SPR model has been found, the user then proceeds to include covariates in the shape to try to improve the model further (i.e., the model is now MPR). This is equivalent to running Algorithm 1 with the first for loop only, to find the best SPR model, followed by running the algorithm again (starting at this SPR model) with the second for loop only.

- **Alternatively**, one may realise that the whole regression space (scale and shape) should be considered at once. However, analogous to standard selection procedures which include covariates one-at-a-time, this user does not implement the simultaneous step (i.e., the third for loop of Algorithm 1). While this procedure is more general than the first-mentioned, it may still miss out on covariates which only show significance when present in both components (as discussed previously).

- A third possibility is a user who only considers simultaneous steps, i.e., covariates are either in the model (in both components) or not in the
model. While this approach does account for the fact that $\hat{\beta}$ and $\hat{\alpha}$ coefficients of the same covariate are highly correlated, it may lead to an over-parametrised model. Conversely, a covariate could be missed using this method if its effect in one component is highly insignificant and, hence, renders the joint component effect statistically insignificant.

Another option for selecting covariates is the popular lasso method (Tibshirani, 1996). While we have not explored the use of this method in MPR models, it seems reasonable to assume that alteration of the penalty term is required in order to consider $(\beta, \alpha)$ pairs simultaneously. It is likely that this can be achieved within the framework of the group lasso (Yuan & Lin, 2006; Yuan et al., 2009; Zhao et al., 2009; Bach et al., 2012) which allows certain covariate structures to be built into the lasso penalty. Whereas the standard lasso defines all terms to be equal, the group lasso has been developed to handle the groupings that arise due to derived (binary) variables representing a categorical covariate and the hierarchy of terms that exists in models with interactions\textsuperscript{11}. In a similar spirit Bien et al. (2013) have developed the hierarchical lasso which may also be useful for our intended purpose in MPR modelling. Furthermore, the Bayesian lasso (Park & Casella, 2008; Hans, 2009, 2010) may lend itself well to our MPR case where, for example, we may wish to put greater prior probability on certain $(\beta, \alpha)$ pairs being jointly non-zero. Further work in this area is required.

**Example 5.7. Log-Logistic MPR Full Covariate Analysis Continued**

In Example 5.4 we carried out a full covariate analysis of the lung cancer data but noted that a more parsimonious model is achievable (see Table 5.4). Thus, we continue our analysis with a view to removing unnecessary covariates from the model. Table 5.8 and Fig. 5.30 show the p-values and 95% confidence regions for the scale, shape and joint effects respectively\textsuperscript{12}.

\textsuperscript{11}When interaction terms are present, it is proper practice to include all lower order terms, e.g., if $A.B$ is in the model then $A$ and $B$ should also be present. This practice is known as respecting marginality and has been discussed by Nelder (1977, 1998) and McCullagh & Nelder (1989). If this is not done then one imposes restrictions on the interpretation of the model which are usually undesirable.

\textsuperscript{12}The p-values and confidence ellipses for the joint effects are calculated using (5.32).
### Table 5.8. Log-Logistic MPR: Significance of Scale and Shape Effects

<table>
<thead>
<tr>
<th></th>
<th>Scale</th>
<th>Shape</th>
<th>Joint</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$</td>
<td>Z</td>
<td>$</td>
</tr>
<tr>
<td><strong>Treatment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palliative</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Surgery</td>
<td>1.86</td>
<td>0.063</td>
<td>0.13</td>
</tr>
<tr>
<td>Chemo</td>
<td>0.77</td>
<td>0.443</td>
<td>1.08</td>
</tr>
<tr>
<td>Radio</td>
<td>4.72</td>
<td>0.000</td>
<td>2.85</td>
</tr>
<tr>
<td>C+R</td>
<td>4.17</td>
<td>0.000</td>
<td>3.34</td>
</tr>
<tr>
<td><strong>Age Group</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 50</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>50 - 60</td>
<td>2.04</td>
<td>0.041</td>
<td>2.51</td>
</tr>
<tr>
<td>60 - 70</td>
<td>2.49</td>
<td>0.013</td>
<td>3.02</td>
</tr>
<tr>
<td>70 - 80</td>
<td>1.92</td>
<td>0.055</td>
<td>2.16</td>
</tr>
<tr>
<td>80 +</td>
<td>1.58</td>
<td>0.115</td>
<td>2.10</td>
</tr>
<tr>
<td><strong>WHO Status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Light Work</td>
<td>0.07</td>
<td>0.942</td>
<td>0.04</td>
</tr>
<tr>
<td>No Work</td>
<td>2.11</td>
<td>0.035</td>
<td>0.66</td>
</tr>
<tr>
<td>&gt; 50% Bed</td>
<td>3.34</td>
<td>0.001</td>
<td>0.45</td>
</tr>
<tr>
<td>Bedbound</td>
<td>3.71</td>
<td>0.000</td>
<td>0.11</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Male</td>
<td>0.80</td>
<td>0.423</td>
<td>0.78</td>
</tr>
<tr>
<td><strong>Smoker</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Yes</td>
<td>0.05</td>
<td>0.964</td>
<td>1.62</td>
</tr>
<tr>
<td>Ex-smoker</td>
<td>0.07</td>
<td>0.944</td>
<td>1.35</td>
</tr>
<tr>
<td>Missing</td>
<td>0.40</td>
<td>0.690</td>
<td>0.15</td>
</tr>
<tr>
<td><strong>Cell Type</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Squamous</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Small</td>
<td>3.20</td>
<td>0.001</td>
<td>0.64</td>
</tr>
<tr>
<td>Adeno.</td>
<td>0.12</td>
<td>0.901</td>
<td>0.93</td>
</tr>
<tr>
<td>Other</td>
<td>1.54</td>
<td>0.122</td>
<td>0.56</td>
</tr>
<tr>
<td><strong>Metastases</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Yes</td>
<td>4.57</td>
<td>0.000</td>
<td>1.87</td>
</tr>
<tr>
<td>Missing</td>
<td>2.33</td>
<td>0.020</td>
<td>1.31</td>
</tr>
<tr>
<td><strong>Sodium</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 136 mmol/l</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>&lt; 136 mmol/l</td>
<td>2.42</td>
<td>0.015</td>
<td>0.06</td>
</tr>
<tr>
<td>Missing</td>
<td>1.94</td>
<td>0.052</td>
<td>2.41</td>
</tr>
<tr>
<td><strong>Albumen</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 35 g/l</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>&lt; 35 g/l</td>
<td>3.68</td>
<td>0.000</td>
<td>0.13</td>
</tr>
<tr>
<td>Missing</td>
<td>1.70</td>
<td>0.089</td>
<td>0.27</td>
</tr>
</tbody>
</table>

\[ l(\hat{\theta}) = -1789.95 \quad AIC = 3683.91 \]

**Note:** A value of $p = 0.000 \Rightarrow p < 0.001$. 

[95x741]148
5.4. HYPOTHESIS TESTING AND VARIABLE SELECTION

Figure 5.30. Joint 95% confidence ellipses (solid) and individual 95% confidence intervals (dash) for the scale and shape coefficients.
Looking at the joint effects, the main conclusions are as follows:

- **Treatment**: The overall effect of chemotherapy is not statistically significant.

- **Age Group**: The two older age groups (80+ and 70 - 80) are not significantly different from the < 50 group. Furthermore, the 50 - 60 group is only just significant.

- **WHO Status**: The light work group is not significantly different from the normal group.

- **Sex**: Males are not significantly different from females.

- **Smoker**: Neither current nor ex-smokers are significantly different from non-smokers.

- **Cell Type**: Adenocarcinoma and other cells are not significantly different from squamous cell.

- All other effects are highly statistically significant (apart from the “missing” groups which we are not interested in anyway).

These conclusions echo our earlier findings which were based on the hazard ratio plots (Figs. 5.11 - 5.19). Of course this is not surprising as the hazard ratios take into account both the scale and shape effects of covariates.

In addition to the above, while WHO status, cell type and albumen level have categories which are statistically significant, the shape effects are all non-significant. Thus, the extra flexibility afforded by shape regression is unnecessary in these cases\(^\text{13}\).

Clearly we can reduce the dimension of the model by dropping non-significant terms. To this end we fitted the 27 reduced models which arise from excluding each of the nine covariates from scale, the shape and simultaneously from the scale and shape. We then compared these models to the

\(^{13}\text{It is worth noting that testing for non-zero shape coefficients in the log-logistic MPR model provides a test of proportional odds of survival (see Section 5.2.3).}\)
full model in order to determine the importance of each covariate in these regression components (using likelihood ratio tests and $AIC$ differences). The results are shown in Table 5.9.

Table 5.9. Likelihood Ratio Tests and $AIC$ differences

<table>
<thead>
<tr>
<th>L.R. Test</th>
<th>$\Delta_{AIC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale</td>
<td>Shape</td>
</tr>
<tr>
<td>Treatment</td>
<td>0.000</td>
</tr>
<tr>
<td>Age Group</td>
<td>0.146</td>
</tr>
<tr>
<td>WHO Status</td>
<td>0.000</td>
</tr>
<tr>
<td>Sex</td>
<td>0.421</td>
</tr>
<tr>
<td>Smoker</td>
<td>0.974</td>
</tr>
<tr>
<td>Cell Type</td>
<td>0.008</td>
</tr>
<tr>
<td>Metastases</td>
<td>0.000</td>
</tr>
<tr>
<td>Sodium</td>
<td>0.004</td>
</tr>
<tr>
<td>Albumen</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Note: A value of $p = 0.000 \Rightarrow p < 0.001$. Furthermore, $\Delta_{AIC} = AIC_{\text{reduced}} - AIC_{\text{full}}$. Therefore, $\Delta_{AIC} > 0$ represents superiority of the full model fit.

Before attempting to reduce the dimension of the model, we first note that the joint $\Delta_{AIC}$ values give us a sense of the relative importance of each covariate. In particular, treatment and WHO status are the most important in determining survival of the patient, followed by the presence of metastases. Next, cell type, sodium and albumen are similar in importance. Finally, age group, sex and smoking status are less important.

Taking into consideration all of the p-values and $AIC$ differences in the above table, we initially removed age group, sex and smoking status from the model (i.e., from both regressions). Furthermore, WHO status, cell type, metastases and albumen were dropped from the shape regression. This gave a model with $AIC = 3660.92$. Removing covariates further did not improve on this $AIC$ value, however the re-inclusion of metastases in the shape regression gave $AIC = 3659.67$. Although this last improvement is small, we accepted it as our final model (see Table 5.10). In any case we have significantly improved on the $AIC$ value of the full model ($3683.91 - 3659.67 = 24.24$).
The overall and average-covariates hazard ratios under this reduced model (shown in Figs. 5.31 - 5.36) are similar to those from the full model (see Example 5.4) but the C.I.s are generally narrower. Of course this analysis concludes that age group, sex and smoking status do not impact survival.

### Table 5.10. Reduced Log-Logistic MPR Model Fit

<table>
<thead>
<tr>
<th>Scale</th>
<th>Shape</th>
<th>Joint</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>S.E.</td>
<td>$</td>
</tr>
<tr>
<td>Intercept</td>
<td>-4.47 (0.44)</td>
<td>——</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palliative</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Surgery</td>
<td>-1.87 (0.83)</td>
<td>2.25</td>
</tr>
<tr>
<td>Chemo</td>
<td>-0.55 (0.41)</td>
<td>1.33</td>
</tr>
<tr>
<td>Radio</td>
<td>-1.23 (0.23)</td>
<td>5.28</td>
</tr>
<tr>
<td>C+R</td>
<td>-5.18 (1.18)</td>
<td>4.39</td>
</tr>
<tr>
<td>WHO Status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Light Work</td>
<td>0.03 (0.26)</td>
<td>0.13</td>
</tr>
<tr>
<td>No Work</td>
<td>0.88 (0.27)</td>
<td>3.30</td>
</tr>
<tr>
<td>&gt; 50% Bed</td>
<td>1.63 (0.29)</td>
<td>5.59</td>
</tr>
<tr>
<td>Bedbound</td>
<td>2.29 (0.47)</td>
<td>4.90</td>
</tr>
<tr>
<td>Cell Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Squamous</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Small</td>
<td>1.09 (0.25)</td>
<td>4.26</td>
</tr>
<tr>
<td>Adeno.</td>
<td>0.19 (0.21)</td>
<td>0.88</td>
</tr>
<tr>
<td>Other</td>
<td>0.30 (0.16)</td>
<td>1.87</td>
</tr>
<tr>
<td>Metastases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Yes</td>
<td>1.69 (0.34)</td>
<td>4.89</td>
</tr>
<tr>
<td>Missing</td>
<td>1.02 (0.37)</td>
<td>2.77</td>
</tr>
<tr>
<td>Sodium</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\geq$ 136 mmol/l</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>&lt; 136 mmol/l</td>
<td>0.46 (0.17)</td>
<td>2.64</td>
</tr>
<tr>
<td>Missing</td>
<td>-0.98 (0.50)</td>
<td>1.95</td>
</tr>
<tr>
<td>Albumen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\geq$ 35 g/l</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>&lt; 35 g/l</td>
<td>0.68 (0.14)</td>
<td>4.67</td>
</tr>
<tr>
<td>Missing</td>
<td>0.50 (0.25)</td>
<td>2.01</td>
</tr>
</tbody>
</table>

$\ell(\hat{\theta}) = -1802.84 \quad AIC = 3659.67$

**Note:** Age group, sex and smoking status were not selected in this reduced model.

A value of $p = 0.000 \Rightarrow p < 0.001$. 

5.4. HYPOTHESIS TESTING AND VARIABLE SELECTION

Figure 5.31. Treatment (reference: palliative care) overall (solid) and average-covariates (dash) hazard ratios with confidence intervals for the reduced log-logistic MPR model.

Figure 5.32. WHO status (reference: normal) overall (solid) and average-covariates (dash) hazard ratios with confidence intervals for the reduced log-logistic MPR model.
**Figure 5.33.** Cell type (reference: squamous) overall (solid) and average-covariates (dash) hazard ratios with confidence intervals for the reduced log-logistic MPR model.

**Figure 5.34.** Metastases (reference: no) overall (solid) and average-covariates (dash) hazard ratios with confidence intervals for the reduced log-logistic MPR model.
5.4. HYPOTHESIS TESTING AND VARIABLE SELECTION

Figure 5.35. Sodium (reference: $\geq 136$ mmol/l) overall (solid) and average-covariates (dash) hazard ratios with confidence intervals for the reduced log-logistic MPR model.

Figure 5.36. Albumen (reference: $\geq 35$ g/l) overall (solid) and average-covariates (dash) hazard ratios with confidence intervals for the reduced log-logistic MPR model.
Example 5.8. MPR Variable Selection in Lung Cancer Data
We now show the results of applying automatic MPR variable selection to the lung cancer data. The procedure used here is more general than Algorithm 1, as it contains backward steps as well as forward steps. We applied this procedure to a variety of MPR and SPR models (in the latter case the procedure reduces to standard selection in one regression component) where the objective was to minimise the AIC value; the resulting selections are shown in Table 5.11. Note that in all cases considered here, starting from the full model gave the same results as starting from the null model\textsuperscript{14}.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Weibull</th>
<th>Log-Logistic</th>
<th>TDL</th>
<th>Gompertz</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MPR</td>
<td>SPR</td>
<td>MPR</td>
<td>SPR</td>
</tr>
<tr>
<td>Treatment</td>
<td>$\beta, \alpha$</td>
<td>$\beta$</td>
<td>$\beta, \alpha$</td>
<td>$\beta$</td>
</tr>
<tr>
<td>Age Group</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>WHO Status</td>
<td>$\beta$</td>
<td>$\beta$</td>
<td>$\beta$</td>
<td>$\beta$</td>
</tr>
<tr>
<td>Sex</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Smoker</td>
<td>$\alpha$</td>
<td>$\beta$</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Cell Type</td>
<td>$\beta$</td>
<td>$\beta$</td>
<td>$\beta$</td>
<td>$\beta$</td>
</tr>
<tr>
<td>Metastases</td>
<td>$\beta, \alpha$</td>
<td>$\beta$</td>
<td>$\beta, \alpha$</td>
<td>$\beta$</td>
</tr>
<tr>
<td>Sodium</td>
<td>$\beta$</td>
<td>$\beta$</td>
<td>$\beta, \alpha$</td>
<td>$\beta$</td>
</tr>
<tr>
<td>Albumen</td>
<td>$\beta, \alpha$</td>
<td>$\beta$</td>
<td>$\beta$</td>
<td>$\beta, \alpha$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Weibull</th>
<th>Log-Logistic</th>
<th>TDL</th>
<th>Gompertz</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ell(\hat{\theta})$</td>
<td>-1809.8</td>
<td>-1839.5</td>
<td>-1802.8</td>
<td>-1822.8</td>
</tr>
<tr>
<td>AIC</td>
<td>3679.7</td>
<td>3723.1</td>
<td>3659.7</td>
<td>3683.7</td>
</tr>
<tr>
<td>$\Delta_{AIC}$</td>
<td>20.0</td>
<td>63.4</td>
<td>0.00</td>
<td>24.0</td>
</tr>
<tr>
<td>dim($\theta$)</td>
<td>30</td>
<td>22</td>
<td>27</td>
<td>19</td>
</tr>
</tbody>
</table>

\textbf{Note:} $\beta$ = “chosen in scale”, $\alpha$ = “chosen in shape” and — = “not chosen”.

\textsuperscript{14}In the lung cancer data the sample size is large (855) with a relatively low level of censoring ($\approx 20\%$) and, furthermore, the number of covariates is not too large. Thus, the information per covariate is quite high which may explain why the starting model did not impact the results; in general, automatic procedures can be much more unstable than in this particular case.
The models are all in agreement that neither age group nor sex affect survival. The log-logistic models (MPR and SPR), however, also remove smoking status. The log-logistic MPR model has a much lower $AIC$ value than all of the other models (the next best is the Weibull MPR model with an $AIC$ value 20 units higher) which we have found in other examples previously. Indeed, the log-logistic model in SPR form has the third lowest $AIC$ of the eight models considered, i.e., it is better even than the TDL and Gompertz MPR models.

It is noteworthy that the covariates selected in the log-logistic MPR model by this automatic procedure are the same as those selected in Example 5.7 by non-automatic means. Thus, even though such procedures typically receive criticism (as we have mentioned previously), this does not imply that they are destined to lead to unreasonable results. For the sake of interest we show the selection path that led to this final, automatically selected, model in Appendix B.5.3.

Comparing each MPR model with its corresponding SPR model, we find that extending to MPR form reduces the $AIC$ significantly in all cases, i.e., the values of $AIC_{SPR} - AIC_{MPR}$ are 43.4 (Weibull), 24.0 (log-logistic), 39.5 (TDL) and 39.3 (Gompertz) respectively. Of course, MPR models can only improve the fit; if the extra flexibility of the MPR extension is not required, then no shape covariates will be selected and the model reduces to SPR.

Finally, recall that the Weibull and Gompertz SPR models are PH models and, therefore, selection of covariates in the shape represents a departure from proportional hazards (see Sections 5.2.1 and 5.2.2). Thus, within the Weibull MPR model, treatment, smoking status, metastases and albumen are deemed non-PH whereas, the Gompertz MPR analysis finds treatment, WHO status, smoking status and albumen to have non-PH effects. Of course, in any case, we would need to investigate the hazard ratio plots to ascertain the degree of non-proportionality (i.e., departure from a straight line) and nature of the effect (increasing / decreasing), however we will not pursue this here.

\footnote{Analogously, shape covariates in the log-logistic MPR model represents departure from proportional odds of survival (see Section 5.2.3). Thus, in this example we find that treatment, metastases and sodium have non-PO effects.}
5.5 Simulation Studies

In the following sections we investigate (by simulation) the correlation structure for the estimated regression coefficients (Section 5.5.1) and also the bias in these estimates (Section 5.5.2). In particular, we compare the standard Weibull MPR model with the orthogonal Weibull MPR model\(^\text{16}\). Moreover, we evaluate the proposed MPR variable selection procedure of Section 5.4 in the standard Weibull MPR case (Section 5.5.3). Some additional simulation studies involving the log-logistic and Gompertz MPR models appear in Appendix B.4.

5.5.1 Correlation Matrix for Estimated Parameters

In Section 5.4 we discussed the implications of correlated estimates in hypothesis testing and variable selection. While we have shown previously that correlation exists across the scale-shape space in some practical examples (see Figs. 5.10 and 5.20), we now investigate the structure of this correlation more formally by simulation.

We simulated data from a standard Weibull MPR model (using the method of Chapter 3) with regression components 
\[ \log \lambda = \beta_0 + x_1 \beta_1 + x_2 \beta_2 \]
and 
\[ \log \gamma = \alpha_0 + x_1 \alpha_1 + x_2 \alpha_2, \]
respectively, where \( x_1 \) and \( x_2 \) are independent binary covariates\(^\text{17}\). We selected 30 different parameter vectors, \( \theta = (\beta_0, \beta_1, \beta_2, \alpha_0, \alpha_1, \alpha_2)^T \), to simulate from - the values of which are given in Appendix B.4 and were chosen based on fitting the Weibull MPR model first to real data (i.e., data-directed simulation). In addition to this we also varied the sample size (\( n = 100, 500 \) and 1000) and censored proportion (\( p = 20\%, 50\% \) and 80\%). Thus, there are \( 30 \times 3 \times 3 = 270 \) scenarios in total; each scenario was repeated 500 times. At the \( j \)th repetition of a particular scenario the Weibull MPR model (i.e., the true model) was fitted to the simulated data producing an m.l.e. vector. Hence, after 500 replicates, we have a \( 500 \times 6 \) matrix of estimates, the columns of which can be plotted against each other to investigate the correlation structure (for that scenario).

\(^{16}\)Appendix C contains details on the orthogonal Weibull model and its MPR extension.

\(^{17}\)The binary variables were both generated using \texttt{rbinom(n, size=1, prob=0.5)} in R. This code creates a sample of \( n \) Bernoulli variables with \( \Pr(X = 1) = \Pr(X = 0) = 0.5 \).
5.5. SIMULATION STUDIES

Figure 5.37 shows three scatter matrices of estimated regression coefficients. We only show results for scenarios with $n = 1000$ as the correlation pattern is virtually identical at the other sample sizes\(^{18}\). Furthermore, we omit all scatter plots involving the intercepts as we are primarily interested in the way coefficients of covariates are correlated with each other. We can see that the “same-covariate-pairs”, $(\hat{\beta}_1, \hat{\alpha}_1)$ and $(\hat{\beta}_2, \hat{\alpha}_2)$ (i.e., pairs corresponding to the same covariate), are highly correlated whereas the other estimates are relatively uncorrelated; this matches the structure previously found in practice (see Figs. 5.10 and 5.20). Furthermore, the same pattern emerges in log-logistic and Gompertz MPR simulation studies (Appendix B.4).

We have found that this particular correlation pattern arises over a range of different sample sizes, censoring levels, MPR models and parameter values. Hence, we have “proved” by simulation that this is indeed a general feature of MPR models which supports the earlier discussion, and proposed extensions, of Section 5.4. Note that in the study presented here (and those in Appendix

\(^{18}\)There are three censoring levels and 30 parameter vectors which, therefore, gives $30 + 30 + 30 = 90$ scenarios with $n = 1000$. Hence, each of the three scatter matrices in Fig. 5.37 is based on 30 scenarios (one for each parameter vector) and so each individual cell contains 30 scatter plots.
B.4) the covariates, $x_1$ and $x_2$, are independent. If they are correlated then, of course, the other regression coefficients will also become correlated (in addition to same-covariate-pairs). However, we have found that these correlations are not at the same level as correlations between same-covariate-pairs. In any case, the methods described in Section 5.4 could be generalised further to handle such correlations if so desired.

In addition to the above simulation study, we also carried out a study involving the orthogonal Weibull MPR model. The study proceeded in exactly the same manner described for the standard Weibull MPR model and, hence, the corresponding scatter matrix plots are given in Fig. 5.38. The orthogonal Weibull model is derived based on the assumption of full information (see Appendix C.4) and, therefore, $\beta$ and $\alpha$ are only truly orthogonal when there is no censoring. It is no surprise then that same-covariate-pairs, $(\hat{\beta}_1, \hat{\alpha}_1)$ and $(\hat{\beta}_2, \hat{\alpha}_2)$, are not so highly correlated when censoring is low but become more correlated as the censored proportion increases. Hence the advantage of orthogonality breaks down in the presence of censoring.

It is worth noting a connection between our work and Lee & Whitmore (2006) who discussed the inverse-Gaussian MPR model (which they refer to as a threshold regression model - see Section 5.1 for further details). These

<table>
<thead>
<tr>
<th>n = 1000</th>
<th>p = 20%</th>
<th>n = 1000</th>
<th>p = 50%</th>
<th>n = 1000</th>
<th>p = 80%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_1$</td>
<td>0.04</td>
<td>0.19</td>
<td>0.03</td>
<td>$\hat{\beta}_1$</td>
<td>0.04</td>
</tr>
<tr>
<td>$\hat{\beta}_2$</td>
<td>0.03</td>
<td>0.17</td>
<td></td>
<td>$\hat{\beta}_2$</td>
<td>0.03</td>
</tr>
<tr>
<td>$\hat{\alpha}_1$</td>
<td>0.05</td>
<td></td>
<td></td>
<td>$\hat{\alpha}_1$</td>
<td>0.04</td>
</tr>
<tr>
<td>$\hat{\alpha}_2$</td>
<td></td>
<td></td>
<td></td>
<td>$\hat{\alpha}_2$</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5.38. Scatter matrices of estimated regression coefficients for the Orthogonal Weibull MPR model. Each cell in the lower triangle contains all 30 scatter plots with least squares lines overlayed (red). The upper triangle shows the corresponding average absolute correlation value.
authors also found that estimated regression coefficients are correlated but suggested that it was a consequence of censoring: “Where (the) models are estimated only from censored survival data, parameter estimators may exhibit significant multicollinearity”. Our simulation work allows us to expand on this statement. While it is true for “orthogonal” models (i.e., those orthogonal in uncensored data), it is not true in general. We have shown that the correlation pattern arises in a variety of models due to an inherent lack of orthogonality across the scale-shape space irrespective of censoring. Lee & Whitmore also state that this correlation “raises some new issues for estimation and inference”. Firstly, we agree that the correlation raises issues for inference and we have discussed solutions to this in Section 5.4. With respect to estimation, however, we disagree. In our simulation work and practical applications, such issues did not arise; the models fitted easily and quickly using the Newton-Raphson method and estimates were stable upon altering starting values. Furthermore, we show that the estimates are unbiased in the following section (see Section 3.6.3 for a related, but simpler, study).

5.5.2 Bias of Estimated Parameters

We now investigate the bias in parameter estimates for the standard and orthogonal Weibull MPR models. In particular, we wish to ascertain the effect of varying the sample size, $n$, and censored proportion, $p$. Each combination of $n$ and $p$ ($3 \times 3 = 9$ in total) comprises 30 scenarios - one for each $\theta$ vector. For a particular $n-p$ combination, let $\theta_{jk}$ be the true $j$th parameter ($j = 1, \ldots, 6$) in the $k$th scenario ($k = 1, \ldots, 30$) and, furthermore, $\hat{\theta}_{jkl}$ is the corresponding estimated value from the $l$th repetition ($l = 1, \ldots, 500$) of this ($k$th) scenario. Thus, we define the average relative bias in the estimate of the $j$th parameter, for a particular $n-p$ combination, as

$$\text{rbias}(\hat{\theta}_j) = \frac{1}{30 \times 500} \sum_{k=1}^{30} \sum_{l=1}^{500} \frac{\hat{\theta}_{jkl} - \theta_{jk}}{|\theta_{jk}|} = \frac{1}{30} \sum_{k=1}^{30} \frac{\bar{\theta}_{jk} - \theta_{jk}}{|\theta_{jk}|}, \quad (5.33)$$

where $\bar{\theta}_{jk} = \sum_{l=1}^{500} \hat{\theta}_{jkl}/500$. The results for the standard and orthogonal Weibull MPR models are given in Tables 5.12 and 5.13 respectively. We can see that the estimates are generally unbiased; the only exception to this is
\( \hat{\alpha}_0 \) when \( n = 100 \) (relative bias = 0.16 - 0.3). Naturally the bias decreases when information increases (i.e., \( n \uparrow \) or \( p \downarrow \)). Note that similar results are found in the log-logistic and Gompertz MPR models (Appendix B.4).

### Table 5.12. Weibull MPR Model: Average Relative Bias

<table>
<thead>
<tr>
<th>( n )</th>
<th>( p )</th>
<th>( \hat{\beta}_0 )</th>
<th>( \hat{\beta}_1 )</th>
<th>( \hat{\beta}_2 )</th>
<th>( \hat{\alpha}_0 )</th>
<th>( \hat{\alpha}_1 )</th>
<th>( \hat{\alpha}_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>80%</td>
<td>-0.086</td>
<td>0.031</td>
<td>-0.001</td>
<td>0.278</td>
<td>-0.020</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>50%</td>
<td>-0.051</td>
<td>-0.013</td>
<td>0.017</td>
<td>0.177</td>
<td>-0.007</td>
<td>0.011</td>
</tr>
<tr>
<td>20%</td>
<td>-0.046</td>
<td>0.088</td>
<td>0.021</td>
<td>0.170</td>
<td>-0.001</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td>80%</td>
<td>-0.014</td>
<td>0.000</td>
<td>0.006</td>
<td>0.051</td>
<td>0.000</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>50%</td>
<td>-0.009</td>
<td>-0.010</td>
<td>0.015</td>
<td>0.032</td>
<td>0.003</td>
<td>-0.002</td>
</tr>
<tr>
<td>20%</td>
<td>-0.009</td>
<td>0.021</td>
<td>-0.009</td>
<td>0.035</td>
<td>-0.007</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>80%</td>
<td>-0.008</td>
<td>0.024</td>
<td>0.006</td>
<td>0.032</td>
<td>-0.010</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>50%</td>
<td>-0.006</td>
<td>0.007</td>
<td>0.012</td>
<td>0.021</td>
<td>-0.002</td>
<td>-0.009</td>
</tr>
<tr>
<td>20%</td>
<td>-0.004</td>
<td>0.005</td>
<td>0.008</td>
<td>0.016</td>
<td>-0.001</td>
<td>-0.005</td>
<td></td>
</tr>
</tbody>
</table>

### Table 5.13. Orthogonal Weibull MPR Model: Average Relative Bias

<table>
<thead>
<tr>
<th>( n )</th>
<th>( p )</th>
<th>( \hat{\beta}_0 )</th>
<th>( \hat{\beta}_1 )</th>
<th>( \hat{\beta}_2 )</th>
<th>( \hat{\alpha}_0 )</th>
<th>( \hat{\alpha}_1 )</th>
<th>( \hat{\alpha}_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>80%</td>
<td>-0.008</td>
<td>-0.038</td>
<td>0.094</td>
<td>0.299</td>
<td>0.044</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>50%</td>
<td>0.006</td>
<td>-0.013</td>
<td>-0.009</td>
<td>0.203</td>
<td>-0.005</td>
<td>-0.019</td>
</tr>
<tr>
<td>20%</td>
<td>0.007</td>
<td>-0.007</td>
<td>0.006</td>
<td>0.166</td>
<td>-0.006</td>
<td>-0.005</td>
<td></td>
</tr>
<tr>
<td>80%</td>
<td>0.000</td>
<td>0.012</td>
<td>-0.013</td>
<td>0.049</td>
<td>-0.009</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>50%</td>
<td>0.001</td>
<td>-0.002</td>
<td>0.009</td>
<td>0.042</td>
<td>-0.004</td>
<td>-0.011</td>
</tr>
<tr>
<td>20%</td>
<td>0.001</td>
<td>0.007</td>
<td>0.002</td>
<td>0.028</td>
<td>0.001</td>
<td>-0.003</td>
<td></td>
</tr>
<tr>
<td>80%</td>
<td>-0.001</td>
<td>-0.009</td>
<td>-0.001</td>
<td>0.029</td>
<td>-0.004</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>50%</td>
<td>0.000</td>
<td>-0.005</td>
<td>-0.001</td>
<td>0.021</td>
<td>-0.004</td>
<td>-0.002</td>
</tr>
<tr>
<td>20%</td>
<td>0.001</td>
<td>0.002</td>
<td>0.004</td>
<td>0.019</td>
<td>-0.003</td>
<td>-0.004</td>
<td></td>
</tr>
</tbody>
</table>
5.5. SIMULATION STUDIES

5.5.3 Variable Selection

In Section 5.4 we discussed MPR variable selection which involves carrying out scale steps, shape steps and simultaneous steps. Here we investigate this procedure by means of simulation. Specifically, we evaluate the performance of an algorithm which starts from the null model, contains both forward and backward steps and uses AIC as the basis of selection (see Appendix B.5.3 for an example of this algorithm applied to the lung cancer data).

We simulated data from the standard Weibull MPR model with

$$\log \lambda = x^T ( -1.5, -1.0, 1.0, 0.5, -0.5, 0.0, 0.0, -0.8, 0.5, 0.0, 0.0 )^T$$

$$\log \gamma = z^T ( 0.5, 0.4, -0.4, 0.2, -0.2, 0.4, -0.2, 0.0, 0.0, 0.0, 0.0 )^T,$$

where $x = z = (1, x_1, \ldots, x_{10})^T$ and each covariate is a binary variable (generated independently using \texttt{rbinom} in R). Thus, the first four covariates, $x_1$, $x_2$, $x_3$ and $x_4$, affect both the scale and the shape, $x_5$ and $x_6$ affect the shape only, $x_7$ and $x_8$ affect the scale only and, finally, $x_9$ and $x_{10}$ have no affect on either component. This setup gives a good variety of scale and shape effects of different strengths where the specific values of the regression coefficients were chosen to lead to realistic survival data. Furthermore, we varied the sample size ($n = 100, 500$ and $1000$) and censored proportion ($p = 20\%, 50\%$ and $80\%$) giving 9 scenarios in total, each of which was repeated 500 times.

At the $j$th repetition of a particular scenario we applied the variable selection algorithm described above and, hence, we let $x^{(j)}$ and $z^{(j)}$ denote the vectors of scale and shape covariates selected in this repetition. The relative frequency of selection of the covariate $x_k$, $k = 1, \ldots, 10$, in a particular scenario was then calculated as

$$\text{rfreq}(x_k \in x) = \frac{1}{500} \sum_{j=1}^{500} 1(x_k \in x^{(j)}),$$

for the scale and

$$\text{rfreq}(x_k \in z) = \frac{1}{500} \sum_{j=1}^{500} 1(x_k \in z^{(j)}),$$

for the shape. The numeric results for each scenario are given in Table 5.14 but are much more clearly visualised in Fig. 5.39.
We can see (from Fig. 5.39) that it is difficult to determine which covariates affect survival when the information available is low, i.e., small sample size \((n = 100)\) with a high censored proportion \((p = 80\%)\). However, even when the sample size is small, the covariates with stronger effects (i.e., those with larger regression coefficients) are still selected 75%-85% of the time if censoring is not so high \((p = 20\%)\). When a moderate level of information is available, \((n, p) = (500, 50\%)\), the covariates with stronger effects are selected 100% of the time while those with weaker effects are all selected over 80% of

### Table 5.14. Relative Frequency of Selections

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>(n=100)</th>
<th>(n=500)</th>
<th>(n=1000)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(x)</td>
<td>80%</td>
<td>50%</td>
<td>20%</td>
</tr>
<tr>
<td>1</td>
<td>(\beta_1 = -1.0)</td>
<td>0.56</td>
<td>0.73</td>
<td>0.80</td>
</tr>
<tr>
<td>2</td>
<td>(\beta_2 = 1.0)</td>
<td>0.64</td>
<td>0.77</td>
<td>0.85</td>
</tr>
<tr>
<td>3</td>
<td>(\beta_3 = 0.5)</td>
<td>0.36</td>
<td>0.56</td>
<td>0.58</td>
</tr>
<tr>
<td>4</td>
<td>(\beta_4 = -0.5)</td>
<td>0.38</td>
<td>0.51</td>
<td>0.58</td>
</tr>
<tr>
<td>5</td>
<td>(\beta_5 = 0.0)</td>
<td>0.28</td>
<td>0.28</td>
<td>0.22</td>
</tr>
<tr>
<td>6</td>
<td>(\beta_6 = 0.0)</td>
<td>0.22</td>
<td>0.23</td>
<td>0.24</td>
</tr>
<tr>
<td>7</td>
<td>(\beta_7 = -0.8)</td>
<td>0.52</td>
<td>0.75</td>
<td>0.82</td>
</tr>
<tr>
<td>8</td>
<td>(\beta_8 = 0.5)</td>
<td>0.36</td>
<td>0.47</td>
<td>0.61</td>
</tr>
<tr>
<td>9</td>
<td>(\beta_9 = 0.0)</td>
<td>0.23</td>
<td>0.19</td>
<td>0.21</td>
</tr>
<tr>
<td>10</td>
<td>(\beta_{10} = 0.0)</td>
<td>0.25</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td>1</td>
<td>(\alpha_1 = 0.4)</td>
<td>0.38</td>
<td>0.57</td>
<td>0.74</td>
</tr>
<tr>
<td>2</td>
<td>(\alpha_2 = -0.4)</td>
<td>0.40</td>
<td>0.58</td>
<td>0.72</td>
</tr>
<tr>
<td>3</td>
<td>(\alpha_3 = 0.2)</td>
<td>0.30</td>
<td>0.42</td>
<td>0.50</td>
</tr>
<tr>
<td>4</td>
<td>(\alpha_4 = -0.2)</td>
<td>0.31</td>
<td>0.46</td>
<td>0.50</td>
</tr>
<tr>
<td>5</td>
<td>(\alpha_5 = 0.4)</td>
<td>0.43</td>
<td>0.73</td>
<td>0.88</td>
</tr>
<tr>
<td>6</td>
<td>(\alpha_6 = -0.2)</td>
<td>0.30</td>
<td>0.42</td>
<td>0.55</td>
</tr>
<tr>
<td>7</td>
<td>(\alpha_7 = 0.0)</td>
<td>0.25</td>
<td>0.28</td>
<td>0.26</td>
</tr>
<tr>
<td>8</td>
<td>(\alpha_8 = 0.0)</td>
<td>0.23</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>9</td>
<td>(\alpha_9 = 0.0)</td>
<td>0.20</td>
<td>0.21</td>
<td>0.22</td>
</tr>
<tr>
<td>10</td>
<td>(\alpha_{10} = 0.0)</td>
<td>0.26</td>
<td>0.20</td>
<td>0.19</td>
</tr>
</tbody>
</table>
the time. Furthermore, as the sample size increases, the relative frequency of selection of any covariate with nonzero effect approaches one.

![Figure 5.39. Relative frequency of selection of each covariate in the scale (black) and shape (red). Solid circles denote covariates with stronger effects (i.e., larger regression coefficients), open circles denote those with weaker effects and asterisks denote those with no effect. The order of terms in the legend matches the order of relative frequencies from the first scenario, \((n, p) = (100, 80\%)\), e.g., \(x_2\) in the scale has the highest frequency whereas \(x_9\) in the shape has the lowest.](image)

We can also see that covariates with no effect are selected approximately 20% of the time in all scenarios. To combat this, the algorithm could be altered so that a new model is only accepted if the AIC reduction is larger than some threshold (Table 2.1 would suggest using \(0 < \Delta_{AIC} \leq 2\)). Of course this will have some impact on the probability of selecting important covariates too - we will not pursue the issue of selecting an optimum threshold value however. Rather, in practical applications, one can track the AIC reductions to ascertain the merit of each step of the algorithm. In other words, automatic procedures are best used to guide us in finding important
covariates; more detailed consideration is advisable in practice. Nonetheless, this study shows the basic performance of the algorithm in its own right.

In addition to applying the selection procedure to the simulated data, the full model was also fitted at each repetition. Figure 5.40 shows boxplots of estimated regression coefficients from this full model over 500 replicates (for the three scenarios with \( p = 50\% \)). The level of uncertainty at different sample sizes is clearly visualised in these boxplots and serves to explain the performance of the selection procedure as discussed above.

![Boxplots of estimated regression coefficients for scenarios with \( p = 50\% \). The true value of each coefficient is denoted by “\( \times \)”.](image)

**Figure 5.40.** Boxplots of estimated regression coefficients for scenarios with \( p = 50\% \). The true value of each coefficient is denoted by “\( \times \)”. 
Chapter 6

Implied Regression Models

6.1 Introduction

Regression models are often constructed so that covariates act linearly on a particular scale (see Sections 4.2 and 4.3 for standard survival regression models); the effects of these covariates, on some quantity of interest, are then directly interpreted via their corresponding regression coefficients. Of course, specifying a (linear) model for one quantity leads to implied regression models for other quantities. However, in terms of these other quantities, covariate effects are unlikely to be described by simple parameters as, typically, the implied regression models are such that covariates appear in a functionally complicated manner. For example, consider the PH model, $\lambda(t \mid x) = \exp(x^T \beta)\lambda_0(t)$, which is linear on the log-hazard scale but implies a model for median survival which is generally difficult to interpret, i.e., $t_{0.5}(x) = S_0^{-1}(0.5^{\exp(-x^T \beta)})$ using the fact that $S(t \mid x) = S_0(t)^{\exp(x^T \beta)}$.

In order to formalise the above, assume a model for survival with density function $f(t \mid x, \theta)$ where $x = (1, x_1, \ldots, x_p)^T$ is a vector of covariates and $\theta = (\theta_1, \ldots, \theta_k)^T$ is the vector of model parameters. Furthermore, let $\theta = (\beta, \zeta)^T$ where $(\beta_0, \beta_1, \ldots, \beta_p)^T$ is the vector of regression coefficients and $\zeta$ is a vector of any remaining parameters such that $\dim(\zeta) = k - p - 1$. Moreover, let $\phi$ be a (possibly time-dependent) quantity which, in our model, depends on covariates via

$$\phi(x) = g^{-1}(x^T \beta \mid \zeta),$$
where $g^{-1}(\cdot)$ is assumed to have a functional form such that

$$g[\phi(x_1 = 1)] - g[\phi(x_1 = 0)] = \beta_1,$$

i.e., $g(\phi)$ is linear in the covariates so that $\beta$ coefficients are directly interpretable; we may say that $g(\phi)$ is the natural scale of the model. Now, let $\psi(x)$ be another quantity which can be derived from the model. However, unlike $\phi(x)$, we assume that the functional form of $\psi(x)$ prevents interpretation of the corresponding covariate effects using simple parameters. Note that we can write

$$\psi(x) = h[\phi(x)],$$

for some function $h(\cdot)$, i.e., it is clear that by specifying a model for $\phi$, one also specifies an implied model for $\psi$.

The aim of this chapter is to evaluate covariate effects in implied regression models. One could argue that if interest lies in the quantity $\psi$ then it should be modelled directly, rather than modelling $\phi$ and attempting to interpret the implied model for $\psi$. This view is reasonable of course - however it may be desirable to interpret implied effects in certain situations such as:

1. **Adjunct interpretation**: Assume the quantity $\phi$ is fundamental in the statistical formulation of the problem and is well understood by the expert. Nonetheless, it may be useful to translate results to the $\psi$-scale if this is more easily understood by a wider (non-specialist) audience.

2. **Difficulty modelling $\psi$ directly**: We are interested in $\psi$ but difficulties in modelling it directly lead us to fit the $\phi$-model; we then interpret this fitted model in terms of $\psi = h(\phi)$ which is an implied regression.

3. **Avoiding multiple models**: Several quantities are of interest ($\phi, \psi_1, \psi_2, \psi_3, \ldots$) but we do not wish to specify/fit a separate model for each one (e.g., for computational reasons). Rather, we fit the $\phi$-model and use it as the basis for interpreting covariate effects on all of the quantities.

In Chapter 5 we developed multi-parameter regression (MPR) models where issues of interpretation are inherent. Recall that the MPR framework is defined by regressing distributional parameters on covariates, i.e., covariate effects are linear on a scale functionally related (often via the log link) to
6.2. APPROXIMATING IMPLIED REGRESSION MODELS

6.2 Approximating Implied Regression Models

In Section 5.3 we described two approaches for evaluating the effect of a particular covariate, \( c \), in an implied regression model (for the hazard), namely:

(i) calculating the effect of \( c \) conditional on a particular set of covariates and
(ii) averaging the effect of \( c \) over the empirical distribution of covariates.

While these approaches may be useful in their own right, the resulting effects are not analogous to those of a directly specified (i.e., non-implied) regression model. The effect determined by method (i) is clearly only valid for a given covariate profile, whereas method (ii) produces a marginal effect. In a directly specified regression model the estimated effects are conditional on (i.e., adjusted for) other covariates in the model without being limited to any given covariate profile.

In light of the above, it seems reasonable to seek estimated effects for an
implied regression model, \( \psi(x) \), which are interpretable in a manner which is analogous to that of a non-implied regression model, \( \phi(x) \). In particular, we will approximate \( \psi(x) \) using a linear model,

\[
\psi(x) \approx \beta_0^* + x_1 \beta_1^* + \ldots + x_p \beta_p^* = x^T \beta^*,
\]

(6.1)

where we use the superscript “*” in order to distinguish these regression coefficients from those which appear in the regression model for \( \phi \). Clearly, if \( \psi(x) \) is well approximated by this linear model then the \( \beta^* \) coefficients will provide us with direct interpretation of covariate effects on the scale of \( \psi \) (analogous to \( \beta \) coefficients in the context of \( \phi \)).

### 6.2.1 Least Squares Solution

Let \( f(t \mid x, \theta) \) be a survival model wherein \( \psi(x, \theta) \) is an implied regression that we aim to approximate with a linear model as in (6.1). Given the true parameter vector \( \theta \), we can compute \( \psi_i = \psi(x_i, \theta) \): the true value of \( \psi \) at a particular set of covariates, \( x_i = (1, x_{1i}, \ldots, x_{pi})^T \). We can repeat this for \( n \) such covariate vectors and then form a system of equations

\[
\begin{align*}
\beta_0^* + x_{11} \beta_1^* + \ldots + x_{p1} \beta_p^* &= \psi_1 \\
&\vdots \\
\beta_0^* + x_{1i} \beta_1^* + \ldots + x_{pi} \beta_p^* &= \psi_i \\
&\vdots \\
\beta_0^* + x_{1n} \beta_1^* + \ldots + x_{pn} \beta_p^* &= \psi_n
\end{align*}
\]

(6.2)

whose solution provides us with the \( \beta^* \) coefficients required for (6.1). Note that (6.2) can be written in the form

\[
X \beta^* = y,
\]

(6.3)

where \( X \) is the \( n \times (p + 1) \) matrix of covariate vectors, \( \beta^* = (\beta_0^*, \beta_1^*, \ldots, \beta_p^*)^T \) and \( y = (\psi_1, \ldots, \psi_n)^T \). Naturally, this system of equations has an approximate solution given by least squares,

\[
\hat{\beta}^* = (X^T X)^{-1} X^T y,
\]

(6.4)
provided that $X^T X$ is invertible and $n > p$.

In a purely theoretical exercise, the parameter values in $\theta$ would be specified and also a grid of covariate values over which some feature of $f(t \mid x, \theta)$, $\psi(x)$, is to be investigated. However this is not our intended use of the method presented above; rather, we consider its application in practice where $\theta$ is not known. In this case we first fit the model $f(t \mid x, \theta)$ to data using maximum likelihood (Chapter 2) giving the m.l.e., $\hat{\theta}$. Thus, we have $\hat{\psi}_i = \psi(x_i, \hat{\theta})$ and $\hat{y} = (\hat{\psi}_1, \ldots, \hat{\psi}_n)^T$ (rather than $\psi_i$ and $y$). Apart from this minor notational modification however, everything remains the same as before. Of course, in such practical applications, it seems most reasonable to use the covariate profiles from the dataset (used to fit $f(t \mid x, \theta)$) when forming the system of $n$ equations given in (6.2); $n$ is therefore the number of individuals in the dataset. Indeed, we have used this approach in our own practical work.

It is worth clarifying how the above procedure differs from fitting a multiple linear regression model to data (where $y$ is a random vector). Our use of least squares here is for the purpose of model interpretation, it is not a data fitting exercise. Note that, here, $y$ is not a vector of random variables - it is a vector of known quantities, $\psi_1, \ldots, \psi_n$, calculated from the model $f(t \mid x, \theta)$ (this is true whether $\theta$ is known or estimated by maximum likelihood). Thus, error arises solely due to the systematic departure of $\psi(x, \theta)$ from linearity, i.e., there is no random error term. Hence, the coefficient of determination, given by

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^{n} (\psi_i - x_i^T \hat{\beta}^*)^2}{\sum_{i=1}^{n} (\psi_i - \frac{\sum_{i=1}^{n} \psi_i}{n})^2},$$

may be interpreted as the degree to which $\psi$ can be approximated by a linear function. If $\psi$ is highly non-linear, then we could improve the approximation using, for example, polynomials, splines or interaction terms; we will not consider this in our work however.

### 6.2.2 Uncertainty

While it is true that $\hat{y}$ is non-random given a fitted model $f(t \mid x, \hat{\theta})$ and set of covariate profiles, $x_1, \ldots, x_n$, as described above, it does, of course, vary from sample to sample just as $\hat{\theta}$ does; indeed $\hat{y}$ is a (vector-valued) function of $\hat{\theta}$. 
Hence, $\hat{\beta}^*$, which is calculated using $\hat{y}$, clearly has sampling variation also and this must be accounted for in our analyses, i.e., the associated standard errors/confidence intervals must be reported.

**Delta Method**

Using the fact that $\hat{y} = y(\hat{\theta})$ is a vector-valued function of $\hat{\theta}$, the *delta method* gives the following estimate of its covariance matrix:

$$
\Sigma(\hat{y}) = \left( \frac{\partial y(\hat{\theta})}{\partial \theta} \right) I_o(\hat{\theta})^{-1} \left( \frac{\partial y(\hat{\theta})}{\partial \theta} \right)^T,
$$

(6.5)

where $I_o(\hat{\theta})$ is the observed information matrix for $\hat{\theta}$, which results from fitting $f(t \mid x, \theta)$ to the data (Section 2.3), and $(\partial y/\partial \theta)$ represents the $n \times k$ matrix of partial derivatives of the vector $y$ with respect to the vector $\theta$, i.e.,

$$
\left( \frac{\partial y}{\partial \theta} \right) = \begin{bmatrix}
\frac{\partial y_1}{\partial \theta_1} & \frac{\partial y_1}{\partial \theta_2} & \cdots & \frac{\partial y_1}{\partial \theta_k} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial y_n}{\partial \theta_1} & \frac{\partial y_n}{\partial \theta_2} & \cdots & \frac{\partial y_n}{\partial \theta_k}
\end{bmatrix}.
$$

Note that this (asymptotic) result arises from a version of the delta method more general than that discussed in Section 2.3.2 (compare (6.5) with (2.25)); see Bishop et al. (2007, sec. 14.6.3) for further details.

The covariance matrix of $\hat{y}$ is not really of interest in itself but is required to estimate the covariance matrix of $\hat{\beta}^*$ - just as $\hat{y}$ is not of interest but is used to estimate $\hat{\beta}^*$. Using the fact that $\hat{\beta}^* = (X^T X)^{-1} X^T \hat{y}$, and the identity $\text{cov}(Ay) = A \text{cov}(y) A^T$, the $p \times p$ covariance matrix for $\hat{\beta}^*$ can be estimated by

$$
\Sigma(\hat{\beta}^*) = (X^T X)^{-1} X^T \Sigma(\hat{y}) X (X^T X)^{-1},
$$

(6.6)

where $\Sigma(\hat{y})$ is given in (6.5). Confidence intervals for the $\beta^*$ coefficients can be calculated in the usual way using the diagonal elements of $\Sigma(\hat{\beta}^*)$.

**M.L.E. Simulation**

Although the delta method is the classical approach for estimating the precision of functions of $\hat{\theta}$, it is not the approach we will be using. As has been
the case throughout this thesis, we prefer to use our proposed method of m.l.e. simulation due to its ease of implementation (see Section 2.3.2); we now describe this method within the current setting. Recall that m.l.e. simulation is based on simulating a sample of $\hat{\theta}$ vectors, $\{\hat{\theta}(1), \ldots, \hat{\theta}(m)\}$, equivalent to a sample arising from $m$ independent datasets. Thus, for the $b$th of these vectors, $\hat{\theta}(b)$, we can compute $\hat{y}(b) = (\psi(x_1, \hat{\theta}(b)), \ldots, \psi(x_n, \hat{\theta}(b)))^T$. Repeating this (for $b = 1, \ldots, m$) produces a sample $\{\hat{y}(1), \ldots, \hat{y}(m)\}$ which can be stored as an $n \times m$ matrix, $\hat{Y}$, i.e., the $b$th column of $\hat{Y}$ is $\hat{y}(b)$. Hence,

$$\hat{B}^* = (X^TX)^{-1}X^T\hat{Y}$$

is the $(p + 1) \times m$ matrix whose $b$th column is the vector $\hat{\beta}^*(b)$. Therefore, row $j + 1$ of this matrix ($j = 0, \ldots, p$) contains the sample $\{\hat{\beta}^*_j(1), \ldots, \hat{\beta}^*_j(m)\}$ which can be used to construct a $100(1 - \alpha)$% confidence interval for $\beta^*_j$ (by selecting the appropriate percentiles) or to calculate $\text{se}(\hat{\beta}^*_j)$.

### 6.3 Life Time

Although the (log-)hazard scale has become the natural scale for incorporating covariates in survival regression models - largely due to the influential paper by Cox (1972) - it may be more intuitive to interpret covariate effects on the scale of $\mu = E(T)$, i.e., life expectancy. Thus, given some parametric (MPR) model, $f(t \mid x, \theta)$, where $\mu(x) = \int_0^\infty t \cdot f(t \mid x, \theta) \, dt$ is an implied regression model, we can use the method of Section 6.2 to interpret covariate effects on this scale by setting $\hat{\psi}(x) = \hat{\mu}(x)$. However, the tail of the distribution, and hence also the mean, can be quite unstable due to censoring. Thus, we consider instead the implied effect of covariates on the restricted mean, $\mu_r = E[\min(T, t^*)]$, which is more stable (Section 6.3.1); the median is also considered briefly in Section 6.3.2.
Before proceeding, we note that relating covariates directly (i.e., in a non-implied regression) to the survival time is also possible. The accelerated failure time (AFT) class of models (mentioned in Section 4.3) is the most prominent approach wherein $E(\log T) = x^T \beta$. Parametric AFT models are quite standard and can be estimated using maximum likelihood (see Lawless (2003, chap. 6) and Kalbfleisch & Prentice (1973, chap. 3)); indeed, it is easy to show that the Weibull and log-logistic SPR models (Section 4.5) are AFT models. Estimation of the semi-parametric AFT model (i.e., unspecified distribution) has also been developed (Miller, 1976; Prentice, 1978; Buckley & James, 1979; Ritov, 1990; Tsiatis, 1990; Ying, 1993) but computational difficulties inhibit application. Furthermore, these procedures require the censoring distribution to be independent of covariates. Hence, the semi-parametric AFT model is not widely used. Another interesting extension of the parametric AFT model is the so-called transformation model - defined by $E[g(T)] = x^T \beta$ where $g(\cdot)$ is an unspecified, but strictly increasing, function (Doksum, 1987; Cuzick, 1988; Cheng et al., 1995; Fine et al., 1998; Chen et al., 2002). As with the semi-parametric AFT model, necessarily complicated estimation procedures have hindered wider use of transformation models.

### 6.3.1 Restricted Mean Life

As mentioned above, the mean life time is generally ill-determined due to censoring. Hence, we will use the restricted mean, $\mu_r = E[\min(T, t^*)]$, which was first suggested by Irwin (1949). Note that we can write

$$\min(T, t^*) = 1(T \leq t^*) T + 1(T > t^*) t^*$$

$$\Rightarrow E[\min(T, t^*)] = \int_0^\infty 1(t \leq t^*) \, t \, f(t) \, dt + \int_0^\infty 1(t > t^*) \, t^* \, f(t) \, dt$$

$$= \int_0^{t^*} t \, f(t) \, dt + t^* \int_{t^*}^\infty f(t) \, dt$$

$$= \int_0^{t^*} t \, f(t) \, dt + t^* S(t^*).$$

\(^{2}\)The situation was improved by Jin et al. (2003), using a linear programming approach, and Tian et al. (2004), using importance sampling.
Integrating \( \int_0^\tau t f(t) \, dt \) by parts \((u = t, \, dv = f(t)dt)\) gives
\[
t^* [1 - S(t^*)] - \int_0^{t^*} [1 - S(t)] \, dt = t^* [1 - S(t^*)] - t^* + \int_0^{t^*} S(t) \, dt
\]

and, therefore,
\[
\mu_{t^*} = \int_0^{t^*} S(t) \, dt.
\] (6.7)

Thus, \( \mu_{t^*} \) is straightforwardly calculated by integrating the survivor curve over the range \([0, t^*]\).

Clearly \( \lim_{t^* \to \infty} \min(T, t^*) = T \Rightarrow \mu_\infty = \mu \), i.e., integrating to time infinity gives the usual unrestricted mean. Thus, if \([0, t^*]\) is the observed time range in a study, it is clear that calculation of \( \mu \) involves extrapolation beyond the data and, hence, the value of \( \mu_{t^*} \) is more relevant given the observed time range.

The interpretation of the restricted mean is the “average life within the first \( t^* \) months/years”. Note that it is common in the literature to report the \( t^* \)-month/year survival probability - the probability of surviving beyond time \( t^* \), i.e., \( S(t^*) \). However, the value of \( \mu_{t^*} \) is more intuitive being on the time scale rather than the probability scale. Moreover, whereas \( S(t^*) \) is simply a snapshot in time (i.e., the survivor curve evaluated at time \( t^* \)), the value of \( \mu_{t^*} \) takes into account the whole distribution up to time \( t^* \). Thus, the restricted mean has much to recommend it.

Once a model \( f(t \mid x) \) has been fitted to data, we can estimate the restricted mean for a given covariate profile
\[
\mu_{t^*}(x, \hat{\theta}) = \int_0^{t^*} S(t \mid x, \hat{\theta}) \, dt,
\] (6.8)

which typically requires numerical integration\(^3\). Hence, using the method in Section 6.2 (i.e., \( \hat{\psi}(x) = \mu_{t^*}(x, \hat{\theta}) \)) we can obtain an (approximate) regression model for the restricted mean,
\[
\hat{\mu}_{t^*}(x) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_p x_p.
\]

\(^3\)In our work we have used the \texttt{integrate} function in \texttt{R} which implements routines from the well-known “quadpack” numerical library (Piessens et al., 1983)).
where the $\hat{\beta}^*$ coefficients are straightforward to interpret, i.e., $\hat{\beta}^*_j$ is the amount by which covariate $x_j$ alters the average lifetime (in the first $t^*$ months/years).

The restricted mean has been considered previously by authors within a regression setting. Karrison (1987) and Zucker (1998) used the implied regression model for the restricted mean (within the PH model) to quantify the difference in two treatment groups; Chen & Tsiatis (2001) considered the same problem but used a more general (non-PH) model. These authors evaluated marginal effects by averaging over the empirical distribution of covariates, i.e., the same approach we used to calculate “overall hazard ratios” in Section 5.3. Andersen et al. (2004) modelled the restricted mean directly (i.e., $E[\min(T, t^*)] = x^T \beta$) via the so called pseudo-observations approach introduced by Andersen et al. (2003) (see Andersen & Perme (2010) and references therein for a review of pseudo-observations and their use in survival analysis). However, a disadvantage of the pseudo-observations estimation procedure is that the censoring distribution is assumed to be independent of covariates.

**Example 6.1. Restricted Mean Regression in the Lung Cancer Data: Log-Logistic MPR Model**

We now estimate the linear model for the restricted mean based on the log-logistic MPR model fit to the lung cancer data. In particular we use the full model (i.e., all covariates) which was previously discussed in Example 5.4. Furthermore, we estimate covariate effects on the 20-month restricted mean, $\mu_{20}$, as this was the largest survival time, i.e., we set $t^* = 20$. Note that Karrison (1987) discusses the issue of choosing an appropriate $t^*$ value (for further related discussion see Royston & Parmar (2011)).

The resulting least squares estimates are shown in Table 6.1 along with the corresponding standard errors and 95% confidence intervals - both calculated using m.l.e. simulation (as described in Section 6.2.2). Furthermore, Z-scores are also shown (to give an indication of significance) even though the confidence intervals may be asymmetric. Before interpreting these results, we note that $R^2 = 0.98$ and, therefore, the linear approximation to the log-logistic MPR restricted mean is excellent.
### 6.3. LIFE TIME

**Table 6.1. Restricted Mean Regression: Log-Logistic MPR Model**

|                  | $\hat{\beta}^*$ | S.E. | $|Z|$ | C.I.               |
|------------------|------------------|------|------|-------------------|
| **Intercept**    | 14.17 (1.57)     |      |      | (10.83, 16.96) *  |
| **Treatment**    |                  |      |      |                   |
| Palliative       | 0.00             |      |      |                   |
| Surgery          | 6.08 (0.91)      | 6.67 | (3.98, 7.54) * |
| Chemo            | 2.02 (1.01)      | 2.00 | (0.02, 3.98) *  |
| Radio            | 1.26 (0.51)      | 2.48 | (0.22, 2.18) *  |
| C+R              | 4.37 (1.05)      | 4.15 | (2.15, 6.30) *  |
| **Age Group**    |                  |      |      |                   |
| < 50             | 0.00             |      |      |                   |
| 50 - 60          | -0.95 (1.22)     | 0.78 | (-3.37, 1.45)  |
| 60 - 70          | -1.29 (1.13)     | 1.14 | (-3.33, 0.99)  |
| 70 - 80          | -0.77 (1.13)     | 0.68 | (-2.89, 1.44)  |
| 80 +             | -1.27 (1.19)     | 1.07 | (-3.66, 1.16)  |
| **WHO Status**   |                  |      |      |                   |
| Normal           | 0.00             |      |      |                   |
| Light Work       | -0.06 (0.85)     | 0.07 | (-1.62, 1.57)  |
| No Work          | -2.59 (0.88)     | 2.94 | (-4.28, -0.86) *|
| > 50% Bed        | -4.24 (0.90)     | 4.68 | (-5.95, -2.43) *|
| Bedbound         | -5.43 (1.07)     | 5.10 | (-7.18, -3.08) *|
| **Sex**          |                  |      |      |                   |
| Female           | 0.00             |      |      |                   |
| Male             | -0.01 (0.37)     | 0.04 | (-0.70, 0.78)  |
| **Smoker**       |                  |      |      |                   |
| Yes              | -1.19 (0.64)     | 1.87 | (-2.36, 0.03)  |
| Ex-smoker        | -0.91 (0.65)     | 1.41 | (-2.21, 0.34)  |
| Missing          | -0.64 (1.31)     | 0.49 | (-3.07, 2.08)  |
| **Cell Type**    |                  |      |      |                   |
| Squamous         | 0.00             |      |      |                   |
| Small            | -2.98 (0.66)     | 4.54 | (-4.15, -1.65) *|
| Adeno.           | -0.94 (0.66)     | 1.43 | (-2.28, 0.33)  |
| Other            | -0.76 (0.53)     | 1.43 | (-1.74, 0.26)  |
| **Metastases**   |                  |      |      |                   |
| No               | 0.00             |      |      |                   |
| Yes              | -2.99 (0.55)     | 5.42 | (-4.00, -1.84) *|
| Missing          | -1.49 (0.62)     | 2.40 | (-2.63, -0.24) *|
| **Sodium**       |                  |      |      |                   |
| ≥ 136 mmol/l     | 0.00             |      |      |                   |
| < 136 mmol/l     | -1.19 (0.36)     | 3.34 | (-1.91, -0.52) *|
| Missing          | -0.53 (0.91)     | 0.58 | (-2.23, 1.28)  |
| **Albumen**      |                  |      |      |                   |
| ≥ 35 g/l         | 0.00             |      |      |                   |
| < 35 g/l         | -1.81 (0.43)     | 4.23 | (-2.63, -0.92) *|
| Missing          | -1.37 (0.73)     | 1.88 | (-2.69, 0.22)  |

**Note:** Symbol "*" indicates that the C.I. does not contain the value zero.
We now summarise the results which, of course, apply to the first 20 months of life:

- **Treatment**: Surgery increases life by just over 6 months compared to palliative care. Similarly, the other treatments also increase life (although not to the same extent): chemotherapy by 2 months, radiotherapy by 1.3 months and the combined treatment (C+R) by 4.4 months.

- **Age group**: Older age groups have a slightly shorter life than the youngest group but this is not statistically significant.

- **WHO status**: As one’s physical state deteriorates so too does one’s lifespan - however those able to carry out light work are not significantly different to the normal group.

- **Sex**: Males are not significantly different from females.

- **Smoker**: Both smokers and ex-smokers live approximately 1 month less than non-smokers. While the effects are not significant, we note that the confidence intervals only just include the value of zero.

- **Cell type**: The small cell type reduces life by 3 months compared with squamous cell whereas other cell types are not significantly different from this reference class.

- **Metastases**: The presence of metastases decreases lifespan by 3 months.

- **Sodium**: Lower sodium levels reduce life slightly (1.2 months).

- **Albumen**: Lower albumen levels reduce life by almost 2 months.

The above results are in line with those previously found in the examples from Chapter 5. However, the analysis presented here is much more intuitive as life time is immediately understood from a physical perspective whereas, for example, a hazard ratio may not be. Indeed, while hazard ratio plots (see Chapter 5) provide useful information, they can sometimes be difficult to interpret. Furthermore, this example shows the utility of our proposed least squares approach to interpreting models. Clearly the method can be applied
6.3. **LIFE TIME**

To any quantity that we can estimate from our fitted model; we consider the median briefly in Section 6.3.2 before considering the hazard in Section 6.4.

### 6.3.2 Median Life

The median, \( t_{0.5} \), is another simple measure of centrality for long-tailed distributions and is often used in survival analysis. Estimation is quite stable provided that censoring is not too heavy (i.e., at least 50% of the distribution must be observed); inverting the Kaplan-Meier curve is the most popular method for estimating \( t_{0.5} \) (see Example 1.1). In spite of the fact that the median survival time is a commonly used quantity, relatively little work has been done in the area of median regression (Ying et al. (1995) and Yang (1999) have developed estimation procedures for semi-parametric median regression).

In our current parametric setting, \( t_{0.5}(x) = S^{-1}(0.5 \mid x, \theta) \) is the implied regression for the median, given the survival model \( f(t \mid \theta) \), where \( S^{-1}(u) \) is the inverse survivor function\(^4\). Again, we wish to approximate this as

\[
\hat{t}_{0.5}(x) = \hat{\beta}_0^* + \hat{\beta}_1^* x_1 + \ldots + \hat{\beta}_p^* x_p.
\]

Clearly, we can investigate covariate effects on other percentiles also but we will only consider the median here.

**Example 6.2. Median Regression in the Lung Cancer Data: Log-Logistic MPR Model**

Replicating Example 5.4, we now estimate the median regression model based on the log-logistic MPR model fit to the lung cancer data. The results are given in Table 6.2 below and are analogous to the restricted mean case; we will not labour the specific details which are qualitatively similar to those from the previous example. Furthermore, we note that \( R^2 = 0.8 \) and, therefore, the linear approximation to the median is very good (although not as good as the restricted mean case which was almost perfect).

\(^4\)Recall that the functional form of \( S^{-1}(u) \) for each distribution we consider is given in Section 1.4 and this function has previously been used for the purposes of simulation (Chapter 3).
Table 6.2. Median Regression: Log-Logistic MPR Model

|                      | $\hat{\beta}$ | S.E.    | $|Z|$  | C.I.           |
|----------------------|---------------|---------|-------|----------------|
| **Intercept**        | 25.98 (17.57) | ——      | ——    | (12.23, 65.18) *|
| **Treatment**        |               |         |       |                |
| Palliative           | 0.00          | ——      | ——    | ——             |
| Surgery              | 21.93 (8.83)  | 2.48    | ——    | (10.67, 42.96) *|
| Chemo                | 1.80 (2.00)   | 0.90    | ——    | (-1.17, 5.97)  |
| Radio                | 1.32 (0.72)   | 1.83    | ——    | (-0.26, 2.63)  |
| C+R                  | 3.66 (2.29)   | 1.60    | ——    | (-0.92, 6.85)  |
| **Age Group**        |               |         |       |                |
| < 50                 | 0.00          | ——      | ——    | ——             |
| 50 - 60              | -8.06 (12.39) | 0.65    | ——    | (-36.57, 1.91) |
| 60 - 70              | -9.37 (12.60) | 0.74    | ——    | (-38.98, 0.47) |
| 70 - 80              | -7.89 (12.57) | 0.63    | ——    | (-36.89, 2.05) |
| 80 +                 | -8.96 (12.89) | 0.70    | ——    | (-37.56, 0.88) |
| **WHO Status**       |               |         |       |                |
| Normal               | 0.00          | ——      | ——    | ——             |
| Light Work           | -1.18 (4.58)  | 0.26    | ——    | (-10.96, 6.32) |
| No Work              | -4.21 (3.95)  | 1.06    | ——    | (-13.27, 1.66) |
| > 50% Bed            | -5.26 (3.99)  | 1.32    | ——    | (-14.12, 0.83) |
| Bedbound             | -6.28 (4.35)  | 1.44    | ——    | (-16.38, 0.02) |
| **Sex**              |               |         |       |                |
| Female               | 0.00          | ——      | ——    | ——             |
| Male                 | -0.09 (0.94)  | 0.09    | ——    | (-2.32, 1.49)  |
| **Smoker**           |               |         |       |                |
| No                   | 0.00          | ——      | ——    | ——             |
| Yes                  | -3.90 (3.79)  | 1.03    | ——    | (-14.09, -0.40) *|
| Ex-smoker            | -3.59 (3.68)  | 0.98    | ——    | (-12.78, 0.11) |
| Missing              | -3.70 (3.90)  | 0.95    | ——    | (-12.58, 1.65) |
| **Cell Type**        |               |         |       |                |
| Squamous             | 0.00          | ——      | ——    | ——             |
| Small                | -2.09 (2.01)  | 1.04    | ——    | (-4.83, 2.47)  |
| Adeno.               | -2.40 (2.17)  | 1.11    | ——    | (-6.77, 1.15)  |
| Other                | -0.51 (1.68)  | 0.31    | ——    | (-3.17, 3.20)  |
| **Metastases**       |               |         |       |                |
| No                   | 0.00          | ——      | ——    | ——             |
| Yes                  | -5.65 (1.87)  | 3.02    | ——    | (-10.20, -3.22) *|
| Missing              | -3.75 (1.50)  | 2.50    | ——    | (-7.47, -1.45) *|
| **Sodium**           |               |         |       |                |
| ≥ 136 mmol/l         | 0.00          | ——      | ——    | ——             |
| < 136 mmol/l         | -1.50 (0.72)  | 2.09    | ——    | (-2.97, -0.38) *|
| Missing              | -0.50 (1.16)  | 0.43    | ——    | (-2.64, 1.86)  |
| **Albumen**          |               |         |       |                |
| ≥ 35 g/l             | 0.00          | ——      | ——    | ——             |
| < 35 g/l             | -2.12 (0.61)  | 3.46    | ——    | (-3.42, -0.99) *|
| Missing              | -1.83 (0.99)  | 1.86    | ——    | (-3.67, 0.36)  |

*Note:* Symbol “*” indicates that the C.I. does not contain the value zero.
6.4 Hazard Ratio

We now return to the hazard function which, as previously discussed, is generally considered to be the key quantity in survival analysis (see Sections 1.2.2 and 2.1.2 for details). While it may be of interest to interpret covariate effects directly on the hazard scale (i.e., additive hazards), recall that in Chapter 5 hazard ratios were used (and are ubiquitous in the literature). For this reason we aim to approximate the log-hazard function (at time \( t \)), i.e.,

\[
\log \lambda(t \mid x_1 = 1, \tilde{x}) - \log \lambda(t \mid x_1 = 0, \tilde{x}) = \hat{\beta}_1^*(t),
\]

using the least squares method described in Section 6.2. Thus,

\[
\lambda(t \mid x_1 = 1, \tilde{x}) / \lambda(t \mid x_1 = 0, \tilde{x}) = \exp[\hat{\beta}_1^*(t)]. \tag{6.9}
\]

In words, exponentiating \( \hat{\beta}_1^*(t) \) gives the hazard ratio (at time \( t \)) for the covariate \( x_1 \) assuming all other covariates are equal. Clearly this has a more standard interpretation than the overall and average-covariate hazard ratios introduced in Section 5.3 (compare (6.9) with (5.28) and (5.29) respectively).

Note that the regression coefficients are time-dependent (unlike those in Section 6.3) and, therefore, we wish to evaluate these coefficients over a range of times. In practice we choose a discrete grid of time points, \( t_1, \ldots, t_s \), at which there are \( s \) implied regression models, \( \log \lambda(t_1 \mid x), \ldots, \log \lambda(t_s \mid x) \), to be approximated. Thus, we can apply least squares separately to each of these implied regression quantities which, collectively, provides us with the values of the regression coefficients over the desired time range. Of course, using matrices, this can be achieved in one step (rather than considering each time point separately). We will not discuss the details here as the approach is essentially identical to that of m.l.e. simulation as discussed in Section 6.2.2 - albeit with the columns of \( \hat{Y} \) corresponding to \( s \) time points rather than \( m \) simulated m.l.e. vectors.
Example 6.3. Log-Hazard Regression in the Lung Cancer Data: Log-Logistic MPR Model

We continue with the full covariate log-logistic MPR analysis of the lung cancer data from Example 5.4. Using the method described above, the log-hazard function was approximated at 100 points between \( t_1 = 0 \) and \( t_{100} = 20 \). At each time point the linear approximation was very good with \( R^2 \) values ranging from 0.92 to 1.00 (and average value of 0.95).

The hazard ratios were calculated by exponentiating the regression coefficients (with 95% confidence intervals constructed using m.l.e. simulation); plots of these hazard ratios are shown in Figs. 6.1 - 6.9 below. Recall that in Example 5.4 we also investigated hazard ratios for this particular log-logistic analysis (see Figs. 5.11 - 5.19). Indeed the hazard ratio plots given below are very similar to those from Example 5.4 and, therefore, we will not reinterpret the results here. Finally, replicating Table 5.5 from Example 5.4, the least squares hazard ratios evaluated at \( t = 1.75 \), \( t = 4.75 \) and \( t = 12 \), respectively, are given in Table 6.3 below.

![Figure 6.1. Treatment (reference: palliative care) hazard ratios with confidence intervals for the log-logistic MPR model.](https://example.com/figure6_1.png)

---

5The fact that the overall, average covariates and least squares hazard ratios are in agreement gives assurance that these approaches produce reasonable estimates. However, the least squares estimate has the most standard interpretation.
Figure 6.2. Age (reference: < 50) hazard ratios with confidence intervals for the log-logistic MPR model.

Figure 6.3. WHO status (reference: normal) hazard ratios with confidence intervals for the log-logistic MPR model.
Figure 6.4. Sex (reference: female) hazard ratio with confidence intervals for the log-logistic MPR model.

Figure 6.5. Smoker (reference: no) hazard ratios with confidence intervals for the log-logistic MPR model.
6.4. HAZARD RATIO

Figure 6.6. Cell type (reference: squamous) hazard ratios with confidence intervals for the log-logistic MPR model.

Figure 6.7. Metastases (reference: no) hazard ratios with confidence intervals for the log-logistic MPR model.
Figure 6.8. Sodium (reference: $\geq 136$ mmol/l) hazard ratios with confidence intervals for the log-logistic MPR model.

Figure 6.9. Albumen (reference: $\geq 35$ g/l) hazard ratios with confidence intervals for the log-logistic MPR model.
## 6.4. HAZARD RATIO

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Treatment Hazard Ratio (C.I.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Palliative</td>
<td>1.00 (---)</td>
</tr>
<tr>
<td>Surgery</td>
<td>0.18 (0.07, 0.54)*</td>
</tr>
<tr>
<td>Chemo</td>
<td>0.66 (0.40, 1.04)*</td>
</tr>
<tr>
<td>Radio</td>
<td>0.58 (0.45, 0.75)*</td>
</tr>
<tr>
<td>C+R</td>
<td>0.05 (0.01, 0.19)*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Hazard Ratio (C.I.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 50</td>
<td>1.00 (---)</td>
</tr>
<tr>
<td>50 - 60</td>
<td>0.82 (0.48, 1.57)</td>
</tr>
<tr>
<td>60 - 70</td>
<td>0.86 (0.52, 1.51)</td>
</tr>
<tr>
<td>70 - 80</td>
<td>0.83 (0.48, 1.45)</td>
</tr>
<tr>
<td>80 +</td>
<td>0.93 (0.54, 1.70)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>WHO Status</th>
<th>Hazard Ratio (C.I.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>1.00 (---)</td>
</tr>
<tr>
<td>Light Work</td>
<td>0.98 (0.54, 1.85)</td>
</tr>
<tr>
<td>No Work</td>
<td>2.21 (1.23, 4.07)*</td>
</tr>
<tr>
<td>&gt; 50% Bed</td>
<td>3.32 (1.80, 6.18)*</td>
</tr>
<tr>
<td>Bedbound</td>
<td>4.14 (1.85, 8.45)*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sex</th>
<th>Male</th>
<th>Female</th>
<th>Hazard Ratio (C.I.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.05 (0.86, 1.25)</td>
<td>1.00 (---)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.14 (0.60, 2.14)</td>
<td>1.00 (---)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smoker</th>
<th>Yes</th>
<th>No</th>
<th>Ex-smoker</th>
<th>Missing</th>
<th>Hazard Ratio (C.I.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.26 (0.93, 1.74)</td>
<td>1.00 (---)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.23 (0.91, 1.71)</td>
<td>1.00 (---)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Squamous</th>
<th>Small</th>
<th>Adeno.</th>
<th>Other</th>
<th>Hazard Ratio (C.I.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00 (---)</td>
<td>1.67 (1.20, 2.32)*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.31 (0.96, 1.82)</td>
<td>1.02 (0.80, 1.29)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metastases</th>
<th>No</th>
<th>Yes</th>
<th>Missing</th>
<th>Hazard Ratio (C.I.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00 (---)</td>
<td>1.79 (1.37, 2.30)*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.29 (0.98, 1.67)</td>
<td>1.07 (0.59, 1.91)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sodium</th>
<th>≥ 136 mmol/l</th>
<th>&lt; 136 mmol/l</th>
<th>Missing</th>
<th>Hazard Ratio (C.I.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00 (---)</td>
<td>1.22 (1.03, 1.46)*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.82 (0.50, 1.44)</td>
<td>1.61 (1.07, 2.37)*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Albumen</th>
<th>≥ 35 g/l</th>
<th>&lt; 35 g/l</th>
<th>Missing</th>
<th>Hazard Ratio (C.I.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00 (---)</td>
<td>1.33 (1.08, 1.61)*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.51 (1.03, 2.10)*</td>
<td>1.07 (0.76, 1.43)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Symbol “∗” indicates that the C.I. does not contain the value one.
Chapter 7

Frailty Models

7.1 Introduction

In Chapters 4 and 5 we discussed regression models as a method for describing heterogeneity in the data through the inclusion of covariates. However, individuals will typically differ in ways that cannot be described entirely using the covariates which have been measured. Indeed it is unrealistic to presume that all sources of variation can be captured within a statistical model - no matter how detailed the model may be. There will always be unobservable features which affect the outcome. Hence, we may ask how much variation remains unexplained by our model and, furthermore, to what extent is this important? In the field of survival analysis frailty modelling - a topic which has received much attention in the literature - is the name given to the study of such variation.

Let $U$ be a random variable representing all unmeasured, or indeed unobservable, features which affect survival; in the literature $U$ is called the frailty term. Furthermore, $\lambda(t \mid u)$ is the hazard function for a given individual (or sub-population) with $U = u$ and is referred to as the conditional hazard$^1$. However, as $U$ is unknown and varies over the population, what we observe is the marginal distribution of $T$ with hazard function denoted by $\lambda_m(t)$ - often referred to as the population hazard - which can be quite different from

$^1$Note that the covariate-dependence has been suppressed for notational convenience but, of course, we can write $\lambda(t \mid x, u)$; see Section 7.4.1.
CHAPTER 7. FRAILTY MODELS

\( \lambda(t \mid u) \). Vaupel & Yashin (1985), who investigated hazard rates in a variety of theoretical scenarios, concluded that:

The observed dynamics at the population level will deviate from the underlying dynamics at the individual level . . . (and) may be surprisingly different (from each other). Researchers interested in uncovering these individual patterns . . . might benefit from an understanding of (unobservable) heterogeneity.

Aalen et al. (2008, sec. 6.5) contains similar developments and conclusions. Aalen & Gjessing (2001) discuss hazard shapes from the perspective of an underlying stochastic process and contrast this with frailty theory.

Note that in the above (and throughout the rest of this chapter) we assume \( U \) is time-constant. However, in some situations there is a need to extend to time-dependent frailty. This is commonly achieved by modelling the hazard using a stochastic process. For example, authors have considered diffusion (Woodbury & Manton, 1977; Myers, 1981; Yashin & Manton, 1997) and Lévy processes (Kebir, 1991; Gjessing et al., 2003); see also Singpurwalla (1995) and Aalen et al. (2008, chap. 11). Another approach is to generalise the piecewise exponential model (of Section 1.4.7) by including a different frailty term in each time-interval (Paik et al., 1994). Moreover, Perperoglou et al. (2006) introduced a model which approximates time-dependent frailty models (given their complexity). Although this is an interesting area for research, time-dependent frailty is beyond the scope of this thesis; we stick to the standard assumption that \( U \) does not depend on time.

7.2 Marginal Functions

If we let \( S(t \mid u) \) be the conditional survivor function, the marginal (or population) survivor function is given by

\[
S_m(t) = \Pr(T > t) = \int \Pr(T > t, U = u) \, du = \int \Pr(T > t \mid U = u) \Pr(U = u) \, du = \int S(t \mid u) g(u) \, du,
\]  

(7.1)
where $g(u)$ is the density function for $U$, and hence we can write

$$S_m(t) = E_U[S(t \mid u)],$$

(7.2)
i.e., the marginal survivor function is an average of conditional survivor functions over the distribution of $U$. Similarly, it is easy to show that the marginal density function is given by

$$f_m(t) = \int f(t \mid u)g(u)\,du = E_U[f(t \mid u)],$$

(7.3)
where $f(t \mid u) = \lambda(t \mid u)S(t \mid u)$ is the conditional density. Note that a model of the form (7.3) is often referred to as a **continuous mixture model** where $g(u)$ is the **mixing distribution** (see Hougaard (2000, sec. 2.2.7) for details on mixture models). Indeed Lancaster (1990, chap. 4) deals with frailty models under the heading “mixture models”.

In light of the above, it is tempting to suggest that the marginal hazard function is simply $E_U[\lambda(t \mid u)]$ - however this is incorrect. Using the relationship $\lambda(t) = f(t)/S(t)$ and equations (7.1) and (7.3) above, we have that

$$\lambda_m(t) = \frac{\lambda(t)}{S_m(t)} = \frac{\int \lambda(t \mid u)S(t \mid u)g(u)\,du}{S_m(t)},$$

but

$$\frac{S(t \mid u)g(u)}{S_m(t)} = \frac{\Pr(T > t, U = u)}{\Pr(T > t)} = \Pr(U = u \mid T > t).$$

Hence,

$$\lambda_m(t) = \int \lambda(t \mid u)\Pr(U = u \mid T > t)\,du$$

$$= E_U[\lambda(t \mid u) \mid T > t].$$

(7.4)
In words, at time $t$, the marginal hazard is an average of conditional hazards over the distribution of $U$ given survival to that time point. Note that an alternative derivation of (7.4) using probability statements is given in Appendix D.2.
7.3 Estimation

We now discuss how to estimate model parameters from a given set of data whilst accounting for frailty. As in Chapter 2, we consider the case of right-censored data where, for the $i$th individual ($i = 1, \ldots, n$), $t_i$ is the observed survival time and $\delta_i$ is the censoring indicator which equals zero if $t_i$ is a censoring time and one otherwise. In addition to the observed pair $(t_i, \delta_i)$, we now also have the unobserved frailty, $u_i$. If the frailties were observed quantities, inference would be based on the full likelihood (also called the complete data likelihood) which is given by

$$L(\theta, \phi) = \prod_{i=1}^{n} L_i(\theta \mid t_i, u_i) \cdot L_i(\phi \mid u_i) \cdot \delta_i S(t_i \mid u_i, \theta) \cdot g(u_i \mid \phi),$$

(7.5)

where $\theta$ is the vector of parameters in the conditional survival model and $\phi$ is the vector of frailty parameters\(^2\). The full log-likelihood is

$$\ell(\theta, \phi) = \sum_{i=1}^{n} \delta_i \log \lambda(t_i \mid u_i, \theta) + \log S(t_i \mid u_i, \theta) + \log g(u_i \mid \phi),$$

(7.6)

which would be maximised in the usual way.

As frailty is unobservable, we cannot use the full likelihood above. Instead, we integrate out the frailty terms to obtain the marginal likelihood

$$L_m(\theta, \phi) = \int \ldots \int L(\theta, \phi) \, du_1 \ldots du_n$$

$$= \prod_{i=1}^{n} \left[ \int \lambda(t_i \mid u_i, \theta)^{\delta_i} S(t_i \mid u_i, \theta) \cdot g(u_i \mid \phi) \, du_i \right].$$

(7.7)

Using the results of the preceding section, it is easy to see that

$$L_m(\theta, \phi) = \prod_{i=1}^{n} \lambda_m(t_i \mid \theta, \phi)^{\delta_i} S_m(t_i \mid \theta, \phi),$$

(7.8)

\(^2\)We have omitted covariate dependence for notational convenience.
which is unsurprising given that the observed data are marginal (as discussed in Section 7.1). The marginal log-likelihood is therefore given by

$$
\ell_m(\theta, \phi) = \sum_{i=1}^{n} \delta_i \log \lambda_m(t_i | \theta, \phi) + \log S_m(t_i | \theta, \phi).
$$

(7.9)

Although we use the marginal likelihood approach, it is not the only possibility. We now discuss some other approaches briefly (see Duchateau & Janssen (2008) for further details).

Viewing $u_i$ as missing data in the full log-likelihood, (7.6), leads naturally to the EM algorithm (Dempster et al., 1977) where, at the E step, all $u_i$ terms (and functions of $u_i$) are replaced by their conditional expectation, given the data $(t_i, \delta_i)$ and current estimates of $(\theta, \phi)$. This is followed by the M step which consists of maximising this likelihood w.r.t. $(\theta, \phi)$ to produce new estimates (the algorithm iterates between these two steps until convergence). Klein (1992), Nielsen et al. (1992) and Andersen et al. (1993, chap. 9) discuss this approach to fitting frailty models. In cases where the E step is intractable, the so-called Monte-Carlo EM (MCEM) algorithm can be used: here the E step is approximated by averaging over simulated samples of $u_i$ (Wei & Tanner, 1990; Vaida & Xu, 2000; Ripatti et al., 2002).

McGilchrist & Aisbett (1991) and McGilchrist (1993) essentially treat $u = (u_1, \ldots, u_n)^T$ as vector of additional parameters and maximise (7.6) w.r.t. $(\theta, u)$ for a given $\phi$ value. The estimated $u$ vector is then used to update the $\phi$ estimate - like EM, the procedure iterates between these two steps. Three versions of this procedure were implemented, namely: the method of best linear unbiased predictor (BLUP), maximum likelihood (ML) and restricted maximum likelihood (REML). Following Breslow & Clayton (1993), Ripatti & Palmgren (2000) applied a Laplace approximation to the integral in (7.7) which led to a procedure equivalent to the ML method of McGilchrist (1993). Note that these procedures may be viewed from a penalised likelihood perspective (Therneau et al., 2003). Hierarchical likelihood (Lee & Nelder, 1996, 2001), or h-likelihood, was developed for estimating generalized linear mixed models using a Laplace integral approximation. Ha et al. (2001, 2002) and Ha & Lee (2003, 2005) used this approach for fitting frailty models and noted equivalence with McGilchrist’s REML method.
7.4 Multiplicative Frailty

Thus far we have not specified how the frailty component may enter the model. However, we now consider the multiplicative frailty model defined by

$$\lambda(t \mid u) = u \lambda(t), \quad (7.10)$$

which is the standard frailty model in the literature (cf. Lancaster (1990), Andersen et al. (1993), Hougaard (2000), Duchateau & Janssen (2008), Aalen et al. (2008) and references therein). Here $\lambda(t)$ is an underlying function common to all individuals, which we will refer to as the basic hazard, and $u$ is a quantity that varies over the population with density function $g(u)$ (as described previously); it is clear (7.10) requires that $u \in (0, \infty)$. Furthermore, as $g(u)$ and $\lambda(t)$ both contain scale parameters, it is standard practice to assume that $E(U) = 1$ in order to eliminate this redundancy (Hougaard, 2000, sec. 7.2). With this assumption we have that

$$\lambda[t \mid E(U)] = E(U) \lambda(t) = \lambda(t), \quad (7.11)$$

i.e., the basic hazard, $\lambda(t)$, corresponds to that of an average or reference individual (the latter phrase appears in Duchateau & Janssen (2008, sec 1.5)).

From (7.4) we have that

$$\lambda_m(t) = \lambda(t) E(U \mid T > t), \quad (7.12)$$

i.e., the basic and marginal hazards differ by the factor $E(U \mid T > t)$ which is easily shown to be decreasing as a function of time, i.e., $dE(U \mid T > t)/dt < 0$ (see Appendix D.2). That $E(U \mid T > t)$ is decreasing intuitively reflects the fact that larger $u$ values imply larger hazards and, hence, the group of survivors is increasingly composed of individuals with smaller $u$ values (i.e., those less likely to experience an event). Moreover, since $E(U \mid T > t)$ decreases from $E(U \mid T > 0) = E(U) = 1$, we have that $E(U \mid T > t) < 1$ for $t > 0$ and, therefore, we can deduce that

$$\lambda_m(t) < \lambda(t),$$

for $t > 0$. Thus, we will underestimate the basic hazard rate if we do not account for the effect of frailty.
The multiplicative frailty model is generally attributed to Vaupel et al. (1979) who introduced the term “frailty” in reference to $u$ - a phrase which has stuck in the literature - however, in the same year, Lancaster (1979) also proposed the model (seemingly independently of Vaupel et al.) and referred to $u$ as an error term. In both cases $u$ was assumed to follow a gamma distribution - still the most common choice of frailty distribution (see Section 7.5). The ubiquity of the multiplicative frailty model is such that we may refer to it simply as the frailty model. One reason for its ubiquity is the obvious connection with the PH model (see Example 7.1 and Lancaster (1990, sec. 4.2)). Another advantage of the multiplicative specification (from a mathematical point of view) is that various quantities of interest can be expressed in terms of the Laplace transform of $g(u)$; we discuss this in Section 7.4.2.

### 7.4.1 Interpreting Covariate Effects in the Multiplicative Frailty Model

In the preceding sections we have suppressed covariate dependence in our notation for convenience. However, it is generally intended that covariates are present in the model. Indeed, a primary role of such analyses is to adjust covariate effects in the presence of frailty\(^3\). Thus, making covariate dependence explicit in our notation, we write

$$\lambda(t \mid x, u) = u \lambda(t \mid x),$$

which is conditional on observed heterogeneity via covariates, $x = (1, x_1, \ldots, x_p)^T$, and unobserved heterogeneity via the multiplicative frailty term, $u$.

We have previously shown how covariates can enter $\lambda(t \mid x)$ in Chapters 4 and 5, i.e., $\lambda(t \mid x)$ may be a single parameter regression (SPR) model or a multi-parameter regression (MPR) model.

As before, we can interpret covariate effects in terms of hazard ratios. The hazard ratio, at time $t$, for an individual with $x_1 = 1$ to an individual with $x_1 = 0$, given that all other things are equal (i.e., frailty and other

\(^3\)Of course the frailty parameters, $\phi$, are also of interest; they describe the degree of unexplained heterogeneity remaining after accounting for observable heterogeneity.
covariates), is given by

\[
\psi(t \mid \tilde{x}, u) = \frac{\lambda(t \mid x_1 = 1, \tilde{x}, u)}{\lambda(t \mid x_1 = 0, \tilde{x}, u)}
\]

\[
= \frac{u \lambda(t \mid x_1 = 1, \tilde{x})}{u \lambda(t \mid x_1 = 0, \tilde{x})}
\]

\[
= \frac{\lambda(t \mid x_1 = 1, \tilde{x})}{\lambda(t \mid x_1 = 0, \tilde{x})}
\]

\[
= \psi(t \mid \tilde{x}),
\]

where \(\tilde{x}\) represents all covariates in \(x\) other than \(x_1\). Thus, the individual hazard ratio does not depend on \(u\); it depends only on the functional form of \(\lambda(t \mid x)\) and can be interpreted in the usual way. For example, if we assume that \(\lambda(t \mid x)\) is a Weibull MPR model, then the hazard ratio is as described in Section 5.2.1.

Although the functional form of the hazard ratio is the same in both the frailty and non-frailty cases (as shown above), it is important to note the difference between the two. Consider the following hazard models:

Frailty: \(u \lambda(t \mid x, \theta), \quad u \sim g(u \mid \phi)\),

Non-Frailty: \(\lambda(t \mid x, \theta)\).

It is true that the hazard ratios in both models have the same functional form, denoted by \(\psi(t \mid \tilde{x}, \theta)\). However, given a set of data, the estimated hazard parameters from the frailty analysis are adjusted for the frailty parameters; we may write \(\hat{\theta}(\hat{\phi})\) to make this explicit. Hence, even though the functional form of \(\psi(t \mid \tilde{x}, \theta)\) is the same in both analyses, its numeric value is not (owing to the fact that \(\hat{\theta}\) will differ numerically in the two analyses - see Example 7.1). The advantage of incorporating a frailty component in the analysis is two-fold: first, the estimated effects of measured covariates may be more robust as other sources of hazard variation have been accounted for and, second, the frailty parameters, \(\phi\), can inform us about this unexplained variation.
Note that the above applies to the ratio of conditional (or individual) hazards; however, the ratio of marginal (or population) hazards is given by

$$\frac{\lambda_m(t \mid x_1 = 1, \tilde{x})}{\lambda_m(t \mid x_1 = 0, \tilde{x})} = \frac{\lambda(t \mid x_1 = 1, \tilde{x}) E(U \mid T > t, x_1 = 1, \tilde{x})}{\lambda(t \mid x_1 = 0, \tilde{x}) E(U \mid T > t, x_1 = 0, \tilde{x})} = \psi(t \mid \tilde{x}) \zeta(t, x),$$

where $\zeta(t, x) = E(U \mid T > t, x_1 = 1, \tilde{x})/E(U \mid T > t, x_1 = 0, \tilde{x})$ and $\psi(t \mid \tilde{x})$ is the individual hazard ratio as before. Thus, failure to account for frailty can lead to incorrect conclusions being drawn about the basic hazard ratio. For example, assume that hazards are proportional at an individual level so that $\psi(t \mid \tilde{x}) = \exp(\beta_1)$. However, as a consequence of frailty, the observed (marginal data) hazard ratio is $\exp(\beta_1) \zeta(t, x)$. This may lead us to reject (wrongly) the assumption of proportional hazards and to then consider more complicated models which can handle time-dependent hazard ratios (e.g., a Weibull MPR model). Furthermore, we could falsely conclude, for example, that a treatment effect diminishes over time. This phenomenon is well known; for further details see Aalen et al. (2008, sec. 6.5).

It is clear that accounting for frailty is important in making more robust inferences about covariate effects. In practice the utility of incorporating frailty in our analysis depends on: (i) the extent to which unobservable heterogeneity can be captured by the multiplicative $u$ term - largely a mathematical convenience - and (ii) the flexibility of the assumed frailty distribution, $g(u)$. Of course, assumptions must be made in any statistical model with the aim of approximating reality and making reasonable conclusions (as discussed in Section 2.2).

### 7.4.2 Laplace Transform

It is easy to show that, in the case of multiplicative frailty, $\Lambda(t \mid u) = u \Lambda(t)$. Hence, the conditional survivor function is

$$S(t \mid u) = \exp[-\Lambda(t \mid u)] = \exp[-u \Lambda(t)],$$
and, from (7.2), the marginal survivor function is

\[ S_m(t) = E\{\exp[-u \Lambda(t)]\} \]
\[ = \int_0^\infty \exp[-u \Lambda(t)]g(u) \, du. \]

Noting that the definition of the Laplace transform of a function \( g(u) \) is

\[ \mathcal{L}(s) = \int_0^\infty \exp(-us)g(u) \, du, \quad (7.13) \]

(see Abramowitz & Stegun (1964, chap. 29)), it is then clear that

\[ S_m(t) = \mathcal{L}[\Lambda(t)], \quad (7.14) \]

i.e., the marginal survivor function is the Laplace transform of \( g(u) \) evaluated at \( s = \Lambda(t) \) (the basic integrated hazard function). Of course, we can also express the marginal hazard in terms of the Laplace transform via

\[ \lambda_m(t) = -\frac{d}{dt} \log S_m(t) \]
\[ = \lambda(t) \frac{-\mathcal{L}'[\Lambda(t)]}{\mathcal{L}[\Lambda(t)]}. \quad (7.15) \]

Comparing this with (7.12) we see that \( E(U \mid T > t) = -\mathcal{L}'[\Lambda(t)]/\mathcal{L}[\Lambda(t)] \).

From (7.9), the marginal log-likelihood is given by

\[ \ell_m(\theta, \phi) = \sum_{i=1}^n \delta_i (\log \lambda(t_i) + \log \{-\mathcal{L}'[\Lambda(t_i)]\} - \log \mathcal{L}[\Lambda(t_i)] + \log \mathcal{L}[\Lambda(t_i)]) \]
\[ = \sum_{i=1}^n \delta_i \log \lambda(t_i) + \delta_i \log \{-\mathcal{L}'[\Lambda(t_i)]\} + (1 - \delta_i) \log \mathcal{L}[\Lambda(t_i)] \]
\[ = \sum_{i=1}^n \delta_i \log \lambda(t_i) + \log \left\{ (-1)^{\delta_i} \mathcal{L}^{(\delta_i)}[\Lambda(t_i)] \right\}, \quad (7.16) \]

where \( \mathcal{L}^{(1)}[\Lambda(t)] = \mathcal{L}'[\Lambda(t)] \) and \( \mathcal{L}^{(0)}[\Lambda(t)] = \mathcal{L}[\Lambda(t)] \). An alternative derivation for the above log-likelihood is given in Appendix D.3.1.
7.5. THE GAMMA FRAILTY MODEL

This connection with the Laplace transform appears to have first been made by Lancaster & Nickell (1980) which, as it turns out, is more than just a mathematical observation. For example, Elbers & Ridder (1982) exploit properties of the Laplace transform in proving that both the conditional survival distribution and the frailty distribution can be identified from marginal data if covariates are included (and $E(U) < \infty$). Heckman & Singer (1984) build on this work and also prove identifiability without the requirement of covariates; these authors instead assume that the conditional survival model is specified up to a finite set of parameters (see also Lancaster (1990, chap. 7)). Hougaard (1984, 1986, 2000) also made use of the Laplace transform in his study of frailty models and, among other things, emphasised the importance of assuming a frailty distribution with tractable Laplace transform for the purpose of straightforward analysis (see Duchateau & Janssen (2008) also); of course, this is clear from (7.14) - (7.16) above. Note that one could start by specifying a functional form for $L(s)$ and, if required, derive the corresponding density, $g(u)$, by means of Fourier inversion (Feller, 1971).

7.5 The Gamma Frailty Model

The results of Section 7.4 apply to any frailty distribution. We now consider the so-called gamma frailty model where the multiplicative frailty, $u$, follows a one-parameter gamma distribution with density function given by

$$g(u) = \frac{1}{\phi^{1/\phi} \Gamma(1/\phi)} u^{1/\phi - 1} \exp(-u/\phi). \quad (7.17)$$

Here $\Gamma(\cdot)$ is the gamma function (Abramowitz & Stegun, 1964, chap. 6) and $\phi > 0$. This distribution fulfills the requirement that $E(U) = 1$ and, furthermore, it is easy to show that

$$\phi = \text{var}(U). \quad (7.18)$$

Thus, the $\phi$ parameter has an intuitive interpretation - it tells us the degree of unobserved heterogeneity in the hazard. In particular, it is clear that as $\phi \to 0$, all frailties are concentrated at $E(U) = 1$ and, hence, $\lambda_m(t) = \lambda(t)$, i.e., there is no unobserved variation in the hazard.
The gamma distribution has Laplace transform given by
\[
    L(s) = (1 + \phi s)^{-1/\phi},
\]  
(7.19)
(see Appendix D.3.2). This simple analytic form makes the gamma frailty model easy to apply in practice; indeed Lancaster (1979) and Vaupel et al. (1979) originally made the assumption of gamma distributed frailty and it continues to be the most popular frailty model in the literature. Furthermore, the derivative of the Laplace transform also has a simple form given by
\[
    L'(s) = -\frac{1}{\phi} (1 + \phi s)^{-1/\phi - 1} \cdot \phi = -(1 + \phi s)^{-(1/\phi+1)}. 
\]  
(7.20)
Hence, the marginal survivor and hazard function are given by
\[
    S_m(t) = L[\Lambda(t)] = [1 + \phi \Lambda(t)]^{-1/\phi},
\]  
(7.21)
and
\[
    \lambda_m(t) = \lambda(t) \frac{-L'[\Lambda(t)]}{L[\Lambda(t)]} = \frac{\lambda(t)}{1 + \phi \Lambda(t)},
\]  
(7.22)
respectively; clearly \(\lim_{\phi \to 0} S_m(t) = \exp[-\Lambda(t)] = S(t)\) and \(\lim_{\phi \to 0} \lambda_m(t) = \lambda(t)\) as expected. Note that, from (7.19) and (7.20), we can write
\[
    L^{(\delta)}(s) = (-1)^\delta (1 + \phi s)^{-(1/\phi+\delta)},
\]
for \(\delta \in \{0, 1\}\) and, therefore, the marginal log-likelihood is given by
\[
    \ell_m(\theta, \phi) = \sum_{i=1}^{n} \delta_i \log \lambda(t_i) + \log \left\{ (-1)^\delta [L^{(\delta_i)}(\Lambda(t_i))] \right\}
\]
\[
    = \sum_{i=1}^{n} \delta_i \log \lambda(t_i) + \log \left\{ (-1)^{\delta_i} (-1)^{\delta_i} (1 + \phi \Lambda(t_i))^{-(1/\phi+\delta_i)} \right\}
\]
\[
    = \sum_{i=1}^{n} \delta_i \log \lambda(t_i) - (1/\phi + \delta_i) \log [1 + \phi \Lambda(t_i)].
\]  
(7.23)
Of course we could also arrive at (7.23) using the fact that, from (7.9), \(\ell_m(\theta, \phi) = \sum \delta_i \log \lambda_m(t_i) + \log S_m(t_i)\) and the above expressions for \(\lambda_m(t)\)
and \( S_m(t) \). The score equations for the above log-likelihood are given by

\[
\frac{\partial \ell_m}{\partial \theta_j} = \sum_{i=1}^{n} \delta_i \frac{\partial}{\partial \theta_j} \log \lambda(t_i) - \frac{1}{1 + \phi \Lambda(t_i)} \frac{\partial \Lambda(t_i)}{\partial \theta_j}
\]

for \( j = 1, \ldots, \dim(\theta) \), and

\[
\frac{\partial \ell_m}{\partial \phi} = \sum_{i=1}^{n} \left( \frac{1}{\phi^2} \right) \log [1 + \phi \Lambda(t_i)] - \frac{1}{1 + \phi \Lambda(t_i)} \frac{\Lambda(t_i)}{1 + \phi \Lambda(t_i)}. \tag{7.25}
\]

Recall from Chapter 2 that the log-likelihood for a standard non-frailty model is given by \( \ell(\theta) = \sum_{i=1}^{n} \delta_i \log \lambda(t_i) - \Lambda(t_i) \) with corresponding score equations given by \( \frac{\partial \ell}{\partial \theta_j} = \sum_{i=1}^{n} \delta_i \frac{\partial \log \lambda(t_i)}{\partial \theta_j} - \frac{\partial \Lambda(t_i)}{\partial \theta_j} \). Thus, for any standard survival model, we will have expressions for \( \lambda(t), \Lambda(t), \frac{\partial \log \lambda(t)}{\partial \theta_j} \) and \( \frac{\partial \Lambda(t)}{\partial \theta_j} \), respectively. Hence, we can easily generalise the estimation procedure to incorporate gamma frailty by plugging these expressions into the above marginal likelihood, (7.23), and score equations, (7.24) and (7.25) (See Example 7.3).

**Example 7.1.** *Weibull-Gamma SPR Analysis of Lung Cancer Data*

We now assume a Weibull SPR model for the hazard regression component, i.e.,

\[
\lambda(t \mid x) = \exp(x^T \beta) \gamma t^{\gamma-1}, \tag{7.26}
\]

which, as shown in Section 4.5.1, is a proportional hazards model so that the hazard ratio for covariate \( x_1 \), for example, is simply \( \exp(\beta_1) \). Thus, the conditional hazard is given by

\[
\lambda(t \mid x, u) = u \exp(x^T \beta) \gamma t^{\gamma-1}
\]

\[
= \exp(x^T \beta + \log u) \gamma t^{\gamma-1}
\]

\[
= \exp(x^T \beta + \epsilon) \gamma t^{\gamma-1} \tag{7.27}
\]

where the last line shows that frailty can be interpreted as an error term in the linear predictor for this model and indeed any proportional hazards
model (see Lancaster (1990, sec. 4.2)). Therefore, multiplicative frailty is quite natural in this setting\(^4\).

From (7.22), the marginal hazard is given by

\[
\lambda_m(t \mid x) = \frac{\exp(x^T \beta) \gamma t^{\gamma-1}}{1 + \phi \exp(x^T \beta) t^{\gamma}},
\]

which is that of a Burr model (see Section 4.5.4). Hence, the Weibull-gamma model can be estimated using the Burr log-likelihood (Appendix B.2.4)\(^5\).

We fitted the above model to the lung cancer data and the corresponding non-frailty Weibull SPR model (i.e., \(\phi = 0\)) for comparison; the results are given in Table 7.1. Firstly we note that the \(AIC\) is much lower in the frailty analysis (\(\Delta_{AIC} = 3732.54 - 3697.02 = 35.52\)) and, therefore, it appears that there is unexplained variation in the hazard with \(\hat{\phi} = 0.73\). However, we also note that, although the Weibull-gamma analysis adjusts for this variation, both analyses are qualitatively similar (in terms of \(\hat{\beta}\) values and their corresponding significance). The main differences are as follows:

1. **Standard Errors:** All standard errors are larger in the Weibull-gamma analysis owing to the fact that other sources of variation have been accounted for.

2. **Smoker:** In the Weibull analysis the effect of being a smoker (relative to a non-smoker) is significant and, furthermore, the effect of being an ex-smoker is near significant. However, in the Weibull-gamma analysis both are non-significant.

3. **Cell Type:** All cell types are significant relative to the reference group (squamous) in the Weibull analysis whereas, in the Weibull-gamma analysis only the small cell type is significant.

We will not provide a detailed discussion of the covariate effects (i.e., hazard ratios) as the results are clear from Table 7.1 and, moreover, a similar, but more general, analysis is the subject of Example 7.4.

\(^4\)Recall that the developments of the preceding sections do not require \(\lambda(t \mid x)\) to have a PH specification but, as shown, this does allow a more specialised interpretation of \(u\).

\(^5\)Clearly the Burr model can be interpreted as the marginal model arising from the Weibull-gamma frailty specification. However, it can also be viewed as a flexible survival model without appealing to its frailty interpretation (as in Chapters 4 and 5 for example).
### 7.5. THE GAMMA FRAILTY MODEL

Table 7.1. Weibull and Weibull-Gamma SPR Full Covariate Analyses

<table>
<thead>
<tr>
<th></th>
<th>Weibull</th>
<th>Weibull-Gamma</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>S.E.</td>
</tr>
<tr>
<td>Frailty: $\phi$</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>Shape: $\gamma$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scale: $\beta$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-3.70</td>
<td>(0.35)</td>
</tr>
<tr>
<td><strong>Treatment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palliative</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Surgery</td>
<td>-1.23</td>
<td>(0.25)</td>
</tr>
<tr>
<td>Chemo</td>
<td>-0.49</td>
<td>(0.21)</td>
</tr>
<tr>
<td>Radio</td>
<td>-0.31</td>
<td>(0.10)</td>
</tr>
<tr>
<td>C+R</td>
<td>-0.92</td>
<td>(0.24)</td>
</tr>
<tr>
<td><strong>Age Group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 50</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>50 - 60</td>
<td>0.13</td>
<td>(0.26)</td>
</tr>
<tr>
<td>60 - 70</td>
<td>0.07</td>
<td>(0.24)</td>
</tr>
<tr>
<td>70 - 80</td>
<td>0.06</td>
<td>(0.25)</td>
</tr>
<tr>
<td>80 +</td>
<td>0.04</td>
<td>(0.27)</td>
</tr>
<tr>
<td><strong>WHO Status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Light Work</td>
<td>0.08</td>
<td>(0.19)</td>
</tr>
<tr>
<td>No Work</td>
<td>0.54</td>
<td>(0.19)</td>
</tr>
<tr>
<td>&gt; 50% Bed</td>
<td>1.03</td>
<td>(0.20)</td>
</tr>
<tr>
<td>Bedbound</td>
<td>1.71</td>
<td>(0.29)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Male</td>
<td>-0.04</td>
<td>(0.09)</td>
</tr>
<tr>
<td><strong>Smoker</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Yes</td>
<td>0.40</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Ex-smoker</td>
<td>0.27</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Missing</td>
<td>0.30</td>
<td>(0.27)</td>
</tr>
<tr>
<td><strong>Cell Type</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Squamous</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Small</td>
<td>0.71</td>
<td>(0.16)</td>
</tr>
<tr>
<td>Adeno.</td>
<td>0.32</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Other</td>
<td>0.21</td>
<td>(0.10)</td>
</tr>
<tr>
<td><strong>Metastases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Yes</td>
<td>0.78</td>
<td>(0.12)</td>
</tr>
<tr>
<td>Missing</td>
<td>0.36</td>
<td>(0.13)</td>
</tr>
<tr>
<td><strong>Sodium</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\geq$ 136 mmol/l</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>$&lt; 136$ mmol/l</td>
<td>0.33</td>
<td>(0.09)</td>
</tr>
<tr>
<td>Missing</td>
<td>-0.08</td>
<td>(0.22)</td>
</tr>
<tr>
<td><strong>Albumen</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\geq$ 35 g/l</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>$&lt; 35$ g/l</td>
<td>0.43</td>
<td>(0.09)</td>
</tr>
<tr>
<td>Missing</td>
<td>0.48</td>
<td>(0.17)</td>
</tr>
</tbody>
</table>

$\ell(\hat{\theta})$      | -1839.27|               |

\[ AIC \]

3732.54

3697.02

Note: H.R. = hazard ratio = $\exp(\beta)$. 
Example 7.2. A Comparison of Weibull Models for the Lung Cancer Data

Let the basic hazard be a Weibull MPR model (Section 5.2.1), i.e.,

$$\lambda(t \mid x, z) = \exp(x^T \beta) \exp(z^T \alpha) t^{\exp(z^T \alpha) - 1},$$

(7.29)

where $x = (1, x_1, \ldots, x_p)^T$ and $z = (1, z_1, \ldots, z_q)^T$ are the scale and shape covariate vectors with corresponding regression coefficients $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T$ and $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_q)^T$ respectively. Hence, with multiplicative gamma frailty, the marginal hazard function is

$$\lambda_m(t \mid x, z) = \frac{\exp(x^T \beta) \exp(z^T \alpha) t^{\exp(z^T \alpha) - 1}}{1 + \phi \exp(x^T \beta) t^{\exp(z^T \alpha)}},$$

(7.30)

which is the Burr MPR hazard (Section 5.2.4) - albeit without $\phi$ depending on covariates.\(^6\)

In Example 7.1 we focussed on the Weibull SPR model (full covariate analysis with and without gamma frailty). Here we compare a variety of models contained in the family defined by (7.30) above (which clearly includes the two models of Example 7.1). In particular we consider models with the following regression structures:

1. **Null Model**: No covariates, i.e., $x^T \beta = \beta_0$, $z^T \alpha = \alpha_0$.

2. **Single Factor SPR**: Scale depends on one factor.\(^7\)

3. **Single Factor MPR**: Scale and shape depend on one factor.

4. **Multi-Factor SPR**: Scale depends on all factors.

5. **Multi-Factor MPR**: Scale and shape depend on all factors.

For each of the above regression structures we consider three frailty specifications:

(a) **No Frailty**: In this case $\phi = 0$ so that (7.30) reduces to (7.29). Therefore, we simply have a standard Weibull (non-frailty) model (Section 5.2.1).

---

\(^6\)Recall that the $\phi$ parameter was denoted by $\rho$ in Section 5.2.4. In the current context of frailty, we consider regressing $\phi$ on covariates in Section 7.7.

\(^7\)We consider the treatment model. Thus $x^T \beta = \beta_0 + x_1 \beta_1 + x_2 \beta_2 + x_3 \beta_3 + x_4 \beta_4$, where $x_1, x_2, x_3$ and $x_4$ are the four binary variables representing treatment.
7.5. THE GAMMA FRAILTY MODEL

(b) **Fixed Frailty:** It is of interest to fix the frailty variance at $\phi = 1$ as (7.30) becomes that of a log-logistic model (Section 5.2.3).

(c) **Frailty:** The Weibull-gamma frailty model with $\phi$ estimated from the data, i.e., the Burr model (Section 5.2.4).

Hence, there are 15 models (1a, 1b, 1c, 2a, ..., 5c); these can be estimated using the Weibull, log-logistic and Burr likelihoods given in Appendix B.2. The models are summarised (in order of complexity) in Table 7.2 below.

Table 7.2. Summary of Fitted Weibull Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Scale</th>
<th>Shape</th>
<th>Frailty</th>
<th>S.E.</th>
<th>dim($\theta$)</th>
<th>$\ell(\hat{\theta})$</th>
<th>AIC</th>
<th>$\Delta_{AIC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a: Null</td>
<td>——</td>
<td>——</td>
<td>$\phi = 0.00$</td>
<td>——</td>
<td>2</td>
<td>-2061.4</td>
<td>4126.8</td>
<td>445.8</td>
</tr>
<tr>
<td>1b: Null-G*</td>
<td>——</td>
<td>——</td>
<td>$\phi = 1.00$</td>
<td>——</td>
<td>2</td>
<td>-2055.3</td>
<td>4114.5</td>
<td>433.5</td>
</tr>
<tr>
<td>1c: Null-G</td>
<td>——</td>
<td>——</td>
<td>$\phi = 0.70$ (0.21)</td>
<td>3</td>
<td>-2054.4</td>
<td>4114.7</td>
<td>433.6</td>
<td></td>
</tr>
<tr>
<td>2a: SF-SPR</td>
<td>Treat.</td>
<td>——</td>
<td>$\phi = 0.00$</td>
<td>——</td>
<td>6</td>
<td>-1960.8</td>
<td>3933.5</td>
<td>252.4</td>
</tr>
<tr>
<td>2b: SF-SPR-G*</td>
<td>Treat.</td>
<td>——</td>
<td>$\phi = 1.00$</td>
<td>——</td>
<td>6</td>
<td>-1943.7</td>
<td>3899.3</td>
<td>218.2</td>
</tr>
<tr>
<td>2c: SF-SPR-G</td>
<td>Treat.</td>
<td>——</td>
<td>$\phi = 0.91$ (0.20)</td>
<td>7</td>
<td>-1943.6</td>
<td>3901.1</td>
<td>220.0</td>
<td></td>
</tr>
<tr>
<td>3a: SF-MPR</td>
<td>Treat.</td>
<td>Treat.</td>
<td>$\phi = 0.00$</td>
<td>——</td>
<td>10</td>
<td>-1938.1</td>
<td>3896.2</td>
<td>215.2</td>
</tr>
<tr>
<td>3b: SF-MPR-G*</td>
<td>Treat.</td>
<td>Treat.</td>
<td>$\phi = 1.00$</td>
<td>——</td>
<td>10</td>
<td>-1927.7</td>
<td>3875.5</td>
<td>194.4</td>
</tr>
<tr>
<td>3c: SF-MPR-G</td>
<td>Treat.</td>
<td>Treat.</td>
<td>$\phi = 0.74$ (0.19)</td>
<td>11</td>
<td>-1926.9</td>
<td>3875.9</td>
<td>194.8</td>
<td></td>
</tr>
<tr>
<td>4a: MF-SPR</td>
<td>All</td>
<td>——</td>
<td>$\phi = 0.00$</td>
<td>——</td>
<td>27</td>
<td>-1839.3</td>
<td>3732.5</td>
<td>51.5</td>
</tr>
<tr>
<td>4b: MF-SPR-G*</td>
<td>All</td>
<td>——</td>
<td>$\phi = 1.00$</td>
<td>——</td>
<td>27</td>
<td>-1821.7</td>
<td>3697.5</td>
<td>16.4</td>
</tr>
<tr>
<td>4c: MF-SPR-G</td>
<td>All</td>
<td>——</td>
<td>$\phi = 0.73$ (0.16)</td>
<td>28</td>
<td>-1820.5</td>
<td>3697.0</td>
<td>15.9</td>
<td></td>
</tr>
<tr>
<td>5a: MF-MPR</td>
<td>All</td>
<td>All</td>
<td>$\phi = 0.00$</td>
<td>——</td>
<td>52</td>
<td>-1798.4</td>
<td>3700.9</td>
<td>19.8</td>
</tr>
<tr>
<td>5b: MF-MPR-G*</td>
<td>All</td>
<td>All</td>
<td>$\phi = 1.00$</td>
<td>——</td>
<td>52</td>
<td>-1790.0</td>
<td>3683.9</td>
<td>2.8</td>
</tr>
<tr>
<td>5c: MF-MPR-G</td>
<td>All</td>
<td>All</td>
<td>$\phi = 0.60$ (0.16)</td>
<td>53</td>
<td>-1787.5</td>
<td>3681.1</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Model naming scheme: Null $\Rightarrow$ no covariates, SF $\Rightarrow$ single factor (i.e., treatment) and MF $\Rightarrow$ multi-factor (i.e., all covariates); SPR $\Rightarrow$ scale only and MPR $\Rightarrow$ scale and shape; G $\Rightarrow$ gamma frailty model with $\phi$ estimated from the data and G* $\Rightarrow$ gamma frailty model with $\phi = 1$.

---

Many of these have appeared in previous examples: Ex. 2.1 (1a, 1b), Ex. 4.2 (2a, 2b), Ex. 5.1 (3a, 3b), Ex. 5.2 (2c, 3c), Ex. 5.4 (5b, alluded to 5a), Ex. 7.1 (4a, 4c).
We can see that the best fitting model (according to AIC) is the most complex model, i.e., the MF-MPR-G model. This is interesting because, in light of the discussion at the end of Section 7.4.1 (second last paragraph), we may question whether the MPR and frailty extensions (of the SPR model) are required simultaneously - both offer explanations for time-dependent hazard ratios. However, if we drop the frailty component, and consider the MF-MPR model, the AIC increases by 19.8 units. Similarly, if we maintain frailty but reduce to SPR (i.e., the MF-SPR-G model) then AIC increases by 15.9 units. Thus, we conclude that the effect of including frailty does not abolish the usefulness of the MPR extension and vice versa.

As $\phi = \text{var}(U)$ describes the degree of unexplained variation, it is not surprising to see that, of the five frailty models (i.e., those of type (c)), $\hat{\phi}$ is lowest for the most complex model (MF-MPR-G). However, the Null-G model - which should have the largest amount of unexplained variation - has the next lowest $\hat{\phi}$ value. Ignoring this null model however, the order of $\hat{\phi}$ values is as we would expect: the more complex the model and, hence, the more variation it explains, the lower its $\hat{\phi}$ value. Still, we notice that the $\hat{\phi}$ values for the SF-MPR-G and MF-SPR-G models are approximately equal; we would not expect this given the large difference in AIC values ($\approx 179$ in favour of the MF-SPR-G model). However, these somewhat anomalous results are less surprising when viewed in conjunction with the corresponding standard errors, i.e., the numerical differences in $\hat{\phi}$ values are small when compared to the precision of these estimates.

Table 7.3 contains some comparisons of different model types. Firstly we can see that extending the model to include frailty reduces the AIC significantly in all cases (24 units on average). However, in this dataset, there is little difference between estimating the frailty variance (i.e., the Burr model) and fixing it at $\phi = 1$ (i.e., the log-logistic model). Comparing the single factor models to the multi-factor models, we can see that the $\Delta_{AIC}$ values are approximately equal to 200. Therefore, even though accounting for

---

$^9$The MPR Weibull model assumes that hazard ratios are time-dependent at an individual level, and, hence, at a population level (Section 5.2.1). The SPR Weibull-gamma model assumes that hazard ratios are constant at an individual level (Section 4.5.1) but time-dependent at a population level due to frailty (Section 7.4.1).
### 7.5. THE GAMMA FRAILTY MODEL

Table 7.3. A Comparison of Weibull Models

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Model 2</th>
<th>$AIC_1$</th>
<th>$AIC_2$</th>
<th>$\Delta_{AIC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Frailty vs Frailty</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Null</td>
<td>Null-G</td>
<td>4126.8</td>
<td>4114.7</td>
<td>12.1</td>
</tr>
<tr>
<td>SF-SPR</td>
<td>SF-SPR-G</td>
<td>3933.5</td>
<td>3901.1</td>
<td>32.4</td>
</tr>
<tr>
<td>SF-MPR</td>
<td>SF-MPR-G</td>
<td>3896.2</td>
<td>3875.9</td>
<td>20.4</td>
</tr>
<tr>
<td>MF-SPR</td>
<td>MF-SPR-G</td>
<td>3732.5</td>
<td>3697.0</td>
<td>35.5</td>
</tr>
<tr>
<td>MF-MPR</td>
<td>MF-MPR-G</td>
<td>3700.9</td>
<td>3681.1</td>
<td>19.8</td>
</tr>
<tr>
<td>$\bar{\Delta}_{AIC} = 24.0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\phi = 1$ vs $\hat{\phi}$

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Model 2</th>
<th>$AIC_1$</th>
<th>$AIC_2$</th>
<th>$\Delta_{AIC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null-G*</td>
<td>Null-G</td>
<td>4114.5</td>
<td>4114.7</td>
<td>-0.2</td>
</tr>
<tr>
<td>SF-SPR-G*</td>
<td>SF-SPR-G</td>
<td>3899.3</td>
<td>3901.1</td>
<td>-1.8</td>
</tr>
<tr>
<td>SF-MPR-G*</td>
<td>SF-MPR-G</td>
<td>3875.5</td>
<td>3875.9</td>
<td>-0.4</td>
</tr>
<tr>
<td>MF-SPR-G*</td>
<td>MF-SPR-G</td>
<td>3697.5</td>
<td>3697.0</td>
<td>0.4</td>
</tr>
<tr>
<td>MF-MPR-G*</td>
<td>MF-MPR-G</td>
<td>3683.9</td>
<td>3681.1</td>
<td>2.8</td>
</tr>
<tr>
<td>$\bar{\Delta}_{AIC} = 0.2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Single Factor vs Multi-Factor

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Model 2</th>
<th>$AIC_1$</th>
<th>$AIC_2$</th>
<th>$\Delta_{AIC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF-SPR</td>
<td>MF-SPR</td>
<td>3933.5</td>
<td>3732.5</td>
<td>201.0</td>
</tr>
<tr>
<td>SF-SPR-G*</td>
<td>MF-SPR-G*</td>
<td>3899.3</td>
<td>3679.5</td>
<td>201.8</td>
</tr>
<tr>
<td>SF-SPR-G</td>
<td>MF-SPR-G</td>
<td>3901.1</td>
<td>3697.0</td>
<td>204.1</td>
</tr>
<tr>
<td>SF-MPR</td>
<td>MF-MPR</td>
<td>3896.2</td>
<td>3700.9</td>
<td>195.4</td>
</tr>
<tr>
<td>SF-MPR-G*</td>
<td>MF-MPR-G*</td>
<td>3875.5</td>
<td>3683.9</td>
<td>191.6</td>
</tr>
<tr>
<td>SF-MPR-G</td>
<td>MF-MPR-G</td>
<td>3875.9</td>
<td>3681.1</td>
<td>194.8</td>
</tr>
<tr>
<td>$\bar{\Delta}_{AIC} = 198.1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SPR vs MPR

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Model 2</th>
<th>$AIC_1$</th>
<th>$AIC_2$</th>
<th>$\Delta_{AIC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF-SPR</td>
<td>SF-MPR</td>
<td>3933.5</td>
<td>3896.2</td>
<td>37.3</td>
</tr>
<tr>
<td>SF-SPR-G*</td>
<td>SF-MPR-G*</td>
<td>3899.3</td>
<td>3875.5</td>
<td>23.8</td>
</tr>
<tr>
<td>SF-SPR-G</td>
<td>SF-MPR-G</td>
<td>3901.1</td>
<td>3875.9</td>
<td>25.3</td>
</tr>
<tr>
<td>MF-SPR</td>
<td>MF-MPR</td>
<td>3732.5</td>
<td>3700.9</td>
<td>31.7</td>
</tr>
<tr>
<td>MF-SPR-G*</td>
<td>MF-MPR-G*</td>
<td>3697.5</td>
<td>3683.9</td>
<td>13.6</td>
</tr>
<tr>
<td>MF-SPR-G</td>
<td>MF-MPR-G</td>
<td>3697.0</td>
<td>3681.1</td>
<td>15.9</td>
</tr>
<tr>
<td>$\bar{\Delta}_{AIC} = 24.6$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Model naming scheme is the same as in Table 7.2. Also, here $\Delta_{AIC} = AIC_1 - AIC_2$. 
frailty (unobserved heterogeneity) reduces the $AIC$, it is not at the same level as accounting for additional important covariates (observed heterogeneity). For a specific example, take the SF-SPR model and add frailty (i.e., SF-SPR-G) - the $AIC$ reduces by 32.4 units. If, instead, we add all additional covariates (i.e., MF-SPR) then the $AIC$ reduces by 201 units. Of course, the fit can still be improved by adding frailty to this multi-factor model (i.e., $AIC_{MF-SPR} - AIC_{MF-SPR-G} =$ 35.5); thus, we may conclude that unexplained variation still remains. Furthermore, although unrelated to the topic of frailty, we also compared the SPR models (scale regression) to the MPR models (scale and shape regression); As in Chapter 5 we find that extending to MPR reduces the $AIC$ significantly in all cases ($\Delta AIC = 24.6$).

**Example 7.3. Log-Logistic Gamma Frailty Model**

In Examples 7.1 and 7.2 we did not show the log-likelihood / score equations required to fit the Weibull-gamma model as it is a Burr model which has been dealt with previously. In this example we consider the case where the basic hazard is that of a log-logistic model. Note that, although that the log-logistic model has a frailty interpretation already (as seen in Example 7.2), there is no reason why we cannot use it as the basic hazard model.

From Section 1.4.4 the basic hazard and integrated hazard functions are given by

$$\lambda(t) = \frac{\lambda \gamma t^{\gamma-1}}{1 + \lambda t^\gamma},$$

and

$$\Lambda(t) = \log(1 + \lambda t^\gamma),$$

respectively. Thus, assuming multiplicative gamma frailty, the marginal hazard function is

$$\lambda_m(t) = \frac{\lambda \gamma t^{\gamma-1}}{1 + \phi \log(1 + \lambda t^\gamma)}, \quad (7.31)$$

which is not a model we have considered previously. In order to fit this model
we need to maximise the marginal log-likelihood which, from (7.23), is

\[
\ell_m(\lambda, \gamma, \phi) = \sum_{i=1}^{n} \delta_i \log \lambda(t_i) - \frac{1}{\phi + \delta_i} \log [1 + \phi \Lambda(t_i)]
\]

\[
= \sum_{i=1}^{n} \delta_i [\log \lambda + \log \gamma + (\gamma - 1) \log t_i - \log(1 + \lambda t_i^\gamma)]
\]

\[- \left( \frac{1}{\phi + \delta_i} \log [1 + \phi \log(1 + \lambda t_i^\gamma)] \right). \tag{7.32}
\]

Recall from (7.24) that the score functions (for \(\lambda\) and \(\gamma\)) are simple extensions of those required for the corresponding non-frailty model. From Appendix B.2.3 we know that:

\[
\frac{\partial}{\partial \lambda} \log \lambda(t) = \frac{1}{\lambda} - \frac{t^\gamma}{1 + \lambda t^\gamma} \quad \frac{\partial}{\partial \lambda} \Lambda(t) = \frac{t^\gamma}{1 + \lambda t^\gamma}
\]

\[
\frac{\partial}{\partial \gamma} \log \lambda(t) = \frac{1}{\gamma} + \log t - \frac{\lambda t^\gamma \log t}{1 + \lambda t^\gamma} \quad \frac{\partial}{\partial \gamma} \Lambda(t) = \frac{\lambda t^\gamma \log t}{1 + \lambda t^\gamma}.
\]

Thus, from (7.24), we have

\[
\frac{\partial \ell_m}{\partial \lambda} = \sum_{i=1}^{n} \delta_i \frac{\partial}{\partial \lambda} \log \lambda(t_i) - \frac{1 + \delta_i \phi}{1 + \phi \Lambda(t_i)} \frac{\partial}{\partial \lambda} \Lambda(t_i)
\]

\[
= \sum_{i=1}^{n} \delta_i \left( \frac{1}{\lambda} - \frac{t_i^\gamma}{1 + \lambda t_i^\gamma} \right) - \frac{1 + \delta_i \phi}{1 + \phi \log(1 + \lambda t_i^\gamma)} \frac{t_i^\gamma}{1 + \lambda t_i^\gamma}, \tag{7.33}
\]

and, similarly,

\[
\frac{\partial \ell_m}{\partial \gamma} = \sum_{i=1}^{n} \delta_i \left( \frac{1}{\gamma} + \log t_i - \frac{\lambda t_i^\gamma \log t_i}{1 + \lambda t_i^\gamma} \right)
\]

\[- \frac{1 + \delta_i \phi}{1 + \phi \log(1 + \lambda t_i^\gamma)} \frac{\lambda t_i^\gamma \log t_i}{1 + \lambda t_i^\gamma}. \tag{7.34}
\]
Furthermore, from (7.25),
\[
\frac{\partial \ell_m}{\partial \phi} = \sum_{i=1}^{n} \left( \frac{1}{\phi^2} \log \left[ 1 + \phi \Lambda(t_i) \right] - \frac{1}{\phi + \delta_i} \frac{\Lambda(t_i)}{1 + \phi \Lambda(t_i)} \right)
\]
\[
= \sum_{i=1}^{n} \left( \frac{1}{\phi^2} \log \left[ 1 + \phi \log(1 + \lambda_{ti}) \right] - \frac{1}{\phi + \delta_i} \frac{\log(1 + \lambda_{ti})}{1 + \phi \log(1 + \lambda_{ti})} \right). \tag{7.35}
\]
Of course we could arrive at (7.33) - (7.35) by simply by differentiating (7.32) directly, but the purpose of this example is to show that, once we have done the work for a particular non-frailty model, the extension to frailty is straightforward.

If we wish to extend the basic hazard model to MPR (which of course includes SPR) we simply write \( \lambda_i = \exp(x_i^T \beta) \) and \( \gamma_i = \exp(z_i^T \alpha) \) (in place of \( \lambda \) and \( \gamma \)) in the above log-likelihood, (7.32), and in the score equation for \( \phi \), (7.35). For the \( \beta \) score equations note that, from the chain rule, \( \frac{\partial \ell_m}{\partial \beta_j} = \sum (\partial \ell_m/\partial \lambda_i \cdot \partial \lambda_i/\partial \beta_j) = \sum (\partial \ell_m/\partial \lambda_i \cdot x_{ji}) \) for \( j = 0, \ldots, p \), i.e., replace \( \lambda \) and \( \gamma \) in (7.33) with \( \lambda_i \) and \( \gamma_i \) and multiply the summand by \( x_{ji} \). Similarly, \( \frac{\partial \ell_m}{\partial \alpha_k} = \sum (\partial \ell_m/\partial \gamma_i \cdot \partial \gamma_i/\partial \alpha_k) = \sum (\partial \ell_m/\partial \gamma_i \cdot z_{ki}) \) for \( k = 0, \ldots, q \).

We fitted the above model to the lung cancer data (not shown) with different regression structures - null model and SPR / MPR models with various covariates - but found in all cases that \( \hat{\phi} \approx 0 \), i.e., the extra flexibility of this parameter is not required. One might assume this to be an identifiability issue but we can safely reject such concerns based on simulation work and fitting the model to other datasets (also not shown).

### 7.6 Multi-Parameter Regression with Error Terms

In Example 7.1 we saw that multiplicative frailty can be interpreted as an error term in the linear predictor for the Weibull SPR model (and all PH models). With this interpretation, the frailty component accounts for
covariates and/or measurement errors in the included covariates (Lancaster, 1990, sec. 4.2)\(^{10}\). Therefore, analogous to this Weibull SPR model, we may wish to extend the MPR models of Chapter 5 to include an error term in each regression component. Although full development of such models is beyond the scope of this thesis, we consider the Weibull MPR case here (upon which future work will be based).

In the spirit of multiplicative frailty (Section 7.4), we extend the Weibull hazard by multiplying each of its parameters (scale and shape) by a frailty term, i.e.,

\[
\lambda(t \mid u, v) = u \lambda v \gamma t^{v \gamma - 1},
\]

such that \(E(U) = E(V) = 1\). Letting scale and shape depend on covariates (as in Chapter 5), the hazard then becomes

\[
\lambda(t \mid x, z, u, v) = u \exp(x^T \beta) v \exp(z^T \alpha) t^{v \exp(z^T \alpha) - 1}
\]

\[
= \exp(x^T \beta + \epsilon) \exp(z^T \alpha + \omega) t^{\exp(z^T \alpha + \omega) - 1},
\]

where \(\epsilon = \log u\) and \(\omega = \log v\) respectively\(^{11}\). Clearly this is a Weibull model with

\[
\log \lambda = x^T \beta + \epsilon, \quad \log \gamma = z^T \alpha + \omega,
\]

which extends the definition of multi-parameter regression, (5.1), to include error terms - albeit in the Weibull case only. Although we could specify distributions for \(\epsilon\) and \(\omega\), we will work in terms of \(u\) and \(v\) in order to preserve the analogy with preceding sections. Furthermore, given the popularity of gamma frailty (Section 7.5), we assume that both \(u\) and \(v\) are gamma distributed, i.e.,

\[
g(u) = \frac{1}{\phi_u^{1/\phi_u} \Gamma(1/\phi_u)} u^{1/\phi_u - 1} \exp(-u/\phi_u),
\]

and

\[
g(v) = \frac{1}{\phi_v^{1/\phi_v} \Gamma(1/\phi_v)} v^{1/\phi_v - 1} \exp(-v/\phi_v),
\]

\(^{10}\)Of course, more generally, multiplicative frailty is viewed as representing all variation (in the hazard) which remains unexplained by the regression component, \(\lambda(t \mid x)\).

\(^{11}\)Note how the hazard in (7.27) has been generalised.
respectively, such that $\phi_u = \text{var}(U)$ and $\phi_v = \text{var}(V)$.

It is worth pointing out that the above model is defined by the following statements:

$$T \sim \text{Weibull}[\lambda = u \cdot \exp(x^T \beta), \, \gamma = v \cdot \exp(z^T \alpha)]$$

$$U \sim \text{Gamma}(\phi_u) \quad (7.39)$$

$$V \sim \text{Gamma}(\phi_v).$$

Of course, this represents but one set of possible assumptions\(^{12}\). Nonetheless, these assumptions follow naturally from the work of the preceding sections and, moreover, appear to be reasonable given the status of both the Weibull model and gamma frailty in survival literature. Indeed, the family contains, and extends, some commonly used models (i.e., null Weibull, PH Weibull, PH Weibull-gamma frailty) as well as models introduced in this thesis (i.e., MPR Weibull, MPR Weibull-gamma frailty).

We can estimate this model using the marginal likelihood approach. From (7.7), we have that

$$L_m(\theta, \phi_u, \phi_v) = \prod_{i=1}^{n} \left[ \int_{0}^{\infty} \int_{0}^{\infty} \lambda(t_i \mid u_i, v_i)^{\delta_i} S(t_i \mid u_i, v_i) g(u_i) du_i g(v_i) dv_i \right], \quad (7.40)$$

as we have two frailty components to be integrated out. From (7.36) we see that $u$ appears multiplicatively in the hazard (for this Weibull case) which allows us to write

$$\lambda(t \mid u, v) = u \lambda(t \mid u = 1, v),$$

and, hence,

$$S(t \mid u, v) = \exp[-u \Lambda(t \mid u = 1, v)].$$

\(^{12}\)We can choose different survival and frailty distributions and, furthermore, there is no reason why $g(u)$ and $g(v)$ must have the same functional form. Alternatively we could work with $\epsilon = \log u$ and $\omega = \log v$ where the (mean-zero) normal distribution may be useful. Moreover, we may assume a bivariate distribution for $(u, v)$ or $(\epsilon, \omega)$.\)
Thus, the double integral in (7.40) can be written as

\[
\int_{0}^{\infty} \lambda(t_i \mid u_i = 1, v_i) \delta_i \int_{0}^{\infty} u_i^\delta_i \exp[-u_i \Lambda(t_i \mid u_i = 1, v_i)] g(u_i) du_i \ g(v_i) dv_i,
\]

where the inner integral, \( I_{u_i} \), is exactly of the form considered in Appendix D.3.1 (i.e., multiplicative frailty). Therefore, we have that

\[
I_{u_i} = (-1)^\delta_i L_u^{(\delta_i)}[\Lambda(t_i \mid u_i = 1, v_i)]
\]

\[
= [1 + \phi_u \Lambda(t_i \mid u_i = 1, v_i)]^{-(1/\phi_u + \delta_i)},
\]

where \( L_u(s) \) is the Laplace transform of \( g(u) \) and the second line comes from the fact that \( L_u^{(\delta_i)}(s) = (-1)^\delta_i (1 + \phi_u s)^{-1/(\phi_u + \delta_i)} \) in the gamma case (see Section 7.5). Putting all of the above together, and using the fact that \( \lambda(t \mid u = 1, v) = \lambda v t^{\gamma - 1} \) and \( \Lambda(t \mid u = 1, v) = \lambda t^{\gamma} \), the marginal likelihood, (7.40), becomes

\[
L_m(\theta, \phi_u, \phi_v)
\]

\[
= \prod_{i=1}^{n} \left[ \int_{0}^{\infty} (\lambda_i v_i t_i t_i^{\gamma_i - 1})^\delta_i [1 + \phi_u \lambda_i t_i^{\gamma_i}]^{-1/(\phi_u + \delta_i)} g(v_i) dv_i \right], \tag{7.41}
\]

where \( \lambda_i = \exp(x_i^T \beta) \) and \( \gamma_i = \exp(z_i^T \alpha) \) respectively.

As the marginal likelihood is a product of \( n \) integrals which are unfortunately non-analytic, numerical integration is required to fit the model (unlike the multiplicative gamma frailty model of Section 7.5). Furthermore, the burden is quite high as these \( n \) numeric integrals are required each time \( L_m(\theta, \phi_u, \phi_v) \) is to be evaluated in the optimisation algorithm\(^\text{13}\). Nonetheless, we have fitted the model to the lung cancer data but, in the cases we have tried, we have found that \( \hat{\phi}_v \approx 0 \), i.e., the model reduces to the MPR Weibull-gamma frailty model considered in Example 7.2. Note that in some preliminary simulation work (not shown) we have found that non-zero estimates of \( \phi_v \) are produced when the true value is indeed non-zero. Thus, we

\(^\text{13}\)Alternatives to marginal likelihood, which avoid integration, exist (e.g., h-likelihood) and these are likely to be useful for fitting MPR models with error components. Some possibilities are discussed in the last paragraph of Section 7.3.
conclude that the extra flexibility of this additional parameter is simply not required in the lung cancer dataset; multiplicative frailty adequately captures the unobserved heterogeneity.

### 7.7 Structured Dispersion

We now return to the gamma frailty model of Section 7.5 which is defined by

\[ \lambda(t | u, x, \theta) = u \lambda(t | x, \theta) \quad \text{where} \quad u \sim \text{Gamma}(\phi), \]

and, in the spirit of multi-parameter regression (Chapter 5), suggest a regression structure for the dispersion parameter, \( \phi \), in addition to the hazard regression model, \( \lambda(t | x, \theta) \)\(^{14} \). Therefore, we have

\[ \lambda(t | u, x, \theta) = u \lambda(t | x, \theta) \quad \text{where} \quad u \sim \text{Gamma}[\exp(w^T \tau)], \quad (7.42) \]

i.e., \( w = (1, w_1, \ldots, w_r)^T \) and \( \tau = (\tau_0, \tau_1, \ldots, \tau_r)^T \) are the vectors of dispersion covariates and regression coefficients. Thus, rather than having a single parameter, \( \phi \), governing the overall hazard dispersion (as in the standard gamma frailty model), we now have a vector, \( \tau \), which describes how this dispersion varies with known covariates. Hence, we refer to (7.42) as a structured dispersion model which, of course, fits under the broad heading of multi-parameter regression\(^{15} \).

The gamma structured dispersion model extends the usual gamma frailty model via \( \phi = \exp(w^T \tau) \) and, therefore, the results of Section 7.5 are generalised by making this substitution. For example, the marginal log-likelihood required to fit the model is given by

\[
\ell_m(\theta, \tau) = \sum_{i=1}^{n} \delta_i \log \lambda(t_i) - \frac{1}{\phi_i + \delta_i} \log \left[ 1 + \phi_i \Lambda(t_i) \right] \\
= \sum_{i=1}^{n} \delta_i \log \lambda(t_i) - \left[ \exp(-w_i^T \tau) + \delta_i \right] \log \left[ 1 + \exp(w_i^T \tau) \Lambda(t_i) \right].
\]

\(^{14}\)We previously alluded to \( \phi \)-regression in Example 7.2 when noting that the Weibull MPR gamma frailty model is a Burr MPR model.

\(^{15}\)References to structured dispersion models in other contexts are given in Section 5.1.
As before, the effect of a covariate on the level of the hazard can be described by the individual hazard ratio (Section 7.4.1). Analogously, we may use the variance ratio (V.R.) to describe the effect of a covariate on the hazard variation, e.g., the ratio of variances for an individual with \( w_1 = 1 \) to an individual with \( w_1 = 0 \) is given by

\[
\frac{\phi(w_1 = 1, \tilde{w})}{\phi(w_1 = 0, \tilde{w})} = \frac{\exp(\tau_1 + \tilde{w}^T \tau)}{\exp(\tilde{w}^T \tau)} = \exp(\tau_1),
\]

where \( \tilde{w}^T \tau = w^T \tau - w_1 \tau_1 \) represents all other terms in the linear predictor.

**Example 7.4. Weibull-Gamma Structured Dispersion Analysis of Lung Cancer Data**

Here we extend the Weibull-gamma model of Example 7.1 to structured dispersion, i.e., we assume the following survival model:

\[
\lambda(t | u, x) = u \exp(x^T \beta) \gamma t^{\gamma-1} \quad \text{where} \quad u \sim \text{Gamma}[\exp(w^T \tau)].
\] (7.43)

A useful feature of this model is that both the hazard and variance ratios are given by exponentiation of \( \beta \) and \( \tau \) coefficients respectively; thus, the model is easily interpreted. Furthermore, the marginal model is a Burr MPR model (with \( \gamma \) constant) and, hence, the log-likelihood given in Appendix B.2.4 is used for estimation.

The model was fitted to the lung cancer data and the results are given in Table 7.4. We now summarise the main findings:

- **Treatment:** Surgery reduces the hazard to 27% of the palliative level, whereas the combined treatment (C+R) reduces the hazard to 16% of the palliative level. Furthermore, while V.R. \( \approx 5 \) for surgery (almost significant), the combined treatment has virtually no variance relative to palliative care (significant). Interestingly, this is the only analysis in this thesis where surgery is not deemed to be the most effective treatment. As for the other treatments (chemotherapy and radiotherapy), the risk is approximately 50% that of palliative care in both cases (the dispersion effects are non-significant).

- **Age group:** The effect of age is not significant apart from the fact that the 50 - 60 group has almost no variation (relative to the < 50 group).
Table 7.4. Weibull-Gamma Structured Dispersion Full Covariate Analysis

<table>
<thead>
<tr>
<th></th>
<th>Hazard: $\gamma$, $\beta$</th>
<th>Dispersion: $\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>S.E.</td>
</tr>
<tr>
<td><strong>Shape:</strong> $\gamma$</td>
<td>1.48 (0.00)</td>
<td>——</td>
</tr>
<tr>
<td><strong>Intercept</strong></td>
<td>-3.88 (0.49)</td>
<td>——</td>
</tr>
<tr>
<td><strong>Treatment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palliative</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Surgery</td>
<td>-1.30 (0.47)</td>
<td>2.74</td>
</tr>
<tr>
<td>Chemo</td>
<td>-0.64 (0.30)</td>
<td>2.15</td>
</tr>
<tr>
<td>Radio</td>
<td>-0.72 (0.16)</td>
<td>4.36</td>
</tr>
<tr>
<td>C+R</td>
<td>-1.81 (0.28)</td>
<td>6.55</td>
</tr>
<tr>
<td><strong>Age Group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$&lt; 50$</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>50 - 60</td>
<td>-0.39 (0.34)</td>
<td>1.15</td>
</tr>
<tr>
<td>60 - 70</td>
<td>-0.09 (0.33)</td>
<td>0.28</td>
</tr>
<tr>
<td>70 - 80</td>
<td>-0.06 (0.34)</td>
<td>0.19</td>
</tr>
<tr>
<td>80 +</td>
<td>0.25 (0.40)</td>
<td>0.63</td>
</tr>
<tr>
<td><strong>WHO Status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Light Work</td>
<td>0.31 (0.21)</td>
<td>1.50</td>
</tr>
<tr>
<td>No Work</td>
<td>0.98 (0.23)</td>
<td>4.28</td>
</tr>
<tr>
<td>$&gt; 50%$ Bed</td>
<td>1.77 (0.27)</td>
<td>6.66</td>
</tr>
<tr>
<td>Bedbound</td>
<td>2.11 (0.41)</td>
<td>5.08</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Male</td>
<td>0.10 (0.13)</td>
<td>0.74</td>
</tr>
<tr>
<td><strong>Smoker</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Yes</td>
<td>-0.03 (0.25)</td>
<td>0.11</td>
</tr>
<tr>
<td>Ex-smoker</td>
<td>0.03 (0.25)</td>
<td>0.11</td>
</tr>
<tr>
<td>Missing</td>
<td>0.22 (0.57)</td>
<td>0.38</td>
</tr>
<tr>
<td><strong>Cell Type</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Squamous</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Small</td>
<td>0.99 (0.23)</td>
<td>4.27</td>
</tr>
<tr>
<td>Adeno.</td>
<td>0.03 (0.21)</td>
<td>0.16</td>
</tr>
<tr>
<td>Other</td>
<td>0.37 (0.18)</td>
<td>2.04</td>
</tr>
<tr>
<td><strong>Metastases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>Yes</td>
<td>0.69 (0.16)</td>
<td>4.40</td>
</tr>
<tr>
<td>Missing</td>
<td>0.27 (0.21)</td>
<td>1.29</td>
</tr>
<tr>
<td><strong>Sodium</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\geq 136$ mmol/l</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>$&lt; 136$ mmol/l</td>
<td>0.32 (0.12)</td>
<td>2.66</td>
</tr>
<tr>
<td>Missing</td>
<td>-0.34 (0.30)</td>
<td>1.12</td>
</tr>
<tr>
<td><strong>Albumen</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\geq 35$ g/l</td>
<td>0.00</td>
<td>——</td>
</tr>
<tr>
<td>$&lt; 35$ g/l</td>
<td>0.80 (0.14)</td>
<td>5.76</td>
</tr>
<tr>
<td>Missing</td>
<td>0.19 (0.24)</td>
<td>0.78</td>
</tr>
</tbody>
</table>

$\ell(\hat{\theta}) = -1795.11 \quad AIC = 3696.23$

**Note:** H.R. = hazard ratio = $\exp(\beta)$ and V.R. = variance ratio = $\exp(\tau)$. 
7.7. STRUCTURED DISPERSION

- **WHO status**: The light work group is not significantly different to the normal group. Otherwise, the hazard increases as physical state deteriorates - dramatically so - with H.R. values of 2.67, 5.84 and 8.22 respectively. The dispersion effects are not significant.

- **Sex**: Males are not significantly different from females in either the level of the hazard or its dispersion.

- **Smoker**: The level of hazard is unaffected by smoking status. However, the hazard variation for smokers is 38% that of non-smokers.

- **Cell type**: The small cell type increases the risk of death by a factor of 2.68 compared with the squamous cell type, the adenocarcinoma cell type is not statistically significant and the “other” cell type is just significant with H.R. = 1.45. In terms of dispersion, only adenocarcinoma is significant (and only just) with approximately 30% the variability of the squamous cell group.

- **Metastases**: Individuals with metastases have 4.4 times the risk (and 56% the variability) of those without metastases.

- **Sodium**: Lower sodium levels increase the hazard by almost 40% but do not alter dispersion.

- **Albumen**: Lower albumen levels more than double the risk of death.

Comparing the above analysis to the earlier Weibull-gamma analysis of Example 7.1, it is clear that they are generally quite similar in terms the estimates/significance of covariate effects on the hazard level. However, the advantage of the current analysis is that we can also determine how covariates affect the variation of the hazard, rather than assuming that it is constant. Furthermore, structuring the dispersion leads to a reduction in $AIC$, albeit small in this particular case ($\Delta_{AIC} = 3697.02 - 3696.23 = 0.79$). However, it is clear from Table 7.4 that covariates can be dropped from the dispersion to reduce the $AIC$ further. Indeed, in the worst case, no dispersion effects are significant and the model reduces to the standard gamma frailty model. Of course the flexibility of multi-parameter regression has been discussed previously in Chapter 5.
7.7.1 Structured Dispersion with Error

Combining the ideas of structured dispersion and regression error terms (Section 7.6), we now briefly consider the following model which extends (7.43):

\[ \lambda(t \mid u, x) = u \exp(x^T \beta) t^{\gamma - 1} \]

\[ u \sim \text{Gamma}[v \exp(w^T \tau)] \]  \hspace{1cm} (7.44)

\[ v \sim \text{Gamma}(\phi_v). \]

This model has an error term, \( u \), for the hazard regression, whose variance is described by covariates, \( w = (1, w_1, \ldots, w_r)^T \), and, in turn, an error term, \( v \), for this dispersion regression.\(^{16}\)

The above model can be fitted using a marginal likelihood approach (similar to that of Section 7.6). Letting \( \lambda_i = \exp(x_i^T \beta) \) and \( \phi_{u,i} = \exp(w_i^T \tau) \), respectively, the marginal likelihood is given by

\[
L_m(\beta, \gamma, \tau, \phi_v) = \prod_{i=1}^n \left[ \int_0^\infty \int_0^\infty \lambda(t_i \mid u_i)^{\delta_i} \exp[-u_i \Lambda(t_i \mid u_i)] g(u_i \mid v_i) \ du_i g(v_i) \ dv_i \right]
\]

\[ = \prod_{i=1}^n \left[ (\lambda_i t_i^{\gamma - 1})^{\delta_i} \int_0^\infty \int_0^\infty u_i^{\delta_i} \exp(-u_i \lambda_i t_i^{\gamma}) g(u_i \mid v_i) \ du_i g(v_i) \ dv_i \right]
\]

\[ = \prod_{i=1}^n \left[ (\lambda_i t_i^{\gamma - 1})^{\delta_i} \int_0^\infty (1 + v_i \phi_{u,i} \lambda_i t_i^{\gamma})^{-(\frac{1}{\phi_{u,i}} + \delta_i)} g(v_i) \ dv_i \right], \]

where the last line comes from the Laplace transform of \( g(u_i \mid v_i) \) which is the same as the usual gamma case with \( \phi_u \) replaced by \( v_i \phi_{u,i} \).

Much like the model of Section 7.6, we have found that \( \hat{\phi}_v \approx 0 \) when we fit model (7.44) to the lung cancer data (various combinations of covariates tried)\(^{17}\). This is not to say that an error term in the dispersion is not useful in other datasets; also note that Lee & Nelder (2006) have considered error terms in dispersion in the context of generalized linear models.

\(^{16}\)We could specify a regression model for \( \phi_v \) also and, then, we then might consider including an error term in this regression. Of course we could continue infinitely in this fashion but it seems logical to assume that higher order terms are of less practical use; even the model considered here, (7.44), is surplus to requirement in the lung cancer dataset.

\(^{17}\)Recall, however, that \( \phi_v \) played a different role in Section 7.6.
Chapter 8

Discussion

In this thesis we have developed multi-parameter regression (MPR) models in the setting of survival analysis; in particular, we have mainly focused on two-component (scale and shape) regression models. Of course, it is clear that the MPR method is widely applicable and not limited to survival analysis or indeed any single area of statistics; it can be used to increase the flexibility of most standard parametric regression models. In spite of this fact, it is quite surprising to find that the MPR approach has not received a great deal of attention in the literature (see Section 5.1). This is particularly true in survival literature, although not so surprising in this case, where much emphasis is placed on non- and semi-parametric methods rather than the development of fully parametric methods.

Our development of these MPR survival models was motivated by asking the following simple question: why should the shape of the distribution be covariate independent? In other words, why assume, a priori, that individuals differ only with respect to the scale? This seems to be an unreasonable limitation as the nature of the time evolution of the hazard function must then be the same for all individuals. However, we can easily think of cases where this is unlikely to hold. For example, an aggressive treatment may cause the hazard to increase initially followed by a decrease, whereas another treatment may simply decrease the hazard (but perhaps it never decreases to level attained by the aggressive treatment). Such phenomena cannot be handled within the framework of standard single parameter regression (SPR)
models. Indeed, the inherent limitations of SPR models are clear (compare the results of Sections 4.5 and 5.2) and, therefore, it is useful to relax the assumptions of these models.

In a field of statistics where practitioners are accustomed to “letting the data speak for themselves”, it seems particularly important to develop more flexible parametric models (as an alternative to non-parametric approaches). Of course, all parametric models impose some structure on the data but, nonetheless, the simple MPR extension generates survival models which support a wider variety of phenomena (than SPR counterparts) such as divergent, convergent or crossing hazards (Section 5.2). Perhaps, in many cases, the data will have little more to say than what is supported by a flexible MPR model, i.e., the fit may be very close to that of a competing non-parametric estimator. Recall that in Example 5.1 the log-logistic MPR treatment model was virtually indistinguishable from the corresponding treatment Kaplan-Meier curves. This is quite noteworthy as the MPR model can provide more insight than Kaplan-Meier curves (through the hazard function for example).

Although we have mainly focused on regression models with two components (scale and shape), further flexibility can be achieved using more general parametric models. To this end, we have also considered the Burr MPR model which has three regression components and unifies the Weibull and log-logistic MPR models. However, the extra flexibility of this model is not required in the lung cancer dataset that we have analysed; the log-logistic model provides the best fit to these data. It is worth noting that the generalised gamma (three parameters) and the generalised F (four parameters) distributions (Kalbfleisch & Prentice, 2002, chap. 2) may also be useful candidates for multi-parameter regression but have not been considered in this thesis (future work).

Irrespective of the specific model, we have shown, through a variety of examples in Chapter 5 (and also Chapter 7), that the MPR extension can lead to significant improvements in fit over the corresponding SPR model. The price paid for this improvement is that MPR models are more difficult to interpret - we might expect this given the aim of capturing more complex structures. Of course, this is not an issue if we simply wish to make predictions based on the fitted model (e.g., predicted survival probabilities) but,
typically, we will want to interpret the individual covariate effects in addition to making predictions. However, while the scale and shape regression coefficients ($\beta$ and $\alpha$ respectively) can give us some insight into covariate effects, we have found that these coefficients are somewhat difficult to interpret directly - further complicated by the fact that scale and shape estimates are correlated (see next paragraph). Thus, we have used the hazard ratio - ubiquitous in survival analysis - to measure the effect of covariates as it takes into account both the $\beta$ and $\alpha$ values. However, apart from the Gompertz and piecewise exponential cases, the hazard ratio typically depends on the other covariates in the model, i.e., these do not cancel out neatly, and, therefore, in Section 5.3 we suggest using the overall hazard ratio and average-covariates hazard ratio respectively: the former is an average hazard ratio over all individuals in the dataset, whereas the latter is the hazard ratio for an average individual. Clearly these have different interpretations but, in our practical work, we have found that the estimates are often numerically close (although the overall hazard ratio has a larger standard error). An alternative estimate comes from using a least squares approximation to the log-hazard function (more on this technique later).

We have found that the following correlation structure exists among estimated regression coefficients in MPR models: given some covariate $c$, which is common to both the scale and shape regression components (i.e., $c \in x, z$), its estimated effects, $\hat{\beta}_c$ and $\hat{\alpha}_c$ say, will be highly correlated with each other. We first discovered this in our practical work (correlation matrices given in Examples 5.4 and 5.5) which motivated the simulation studies of Section 5.5 and Appendix B.4. These studies show that the correlation structure exists over a range of sample sizes, censoring levels, models and parameter values; it is a general feature of MPR models which arises due to a lack of orthogonality across the scale-shape space. Of course, we could abolish this by using a model where the scale and shape parameters are orthogonal but recall that the definition of orthogonality does not extend to censored data (Cox & Reid, 1987) and, indeed, we have found that such “full-information-orthogonality”, if it exists, breaks down in the presence of censoring (Section 5.5). Furthermore, although it is not something that we have pursued, extending Cox & Reid’s orthogonality equations (Appendix C.4) to generate orthogonal models
for censored data seems an undesirable task for two reasons, namely: (i) we would have to specify the censoring distribution (or process) and, therefore, orthogonality would only hold in this specific case, and (ii) these partial differential (orthogonality) equations are generally not straightforward to solve even in the usual full information case. Therefore, regarding orthogonality as a special case, we may say that the regression components in MPR models are typically non-orthogonal which leads to correlated estimates.

Naturally, it is of interest to discover if the scale or shape effects of covariates are statistically significant; this can be achieved via hypothesis testing or through variable selection. However, we must account for the correlation (described above) between estimated regression coefficients when carrying out such procedures (see Section 5.4). For example, given the covariate $c \in x$, it seems unreasonable to investigate its scale effect if $c$ is absent from the shape regression, i.e., $c \notin z$. Indeed, it may be the case that $c$ only becomes significant when present in both the scale and shape components, i.e., $c \in x, z$. Thus, we suggest that $c$ should be present in both regression components so that its scale effect, $\hat{\beta}_c$, is then adjusted for its shape effect, $\hat{\alpha}_c$, and vice versa (due to the fact that these estimates are correlated). Furthermore, we suggest testing the three hypotheses: (i) $H_0 : \beta_c = 0$, (ii) $H_0 : \alpha_c = 0$ and (iii) $H_0 : \beta_c = \alpha_c = 0$. Note that we have implemented a variable selection procedure which includes scale, shape and simultaneous selection steps (see Section 5.4 and Appendix B.5.3). Although this procedure has performed reasonably well in simulation (Section 5.5.3), it is of interest to explore the use of the lasso method (Tibshirani, 1996) within this MPR context (future work - for details see the last paragraph of Section 5.4).

When regression coefficients are highly correlated, stability of the estimation procedure may be of concern. For example, Lee & Whitmore (2006), in the case of the inverse Gaussian MPR model, observed that estimated regression coefficients are correlated and suggested that it “raises some new issues for estimation and inference”. We agree that this correlation raises new issues for inference, and have discussed our solution to this in the preceding paragraph, however, we have not come across any issues as far as estimation is concerned. We have found in all of our practical work to date that estimation of MPR models is both stable and fast using the Newton-Raphson
method with different starting values (see Appendix B.5 for our implementation which uses the \texttt{nlm} function). Furthermore, in simulation (Section 5.5.2 and Appendix B.4) we have found that the estimates are generally unbiased (except when the sample size is small with a high level of censoring).

While the primary focus of this thesis is MPR modelling, our work in this area has led us to develop other techniques that have been used to complement the main line of research, namely:

- M.I.E. simulation (Section 2.3.2).
- Simulating survival data (Chapter 3).
- Least squares approximation to implied regression models (Chapter 6).
- Frailty modelling (Chapter 7).

We now discuss each of these topics briefly.

In Section 2.3.2, we introduced the method of \textit{m.l.e. simulation} which is used to calculate the standard error and confidence intervals for functions of model parameters. The method is \textit{akin to bootstrapping} in the sense that it produces a sample of m.I.E. vectors. However, it is \textit{much less computationally expensive} as it does not require fitting of the model to replicate datasets; rather, m.I.E. vectors are \textit{simulated directly} from the multivariate normal distribution $N(\hat{\theta}, \hat{\Sigma})$. We have shown in practice (Example 2.2), and through a simulation study (Section 3.6.4), that m.I.E. simulation performs comparably to standard methods, i.e., the delta method and bootstrapping. However, m.I.E. simulation is \textit{both easy to implement and computationally inexpensive}. Hence, we favour this method and have used it throughout our work (see Chapters 4 - 6), e.g., in producing confidence intervals for hazard ratios. We note that m.I.E. simulation is particularly useful when developing novel methods (such as MPR) where the primary focus is unlikely to be the tedious calculation/programming of analytic derivatives required for the delta method - a task in which room for error can be high. However, in developing a software package for mainstream use, it might be preferable to implement the delta method as it is analytic and, therefore, instantaneous.

Clearly \textit{simulation studies} have played an important role in this thesis, e.g., in exploring the properties of MPR models and the performance of
m.l.e. simulation (as discussed previously). The method of simulation we have used was originally proposed by MacKenzie (1994) and allows the censored proportion to be controlled (see Chapter 3). We have extended this method to handle cure rate models (Section 3.5) which was motivated by an interest in such models for practical application. However, our simulation work in this area (Sections 3.6.5 and 3.6.6) has deterred us from pursuing cure rate modelling in any great detail due to inherent difficulties in estimating the cured proportion reliably (although Example 5.6 does contain a cure rate analysis). In particular, we have found that estimates of the cured proportion are subject to appreciable bias and imprecision when the censored proportion (of non-cured individuals) is high and, furthermore, it may be difficult to distinguish between decreasing hazards and the presence of cured individuals (particularly when censoring is high). Thus, it seems that one should exercise great caution in reporting such estimates in practice. Indeed, this is the general consensus in the literature and has been confirmed more formally by our simulation work.

In Chapter 6 we proposed the use of least squares to approximate a covariate-dependent model quantity, \( \psi(x) \) say, in which covariate effects cannot be interpreted straightforwardly, i.e., we assume that the implied regression model, \( \psi(x) \), has a complicated functional form. Thus, applying the linear approximation allows direct interpretation of covariate effects via \( \psi(x) \approx x^T \beta^* \). This easily applied method can be used, for example, to aid interpretation of complicated models by translating results onto some conceptually intuitive quantity. Indeed, the difficulties of interpreting MPR models (as discussed previously) motivated our development of this method. In particular, we used this approximation to interpret the log-logistic MPR model in terms of the restricted mean, \( \mu^* = E[\min(T, t^*)] \), which we found to be very intuitive (Section 6.3.1). Furthermore, in Section 6.4, we approximated the log-hazard function, i.e., \( \log \Lambda(t \mid x) \approx x^T \beta^*(t) \), so that the hazard ratio for a particular covariate is given by exponentiating the relevant \( \beta^* \) coefficient. In practice, we found that this least squares estimate of the hazard ratio was numerically close to the overall and average-covariates hazard ratios (discussed previously) providing us with some assurance that all three are quite reasonable estimates.
Frailty modelling is the study of unobserved heterogeneity in survival data. Our work in this area is largely based on the multiplicative gamma frailty model (Section 7.5) - the most commonly used frailty model. Among other things, Example 7.2 shows that both the MPR and (gamma) frailty extensions of the Weibull SPR model are simultaneously supported by the lung cancer data, i.e., the best fitting model is the Weibull MPR gamma frailty model. This is noteworthy as it is well known that frailty causes time-dependent effects (hazard ratios) at a population level even though the effects may be proportional at an individual level (Section 7.4.1). Therefore, one may question the need for the MPR extension of the Weibull SPR model (which generates a time-dependent effects model) if we have accounted for frailty. Thus, finding that the usefulness of MPR is not abolished by the presence of a frailty component - albeit in a single example - adds credence to MPR in this setting.

Apart from considering MPR in the context of the standard multiplicative gamma frailty model we have also considered extensions which merge these two concepts. In Section 7.6, motivated by the Weibull-gamma model, we propose a general class of MPR models with error terms which extends the standard MPR class of Chapter 5 via \( g(\lambda) = x^T \beta + \epsilon \) and \( h(\gamma) = z^T \alpha + \omega \) (compare with (5.1)). Currently we have only considered the Weibull MPR error model and, therefore, further work is needed in this area. In particular, the first challenge is computational as the marginal likelihood is non-analytic; h-likelihood (Lee & Nelder, 1996, 2001) may provide a solution. An alternative extension of the gamma frailty model is the structured dispersion (gamma frailty) model (Section 7.7). This involves modelling the frailty variance parameter as a function of covariates simultaneously with any hazard regression components. Of course, this is simply an application of the MPR method but, in the context of frailty, it would be of interest to practitioners to understand how the hazard variation depends on covariates.

It is clear from the above that there was a need to explore the consequences of MPR survival modelling and, to this end, we feel that this thesis represents a useful contribution to the area. While many interesting research possibilities still remain, inevitably it is impossible to address them all within the confines of a doctoral thesis. However, we have dealt the main issues of
inference and interpretation which, we feel, are fundamental in the use of such MPR models (not only in survival analysis). Furthermore, we have shown that MPR models can greatly improve on SPR counterparts in terms of their flexibility and, hence, their ability to adapt to a wider variety of data. Indeed, practitioners might find themselves relying far less heavily on non-parametric approaches if their parametric toolkit contained more useful tools. To conclude, we have certainly found that MPR models are useful for analysing survival data and suggest that the adoption of flexible parametric approaches can improve current practice.
Appendix A

Lung Cancer Data

All of the models developed in this thesis have been applied, throughout, to a particular lung cancer dataset (see examples); we describe this dataset here in this appendix. Firstly, the dataset contains information on all individuals, resident in Northern Ireland, who were diagnosed with lung cancer during the one-year period October 1st 1991 – September 30th 1992. These data, originally collected/analysed by Wilkinson (1995), and later re-analysed by MacKenzie (1996), were accumulated from a number of different sources, namely: general practitioners, hospital physicians/surgeons, radiotherapists, pathology laborites, the general register office and the Northern Ireland Cancer Registry. Only cases of primary lung cancer were included; cases where lung cancer was a secondary cancer were excluded.

In total there were 855 individuals recruited during the aforementioned period and these individuals were followed-up until the study end date, May 30th 1993. The survival time was taken to be the number of months from the date of diagnosis until the earlier of death or the study end date. Individuals who were still alive at the end of the study, who had died from other causes or who had dropped out of the study were taken to be censored; there were 182 cases where this occurred, i.e., approximately 20% of individuals had survival times which were censored.

In addition to the survival times and corresponding censoring indicators, a number of categorical covariates were also recorded. Table A.1 contains a summary of these covariates and, although this table is generally quite
Table A.1. Lung Cancer Covariates

<table>
<thead>
<tr>
<th>Variable</th>
<th>Category</th>
<th>Individuals</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment</strong></td>
<td>Palliative Care</td>
<td>441</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>Surgery</td>
<td>79</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>Chemotherapy</td>
<td>45</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>Radiotherapy</td>
<td>256</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>Chemo. + Radio.</td>
<td>34</td>
<td>0.04</td>
</tr>
<tr>
<td><strong>Age Group</strong></td>
<td>&lt; 50</td>
<td>32</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>50 - 60</td>
<td>89</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>60 - 70</td>
<td>311</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>70 - 80</td>
<td>299</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>80 +</td>
<td>124</td>
<td>0.15</td>
</tr>
<tr>
<td><strong>WHO Status</strong></td>
<td>Normal</td>
<td>78</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>Light Work Only</td>
<td>278</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>No Work</td>
<td>286</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>&gt; 50% in Bed</td>
<td>191</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>Bedbound</td>
<td>22</td>
<td>0.03</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td>Female</td>
<td>291</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>564</td>
<td>0.66</td>
</tr>
<tr>
<td><strong>Smoker</strong></td>
<td>No</td>
<td>88</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>416</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>Ex-smoker</td>
<td>330</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>Missing</td>
<td>21</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>Cell Type</strong></td>
<td>Squamous</td>
<td>247</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>Small</td>
<td>121</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>Adenocarcinoma</td>
<td>108</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>379</td>
<td>0.44</td>
</tr>
<tr>
<td><strong>Metastases</strong></td>
<td>No</td>
<td>188</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>428</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>Missing</td>
<td>239</td>
<td>0.28</td>
</tr>
<tr>
<td><strong>Sodium Level</strong></td>
<td>≥ 136 mmol/l</td>
<td>505</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>&lt; 136 mmol/l</td>
<td>310</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>Missing</td>
<td>40</td>
<td>0.05</td>
</tr>
<tr>
<td><strong>Albumen Level</strong></td>
<td>≥ 35 g/l</td>
<td>458</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>&lt; 35 g/l</td>
<td>315</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>Missing</td>
<td>82</td>
<td>0.10</td>
</tr>
</tbody>
</table>
self explanatory, we make note the following points:

- **Palliative Care** is a non-curative treatment providing pain relief only.
- **WHO Status** stands for “World Health Organization Status” and is a commonly used scale indicating an individual’s physical state.
- **Cell type** describes the form of the lung cancer.
- **Metastases** means that cancer has spread from the lung to another organ.
- **Sodium Level** is the concentration of sodium in the blood (measured in millimoles per litre).
- **Albumen Level** is the concentration of albumen - a protein made by the liver - in the blood (measured in grams per litre).
- There is missingness in the following covariates: *smoker, metastases, sodium level* and *albumen level*. In the original coding of the data, each of these covariates was given an additional category - called “Missing” - in which individuals without a record were placed. Analysis then proceeded in a routine manner, treating this category just as any other. In our analyses we have followed the same approach, that is, we have not attempted to impute these values as such procedures are beyond the scope of this thesis.
Appendix B

Multi-Parameter Regression

B.1 Introduction

Multi-parameter regression (MPR) models are developed in Chapter 5; this appendix provides supplementary material. In the models we consider there are typically two components, scale and shape\(^1\), which are modelled via

\[ g(\lambda) = x^T \beta, \quad h(\gamma) = z^T \alpha, \quad (B.1) \]

where \(x = (1, x_1, \ldots, x_p)^T\) and \(z = (1, z_1, \ldots, z_q)^T\) are covariate vectors, \(\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T\) and \(\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_q)^T\) are the corresponding regression coefficients and \(g(\cdot)\) and \(h(\cdot)\) are appropriate link functions (often the log link in cases we consider). Thus, the vector of parameters to be estimated from the data is \(\theta = (\beta, \alpha) \in \mathbb{R}^{p+q+2}\).

B.2 Likelihood and Score Functions

In this section we give the form of the likelihood function and score equations which are used in estimating the MPR models considered in Chapter 5. First we introduce some notation. The scale and shape for the \(i\)th individual, \(i = 1, \ldots, n\), will be denoted by \(\lambda_i\) and \(\gamma_i\), respectively, where the dependence

\(^1\)The Burr MPR model has a third regression component given by \(\log(\rho) = w^T \tau\) where \(w = (1, w_1, \ldots, w_r)^T\) and \(\tau = (\tau_0, \tau_1, \ldots, \tau_r)^T\).
on covariates is implicitly assumed, i.e.,
\[
\lambda_i = g^{-1}(x_i^T \beta), \quad \gamma_i = h^{-1}(z_i^T \alpha),
\]
and \( x_i = (1, x_{i1}, \ldots, x_{ip})^T \) and \( z_i = (1, z_{i1}, \ldots, z_{iq})^T \) are the covariate vectors for the \( i \)th individual. As is standard in regression literature we let \( X \) and \( Z \) represent the covariate matrices, of dimension \( n \times (p + 1) \) and \( n \times (q + 1) \) respectively, whose \( i \)th rows are the vectors \( x_i \) and \( z_i \).

It is notationally convenient to write the likelihood function in terms of \( \lambda = (\lambda_1, \ldots, \lambda_n)^T \) and \( \gamma = (\gamma_1, \ldots, \gamma_n)^T \). Of course these are not parameter vectors (otherwise there would be \( 2n \) parameters!), but rather (vector-valued) functions of \( \beta \) and \( \alpha \) where we assume \( p + q + 2 \ll 2n \). Thus, given that \( \ell(\theta) = \ell(\lambda(\beta), \gamma(\alpha)) \), the score functions (for \( \beta \) and \( \alpha \)) are obtained using the chain rule,
\[
\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \ell}{\partial \lambda_i} \frac{\partial \lambda_i}{\partial \beta_j} = \nabla_{\lambda}^T \ell \frac{\partial \lambda}{\partial \beta_j},
\]
\[
\frac{\partial \ell}{\partial \alpha_k} = \sum_{i=1}^{n} \frac{\partial \ell}{\partial \gamma_i} \frac{\partial \gamma_i}{\partial \alpha_k} = \nabla_{\gamma}^T \ell \frac{\partial \gamma}{\partial \alpha_k},
\]
for \( j = 0, \ldots, p \) and \( k = 0 \ldots, q \), where
\[
\nabla_{\lambda} = \left( \frac{\partial}{\partial \lambda_1}, \ldots, \frac{\partial}{\partial \lambda_n} \right)^T, \quad \frac{\partial \lambda}{\partial \beta_j} = \left( \frac{\partial \lambda_1}{\partial \beta_j}, \ldots, \frac{\partial \lambda_n}{\partial \beta_j} \right)^T,
\]
\[
\nabla_{\gamma} = \left( \frac{\partial}{\partial \gamma_1}, \ldots, \frac{\partial}{\partial \gamma_n} \right)^T, \quad \frac{\partial \gamma}{\partial \alpha_k} = \left( \frac{\partial \gamma_1}{\partial \alpha_k}, \ldots, \frac{\partial \gamma_n}{\partial \alpha_k} \right)^T.
\]

We will also require the use of Hadamard multiplication which we denote by the “\( \odot \)” operator. Hadamard multiplication is element-wise multiplication of matrices or vectors, e.g., \( (a_1, a_2, a_3) \odot (b_1, b_2, b_3) = (a_1b_1, a_2b_2, a_3b_3) \). Finally, we let \( a_{r \times c} \) represent a matrix of dimension \( r \times c \) whose elements are all equal to the constant \( a \). In particular \( 0_{r \times 1} \) is a column of \( r \) zeros and \( 1_{r \times 1} \) is a column of \( r \) ones.
B.2. LIKELIHOOD AND SCORE FUNCTIONS

B.2.1 Weibull

In the Weibull MPR model (Section 5.2.1) the hazard and survivor functions for the $i$th individual ($i = 1, \ldots, n$) are given by

$$
\lambda(t \mid x_i, z_i) = \lambda_i \gamma_i t_{\gamma_i}^{\gamma_i-1} \quad S(t \mid x_i, z_i) = \exp(-\lambda_i t_{\gamma_i}),
$$

where $\lambda_i = \exp(x_i^T \beta)$ and $\gamma_i = \exp(z_i^T \alpha)$. Thus, the log-likelihood function is given by

$$
\ell(\theta) = \sum_{i=1}^{n} \delta_i \left[ \log \lambda_i + \log \gamma_i + (\gamma_i - 1) \log t_i \right] - \lambda_i t_i^{\gamma_i},
$$

(B.2)

where $\theta = (\beta, \alpha) \in \mathbb{R}^{p+q+2}$. Using the derivatives

$$
\frac{\partial \ell}{\partial \lambda_i} = \frac{\delta_i}{\lambda_i} - t_i^{\gamma_i}, \quad \frac{\partial \lambda_i}{\partial \beta_j} = x_{ji} \lambda_i
$$

$$
\frac{\partial \ell}{\partial \gamma_i} = \delta_i \left( \frac{1}{\gamma_i} + \log t_i \right) - \lambda_i t_i^{\gamma_i} \log t_i, \quad \frac{\partial \gamma_i}{\partial \alpha_k} = z_{ki} \gamma_i,
$$

and the chain rule, the score functions are given by

$$
\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^{n} (\delta_i - \lambda_i t_i^{\gamma_i}) x_{ji}
$$

(B.3)

$$
\frac{\partial \ell}{\partial \alpha_k} = \sum_{i=1}^{n} [\delta_i (1 + \gamma_i \log t_i) - \lambda_i \gamma_i t_i^{\gamma_i} \log t_i] z_{ki},
$$

(B.4)

for $j = 0, \ldots, p$ and $k = 0, \ldots, q$, which must be set equal to zero and solved (numerically) in order to find the maximum likelihood estimates. These $m = p + q + 2$ score equations can be written compactly as

$$
\begin{pmatrix}
X^T [\lambda \odot \nabla_\lambda \ell(\theta)] \\
Z^T [\gamma \odot \nabla_\gamma \ell(\theta)]
\end{pmatrix} = 0_{m \times 1},
$$

and are also used to fit the Weibull SPR model (Section 4.5.1), by setting $Z = 1_{n \times 1}$, and the basic Weibull model (Section 1.4.2), by setting $X = Z = 1_{n \times 1}$. 

B.2.2 Gompertz

In the Gompertz MPR model (Section 5.2.2) the hazard and survivor functions for the $i$th individual ($i = 1, \ldots, n$) are given by

$$\lambda(t \mid x_i, z_i) = \lambda_i \exp(\gamma_i t) \quad S(t \mid x_i, z_i) = \exp\left\{ -\frac{\lambda_i}{\gamma_i} \left[ \exp(\gamma_i t) - 1 \right] \right\},$$

where $\lambda_i = \exp(x_i^T \beta)$ and $\gamma_i = z_i^T \alpha$. Thus, the log-likelihood function is given by

$$\ell(\theta) = \sum_{i=1}^{n} \delta_i \left( \log \lambda_i + \gamma_i t_i - \frac{\lambda_i}{\gamma_i} \left[ \exp(\gamma_i t_i) - 1 \right] \right), \quad \text{(B.5)}$$

where $\theta = (\beta, \alpha) \in \mathbb{R}^{p+q+2}$. Using the derivatives

$$\frac{\partial \ell}{\partial \lambda_i} = \frac{\delta_i}{\lambda_i} - \frac{1}{\gamma_i} \left[ \exp(\gamma_i t_i) - 1 \right] \quad \frac{\partial \lambda_i}{\partial \beta_j} = x_{ji} \lambda_i$$

$$\frac{\partial \ell}{\partial \gamma_i} = \delta_i t_i - \frac{\lambda_i}{\gamma_i^2} \left[ (\gamma_i t_i - 1) \exp(\gamma_i t_i) + 1 \right] \quad \frac{\partial \gamma_i}{\partial \alpha_k} = z_{ki},$$

and the chain rule, the score functions are given by

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^{n} \left\{ \delta_i - \frac{\lambda_i}{\gamma_i} \left[ \exp(\gamma_i t_i) - 1 \right] \right\} x_{ji} \quad \text{(B.6)}$$

$$\frac{\partial \ell}{\partial \alpha_k} = \sum_{i=1}^{n} \left\{ \delta_i t_i - \frac{\lambda_i}{\gamma_i^2} (\gamma_i t_i - 1) \exp(\gamma_i t_i) + 1 \right\} z_{ki}, \quad \text{(B.7)}$$

for $j = 0, \ldots, p$ and $k = 0, \ldots, q$, which must be set equal to zero and solved (numerically) in order to find the maximum likelihood estimates. These $m = p + q + 2$ score equations can be written compactly as

$$\begin{pmatrix} X^T \left[ \lambda \odot \nabla_\lambda \ell(\theta) \right] \\ Z^T \nabla_\gamma \ell(\theta) \end{pmatrix} = 0_{m \times 1},$$

and are also used to fit the Gompertz SPR model (Section 4.5.2), by setting $Z = 1_{1 \times 1}$, and the basic Gompertz model (Section 1.4.3), by setting $X = Z = 1_{1 \times 1}$.
B.2.3 Log-Logistic

In the log-logistic MPR model (Section 5.2.3) the hazard and survivor functions for the $i$th individual ($i = 1, \ldots, n$) are given by

$$\lambda(t \mid x_i, z_i) = \frac{\lambda_i \gamma_i t_i^{\gamma_i - 1}}{1 + \lambda_i t_i^{\gamma_i}}$$
$$S(t \mid x_i, z_i) = \frac{1}{1 + \lambda_i t_i^{\gamma_i}},$$

where $\lambda_i = \exp(x_i^T \beta)$ and $\gamma_i = \exp(z_i^T \alpha)$. Thus, the log-likelihood function is given by

$$\ell(\theta) = \sum_{i=1}^{n} \delta_i \left[ \log \lambda_i + \log \gamma_i + (\gamma_i - 1) \log t_i \right] - (\delta_i + 1) \log(1 + \lambda_i t_i^{\gamma_i}),$$

(B.8)

where $\theta = (\beta, \alpha) \in \mathbb{R}^{p+q+2}$. Using the derivatives

$$\frac{\partial \ell}{\partial \lambda_i} = \frac{\delta_i}{\lambda_i} - (\delta_i + 1) \frac{\lambda_i t_i^{\gamma_i}}{1 + \lambda_i t_i^{\gamma_i}} = x_i \lambda_i$$
$$\frac{\partial \ell}{\partial \beta_j} = z_i \lambda_i$$

$$\frac{\partial \ell}{\partial \gamma_i} = \delta_i \left( \frac{1}{\gamma_i} + \log t_i \right) - (\delta_i + 1) \frac{\lambda_i t_i^{\gamma_i} \log t_i}{1 + \lambda_i t_i^{\gamma_i}} = z_i \gamma_i,$$

and the chain rule, the score functions are given by

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^{n} \left[ \delta_i - (\delta_i + 1) \frac{\lambda_i t_i^{\gamma_i}}{1 + \lambda_i t_i^{\gamma_i}} \right] x_{ji}$$

(B.9)

$$\frac{\partial \ell}{\partial \alpha_k} = \sum_{i=1}^{n} \left[ \delta_i (1 + \gamma_i \log t_i) - (\delta_i + 1) \frac{\lambda_i \gamma_i t_i^{\gamma_i} \log t_i}{1 + \lambda_i t_i^{\gamma_i}} \right] z_{ki},$$

(B.10)

for $j = 0, \ldots, p$ and $k = 0, \ldots, q$, which must be set equal to zero and solved (numerically) in order to find the maximum likelihood estimates. These $m = p + q + 2$ score equations can be written compactly as

$$\begin{pmatrix} X^T [\lambda \odot \nabla \lambda \ell(\theta)] \\ Z^T [\gamma \odot \nabla \gamma \ell(\theta)] \end{pmatrix} = 0_{m \times 1},$$

and are also used to fit the log-logistic SPR model (Section 4.5.3), by setting $Z = 1_{n \times 1}$, and the basic log-logistic model (Section 1.4.4), by setting $X = Z = 1_{n \times 1}$. 
B.2.4 Burr

In the Burr MPR model (Section 5.2.4) the hazard and survivor functions for the $i$th individual ($i = 1, \ldots, n$) are given by

$$
\lambda(t \mid x_i, z_i, w_i) = \frac{\lambda_i \gamma_i t_i^{\gamma_i - 1}}{1 + \lambda_i \rho_i t_i^{\gamma_i}},
$$

$$
S(t \mid x_i, z_i, w_i) = (1 + \lambda_i \rho_i t_i^{\gamma_i})^{-1/\rho_i},
$$

where $\lambda_i = \exp(x_i^T \beta)$, $\gamma_i = \exp(z_i^T \alpha)$ and $\rho_i = \exp(w_i^T \tau)$. Thus, the log-likelihood function is given by

$$
\ell(\theta) = \sum_{i=1}^n \delta_i \left[ \log \lambda_i + \log \gamma_i + (\gamma_i - 1) \log t_i \right] - \left( \delta_i + \frac{1}{\rho_i} \right) \log(1 + \lambda_i \rho_i t_i^{\gamma_i}),
$$

(B.11)

where $\theta = (\beta, \alpha, \tau) \in \mathbb{R}^{p+q+r+3}$. Using the derivatives

$$
\frac{\partial \ell}{\partial \lambda_i} = \frac{\delta_i - \left( \delta_i + \frac{1}{\rho_i} \right) \frac{\lambda_i \rho_i t_i^{\gamma_i}}{1 + \lambda_i \rho_i t_i^{\gamma_i}}}{x_{ji} \lambda_i}
$$

$$
\frac{\partial \ell}{\partial \beta_j} = x_{ji} \lambda_i
$$

$$
\frac{\partial \ell}{\partial \gamma_i} = \frac{\delta_i \left( \frac{1}{\gamma_i} + \log t_i \right) - \left( \delta_i + \frac{1}{\rho_i} \right) \frac{\lambda_i \rho_i t_i^{\gamma_i} \log t_i}{1 + \lambda_i \rho_i t_i^{\gamma_i}}}{z_{ki} \gamma_i}
$$

$$
\frac{\partial \ell}{\partial \alpha_k} = \frac{\delta_i \left( \frac{1}{\gamma_i} + \log t_i \right) - \left( \delta_i + \frac{1}{\rho_i} \right) \frac{\lambda_i \rho_i t_i^{\gamma_i} \log t_i}{1 + \lambda_i \rho_i t_i^{\gamma_i}}}{w_{ki} \rho_i},
$$

and the chain rule, the score functions are given by

$$
\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^n \left[ \delta_i - \left( \delta_i + \frac{1}{\rho_i} \right) \frac{\lambda_i \rho_i t_i^{\gamma_i}}{1 + \lambda_i \rho_i t_i^{\gamma_i}} \right] x_{ji}
$$

(B.12)

$$
\frac{\partial \ell}{\partial \alpha_k} = \sum_{i=1}^n \left[ \delta_i \left( 1 + \gamma_i \log t_i \right) - \left( \delta_i + \frac{1}{\rho_i} \right) \frac{\lambda_i \gamma_i \rho_i t_i^{\gamma_i} \log t_i}{1 + \lambda_i \rho_i t_i^{\gamma_i}} \right] z_{ki}
$$

(B.13)

$$
\frac{\partial \ell}{\partial \tau_l} = \sum_{i=1}^n \left[ \frac{1}{\rho_i} \log(1 + \lambda_i \rho_i t_i^{\gamma_i}) - \left( \delta_i + \frac{1}{\rho_i} \right) \frac{\lambda_i \rho_i t_i^{\gamma_i}}{1 + \lambda_i \rho_i t_i^{\gamma_i}} \right] w_{li},
$$

(B.14)
B.2. LIKELIHOOD AND SCORE FUNCTIONS

for \( j = 0, \ldots, p, k = 0, \ldots, q \) and \( l = 0, \ldots, r \), which must be set equal to zero and solved (numerically) in order to find the maximum likelihood estimates. These \( m = p + q + r + 3 \) score equations can be written compactly as

\[
\begin{pmatrix}
X^T [\lambda \odot \nabla_\lambda \ell(\theta)] \\
Z^T [\gamma \odot \nabla_\gamma \ell(\theta)] \\
W^T [\rho \odot \nabla_\rho \ell(\theta)]
\end{pmatrix} = 0_{m \times 1},
\]

and are also used to fit the Burr SPR model (Section 4.5.4), by setting \( Z = W = 1_{n \times 1} \), and the basic Burr model (Section 1.4.5), by setting \( X = Z = W = 1_{n \times 1} \).

B.2.5 Time-Dependent Logistic

In the time-dependent logistic MPR model (Section 5.2.5) the hazard and survivor functions for the \( i \)th individual \((i = 1, \ldots, n)\) are given by

\[
\lambda(t \mid x_i, z_i) = \frac{\exp(\gamma_i t + \lambda_i)}{1 + \exp(\gamma_i t + \lambda_i)} \quad S(t \mid x_i, z_i) = \left[ \frac{1 + \exp(\gamma_i t + \lambda_i)}{1 + \exp(\lambda_i)} \right]^{-1/\gamma_i},
\]

where \( \lambda_i = x_i^T \beta \) and \( \gamma_i = z_i^T \alpha \). Thus, the log-likelihood function is given by

\[
\ell(\theta) = \sum_{i=1}^{n} \delta_i (\gamma_i t_i + \lambda_i) - \left( \delta_i + \frac{1}{\gamma_i} \right) \log[1 + \exp(\gamma_i t_i + \lambda_i)] \\
+ \frac{1}{\gamma_i} \log[1 + \exp(\lambda_i)], \quad (B.15)
\]

where \( \theta = (\beta, \alpha) \in \mathbb{R}^{p+q+2} \). Using the derivatives

\[
\frac{\partial \ell}{\partial \lambda_i} = \delta_i - \left( \delta_i + \frac{1}{\gamma_i} \right) \frac{\exp(\gamma_i t_i + \lambda_i)}{1 + \exp(\gamma_i t_i + \lambda_i)} + \frac{1}{\gamma_i} \frac{\exp(\lambda_i)}{1 + \exp(\lambda_i)}
\]

\[
\frac{\partial \ell}{\partial \gamma_i} = \delta_i t_i - \left( \delta_i + \frac{1}{\gamma_i} \right) \frac{t_i \exp(\gamma_i t_i + \lambda_i)}{1 + \exp(\gamma_i t_i + \lambda_i)} + \frac{1}{\gamma_i^2} \log \left[ \frac{1 + \exp(\gamma_i t_i + \lambda_i)}{1 + \exp(\lambda_i)} \right],
\]
(and $\partial \lambda_i / \partial \beta_j = x_{ji}$, $\partial \gamma_i / \partial \alpha_k = z_{ki}$) and the chain rule, the score functions are given by

$$\frac{\partial \ell}{\partial \beta_j} = n \sum_{i=1}^{n} \left[ \delta_i \left( \delta_i + 1 \right) \frac{\exp(\gamma_i t_i + \lambda_i)}{1 + \exp(\gamma_i t_i + \lambda_i)} + \frac{1}{\gamma_i} \frac{\exp(\lambda_i)}{1 + \exp(\lambda_i)} \right] x_{ji}, \quad (B.16)$$

$$\frac{\partial \ell}{\partial \alpha_k} = n \sum_{i=1}^{n} \left\{ \delta_i t_i \left( \delta_i + 1 \right) \frac{t_i \exp(\gamma_i t_i + \lambda_i)}{1 + \exp(\gamma_i t_i + \lambda_i)} + \frac{1}{\gamma_i^2} \log \left[ 1 + \exp(\gamma_i t_i + \lambda_i) \right] \right\} z_{ki}, \quad (B.17)$$

for $j = 0, \ldots, p$ and $k = 0, \ldots, q$, which must be set equal to zero and solved (numerically) in order to find the maximum likelihood estimates. These $m = p + q + 2$ score equations can be written compactly as

$$\begin{pmatrix} X^T \nabla_\lambda \ell(\theta) \\ Z^T \nabla_\gamma \ell(\theta) \end{pmatrix} = 0_{m \times 1},$$

and are also used to fit the time-dependent logistic SPR model (Section 4.5.5), by setting $Z = 1_{n \times 1}$, and the basic time-dependent logistic model (Section 1.4.6), by setting $X = Z = 1_{n \times 1}$.

### B.2.6 Piecewise Exponential

In the piecewise exponential MPR model (Section 5.2.6) the hazard function for the $i$th individual ($i = 1, \ldots, n$) is given by

$$\lambda(t \mid x_i) = a(t)^T (\lambda_{i1}, \ldots, \lambda_{mi})^T,$$

where $\lambda_{ji} = \exp(x_{ji}^T \beta_j)$ (for $j = 1, \ldots, m$), $a(t)$ is an $m$-dimensional vector indicating which interval $t$ lies in (e.g., if $m = 4$ and $t \in I_3$ then $a(t) = (0, 0, 1, 0)^T$), $x = (1, x_{1i}, \ldots, x_{pi})^T$ is the vector of covariates for the $i$th individual and $\beta_j = (\beta_{0j}, \beta_{1j}, \ldots, \beta_{pj})^T$ is the corresponding vector of regression coefficients in the $j$th time interval. The survivor function is given by

$$S(t \mid x_i) = \exp \left[ -d(t)^T (\lambda_{i1}, \ldots, \lambda_{mi})^T \right],$$
where \(d(t)\) represents the time spent in each interval for a particular value of \(t\), e.g., if \(m = 4\) and \(t \in I_3\) then \(d(t) = (t(1) - t(0), t(2) - t(1), t - t(2), 0)^T\).

Thus, the log-likelihood function is given by

\[
\ell(\theta) = \sum_{i=1}^{n} \delta_i \log[a(t_i)^T(\lambda_{1i}, \ldots, \lambda_{mi})^T] - d(t_i)^T(\lambda_{1i}, \ldots, \lambda_{mi})^T
\]

where \(\theta = (\beta_1^T, \ldots, \beta_m^T)^T \in \mathbb{R}^{m(p+1)}\), \(a_{ji}\) is the \(j\)th element of \(a(t_i) = (a_1(t_i), \ldots, a_m(t_i))^T\) and \(d_{ji}\) is the \(j\)th element of \(d(t_i) = (d_1(t_i), \ldots, d_m(t_i))^T\). Furthermore, as

\[
a_{ji} = \begin{cases} 
1 & \text{if } t_i \in I_j, \\
0 & \text{otherwise},
\end{cases}
\]

the term \(\log(a_{1i}\lambda_{1i} + \ldots + a_{mi}\lambda_{mi}) = \log(\lambda_{ji})\) for \(t_i \in I_j\). Thus

\[
\log(a_{1i}\lambda_{1i} + \ldots + a_{mi}\lambda_{mi}) = a_{1i} \log \lambda_{1i} + \ldots + a_{mi} \log \lambda_{mi}.
\]

The log-likelihood therefore becomes

\[
\ell(\theta) = \sum_{i=1}^{n} \delta_i (a_{1i} \log \lambda_{1i} + \ldots + a_{mi} \log \lambda_{mi}) - \sum_{i=1}^{n} d_{ji} \lambda_{ji} \lambda_{ji}.
\]

Using the derivatives

\[
\frac{\partial \ell}{\partial \lambda_{ji}} = \delta_i a_{ji} - d_{ji} \lambda_{ji}, \quad \frac{\partial \lambda_{ji}}{\partial \beta_{kj}} = x_{ki} \lambda_{ji},
\]

and the chain rule, the score functions are given by

\[
\frac{\partial \ell}{\partial \beta_{kj}} = \sum_{i=1}^{n} (\delta_i a_{ji} - d_{ji} \lambda_{ji}) x_{ki},
\]
for $k = 0, \ldots, p$ and $j = 1, \ldots, m$, which must be set equal to zero and solved (numerically) in order to find the maximum likelihood estimates. These $m(p+1)$ score equations can be written compactly as

$$
\begin{bmatrix}
X^T [\lambda_1 \odot \nabla \lambda_1 \ell(\theta)] \\
\vdots \\
X^T [\lambda_m \odot \nabla \lambda_m \ell(\theta)]
\end{bmatrix} = 0_{m(p+1) \times 1},
$$

where $\nabla \lambda_j = (\partial/\partial \lambda_{j1}, \ldots, \partial/\partial \lambda_{jn})^T$. These score equations are also used to fit the PH piecewise exponential model (Section 4.5.6), by imposing the constraint $\beta_{k1} = \cdots = \beta_{km}$ (for $k = 0, \ldots, p$), and the basic piecewise exponential model (Section 1.4.7), by setting $\beta_{k1} = \cdots = \beta_{km} = 0$.

Equation (B.19) deserves some further discussion. If $x_k$ is categorical (or binary), only individuals with $x_{ki} = 1$ contribute to the sum, i.e., those who are members of that group. Thus, the score equation for $\beta_{kj}$ becomes

$$
\sum_{i \mid x_{ki} = 1} \delta_i a_{ji} - d_{ji} \lambda_{ji} = 0
$$

$$
\sum_{i \mid x_{ki} = 1} d_{ji} \lambda_{ji} = \sum_{i \mid x_{ki} = 1} \delta_i a_{ji}.
$$

Furthermore, we can write

$$
\lambda_{ji} = \exp(\beta_{0j} + x_{1i}\beta_{1j} + \ldots + x_{k-1i}\beta_{k-1j} + x_{ki}\beta_{kj} + x_{k+1i}\beta_{k+1j} + \ldots + x_{mi}\beta_{kj})
$$

$$
= \exp(x_{ki}\beta_{kj}) \exp(\beta_{0j} + x_{1i}\beta_{1j} + \ldots + x_{k-1i}\beta_{k-1j} + x_{k+1i}\beta_{k+1j} + \ldots + x_{mi}\beta_{kj})
$$

$$
= \exp(x_{ki}\beta_{kj}) \exp(\tilde{x}_i^T \beta_j),
$$

so that

$$
\sum_{i \mid x_{ki} = 1} d_{ji} \exp(x_{ki}\beta_{kj}) \exp(\tilde{x}_i^T \beta_j) = \sum_{i \mid x_{ki} = 1} \delta_i a_{ji}
$$

$$
\sum_{i \mid x_{ki} = 1} d_{ji} \exp(1 \cdot \beta_{kj}) \exp(\tilde{x}_i^T \beta_j) = \sum_{i \mid x_{ki} = 1} \delta_i a_{ji}
$$

$$
\exp(\beta_{kj}) \sum_{i \mid x_{ki} = 1} d_{ji} \exp(\tilde{x}_i^T \beta_j) = \sum_{i \mid x_{ki} = 1} \delta_i a_{ji}.
$$
B.3. PROPERTIES OF THE HAZARD RATIO

Therefore,

\[
\exp(\hat{\beta}_{kj}) = \frac{\sum_{i|x_{ki}=1} \delta_i a_{ji}}{\sum_{i|x_{ki}=1} d_{ji} \exp(\hat{x}_i^T \hat{\beta}_j)}
\]

\[
\Rightarrow \hat{\beta}_{kj} = \log \left[ \frac{\sum_{i|x_{ki}=1} \delta_i a_{ji}}{\sum_{i|x_{ki}=1} d_{ji} \exp(\hat{x}_i^T \hat{\beta}_j)} \right].
\]

The numerator of the above fraction equals the number of individuals in the group \(x_{ki} = 1\) who experience an event \((\delta_i = 1)\) in the \(j\)th time interval \((a_{ji} = 1)\). Thus, if no individual is a member of this particular combination, then \(\hat{\beta}_{kj} = \log(0) = -\infty\). Karrison (1987) discussed this in a simpler version of the model considered here. Furthermore, if all individuals in the group \(x_{ki} = 1\) experience an event before the \(j\)th time interval then all \(d_{ji} = 0\) (as nobody has passed through this time interval) and so \(\hat{\beta}_{kj} = \log(0/0)\) which is not defined. In fact, in this latter case, it is clear that any value of \(\hat{\beta}_{kj}\) will satisfy (B.19). It is clear that as the number of time intervals, \(m\), is increased, so too is the frequency of such cases. Thus, it is undesirable for \(m\) to be too large or we will have many inestimable parameters.

### B.3 Properties of the Hazard Ratio

In Section 5.2 we explored the form of the hazard ratio, \(\psi(t)\), for each of the MPR models, by examining its time-derivative and limits (at \(t = 0\) and \(t = \infty\)). In this section we produce the analytic work underpinning those results.

First we reintroduce the notation used in Section 5.2. Let \(c\) be a binary covariate whose effect we wish to determine. We will assume that \(c\) is common to both the scale and shape regressions and, furthermore, that it is the first covariate in the vectors \(x\) and \(z\), i.e., \(x_1 = z_1 = c\). Thus we have

\[
x^T \beta = \beta_0 + c\beta_1 + \ldots + x_p\beta_p \quad z^T \alpha = \alpha_0 + c\alpha_1 + \ldots + z_q\alpha_q
\]

\[
= c\beta_1 + \beta_0 + \ldots + x_p\beta_p \quad = c\alpha_1 + \alpha_0 + \ldots + z_q\alpha_q
\]

\[
= c\beta_1 + \hat{x}^T \beta, \quad = c\alpha_1 + \hat{z}^T \alpha,
\]
where $\tilde{x} = (1, x_2, \ldots, x_p)^T$ and $\tilde{z} = (1, z_2, \ldots, z_q)^T$. The hazard ratio, at time $t$, for the covariate $c = x_1 = z_1$ is then given by

$$\psi(t) = \frac{\lambda(t \mid c = 1)}{\lambda(t \mid c = 0)},$$

which we will derive for each model. In addition to this we will derive $\psi'(t) = d\psi(t)/dt$, $\psi(0) = \lim_{t \to 0} \psi(t)$ and $\psi(\infty) = \lim_{t \to \infty} \psi(t)$.

### B.3.1 Weibull

The hazard function for the Weibull MPR model (Section 5.2.1) is given by

$$\lambda(t \mid x, z) = \exp(x^T \beta + z^T \alpha) t^{\exp(z^T \alpha) - 1} = \exp(c^T \beta + \tilde{x}^T \beta + \alpha + \tilde{z}^T \alpha) t^{\exp(c^T \tilde{z} \alpha) - 1}.$$ 

Thus, the hazard ratio for $c$, at time $t$, is given by

$$\psi(t) = \frac{\lambda(t \mid c = 1)}{\lambda(t \mid c = 0)}$$

$$= \frac{\exp(\beta_1 + \tilde{x}^T \beta + \alpha_1 + \tilde{z}^T \alpha) t^{\exp(\alpha_1 + \tilde{z}^T \alpha) - 1}}{\exp(\tilde{z}^T \alpha) t^{\exp(\alpha_1 + \tilde{z}^T \alpha) - 1}}$$

$$= \exp(\beta_1 + \alpha_1) t^{\exp(\alpha_1 + \tilde{z}^T \alpha) - \exp(\tilde{z}^T \alpha)}$$

$$= \exp(\beta_1 + \alpha_1) t^{\exp(\tilde{z}^T \alpha)} \left[ \exp(\alpha_1) - 1 \right], \quad (B.20)$$

and differentiating with respect to $t$ gives

$$\frac{d\psi(t)}{dt} = \exp(\tilde{z}^T \alpha) \left[ \exp(\alpha_1) - 1 \right] \exp(\beta_1 + \alpha_1) t^{\exp(\alpha_1 + \tilde{z}^T \alpha) - \exp(\tilde{z}^T \alpha) - 1}$$

$$= \exp(\tilde{z}^T \alpha) \left[ \exp(\alpha_1) - 1 \right] \frac{\psi(t)}{t}. \quad (B.21)$$

Clearly $\text{sgn}[\psi'(t)] = \text{sgn}(\alpha_1)$. Furthermore, it is easy to see that

$$\psi(0) = \begin{cases} 0 & \text{if } \alpha_1 > 0, \\ \exp(\beta_1) & \text{if } \alpha_1 = 0, \\ \infty & \text{if } \alpha_1 < 0, \end{cases} \quad \psi(\infty) = \begin{cases} \infty & \text{if } \alpha_1 > 0, \\ \exp(\beta_1) & \text{if } \alpha_1 = 0, \\ 0 & \text{if } \alpha_1 < 0. \end{cases} \quad (B.22)$$
B.3. PROPERTIES OF THE HAZARD RATIO

B.3.2 Gompertz

The hazard function for the Gompertz MPR model (Section 5.2.2) is given by

\[
\lambda(t \mid x, z) = \exp(x^T \beta + z^T \alpha t) = \exp(c\beta_1 + \tilde{x}^T \beta + (c\alpha_1 + \tilde{z}^T \alpha)t).
\]

Thus, the hazard ratio for \(c\), at time \(t\), is given by

\[
\psi(t) = \frac{\lambda(t \mid c = 1)}{\lambda(t \mid c = 0)} = \frac{\exp(c\beta_1 + \tilde{x}^T \beta + (c\alpha_1 + \tilde{z}^T \alpha)t)}{\exp(c\beta_1 + \tilde{x}^T \beta + \tilde{z}^T \alpha t)} = \exp(\beta_1 + \alpha_1 t),
\]

and differentiating with respect to \(t\) gives

\[
\frac{d\psi(t)}{dt} = \alpha_1 \exp(\beta_1 + \alpha_1 t) = \alpha_1 \psi(t).
\]

Clearly \(\text{sgn}[\psi'(t)] = \text{sgn}(\alpha_1)\). Furthermore, it is easy to see that

\[
\psi(0) = \exp(\beta_1) \quad \forall \alpha_1, \quad \psi(\infty) = \begin{cases} \infty & \text{if } \alpha_1 > 0, \\
\exp(\beta_1) & \text{if } \alpha_1 = 0, \\
0 & \text{if } \alpha_1 < 0. \end{cases}
\]

B.3.3 Log-Logistic

The hazard function for the log-logistic MPR model (Section 5.2.3) is given by

\[
\lambda(t \mid x, z) = \frac{\exp(x^T \beta + z^T \alpha t)^{\exp(z^T \alpha) - 1}}{1 + \exp(x^T \beta)^{\exp(z^T \alpha)}} = \frac{\exp(c\beta_1 + \tilde{x}^T \beta + c\alpha_1 + \tilde{z}^T \alpha t)^{\exp(c\alpha_1 + \tilde{z}^T \alpha) - 1}}{1 + \exp(c\beta_1 + \tilde{x}^T \beta t)^{\exp(c\alpha_1 + \tilde{z}^T \alpha) - 1}}.
\]
Thus, the hazard ratio for $c$, at time $t$, is given by

$$\psi(t) = \frac{\lambda(t \mid c = 1)}{\lambda(t \mid c = 0)}$$

$$= \frac{\exp(\beta_1 + \tilde{x}^T \beta + \alpha_1 + \tilde{z}^T \alpha) t^{\exp(\alpha_1 + \tilde{z}^T \alpha) - 1}}{\exp(\tilde{x}^T \beta + \tilde{z}^T \alpha) t^{\exp(\tilde{z}^T \alpha) - 1} \left[1 + \exp(\tilde{x}^T \beta) t^{\exp(\tilde{z}^T \alpha)}\right]}$$

$$= \exp(\beta_1 + \alpha_1) t^{\exp(\tilde{z}^T \alpha) [\exp(\alpha_1) - 1]} \frac{1 + \exp(\tilde{x}^T \beta) t^{\exp(\tilde{z}^T \alpha)}}{1 + \exp(\beta_1 + \tilde{x}^T \beta) t^{\exp(\alpha_1 + \tilde{z}^T \alpha)}}.$$  \hspace{1cm} (B.26)

and therefore

$$\log \psi(t) = \beta_1 + \alpha_1 + \exp(\tilde{z}^T \alpha) [\exp(\alpha_1) - 1] \log t + \log[1 + \exp(\tilde{x}^T \beta) t^{\exp(\tilde{z}^T \alpha)}]$$

$$- \log[1 + \exp(\beta_1 + \tilde{x}^T \beta) t^{\exp(\alpha_1 + \tilde{z}^T \alpha)}].$$

Differentiating with respect to $t$ gives

$$\frac{1}{\psi(t)} \frac{d\psi(t)}{dt} = \exp(\tilde{z}^T \alpha) [\exp(\alpha_1) - 1] \frac{1}{t} + \frac{\exp(\tilde{x}^T \beta + \tilde{z}^T \alpha) t^{\exp(\tilde{z}^T \alpha) - 1}}{1 + \exp(\tilde{x}^T \beta) t^{\exp(\tilde{z}^T \alpha)}}$$

$$- \frac{\exp(\tilde{z}^T \alpha) [\exp(\alpha_1) - 1] \frac{1}{t}}{1 + \exp(\beta_1 + \tilde{x}^T \beta) t^{\exp(\alpha_1 + \tilde{z}^T \alpha)}}$$

$$= \exp(\tilde{z}^T \alpha) [\exp(\alpha_1) - 1] \frac{1}{t} + \lambda(t \mid c = 0) - \lambda(t \mid c = 1)$$

$$= \exp(\tilde{z}^T \alpha) [\exp(\alpha_1) - 1] \frac{1}{t} + \lambda(t \mid c = 0)[1 - \psi(t)]$$

$$\Rightarrow \frac{d\psi(t)}{dt} = \left\{ \exp(\tilde{z}^T \alpha) [\exp(\alpha_1) - 1] \frac{1}{t} + \lambda(t \mid c = 0)[1 - \psi(t)] \right\} \psi(t)$$

$$= [a(t) + b(t)] \psi(t),$$ \hspace{1cm} (B.27)

where $a(t) = \exp(\tilde{z}^T \alpha) [\exp(\alpha_1) - 1]/t$ and $b(t) = \lambda(t \mid c = 0)[1 - \psi(t)]$. Clearly $\text{sgn}[a(t)] = \text{sgn}(\alpha_1)$ whereas $\text{sgn}[b(t)] = \text{sgn}[1 - \psi(t)]$. Thus, in the case where
\[ \alpha_1 = 0, \ sgn[\psi'(t)] = sgn[b(t)] = sgn[1 - \psi(t)] \] and

\[
1 - \psi(t \mid \alpha_1 = 0) = 1 - \frac{\exp(\beta_1) + \exp(\beta_1 + \tilde{x}^T \beta) t^{\exp(\tilde{z}^T \alpha)}}{1 + \exp(\beta_1 + \tilde{x}^T \beta) t^{\exp(\tilde{z}^T \alpha)}}
\]

\[
= \frac{1 + \exp(\beta_1 + \tilde{x}^T \beta)^{\exp(\tilde{z}^T \alpha)} - \exp(\beta_1) - \exp(\beta_1 + \tilde{x}^T \beta)^{\exp(\tilde{z}^T \alpha)}}{1 + \exp(\beta_1 + \tilde{x}^T \beta)^{\exp(\tilde{z}^T \alpha)}}
\]

\[
= \frac{1 - \exp(\beta_1)}{1 + \exp(\beta_1 + \tilde{x}^T \beta)^{\exp(\tilde{z}^T \alpha)}}.
\]

Thus, \( sgn[\psi'(t \mid \alpha_1 = 0)] = -sgn(\beta_1) \Rightarrow \psi(t \mid \alpha_1 = 0) \) increases/decreases monotonically (note that when \( \alpha_1 = \beta_1 = 0, \psi(t) = 1 \)). In the more general case, where \( \alpha_1 \neq 0, \ sgn[\psi'(t)] \) cannot be determined as straightforwardly but we can still get some insight. Let’s consider the case where \( \alpha_1 > 0 \). Inspection of the limits (calculated below) shows that \( \psi(t) \) increases from zero to \( \exp(\alpha_1) > 1 \). As \( \psi(t) \) crosses unity at some time, say, \( t^* \), \( b(t) \) changes sign at this point: \( sgn[b(t)] = 1 \) for \( t < t^* \), \( sgn[b(t)] = 0 \) for \( t = t^* \) and \( sgn[b(t)] = -1 \) for \( t > t^* \). Due to the fact that \( sgn[a(t)] = 1 \forall t \Rightarrow sgn[\psi'(t)] = 1 \) for \( t \leq t^* \). For \( t > t^* \), \( \psi'(t) \) will change sign if \( |b(t)| > |a(t)| \) for some \( t > t^* \). Therefore, \( \psi(t) \) may increase monotonically or non-monotonically when \( \alpha_1 > 0 \). Furthermore, using arguments similar to the above, we find that \( \psi(t) \) may decrease monotonically or non-monotonically when \( \alpha_1 < 0 \).

We now investigate the limits of \( \psi(t) \):

\[
\psi(0) = \lim_{t \to 0} \exp(\beta_1 + \alpha_1) t^{\exp(\tilde{z}^T \alpha)[\exp(\alpha_1) - 1]} \frac{1 + \exp(\tilde{x}^T \beta)^{\exp(\tilde{z}^T \alpha)}}{1 + \exp(\beta_1 + \tilde{x}^T \beta)^{\exp(\alpha_1 + \tilde{z}^T \alpha)}}
\]

\[
= \lim_{t \to 0} \exp(\beta_1 + \alpha_1) t^{\exp(\tilde{z}^T \alpha)[\exp(\alpha_1) - 1]} \frac{1 + 0}{1 + 0}
\]

\[
= \lim_{t \to 0} \exp(\beta_1 + \alpha_1) t^{\exp(\tilde{z}^T \alpha)[\exp(\alpha_1) - 1]}
\]

\[
= \begin{cases} 
0 & \text{if } \alpha_1 > 0, \\
\exp(\beta_1) & \text{if } \alpha_1 = 0, \\
\infty & \text{if } \alpha_1 < 0, 
\end{cases}
\]

(B.28)
and

$$
\psi(\infty) = \lim_{t \to \infty} \exp(\beta_1 + \alpha_1) t^{\exp(\tilde{z}^T \alpha) | \exp(\alpha_1) - 1} \frac{1 + \exp(\tilde{x}^T \beta) t^{\exp(\tilde{z}^T \alpha)}}{1 + \exp(\beta_1 + \tilde{x}^T \beta) t^{\exp(\alpha_1 + \tilde{z}^T \alpha)}}
$$

$$
= \lim_{t \to \infty} \exp(\beta_1 + \alpha_1) t^{\exp(\tilde{z}^T \alpha) | \exp(\alpha_1) - 1} \frac{\exp(\tilde{x}^T \beta) t^{\exp(\tilde{z}^T \alpha)}}{\exp(\beta_1 + \tilde{x}^T \beta) t^{\exp(\alpha_1 + \tilde{z}^T \alpha)}}
$$

$$
= \lim_{t \to \infty} \exp(\beta_1 + \alpha_1) t^{\exp(\tilde{z}^T \alpha) | \exp(\alpha_1) - 1} \exp(-\beta_1) t^{-\exp(\tilde{z}^T \alpha) | \exp(\alpha_1) - 1}
$$

$$
= \lim_{t \to \infty} \exp(\alpha_1) \quad \forall \alpha_1.
$$

(B.29)

### B.3.4 Burr

The hazard function for the Burr MPR model (Section 5.2.4) is given by

$$
\lambda(t \mid x, z, w) = \exp(x^T \beta + z^T \alpha) t^{\exp(z^T \alpha) - 1} \frac{1 + \exp(\tilde{x}^T \beta) t^{\exp(\tilde{z}^T \alpha)}}{1 + \exp(\beta_1 + \tilde{x}^T \beta + c \tau_1 + \tilde{w}^T \tau) t^{\exp(\alpha_1 + \tilde{z}^T \alpha)}}
$$

Thus, the hazard ratio for $c$, at time $t$, is given by

$$
\psi(t) = \frac{\lambda(t \mid c = 1)}{\lambda(t \mid c = 0)}
$$

$$
= \frac{\exp(\beta_1 + \tilde{x}^T \beta + c \alpha_1 + \tilde{z}^T \alpha) t^{\exp(\alpha_1 + \tilde{z}^T \alpha) - 1} \exp(\tilde{x}^T \beta + \tilde{w}^T \tau) t^{\exp(\tilde{z}^T \alpha)}}{1 + \exp(\beta_1 + \tilde{x}^T \beta + c \tau_1 + \tilde{w}^T \tau) t^{\exp(\alpha_1 + \tilde{z}^T \alpha)}}
$$

$$
= \frac{\exp(\beta_1 + \alpha_1) t^{\exp(\tilde{z}^T \alpha) | \exp(\alpha_1) - 1} \exp(\tilde{x}^T \beta + \tilde{w}^T \tau) t^{\exp(\tilde{z}^T \alpha)}}{1 + \exp(\beta_1 + \tilde{x}^T \beta + \tilde{w}^T \tau) t^{\exp(\alpha_1 + \tilde{z}^T \alpha)}}
$$

(B.30)

and therefore

$$
\log \psi(t) = \beta_1 + \alpha_1 + \exp(\tilde{z}^T \alpha) | \exp(\alpha_1) - 1 \log t + \log[1 + \exp(\tilde{x}^T \beta + \tilde{w}^T \tau) t^{\exp(\tilde{z}^T \alpha)}]
$$

$$
- \log[1 + \exp(\beta_1 + \tilde{x}^T \beta + \tilde{w}^T \tau) t^{\exp(\alpha_1 + \tilde{z}^T \alpha)}].
$$
Differentiating with respect to \( t \) gives

\[
\frac{1}{\psi(t)} \frac{d\psi(t)}{dt} = \exp(\tilde{z}^T \alpha)[\exp(\alpha_1) - 1] \frac{1}{t} + \frac{\exp(\tilde{x}^T \beta + \tilde{z}^T \alpha + \tilde{w}^T \tau) t^{\exp(\tilde{z}^T \alpha) - 1}}{1 + \exp(\tilde{x}^T \beta + \tilde{w}^T \tau) t^{\exp(\tilde{z}^T \alpha)}} \cdot \frac{- \exp(\beta_1 + \tilde{x}^T \beta + \alpha_1 + \tilde{z}^T \alpha + \tau_1 + \tilde{w}^T \tau) t^{\exp(\alpha_1 + \tilde{z}^T \alpha) - 1}}{1 + \exp(\beta_1 + \tilde{x}^T \beta + \tau_1 + \tilde{w}^T \tau) t^{\exp(\alpha_1 + \tilde{z}^T \alpha)}}
\]

\[
= \exp(\tilde{z}^T \alpha)[\exp(\alpha_1) - 1] \frac{1}{t} + \exp(\tilde{w}^T \tau) \lambda(t \mid c = 0) - \exp(\tau_1 + \tilde{w}^T \tau) \lambda(t \mid c = 1)
\]

\[
= \exp(\tilde{z}^T \alpha)[\exp(\alpha_1) - 1] \frac{1}{t} + \exp(\tilde{w}^T \tau) \lambda(t \mid c = 0)[1 - \exp(\tau_1)\psi(t)]
\]

\[
\Rightarrow \frac{d\psi(t)}{dt} = \left\{ \exp(\tilde{z}^T \alpha)[\exp(\alpha_1) - 1] \frac{1}{t} + \exp(\tilde{w}^T \tau) \lambda(t \mid c = 0)[1 - \exp(\tau_1)\psi(t)] \right\} \psi(t)
\]

\[
= [a(t) + b(t)]\psi(t), \quad \text{(B.31)}
\]

where \( a(t) = \exp(\tilde{z}^T \alpha)[\exp(\alpha_1) - 1]/t \) and \( b(t) = \exp(\tilde{w}^T \tau) \lambda(t \mid c = 0)[1 - \exp(\tau_1)\psi(t)] \). In the case where \( \alpha_1 = 0 \), \( \text{sgn}[\psi'(t)] = \text{sgn}[b(t)] = \text{sgn}[1 - \exp(\tau_1)\psi(t)] \) and

\[
1 - \exp(\tau_1)\psi(t \mid \alpha_1 = 0) = 1 - \frac{\exp(\beta_1 + \tau_1) + \exp(\beta_1 + \tilde{x}^T \beta + \tau_1 + \tilde{w}^T \tau) t^{\exp(\tilde{z}^T \alpha)}}{1 + \exp(\beta_1 + \tilde{x}^T \beta + \tau_1 + \tilde{w}^T \tau) t^{\exp(\tilde{z}^T \alpha)}}
\]

\[
= \frac{1 - \exp(\beta_1 + \tau_1)}{1 + \exp(\beta_1 + \tilde{x}^T \beta + \tau_1 + \tilde{w}^T \tau) t^{\exp(\tilde{z}^T \alpha)}}.
\]

Thus, \( \text{sgn}[\psi'(t \mid \alpha_1 = 0)] = -\text{sgn}(\beta_1 + \tau_1) \Rightarrow \psi(t \mid \alpha_1 = 0) \) can increase/decrease monotonically or remain constant. As was the case with the log-logistic model, when \( \alpha_1 \neq 0 \) in the Burr model, \( \psi(t) \) can increase or decrease either monotonically or non-monotonically; this depends on the signs of \( a(t) \) and \( b(t) \) and which is larger in magnitude over time.
We now investigate the limits of $\psi(t)$:

$$\psi(0) = \lim_{t \to 0} \frac{1 + \exp(\tilde{x}^T \beta + \tilde{w}^T \tau) t^{\exp(\tilde{z}^T \alpha)}}{1 + \exp(\beta_1 + \tilde{x}^T \beta + \tau_1 + \tilde{w}^T \tau) t^{\exp(\alpha_1 + \tilde{z}^T \alpha)}}$$

$$= \lim_{t \to 0} \frac{1 + 0}{1 + 0}$$

$$= \lim_{t \to 0} \frac{1 + \exp(\tilde{x}^T \beta + \tilde{w}^T \tau) t^{\exp(\tilde{z}^T \alpha)}}{1 + \exp(\beta_1 + \tilde{x}^T \beta + \tau_1 + \tilde{w}^T \tau) t^{\exp(\alpha_1 + \tilde{z}^T \alpha)}}$$

$$= \left\{ \begin{array}{ll} 0 & \text{if } \alpha_1 > 0, \\
\exp(\beta_1) & \text{if } \alpha_1 = 0, \\
\infty & \text{if } \alpha_1 < 0, \end{array} \right. \quad \text{(B.32)}$$

and

$$\psi(\infty) = \lim_{t \to \infty} \frac{1 + \exp(\tilde{x}^T \beta + \tilde{w}^T \tau) t^{\exp(\tilde{z}^T \alpha)}}{1 + \exp(\beta_1 + \tilde{x}^T \beta + \tau_1 + \tilde{w}^T \tau) t^{\exp(\alpha_1 + \tilde{z}^T \alpha)}}$$

$$= \lim_{t \to \infty} \frac{\exp(-\beta_1 - \tau_1) t^{-\exp(\tilde{z}^T \alpha) [\exp(\alpha_1 - 1]}}{1 + \exp(\alpha_1 - \tau_1)}$$

$$\forall \alpha_1. \quad \text{(B.33)}$$

### B.3.5 Time-Dependent Logistic

The hazard function for the time-dependent logistic MPR model (Section 5.2.5) is given by

$$\lambda(t \mid x, z) = \frac{\exp(\tilde{z}^T \alpha t + x^T \beta)}{1 + \exp(\tilde{z}^T \alpha t + x^T \beta)}$$

$$= \frac{\exp((c \alpha_1 + \tilde{z}^T \alpha) t + c \beta_1 + \tilde{x}^T \beta)}{1 + \exp((c \alpha_1 + \tilde{z}^T \alpha) t + c \beta_1 + \tilde{x}^T \beta)}$$
Thus, the hazard ratio for \( c \), at time \( t \), is given by
\[
\psi(t) = \frac{\lambda(t \mid c = 1)}{\lambda(t \mid c = 0)}
\]
\[
= \frac{\exp[(\alpha_1 + \tilde{z}^T \alpha)t + \beta_1 + \tilde{x}^T \beta]}{1 + \exp(\tilde{z}^T \alpha t + \tilde{x}^T \beta)} \cdot \lambda(t \mid c = 0)
\]
\[
= \exp(\alpha_1 t + \beta_1) \frac{1 + \exp(\tilde{z}^T \alpha t + \tilde{x}^T \beta)}{1 + \exp((\alpha_1 + \tilde{z}^T \alpha)t + \beta_1 + \tilde{x}^T \beta)},
\]
\[
(B.34)
\]
and therefore
\[
\log \psi(t) = \alpha_1 t + \beta_1 + \log[1 + \exp(\tilde{z}^T \alpha t + \tilde{x}^T \beta)]
\]
\[
- \log\{1 + \exp((\alpha_1 + \tilde{z}^T \alpha)t + \beta_1 + \tilde{x}^T \beta)\}.
\]

Differentiating with respect to \( t \) gives
\[
\frac{1}{\psi(t)} \frac{d\psi(t)}{dt} = \alpha_1 + \tilde{z}^T \alpha \frac{\exp(\tilde{z}^T \alpha t + \tilde{x}^T \beta)}{1 + \exp(\tilde{z}^T \alpha t + \tilde{x}^T \beta)}
\]
\[
- (\alpha_1 + \tilde{z}^T \alpha) \frac{\exp((\alpha_1 + \tilde{z}^T \alpha)t + \beta_1 + \tilde{x}^T \beta)}{1 + \exp((\alpha_1 + \tilde{z}^T \alpha)t + \beta_1 + \tilde{x}^T \beta)}
\]
\[
= \alpha_1 + \tilde{z}^T \alpha \lambda(t \mid c = 0) - (\alpha_1 + \tilde{z}^T \alpha)\lambda(t \mid c = 1)
\]
\[
\Rightarrow \frac{d\psi(t)}{dt} = [\alpha_1 + \tilde{z}^T \alpha \lambda(t \mid c = 0) - (\alpha_1 + \tilde{z}^T \alpha)\lambda(t \mid c = 1)] \psi(t).
\]
\[
(B.35)
\]
We can see that sign and magnitude of both \( \alpha_1 \) and \( \tilde{z}^T \alpha \) play a central role in characterising the time evolution of \( \psi(t) \).

We now investigate the limits of \( \psi(t) \). Firstly
\[
\psi(0) = \exp(0 + \beta_1) \frac{1 + \exp(0 + \tilde{x}^T \beta)}{1 + \exp(0 + \beta_1 + \tilde{x}^T \beta)}
\]
\[
= \exp(\beta_1) \frac{1 + \exp(\tilde{x}^T \beta)}{1 + \exp(\beta_1 + \tilde{x}^T \beta)},
\]
\[
(B.36)
\]
Calculating \( \psi(\infty) \) is a little bit more involved. It is helpful to express \( \psi(t) \) in the following way:
\[
\psi(t) = \exp(\beta_1) \frac{\exp(\alpha_1 t) + \exp((\alpha_1 + \tilde{z}^T \alpha)t + \tilde{x}^T \beta)}{1 + \exp((\alpha_1 + \tilde{z}^T \alpha)t + \beta_1 + \tilde{x}^T \beta)},
\]
\[
(B.37)
\]
where it is clear that $\psi(\infty)$ depends on the sign and magnitude of both $\alpha_1$ and $\hat{z}^T \alpha$. There are a variety of possibilities to be considered:

<table>
<thead>
<tr>
<th>$\text{sgn} (\alpha_1)$</th>
<th>$\text{sgn} (\hat{z}^T \alpha)$</th>
<th>$\text{sgn} (\alpha_1 + \hat{z}^T \alpha)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td>$</td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
<td>$</td>
</tr>
<tr>
<td>5</td>
<td>-1</td>
<td>$</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>9</td>
<td>-1</td>
<td>$</td>
</tr>
<tr>
<td>10</td>
<td>-1</td>
<td>$</td>
</tr>
<tr>
<td>11</td>
<td>-1</td>
<td>$</td>
</tr>
<tr>
<td>12</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

We now calculate $\psi(\infty)$, using $\psi(t)$ in the form given in (B.37), for each of the above possibilities.

Cases 1, 2 and 3: $\alpha_1 > 0$, $\alpha_1 + \hat{z}^T \alpha > 0$

$$
\psi(\infty) = \lim_{t \to \infty} \exp(\beta_1) \frac{\exp(\alpha_1 t) + \exp[(\alpha_1 + \hat{z}^T \alpha) t + \hat{x}^T \beta]}{1 + \exp[(\alpha_1 + \hat{z}^T \alpha) t + \beta_1 + \hat{x}^T \beta]} \\
= \lim_{t \to \infty} \exp(\beta_1) \frac{\exp(\alpha_1 t) + \exp[(\alpha_1 + \hat{z}^T \alpha) t + \hat{x}^T \beta]}{\exp[(\alpha_1 + \hat{z}^T \alpha) t + \beta_1 + \hat{x}^T \beta]} \\
= \lim_{t \to \infty} \exp(\beta_1)[\exp(-\hat{z}^T \alpha t - \beta_1 - \hat{x}^T \beta) + \exp(-\beta_1)] \\
= \begin{cases} 
1 & \text{if } \hat{z}^T \alpha > 0, \\
\frac{1 + \exp(\hat{x}^T \beta)}{\exp(\hat{x}^T \beta)} & \text{if } \hat{z}^T \alpha = 0, \\
\infty & \text{if } -\alpha_1 < \hat{z}^T \alpha < 0.
\end{cases}
$$
B.3. PROPERTIES OF THE HAZARD RATIO

Case 4: \( \alpha_1 > 0, \ \alpha_1 + z^T \alpha = 0 \)

\[
\psi(\infty) = \lim_{t \to \infty} \exp(\beta_1) \frac{\exp(\alpha_1 t) + \exp(\tilde{x}^T \beta)}{1 + \exp(\beta_1 + \tilde{x}^T \beta)}
\]

= \infty.

Case 5: \( \alpha_1 > 0, \ \alpha_1 + z^T \alpha < 0 \)

\[
\psi(\infty) = \lim_{t \to \infty} \exp(\beta_1) \frac{\exp(\alpha_1 t) + \exp[(\alpha_1 + z^T \alpha)t + \tilde{x}^T \beta]}{1 + \exp[(\alpha_1 + z^T \alpha)t + \beta_1 + \tilde{x}^T \beta]}
\]

= \lim_{t \to \infty} \exp(\beta_1) \frac{\exp(\alpha_1 t) + 0}{1 + 0}

= \infty.

Cases 6, 7, 8: \( \alpha_1 = 0 \)

\[
\psi(\infty) = \lim_{t \to \infty} \exp(\beta_1) \frac{1 + \exp(\tilde{z}^T \alpha t + \tilde{x}^T \beta)}{1 + \exp(\tilde{z}^T \alpha t + \beta_1 + \tilde{x}^T \beta)}
\]

= \begin{cases} 
1 & \text{if } \tilde{z}^T \alpha > 0, \\
\exp(\beta_1) \frac{1+\exp(\tilde{x}^T \beta)}{1+\exp(\beta_1+\tilde{x}^T \beta)} & \text{if } \tilde{z}^T \alpha = 0, \\
\exp(\beta_1) & \text{if } \tilde{z}^T \alpha < 0.
\end{cases}
Case 9: \( \alpha_1 < 0, \quad \alpha_1 + \mathbf{z}^T \mathbf{\alpha} > 0 \)

\[
\psi(\infty) = \lim_{t \to \infty} \exp(\beta_1) \frac{\exp(\alpha_1 t) + \exp((\alpha_1 + \mathbf{z}^T \mathbf{\alpha}) t + \mathbf{x}^T \mathbf{\beta})}{1 + \exp((\alpha_1 + \mathbf{z}^T \mathbf{\alpha}) t + \beta_1 + \mathbf{x}^T \mathbf{\beta})}
\]

\[
= \lim_{t \to \infty} \exp(\beta_1) \frac{\exp((\alpha_1 + \mathbf{z}^T \mathbf{\alpha}) t + \mathbf{x}^T \mathbf{\beta})}{\exp((\alpha_1 + \mathbf{z}^T \mathbf{\alpha}) t + \beta_1 + \mathbf{x}^T \mathbf{\beta})}
\]

\[
= 1.
\]

Case 10: \( \alpha_1 < 0, \quad \alpha_1 + \mathbf{z}^T \mathbf{\alpha} = 0 \)

\[
\psi(\infty) = \lim_{t \to \infty} \exp(\beta_1) \frac{\exp(\alpha_1 t) + \exp(\mathbf{x}^T \mathbf{\beta})}{1 + \exp(\beta_1 + \mathbf{x}^T \mathbf{\beta})}
\]

\[
= \frac{\exp(\beta_1 + \mathbf{x}^T \mathbf{\beta})}{1 + \exp(\beta_1 + \mathbf{x}^T \mathbf{\beta})}.
\]

Cases 11, 12, 13: \( \alpha_1 < 0, \quad \alpha_1 + \mathbf{z}^T \mathbf{\alpha} < 0 \)

\[
\psi(\infty) = \lim_{t \to \infty} \exp(\beta_1) \frac{\exp(\alpha_1 t) + \exp((\alpha_1 + \mathbf{z}^T \mathbf{\alpha}) t + \mathbf{x}^T \mathbf{\beta})}{1 + \exp((\alpha_1 + \mathbf{z}^T \mathbf{\alpha}) t + \beta_1 + \mathbf{x}^T \mathbf{\beta})}
\]

\[
= \frac{\exp(\beta_1) 0 + 0}{1 + 0}
\]

\[
= 0.
\]
Thus, putting all of the above together,

$$\psi(\infty) = \begin{cases} 
1 & \text{if } \alpha_1 \geq 0 \text{ and } \tilde{z}^T \alpha > 0, \\
\frac{1+\exp(\tilde{x}^T \beta)}{\exp(\tilde{x}^T \beta)} & \text{if } \alpha_1 > 0 \text{ and } \tilde{z}^T \alpha = 0, \\
\infty & \text{if } \alpha_1 > 0 \text{ and } \tilde{z}^T \alpha < 0, \\
\exp(\beta_1) & \text{if } \alpha_1 = 0 \text{ and } \tilde{z}^T \alpha = 0, \\
\exp(\tilde{z}^T \alpha) & \text{if } \alpha_1 < 0 \text{ and } \tilde{z}^T \alpha = -\alpha_1, \\
0 & \text{if } \alpha_1 < 0 \text{ and } \tilde{z}^T \alpha < -\alpha_1. 
\end{cases}$$

(B.38)

### B.4 MPR Simulation Studies

In Section 5.5, we presented simulation studies involving the standard and orthogonal Weibull MPR models; data were simulated using 30 different parameter vectors for each model. The values of the parameters in these vectors are shown in Tables B.1 and B.2. These values were selected based on fitting the models to a variety of real datasets first in order to create realistic simulated data, i.e., data-directed simulation. Similarly, Tables B.3 and B.4 show the parameter vectors for log-logistic and Gompertz MPR simulations; the results of these supplementary studies (alluded to in Section 5.5) are also given in this appendix.

Recall that in Sections 5.5.1 and 5.5.2 we studied the correlation structure for estimated parameters and the bias in these estimates (for the standard and orthogonal Weibull MPR models). Here the scatter matrix plots arising from the log-logistic and Gompertz simulation studies are given in Figs. B.1 and B.2. Moreover, Tables B.5 and B.6 show the bias in estimation for these two models. The results of these supplementary simulation studies mirror those from Sections 5.5.1 and 5.5.2 apart from the fact that the Gompertz MPR model has some very large estimation bias when the sample size is small with a high censored proportion.
### Table B.1. Weibull MPR Simulation Settings

<table>
<thead>
<tr>
<th>No.</th>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \alpha_0 )</th>
<th>( \alpha_1 )</th>
<th>( \alpha_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-3.4</td>
<td>0.0</td>
<td>3.0</td>
<td>0.1</td>
<td>0.0</td>
<td>-0.3</td>
</tr>
<tr>
<td>2</td>
<td>-2.2</td>
<td>0.4</td>
<td>0.0</td>
<td>-0.5</td>
<td>-0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>3</td>
<td>-4.3</td>
<td>1.0</td>
<td>2.1</td>
<td>0.2</td>
<td>-0.5</td>
<td>-0.4</td>
</tr>
<tr>
<td>4</td>
<td>-4.4</td>
<td>0.9</td>
<td>1.7</td>
<td>-0.1</td>
<td>0.1</td>
<td>-0.2</td>
</tr>
<tr>
<td>5</td>
<td>-5.5</td>
<td>0.3</td>
<td>1.7</td>
<td>0.4</td>
<td>-0.2</td>
<td>-0.6</td>
</tr>
<tr>
<td>6</td>
<td>-3.5</td>
<td>-0.1</td>
<td>0.7</td>
<td>0.1</td>
<td>-0.1</td>
<td>-0.1</td>
</tr>
<tr>
<td>7</td>
<td>-3.0</td>
<td>-0.2</td>
<td>-2.3</td>
<td>-0.2</td>
<td>0.1</td>
<td>0.5</td>
</tr>
<tr>
<td>8</td>
<td>-5.3</td>
<td>0.8</td>
<td>0.9</td>
<td>-0.2</td>
<td>-0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>9</td>
<td>-5.0</td>
<td>0.9</td>
<td>0.6</td>
<td>0.0</td>
<td>-0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>10</td>
<td>-4.0</td>
<td>-0.3</td>
<td>-0.6</td>
<td>-0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>11</td>
<td>-5.5</td>
<td>-0.2</td>
<td>0.9</td>
<td>-0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>12</td>
<td>-4.1</td>
<td>-0.4</td>
<td>-1.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>13</td>
<td>-4.2</td>
<td>-0.3</td>
<td>-0.2</td>
<td>-0.1</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>14</td>
<td>-6.7</td>
<td>0.1</td>
<td>1.2</td>
<td>0.2</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>15</td>
<td>-7.5</td>
<td>0.5</td>
<td>4.6</td>
<td>0.5</td>
<td>-0.2</td>
<td>-0.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No.</th>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \alpha_0 )</th>
<th>( \alpha_1 )</th>
<th>( \alpha_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>-12.3</td>
<td>0.5</td>
<td>7.7</td>
<td>0.9</td>
<td>0.0</td>
<td>-0.9</td>
</tr>
<tr>
<td>17</td>
<td>-1.7</td>
<td>-0.8</td>
<td>0.4</td>
<td>-0.2</td>
<td>0.5</td>
<td>-0.2</td>
</tr>
<tr>
<td>18</td>
<td>-2.9</td>
<td>-0.4</td>
<td>0.2</td>
<td>-0.2</td>
<td>-0.2</td>
<td>-0.1</td>
</tr>
<tr>
<td>19</td>
<td>-2.7</td>
<td>0.3</td>
<td>0.2</td>
<td>-0.4</td>
<td>0.0</td>
<td>-0.1</td>
</tr>
<tr>
<td>20</td>
<td>-3.1</td>
<td>0.4</td>
<td>0.3</td>
<td>-0.1</td>
<td>-0.3</td>
<td>-0.3</td>
</tr>
<tr>
<td>21</td>
<td>-2.7</td>
<td>0.1</td>
<td>0.2</td>
<td>-0.5</td>
<td>0.0</td>
<td>0.2</td>
</tr>
<tr>
<td>22</td>
<td>-3.9</td>
<td>-0.6</td>
<td>3.3</td>
<td>1.3</td>
<td>0.3</td>
<td>-1.5</td>
</tr>
<tr>
<td>23</td>
<td>-3.9</td>
<td>3.2</td>
<td>0.1</td>
<td>1.2</td>
<td>-1.4</td>
<td>0.1</td>
</tr>
<tr>
<td>24</td>
<td>-3.5</td>
<td>3.2</td>
<td>-1.1</td>
<td>1.4</td>
<td>-1.5</td>
<td>0.1</td>
</tr>
<tr>
<td>25</td>
<td>-3.9</td>
<td>0.7</td>
<td>-0.8</td>
<td>-0.2</td>
<td>-0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>26</td>
<td>-3.9</td>
<td>1.5</td>
<td>-0.9</td>
<td>1.1</td>
<td>-1.2</td>
<td>-0.2</td>
</tr>
<tr>
<td>27</td>
<td>-1.0</td>
<td>0.6</td>
<td>-1.1</td>
<td>0.9</td>
<td>-1.3</td>
<td>0.0</td>
</tr>
<tr>
<td>28</td>
<td>-2.6</td>
<td>-0.2</td>
<td>-0.9</td>
<td>-0.3</td>
<td>-0.1</td>
<td>0.4</td>
</tr>
<tr>
<td>29</td>
<td>-2.7</td>
<td>0.1</td>
<td>0.0</td>
<td>-0.2</td>
<td>-0.3</td>
<td>0.1</td>
</tr>
<tr>
<td>30</td>
<td>-1.4</td>
<td>-0.5</td>
<td>-0.7</td>
<td>-0.8</td>
<td>0.3</td>
<td>0.2</td>
</tr>
</tbody>
</table>

### Table B.2. Orthogonal Weibull MPR Simulation Settings

<table>
<thead>
<tr>
<th>No.</th>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \alpha_0 )</th>
<th>( \alpha_1 )</th>
<th>( \alpha_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-3.4</td>
<td>0.0</td>
<td>2.5</td>
<td>0.2</td>
<td>-0.1</td>
<td>-0.3</td>
</tr>
<tr>
<td>2</td>
<td>-4.1</td>
<td>0.0</td>
<td>0.4</td>
<td>-0.5</td>
<td>-0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>3</td>
<td>-5.9</td>
<td>1.6</td>
<td>1.7</td>
<td>-0.3</td>
<td>0.5</td>
<td>-0.1</td>
</tr>
<tr>
<td>4</td>
<td>-3.8</td>
<td>-0.7</td>
<td>-1.0</td>
<td>0.4</td>
<td>-0.3</td>
<td>-0.4</td>
</tr>
<tr>
<td>5</td>
<td>-3.7</td>
<td>-0.4</td>
<td>0.1</td>
<td>-0.1</td>
<td>-0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>6</td>
<td>-4.0</td>
<td>0.3</td>
<td>0.1</td>
<td>-0.1</td>
<td>0.0</td>
<td>0.5</td>
</tr>
<tr>
<td>7</td>
<td>-6.5</td>
<td>0.0</td>
<td>1.9</td>
<td>0.0</td>
<td>-0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>8</td>
<td>-5.2</td>
<td>0.1</td>
<td>0.5</td>
<td>0.1</td>
<td>-0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>9</td>
<td>-4.8</td>
<td>0.3</td>
<td>-0.3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>10</td>
<td>-6.6</td>
<td>0.3</td>
<td>1.4</td>
<td>-0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>11</td>
<td>-4.5</td>
<td>-0.2</td>
<td>-1.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>12</td>
<td>-5.0</td>
<td>-0.2</td>
<td>0.3</td>
<td>-0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>13</td>
<td>-5.7</td>
<td>-0.1</td>
<td>0.8</td>
<td>0.2</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>14</td>
<td>-5.0</td>
<td>-0.3</td>
<td>1.8</td>
<td>0.2</td>
<td>0.2</td>
<td>-0.6</td>
</tr>
<tr>
<td>15</td>
<td>-4.3</td>
<td>0.5</td>
<td>-0.3</td>
<td>-0.8</td>
<td>0.3</td>
<td>0.2</td>
</tr>
</tbody>
</table>
### Table B.3. Log-Logistic MPR Simulation Settings

<table>
<thead>
<tr>
<th>No.</th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\alpha_0$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>No.</th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\alpha_0$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-3.7</td>
<td>0.2</td>
<td>3.8</td>
<td>0.3</td>
<td>0.0</td>
<td>0.0</td>
<td>16</td>
<td>-13.0</td>
<td>0.9</td>
<td>7.8</td>
<td>0.9</td>
<td>0.0</td>
<td>-0.8</td>
</tr>
<tr>
<td>2</td>
<td>-2.8</td>
<td>0.8</td>
<td>-0.3</td>
<td>0.1</td>
<td>-0.2</td>
<td>0.2</td>
<td>17</td>
<td>-1.7</td>
<td>-1.3</td>
<td>0.6</td>
<td>0.1</td>
<td>0.5</td>
<td>-0.2</td>
</tr>
<tr>
<td>3</td>
<td>-4.6</td>
<td>0.9</td>
<td>2.2</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.2</td>
<td>18</td>
<td>-3.2</td>
<td>-0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>-0.3</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>-4.2</td>
<td>0.5</td>
<td>1.4</td>
<td>-0.1</td>
<td>0.2</td>
<td>0.1</td>
<td>19</td>
<td>-2.9</td>
<td>0.2</td>
<td>0.2</td>
<td>-0.2</td>
<td>0.1</td>
<td>-0.1</td>
</tr>
<tr>
<td>5</td>
<td>-6.0</td>
<td>2.4</td>
<td>-1.2</td>
<td>0.6</td>
<td>-0.6</td>
<td>0.1</td>
<td>20</td>
<td>-3.7</td>
<td>0.6</td>
<td>0.6</td>
<td>0.3</td>
<td>-0.4</td>
<td>-0.4</td>
</tr>
<tr>
<td>6</td>
<td>-4.4</td>
<td>0.1</td>
<td>1.0</td>
<td>0.4</td>
<td>-0.1</td>
<td>-0.2</td>
<td>21</td>
<td>-2.8</td>
<td>0.2</td>
<td>0.0</td>
<td>-0.3</td>
<td>0.0</td>
<td>0.3</td>
</tr>
<tr>
<td>7</td>
<td>-3.7</td>
<td>-0.4</td>
<td>-2.4</td>
<td>0.2</td>
<td>0.2</td>
<td>0.4</td>
<td>22</td>
<td>-3.9</td>
<td>-0.8</td>
<td>3.7</td>
<td>1.5</td>
<td>0.2</td>
<td>-1.5</td>
</tr>
<tr>
<td>8</td>
<td>-5.6</td>
<td>1.0</td>
<td>0.3</td>
<td>-0.1</td>
<td>-0.1</td>
<td>0.4</td>
<td>23</td>
<td>-3.9</td>
<td>3.5</td>
<td>0.1</td>
<td>1.3</td>
<td>-1.4</td>
<td>0.2</td>
</tr>
<tr>
<td>9</td>
<td>-5.5</td>
<td>1.0</td>
<td>0.3</td>
<td>0.2</td>
<td>-0.1</td>
<td>0.1</td>
<td>24</td>
<td>-3.4</td>
<td>3.7</td>
<td>-1.4</td>
<td>1.6</td>
<td>-1.5</td>
<td>0.1</td>
</tr>
<tr>
<td>10</td>
<td>-4.7</td>
<td>-0.4</td>
<td>-0.4</td>
<td>0.2</td>
<td>0.2</td>
<td>0.0</td>
<td>25</td>
<td>-0.6</td>
<td>0.2</td>
<td>-0.2</td>
<td>-0.1</td>
<td>0.2</td>
<td>-0.1</td>
</tr>
<tr>
<td>11</td>
<td>-5.7</td>
<td>-0.5</td>
<td>0.7</td>
<td>0.0</td>
<td>0.2</td>
<td>0.2</td>
<td>26</td>
<td>-0.6</td>
<td>1.9</td>
<td>-1.1</td>
<td>1.2</td>
<td>-1.1</td>
<td>-0.2</td>
</tr>
<tr>
<td>12</td>
<td>-4.9</td>
<td>-0.3</td>
<td>-1.3</td>
<td>0.3</td>
<td>0.0</td>
<td>-0.1</td>
<td>27</td>
<td>-0.7</td>
<td>0.7</td>
<td>-1.3</td>
<td>1.1</td>
<td>-1.3</td>
<td>0.0</td>
</tr>
<tr>
<td>13</td>
<td>-4.7</td>
<td>-0.3</td>
<td>-0.5</td>
<td>0.1</td>
<td>0.0</td>
<td>0.2</td>
<td>28</td>
<td>-2.7</td>
<td>-0.2</td>
<td>-1.0</td>
<td>-0.1</td>
<td>-0.1</td>
<td>0.4</td>
</tr>
<tr>
<td>14</td>
<td>-7.1</td>
<td>0.3</td>
<td>0.8</td>
<td>0.3</td>
<td>-0.1</td>
<td>0.1</td>
<td>29</td>
<td>-2.9</td>
<td>0.4</td>
<td>-0.2</td>
<td>0.1</td>
<td>-0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>15</td>
<td>-7.8</td>
<td>0.0</td>
<td>4.5</td>
<td>0.6</td>
<td>-0.1</td>
<td>-0.2</td>
<td>30</td>
<td>-1.3</td>
<td>-0.6</td>
<td>-0.8</td>
<td>-0.6</td>
<td>0.3</td>
<td>0.2</td>
</tr>
</tbody>
</table>

### Table B.4. Gompertz MPR Simulation Settings

<table>
<thead>
<tr>
<th>No.</th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\alpha_0$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>No.</th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\alpha_0$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-3.1</td>
<td>0.0</td>
<td>2.8</td>
<td>0.0</td>
<td>0.0</td>
<td>-0.1</td>
<td>16</td>
<td>-0.3</td>
<td>-0.4</td>
<td>0.5</td>
<td>-0.5</td>
<td>-0.2</td>
<td>1.9</td>
</tr>
<tr>
<td>2</td>
<td>-2.0</td>
<td>0.0</td>
<td>0.7</td>
<td>-0.1</td>
<td>0.1</td>
<td>-0.1</td>
<td>17</td>
<td>-2.4</td>
<td>-0.4</td>
<td>2.0</td>
<td>1.6</td>
<td>-0.6</td>
<td>-2.0</td>
</tr>
<tr>
<td>3</td>
<td>-2.5</td>
<td>0.1</td>
<td>-0.6</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>18</td>
<td>-0.3</td>
<td>-0.9</td>
<td>0.1</td>
<td>-0.7</td>
<td>0.6</td>
<td>-0.7</td>
</tr>
<tr>
<td>4</td>
<td>-2.4</td>
<td>0.0</td>
<td>0.0</td>
<td>-0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>19</td>
<td>-4.3</td>
<td>4.5</td>
<td>-1.3</td>
<td>1.8</td>
<td>-2.8</td>
<td>0.8</td>
</tr>
<tr>
<td>5</td>
<td>-4.8</td>
<td>1.8</td>
<td>-0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>20</td>
<td>-0.4</td>
<td>0.7</td>
<td>-0.2</td>
<td>-1.9</td>
<td>-1.0</td>
<td>1.3</td>
</tr>
<tr>
<td>6</td>
<td>-4.0</td>
<td>1.0</td>
<td>-1.2</td>
<td>0.0</td>
<td>-0.1</td>
<td>0.1</td>
<td>21</td>
<td>-0.6</td>
<td>0.6</td>
<td>0.3</td>
<td>-0.4</td>
<td>-0.3</td>
<td>-2.0</td>
</tr>
<tr>
<td>7</td>
<td>-3.5</td>
<td>-0.8</td>
<td>1.5</td>
<td>0.1</td>
<td>0.0</td>
<td>-0.1</td>
<td>22</td>
<td>-0.4</td>
<td>0.7</td>
<td>-0.3</td>
<td>-1.3</td>
<td>-0.7</td>
<td>1.0</td>
</tr>
<tr>
<td>8</td>
<td>-4.4</td>
<td>0.5</td>
<td>-0.7</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>23</td>
<td>0.1</td>
<td>0.0</td>
<td>-1.4</td>
<td>-2.2</td>
<td>1.2</td>
<td>0.8</td>
</tr>
<tr>
<td>9</td>
<td>-5.2</td>
<td>-0.9</td>
<td>1.5</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>24</td>
<td>-0.5</td>
<td>0.0</td>
<td>0.4</td>
<td>-0.4</td>
<td>-0.1</td>
<td>-2.0</td>
</tr>
<tr>
<td>10</td>
<td>-7.0</td>
<td>-0.4</td>
<td>2.9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>25</td>
<td>-3.0</td>
<td>0.3</td>
<td>2.3</td>
<td>2.5</td>
<td>-0.7</td>
<td>-2.7</td>
</tr>
<tr>
<td>11</td>
<td>-6.1</td>
<td>1.2</td>
<td>1.7</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>26</td>
<td>0.2</td>
<td>0.3</td>
<td>-1.2</td>
<td>-1.6</td>
<td>-0.8</td>
<td>1.5</td>
</tr>
<tr>
<td>12</td>
<td>-2.1</td>
<td>1.3</td>
<td>0.7</td>
<td>0.0</td>
<td>-0.1</td>
<td>0.2</td>
<td>27</td>
<td>-0.1</td>
<td>0.5</td>
<td>-1.3</td>
<td>-0.5</td>
<td>-2.0</td>
<td>0.7</td>
</tr>
<tr>
<td>13</td>
<td>-1.5</td>
<td>-0.8</td>
<td>0.5</td>
<td>-0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>28</td>
<td>0.3</td>
<td>-0.1</td>
<td>-1.1</td>
<td>-1.9</td>
<td>0.7</td>
<td>1.0</td>
</tr>
<tr>
<td>14</td>
<td>-1.4</td>
<td>-0.8</td>
<td>1.0</td>
<td>-0.3</td>
<td>0.2</td>
<td>-0.3</td>
<td>29</td>
<td>-3.0</td>
<td>-0.5</td>
<td>-0.6</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>15</td>
<td>-0.1</td>
<td>-0.4</td>
<td>-0.2</td>
<td>-1.3</td>
<td>-0.5</td>
<td>0.9</td>
<td>30</td>
<td>-2.4</td>
<td>-0.8</td>
<td>-0.7</td>
<td>-0.1</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Figure B.1. Scatter matrices of estimated regression coefficients for the Log-logistic MPR model. Each cell in the lower triangle contains all 30 scatter plots with least squares lines overlayed (red). The upper triangle shows the corresponding average absolute correlation value.

Figure B.2. Scatter matrices of estimated regression coefficients for the Gompertz MPR model. Each cell in the lower triangle contains all 30 scatter plots with least squares lines overlayed (red). The upper triangle shows the corresponding average absolute correlation value.
Table B.5. Log-Logistic MPR Model: Average Relative Bias

<table>
<thead>
<tr>
<th>n</th>
<th>p</th>
<th>$\hat{\beta}_0$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
<th>$\hat{\alpha}_0$</th>
<th>$\hat{\alpha}_1$</th>
<th>$\hat{\alpha}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>80%</td>
<td>-0.094</td>
<td>0.064</td>
<td>-0.010</td>
<td>0.308</td>
<td>0.014</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>-0.056</td>
<td>0.022</td>
<td>0.008</td>
<td>0.170</td>
<td>0.011</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>-0.044</td>
<td>0.025</td>
<td>-0.015</td>
<td>0.151</td>
<td>-0.003</td>
<td>-0.003</td>
</tr>
<tr>
<td>500</td>
<td>80%</td>
<td>-0.014</td>
<td>0.009</td>
<td>-0.009</td>
<td>0.050</td>
<td>0.000</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>-0.009</td>
<td>0.007</td>
<td>-0.017</td>
<td>0.034</td>
<td>0.004</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>-0.008</td>
<td>0.007</td>
<td>0.013</td>
<td>0.028</td>
<td>-0.003</td>
<td>-0.007</td>
</tr>
<tr>
<td>1000</td>
<td>80%</td>
<td>-0.007</td>
<td>0.014</td>
<td>-0.001</td>
<td>0.027</td>
<td>0.006</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>-0.005</td>
<td>-0.007</td>
<td>-0.003</td>
<td>0.019</td>
<td>0.002</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>-0.004</td>
<td>-0.008</td>
<td>-0.006</td>
<td>0.011</td>
<td>0.004</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Table B.6. Gompertz MPR Model: Average Relative Bias

<table>
<thead>
<tr>
<th>n</th>
<th>p</th>
<th>$\hat{\beta}_0$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
<th>$\hat{\alpha}_0$</th>
<th>$\hat{\alpha}_1$</th>
<th>$\hat{\alpha}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>80%</td>
<td>-1.388</td>
<td>-0.114</td>
<td>-0.147</td>
<td>-2.181</td>
<td>-3.939</td>
<td>0.159</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>-0.116</td>
<td>0.029</td>
<td>0.020</td>
<td>-0.080</td>
<td>-0.055</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>-0.013</td>
<td>0.005</td>
<td>-0.009</td>
<td>-0.036</td>
<td>-0.036</td>
<td>0.032</td>
</tr>
<tr>
<td>500</td>
<td>80%</td>
<td>-0.028</td>
<td>0.000</td>
<td>-0.010</td>
<td>-0.200</td>
<td>-0.101</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>-0.009</td>
<td>0.001</td>
<td>-0.002</td>
<td>-0.020</td>
<td>-0.001</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>0.003</td>
<td>0.002</td>
<td>0.004</td>
<td>-0.007</td>
<td>0.002</td>
<td>0.010</td>
</tr>
<tr>
<td>1000</td>
<td>80%</td>
<td>-0.016</td>
<td>-0.002</td>
<td>-0.005</td>
<td>-0.082</td>
<td>-0.062</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>-0.001</td>
<td>0.004</td>
<td>-0.009</td>
<td>-0.013</td>
<td>0.010</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>-0.004</td>
<td>0.007</td>
<td>0.003</td>
<td>-0.003</td>
<td>0.000</td>
<td>0.004</td>
</tr>
</tbody>
</table>
B.5 The mpr Class

In Appendix B.2 we derived the likelihood function and score equations for each of the MPR models developed in Chapter 5 (which of course includes the SPR models of Chapter 4). We have created the mpr class in R which contains all of these models (except the piecewise exponential model - see Appendix B.5.4). In this section we discuss samples of our code. Although not yet ready for public use, this code forms the basis of an R package (in development) - the mpr package.

B.5.1 mprfit

In order to fit an MPR model, one uses the mprfit function:

```r
mprfit <- function(forms, data, family="Weibull", init=NA, hessian=F, iterlim=1000, varnames=T, ...){
    mprfitnum <- switch(family, Weibull=mprfit2, Gompertz=mprfit2, Loglogistic=mprfit2,
                        TDL=mprfit2, Burr=mprfit3,)
    mprfitnum(forms, data, family, init, hessian, iterlim, varnames, ...)
}
```

Here `forms` is a list (of formula objects), whose length is equal to the number of regression components in the specified model family (e.g., Weibull), and `data` is a data.frame containing the survival data. Presently, we require that the first two columns of `data` are the survival times and censoring indicators, respectively; the remaining columns contain any covariates. In other words, we have not used the Surv construct from the survival package to store the survival time and censoring indicator. Thus, the three models in Example 5.1 were fitted as follows:

```r
forma <- formb <- formula(~treat)
mprfit(list(forma,formb), data, family="Weibull")
mprfit(list(forma,formb), data, family="Loglogistic")
mprfit(list(forma,formb), data, family="TDL")
```

Note that the shape formula comes first, `forma`, followed by the scale, `formb`. 
We can see from the above that \texttt{mprfit} is a \textit{wrapper function} whose purpose is to select \texttt{mprfit2} for two-component models and \texttt{mprfit3} for the Burr model (i.e., the only three-component model we have considered). The \texttt{mprfit2} function is given below (\texttt{mprfit3} is not shown but is very similar). The output of this function is an \texttt{mpr} object which, among other things, stores the log-likelihood value, the m.l.e.s and the hessian matrix.

```r
# ---------------------------------------------------------------------------------------- #
mprfit2 <- function(forms,data,family="Weibull",init=NA,hessian=F,iterlim=1000,varnames=T){
  forma <- forms[[1]]
  formb <- forms[[2]]

  alpha.xis <- model.matrix(forma,data=data)
  beta.xis <- model.matrix(formb,data=data)

  k.alp <- dim(alpha.xis)[2]
  k.beta <- dim(beta.xis)[2]

  surtim <- data[,1]
  cen <- data[,2]
  surdat <- cbind(surtim, cen, alpha.xis, beta.xis)

  if(any(is.na(init))){
    init <- c(-0.2, rep(0.01, k.alp - 1), 0.5, rep(0.01, k.beta - 1))
  }else{
    if(any(init == "random")){
      init <- rnorm(k.alp+k.beta,0,0.2)
    }
  }

  loglik.res <- nlm(mprloglike,init,surdat=surdat,k=k.alp,family=family,
                     hessian=hessian,print.level=0,iterlim=iterlim)

  loglik <- -loglik.res$min
  npar <- length(init)
  aic <- 2*npar - 2*loglik
  bic <- log(dim(surdat)[1])*npar - 2*loglik
  stopcode <- loglik.res$code
  gradient <- loglik.res$grad
  iter <- loglik.res$iter

  mles <- loglik.res$estimate
  mles.a <- mles[1:k.alp]
  mles.b <- mles[-(1:k.alp)]

  hessian <- loglik.res$hess

  aname <- paste(colnames(alpha.xis), ".a",sep="")
  bname <- paste(colnames(beta.xis), ".b",sep="")
```

APPENDIX B. MULTI-PARAMETER REGRESSION

if(varnames){
  names(mles.a) <- aname
  names(mles.b) <- bname
  names(gradient) <- c(aname, bname)
  if(! is.null(hessian)){
    rownames(hessian) <- colnames(hessian) <- c(aname, bname)
  }
}

abterms <- union(attr(terms(forma), "term.labels"), attr(terms(formb), "term.labels"))
if(length(abterms) == 0){
  formab <- formula("~1")
} else{
  formab <- formula(paste("~", paste(abterms, collapse="+"), sep=""))
}
ab.mf <- model.frame(formab, data=data)
xlevels <- .getXlevels(terms(formab), ab.mf)
xvars <- sapply(attr(terms(formab), "variables"), deparse, width.cutoff = 500)[-1L]
out <- list(fit=data.frame(loglik,aic,bic,npar,stopcode,iter,family=family,stringsAsFactors=F),
              mles.a=mles.a, mles.b=mles.b, gradient=gradient, hessian=hessian,
              forms=forms, xlevels=xlevels, xvars=xvars, call=match.call())
class(out) <- "mpr"
out

# Note that mprfit2 uses the nlm function - a standard R implementation of the Newton-Raphson algorithm - to maximise the log-likelihood function (or rather minimise its negation). The appropriate log-likelihood function is chosen by passing family into the wrapper function mprloglike:

# ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

mprloglike <- function(param,surdat,k,family="Weibull",...){
  famloglike <- switch(family,
    Weibull=loglikeweibmpr,
    Gompertz=loglikegompmpr,
    Loglogistic=loglikeloglogistmpr,
    TDL=logliketdlmpr,
    Burr=loglikeburrmpr,
  )
  famloglike(param,surdat,k,...)
}
# ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Here param is the vector of shape and scale parameters (in the two-component case), i.e., \( \theta = (\alpha, \beta)^T \), and surdat is a data.frame whose first two columns are the survival times and censoring indicators, whose next \( k \) columns are the shape covariates and whose remaining columns are the scale covariates.
(recall that the shape and scale may have covariates in common so that there may be duplicate columns in `surdat`).

The log-likelihood functions (and corresponding score functions) which `mprloglike` selects from are given below, i.e., the R implementation of the work in Appendix B.2. In particular, the Weibull case, `loglikeweibmpr`, is shown in its entirety whereas the other cases are only partially shown as much of the code overlaps with `loglikeweibmpr`.

```r
# ---------------------------------------------------------------------------------------- #
loglikeweibmpr <- function(param,surdat,k){
  p <- length(param)
  q <- dim(surdat)[2]
  alphas <- param[1:k]
  betas <- param[(k+1):p]

ti <- surdat[,1]
deltai <- surdat[,2]

xi1 <- as(surdat[,3:(k+2)],"matrix")
xi2 <- as(surdat[,,(k+3):q],"matrix")

xialphas <- xi1%*%alphas
xibetas <- xi2%*%betas

gam <- exp(xialphas)
lan <- exp(xibetas)

dldalp <- t( deltai*(1+gam*log(ti)) - lam*gam*(ti^gam)*log(ti) )%*%xi1
dldbet <- t( deltai-lam*(ti^gam) )%*%xi2

loglike <- -sum(deltai*log(gam*lan*(ti^(gam-1)))-lam*(ti^gam))
attr(loglike, "gradient") <- -c(dldalp, dldbet)

loglike
}
# ---------------------------------------------------------------------------------------- #
loglikegompmpr <- function(param,surdat,k){
  .
  gam <- xialphas
  lan <- exp(xibetas)

dldalp <- t( deltai*ti + (lam/(gam^2))*(exp(gam*ti)-1) - (lam/gam)*exp(gam*ti)*ti )%*%xi1
  dldbet <- t( deltai - (lam/gam)*exp(gam*ti)-1 )%*%xi2

  loglike <- -sum(deltai*(log(lam*exp(gam*ti)))-(lam/gam)*(exp(gam*ti)-1))
  attr(loglike, "gradient") <- -c(dldalp, dldbet)

  loglike
}
# ---------------------------------------------------------------------------------------- #
```

```
APPENDIX B. MULTI-PARAMETER REGRESSION

loglikeloglistmpr <- function(param, surdat, k){
  gam <- exp(xialphas)
  lam <- exp(xibetas)
  dldalp <- t( deltai*(1+gam*log(ti)) - (deltai+1)*(lam*gam*(ti^gam)*log(ti))/(1+lam*(ti^gam)) )%*%xi1
  dldbet <- t( deltai - (deltai+1)*(lam*ti^gam)/(1+lam*(ti^gam)) )%*%xi2
  loglike <- -sum( deltai*log((lam*gam*(ti^(gam-1)))/(1+lam*(ti^gam))) - log(1+lam*(ti^gam)) )
  attr(loglike, "gradient") <- -c(dldalp, dldbet)
  loglike
}

logliketdlmpr <- function(param, surdat, k){
  gam <- xialphas
  lam <- xibetas
  zi <- exp(gam*ti+lam)
  wi <- exp(lam)
  dldalp <- t( deltai*ti - (deltai+(1/gam))*(ti*zi/(1+zi)) + (1/(gam^2))*log((1+zi)/(1+wi)) )%*%xi1
  dldbet <- t( deltai - (deltai+(1/gam))*(zi/(1+zi)) + (1/gam)*(wi/(1+wi)) )%*%xi2
  loglike <- -sum(deltai*log(zi/(1+zi))-(1/gam)*log((1+zi)/(1+wi)))
  attr(loglike, "gradient") <- -c(dldalp, dldbet)
  loglike
}

loglikeburrmpr <- function(param, surdat, k){
  p <- length(param)
  q <- dim(surdat)[2]
  k1 <- k[1]
  k2 <- k[2]
  taus <- param[1:k1]
  alphas <- param[(k1+1):(k1+k2)]
  betas <- param[(k1+k2+1):p]
  ti <- surdat[,1]
  deltai <- surdat[,2]
  # We also show the Burr case below in its entirety which, of course, is used within mprfit3 (not shown). Note that, here, k is a vector.
}
B.5. THE MPR CLASS

\[
\begin{align*}
\xi_1 & \leftarrow \text{as(surdat[,3:(k1+2)],"matrix")} \\
\xi_2 & \leftarrow \text{as(surdat[,,(k1+3):(k1+k2+2)],"matrix")} \\
\xi_3 & \leftarrow \text{as(surdat[,,(k1+k2+3):q],"matrix")} \\
\end{align*}
\]

\[
\begin{align*}
\xi_{\text{taus}} & \leftarrow \xi_1 \%\times \text{taus} \\
\xi_{\text{alphas}} & \leftarrow \xi_2 \%\times \text{alphas} \\
\xi_{\text{betas}} & \leftarrow \xi_3 \%\times \text{betas} \\
\rho & \leftarrow \exp(\xi_{\text{taus}}) \\
\gamma & \leftarrow \exp(\xi_{\text{alphas}}) \\
\lambda & \leftarrow \exp(\xi_{\text{betas}}) \\
\end{align*}
\]

\[
\begin{align*}
d\ell/d\tau & \leftarrow (1/\rho) \log(1+\lambda \rho (\tilde{t})^\gamma) - \frac{d\delta}{d\tau} \frac{1}{1+\lambda \rho (\tilde{t})^\gamma} \\
d\ell/d\alpha & \leftarrow \delta (1+\gamma \log(\tilde{t})) - \frac{d\delta}{d\tau} \frac{\lambda \gamma \rho (\tilde{t})^\gamma \log(\tilde{t})}{1+\lambda \rho (\tilde{t})^\gamma} \\
d\ell/d\beta & \leftarrow \delta - \frac{d\delta}{d\tau} \frac{\lambda \rho (\tilde{t})^\gamma}{1+\lambda \rho (\tilde{t})^\gamma} \\
\text{loglike} & \leftarrow -\sum \delta \log \left( \frac{\lambda \gamma (\tilde{t})^{(\gamma-1)}}{1+\lambda \rho (\tilde{t})^\gamma} \right) \\
\end{align*}
\]

\[
\begin{align*}
\text{loglike} & \leftarrow \text{c(d\ell/d\tau, d\ell/d\alpha, d\ell/d\beta)} \\
\end{align*}
\]

B.5.2 predict.mpr

Once we have fitted a particular MPR model using mprfit, we have an mpr object (as shown above). Naturally we would like to make predictions based on our fitted model. This can be done using predict.mpr which works just as other predict functions in R (e.g., predict.lm), i.e., we call it simply using “predict” - R knows that predict.mpr must be used when predict is applied to an mpr object.

Like the mprfit function, predict.mpr is in fact a wrapper function:

\[
\begin{align*}
\text{predict.mpr} & \leftarrow \text{function(object,newdata,ti,type="survivor",m=0,MLES=NA,tp=0.5,tstr=Inf)} \{ \\
\text{family} & \leftarrow \text{as.character(object$fit[[7]])} \\
\text{predmprnum} & \leftarrow \text{switch(family, Weibull=predmpr2, Gompertz=predmpr2, Loglogistic=predmpr2, TDL=predmpr2, Burr=predmpr3, )} \\
\text{predmprnum(object,newdata,ti,type,m,MLES,tp,tstr)} \\
\end{align*}
\]
The above function calls either `predmpr2` (shown below) or, in the case of a three component model (i.e., the Burr), `predmpr3` (not shown).

```r
# ---------------------------------------------------
# predmpr2 <- function(object,newdata,ti,type="survivor",n=0,MLES=NA,tp=0.5,tstr=Inf){
#      family <- as.character(object$fit[[7]])
#
#      mles.a <- object$mles.a
#      mles.b <- object$mles.b
#
#      if(m!=0 | is.list(MLES)) {
#         if(!is.list(MLES)){
#            MLES <- nsim2(object, m)
#         }
#         mles.a <- MLES$MLES.a
#         mles.b <- MLES$MLES.b
#      }
#
#      forms <- object$forms
#      forma <- forms[[1]]
#      formb <- forms[[2]]
#
#      xvars <- object$xvars
#      xlevels <- object$xlevels
#      xfac <- names(object$xlevels)
#      newnam <- colnames(newdata)
#
#      mvars <- match(xvars, newnam)
#      vna <- is.na(mvars)
#
#      if(any(vna)){
#         errmess <- paste("The following variables not found:",
#                            paste(xvars[vna], collapse=",")
#         stop(errmess)
#      }
#
#      nums <- match(setdiff(xvars,xfac), newnam)
#      facs <- match(xfac, newnam)
#
#      if(length(nums) > 0){
#         for(i in 1:length(nums)){
#            newdata[,nums[i]] <- as.numeric(newdata[,nums[i]])
#         }
#      }
#
#      if(length(facs) > 0){
#         for(i in 1:length(facs)){
#            newdata[,facs[i]] <- as.factor(newdata[,facs[i]])
#         }
#      }
# }
B.5. THE MPR CLASS

alevels <- xlevels[match(attr(terms(forma), "term.labels"), xfac)]
blevels <- xlevels[match(attr(terms(formb), "term.labels"), xfac)]

if(!is.null(alevels)){ alevels <- alevels[!is.na(names(alevels))] }
if(!is.null(blevels)){ blevels <- blevels[!is.na(names(blevels))] }

alpha.xis <- model.matrix(terms(forma), data=newdata, xlev=alevels)
beta.xis <- model.matrix(terms(formb), data=newdata, xlev=blevels)

alpha <- as.vector(alpha.xis%*%mles.a)
beta <- as.vector(beta.xis%*%mles.b)

parmat <- cbind(alpha, beta)
# list(alpha.xis, beta.xis, alpha, beta)

out <- switch(type,
  survivor = mprsurv(parmat,ti,family),
  hazard = mprhaz(parmat,ti,family),
  time = mprsim(1-tp,parmat,family),
  resmean = mprresmean(tstr,parmat,family),
)

out

# ---------------------------------------------------------------------------------------- #

Thus, predmpr2 uses the mpr object and a data.frame called newdata containing rows of individual covariate profiles at which predictions are to be made; of course, these individuals may be real or hypothetical.

The type of prediction must also be specified - there are four options, namely: survivor, hazard, time (percentile) and resmean (restricted mean - see Section 6.3.1). As can be seen from the last few lines of the above code, the type option is used to select between the functions mprsurv, mprhaz, mprsim (inverse survivor function) and mprresmean, respectively. These are shown below.

# ---------------------------------------------------------------------------------------- #

mprsurv <- function(parmat,ti,family="Weibull"){

  if(is.null(dim(parmat))){
    parmat <- matrix(parmat, nrow=1)
  }

  fansurv <- switch(family, Weibull=weibsurv, Gompertz=gomp surv, Loglogistic=loglogistsurv, TDL=tdlsurv, Burr=burrsurv, )

  fansurv(parmat,ti)
}
Clearly the first three of these functions (mprsurv, mprhaz and mprsim) are wrapper functions, which select the appropriate survivor, hazard or inverse survivor function for the given family, whereas, mprresmean integrates over the mprsurv function (recall from Section 6.3.1 that the restricted mean is calculated by integrating over the survivor function).

The code for the survivor, hazard and inverse survivor functions for each of the models considered in this thesis is shown below (based on Section 1.4).
# B.5. THE MPR CLASS

```r
def weibsurv(parmat, ti) {
  gam <- exp(parmat[,1])
  lam <- exp(parmat[,2])
  exp(-lam*(ti^gam))
}

def weibhaz(parmat, ti) {
  gam <- exp(parmat[,1])
  lam <- exp(parmat[,2])
  lam*gam*ti^(gam-1)
}

def weibsim(u, parmat) {
  gam <- exp(parmat[,1])
  lam <- exp(parmat[,2])
  (-log(u)/lam)^(1/gam)
}

#

def gompsurv(parmat, ti) {
  gam <- parmat[,1]
  lam <- exp(parmat[,2])
  ifelse(gam == 0, exp(-lam*ti), exp(-(lam/gam)*(exp(gam*ti)-1)))
}

def gomphaz(parmat, ti) {
  gam <- parmat[,1]
  lam <- exp(parmat[,2])
  lam*exp(gam*ti)
}

def gompsim(u, parmat) {
  gam <- parmat[,1]
  lam <- exp(parmat[,2])
  ifelse(gam == 0, -log(u)/lam, (1/gam)*log(-(gam/lam)*log(u)+1))
}

loglogistsurv(parmat, ti) {
  gam <- exp(parmat[,1])
  lam <- exp(parmat[,2])
  1/(1+lam*(ti^gam))
}
```

loglogisthaz <- Vectorize(function(parmat,ti){
gam <- exp(parmat[,1])
lam <- exp(parmat[,2])

(lam*gam*((ti^((gam-1))))/((1+lam*ti^gam))
}, vectorize.arg="ti")

loglogistsim <- function(u,parmat){
gam <- exp(parmat[,1])
lam <- exp(parmat[,2])

(((1/u)-1)/lam)^(1/gam)
}

# ---------------------------------------------------------------------------------------- #
tdslsurv <- Vectorize(function(parmat,ti){
gam <- parmat[,1]
lam <- parmat[,2]

ifelse(gam == 0, exp(-(-exp(lam)/(1+exp(lam)))*ti),

(((1+exp(gam*ti+lam))/(1+exp(lam))))^(-1/gam) )
}, vectorize.arg="ti")

tdlsim <- function(u, parmat){
gam <- parmat[,1]
lam <- parmat[,2]

ifelse(gam == 0, -log(u)/(-exp(lam)/(1+exp(lam))),

(1/gam)*log((u^(-gam))*(1+exp(lam))-1)-lam) )
}

# ---------------------------------------------------------------------------------------- #
burrsurv <- Vectorize(function(parmat,ti){
rho <- exp(parmat[,1])
gam <- exp(parmat[,2])
lam <- exp(parmat[,3])

((1+lam*rho*(ti^gam))"(-1/rho)
}, vectorize.arg="ti")

burrrhaz <- Vectorize(function(parmat,ti){
rho <- exp(parmat[,1])
gam <- exp(parmat[,2])
lam <- exp(parmat[,3])

(lam*gam*((ti^((gam-1))))/((1+lam*rho*ti^gam))

})
B.5. THE MPR CLASS

burrsim <- function(u, parmat) {
    rho <- exp(parmat[,1])
    gam <- exp(parmat[,2])
    lam <- exp(parmat[,3])
    (((u^(-rho))-1)/(lam*rho))^(1/gam)
}

Finally, we return to the predmpr2 function and point out that it makes use of a function called msim2. This is our implementation of m.l.e. simulation (Section 2.3.2) which uses the mvrnorm function from the MASS package - see below. Like mprfit3 and predmpr3, the msim3 function is not shown.

msim2 <- function(object, m = 1000) {
    mles.a <- object$mles.a
    mles.b <- object$mles.b
    alen <- length(mles.a); blen <- length(mles.b)
    hess <- object$hessian
    varcov <- solve(hess)
    MLES <- matrix(mvrnorm(m, c(mles.a, mles.b), varcov), nrow=m)
    MLES.a <- matrix(t(MLES[,1:alen]), nrow=alen)
    MLES.b <- matrix(t(MLES[-(1:alen)]), nrow=blen)
    list(MLES.a=MLES.a, MLES.b=MLES.b)
}

B.5.3 mprstep

In Section 5.4 we discussed MPR stepwise variable selection. Here we present the R code underpinning our selection procedure (mprstep). This is followed by Example B.1 which shows an application of mprstep to the lung cancer data.

At its most basic level, mprstep consists of repeated applications of the addvar and dropvar functions which, given a formula object, are used to add or remove variables.

addvar <- function(newvar, oldform) {
    newform <- update(oldform, paste("~ . +", newvar))
    newform
}

More specifically, the `add.terms` function (below) repeatedly calls `addvar`. Given a particular set of candidate covariates (`newterms`), this function fits all new models, i.e., the old model with the addition of one candidate covariate. Of course, as discussed in Section 5.4, we must specify whether the covariates are to be added to the shape (`whichpar="a"`), the scale (`whichpar="b"`) or to both simultaneously (`whichpar="both"`).

```r
add.terms <- function(newterms, forms, data, family="Weibull", init=NA, mles=NA, whichpar) {
    if(whichpar == "a"){
        forms <- list(addvar(newterms[[1]], forma), formb)
    } else if(whichpar == "b"){
        forms <- list(forma, addvar(newterms[[1]], formb))
    } else if(whichpar == "both"){
        forms <- list(addvar(newterms[[1]], forma), addvar(newterms[[1]], formb))
    }
    res <- mprfit(forms, data, family, init, hessian=F, iterlim=1000)
    results[i,1:3] <- c(whichpar, "+", newterms[[i]])
    results[i,4:10] <- res$fit
    results[i,11] <- as.character(forms[[1]][2])
    results[i,12] <- as.character(forms[[2]][2])
    mleslist[[i]] <- c(res$mles.a, res$mles.b)
}
```

Analogous to the `add.terms` function, we also have the `drop.terms` function (below) which calls `dropvar`:
B.5. THE MPR CLASS

#----------------------------------------------------------------------------------------#
drop.terms <- function(oldterms,forms,data,family="Weibull",init=NA,mles=NA,whichpar){
  forma <- forms[[1]]
  formb <- forms[[2]]

  results <- data.frame(matrix(NA, length(oldterms), 12))
  mleslist <- as.list(rep(NA,length(oldterms)))

  for(i in 1:length(oldterms)){
    if(whichpar == "a"){
      forms <- list(dropvar(oldterms[i], forma),formb)
    }else{
      if(whichpar == "b"){
        forms <- list(forma,dropvar(oldterms[i], formb))
      }else{
        if(whichpar == "both"){
          forms <- list(dropvar(oldterms[i], forma),dropvar(oldterms[i], formb))
        }
      }
    }
    res <- mprfit(forms,data,family,init,hessian=F,iterlim=1000)
    results[i,1:3] <- c(whichpar, "-", oldterms[i])
    results[i,4:10] <- res$fit
    results[i,11] <- as.character(forms[[1]])[2]
    results[i,12] <- as.character(forms[[2]])[2]
    mleslist[[i]] <- c(res$mles.a,res$mles.b)
  }
  list(results,mleslist)
}
#----------------------------------------------------------------------------------------#

Thus, the stepmpr function (below) uses add.terms and drop.terms to fit all possible new models in a given iteration of the while loop and, then, the best model (based on AIC or BIC) is chosen at this iteration. The while loop continues in this fashion until no improvement can be made and stopnow is set to TRUE.

#----------------------------------------------------------------------------------------#
stepmpr <- function(forms,formup,formlo=formula(~1),select=c(T,T,T),
direction="both",printout=T,bic=F,init=NA,mles=F,data,family="Weibull"){

  # First Step
  k <- 0
  forma <- forms[[1]]
  formb <- forms[[2]]
  names(forms) <- c("a", "b")

  res <- mprfit(forms,data,family,init,hessian=F,iterlim=1000)
  modcur <- res$fit
mlescur <- c(res$mles.a, res$mles.b)
mleslist <- list(mlescur)

stopnow <- FALSE

endresult <- data.frame(matrix(NA, 1, 12))
names(endresult) <- c("which", "pm", "var", "loglike", "aic", "bic", "npar", "stopcode", "iter", "family", "forma", "formb")
rownames(endresult) <- k
endresult[1,1:3] <- c("both", "+/-", ["none"])
endresult[1,4:10] <- modcur
endresult[1,11] <- as.character(forma)[2]
endresult[1,12] <- as.character(formb)[2]

if(mles==F){
  mles <- NA
}

while(! stopnow){
  k <- k + 1
  midresult <- data.frame(matrix(NA, 1, 12))
  midresult[1,1:3] <- c("both", "+/-", ["none"])
  midresult[1,4:10] <- modcur
  midresult[1,11] <- as.character(forma)[2]
  midresult[1,12] <- as.character(formb)[2]
  if(any(!is.na(mles))){
    mles <- mlescur
  }

  ### shape
  if(select[1] == TRUE){
    aterms <- list(add.scope(forma, formup), drop.scope(forma, formlo))
    if(length(aterms[[1]]) > 0 && (direction=="forward" || direction=="both")){
      resterms <- add.terms(aterms[[1]], forms, data=data, family=family, whichpar="a")
      results1 <- resterms[[1]]
      midresult <- rbind(midresult, results1)
      mleslist <- lrbind(mleslist, resterms[[2]])
    }
    if(length(aterms[[2]]) > 0 && (direction=="backward" || direction=="both")){
      resterms <- drop.terms(aterms[[2]], forms, data=data, family=family, mles=mles, whichpar="a")
      results1 <- resterms[[1]]
      midresult <- rbind(midresult, results1)
      mleslist <- lrbind(mleslist, resterms[[2]])
    }
  }
}
### scale
if(select[2] == TRUE){

    bterms <- list(add.scope(formb, formup), drop.scope(formb, formlo))

    if(length(bterms[[1]]) > 0 & (direction=="forward" || direction=="both")){
        resterms <- add.terms(bterms[[1]], forms, data=data, family=family, mles=mles, whichpar="b")
        results1 <- resterms[[1]]
        midresult <- rbind(midresult, results1)
        mleslist <- lrbind(mleslist, resterms[[2]])
    }

    if(length(bterms[[2]]) > 0 & (direction=="backward" || direction=="both")){
        resterms <- drop.terms(bterms[[2]], forms, data=data, family=family, mles=mles, whichpar="b")
        results1 <- resterms[[1]]
        midresult <- rbind(midresult, results1)
        mleslist <- lrbind(mleslist, resterms[[2]])
    }
}

### both simultaneously
if(select[3] == TRUE){

    abterms <- list(intersect(aterms[[1]], bterms[[1]]), intersect(aterms[[2]], bterms[[2]]))

    if(length(abterms[[1]]) > 0 & (direction=="forward" || direction=="both")){
        resterms <- add.terms(abterms[[1]], forms, data=data, family=family, mles=mles, whichpar="both")
        results1 <- resterms[[1]]
        midresult <- rbind(midresult, results1)
        mleslist <- lrbind(mleslist, resterms[[2]])
    }

    if(length(abterms[[2]]) > 0 & (direction=="backward" || direction=="both")){
        resterms <- drop.terms(abterms[[2]], forms, data=data, family=family, mles=mles, whichpar="both")
        results1 <- resterms[[1]]
        midresult <- rbind(midresult, results1)
        mleslist <- lrbind(mleslist, resterms[[2]])
    }
}

names(midresult) <- c("which", "pm", "var", "loglike", "aic", "npar", "stopcode", "iter", "family", "forma", "formb")

if(bic){
    ordering <- order(midresult[,6], decreasing=F)
} else{
    ordering <- order(midresult[,5], decreasing=F)
}

midresult <- midresult[ordering,]

mleslist <- mleslist[ordering][1]
mlescur <- mleslist[[1]]
APPENDIX B. MULTI-PARAMETER REGRESSION

```r
if(midresult[1,3]=="[none]"咥
   stopnow <- TRUE
else{
   modcur <- midresult[1,4:10]
   forms[[1]] <- forma <- as.formula(paste("~",midresult[1,11]))
   forms[[2]] <- formb <- as.formula(paste("~",midresult[1,12]))
   endresult1 <- midresult[1,]
   rownames(endresult1) <- k
   endresult <- rbind(endresult,endresult1)
}

if(printout){
   print(midresult[,1:10])
   cat("a: ",midresult[1,11],"\n")
   cat("b: ",midresult[1,12],"\n\n")
   cat("iter",k,"":midresult[1,1],midresult[1,2],midresult[1,3],"\n\n")
   cat("---------------------------------------------------------------","\n\n")
}

out <- list(endresult, mlescur)
# ---------------------------------------------------------------------------------------- #
```

Note that the `stepmpr` function uses `lrbind` - a function for combining two lists.

```r
lrbind <- function(x, y){
  lx <- length(x)
  ly <- length(y)
  z <- as.list(rep(NA,(lx+ly)))
  z[1:lx] <- x
  z[-(1:lx)] <- y
  z
}
```

# ---------------------------------------------------------------------------------------- # 

Example B.1. Log-Logistic MPR: Variable Selection Output

In Example 5.8 we gave the results of applying MPR variable selection (i.e., `mprstep`) to the lung cancer data. The procedure was carried out within a variety of different models and the log-logistic MPR model had the lowest AIC of all final models. Here we show the corresponding path of selection (starting from the null model) that led to the final log-logistic model. This serves to elucidate both the discussion of MPR variable selection given in Section 5.4 and the `mprstep` code given in this appendix.

The output below comes from the `mprstep` function but has been edited slightly for presentation. Note first that `which` specifies the component of
the model in which selection has taken place: \( a \) = shape, \( b \) = scale and \( \text{both} \) = simultaneous step. The next column, \( \text{pm} \), contains a plus or a minus indicating whether the algorithm is adding or removing a covariate; \( \text{var} \) gives the name of covariate in question. Finally, \( \text{loglike} \), \( \text{aic} \) and \( \text{npar} \) give the log-likelihood value, the \( AIC \) value and the number of parameters in the model respectively. At each iteration of the algorithm, all candidate models are ranked by their \( AIC \) value (in ascending order). The outcome of each iteration is given at the bottom of this table of candidate models, i.e., the model chosen at each step.

\[
\begin{array}{cccccc}
\text{which} & \text{pm} & \text{var} & \text{loglike} & \text{aic} & \text{npar} \\
\hline
b & + & \text{who} & -1923.637 & 3859.274 & 6 \\
\text{both} & + & \text{who} & -1921.571 & 3863.143 & 10 \\
\text{both} & + & \text{treat} & -1927.737 & 3875.474 & 10 \\
b & + & \text{treat} & -1943.658 & 3899.317 & 6 \\
\text{both} & + & \text{net} & -1985.630 & 3983.261 & 6 \\
b & + & \text{net} & -1988.262 & 3984.524 & 4 \\
b & + & \text{alb} & -1998.763 & 4005.525 & 4 \\
\text{both} & + & \text{alb} & -1997.663 & 4007.326 & 6 \\
a & + & \text{treat} & -1997.921 & 4007.842 & 6 \\
a & + & \text{who} & -2007.928 & 4027.855 & 6 \\
a & + & \text{net} & -2020.869 & 4049.738 & 4 \\
b & + & \text{cell} & -2026.063 & 4062.126 & 5 \\
\text{both} & + & \text{cell} & -2024.593 & 4065.186 & 8 \\
\text{both} & + & \text{sod} & -2027.288 & 4066.576 & 6 \\
b & + & \text{sod} & -2031.064 & 4070.128 & 4 \\
a & + & \text{alb} & -2032.572 & 4073.143 & 4 \\
a & + & \text{sod} & -2042.054 & 4092.107 & 4 \\
\text{both} & + & \text{agegrp} & -2037.227 & 4094.455 & 10 \\
b & + & \text{agegrp} & -2044.017 & 4100.035 & 6 \\
a & + & \text{cell} & -2045.442 & 4100.884 & 5 \\
a & + & \text{agegrp} & -2050.115 & 4112.231 & 6 \\
\text{both} & + & \text{agegrp} & -2055.274 & 4116.549 & 2 \\
b & + & \text{sex} & -2055.253 & 4116.506 & 3 \\
a & + & \text{sex} & -2055.264 & 4116.529 & 3 \\
b & + & \text{smoking} & -2053.434 & 4116.868 & 5 \\
a & + & \text{smoking} & -2053.764 & 4117.528 & 5 \\
\text{both} & + & \text{sex} & -2055.253 & 4118.506 & 4 \\
\text{both} & + & \text{smoking} & -2052.296 & 4120.591 & 8 \\
\end{array}
\]

\[a: 1\]
\[b: \text{who}\]

iter 1 : b + who
## APPENDIX B. MULTI-PARAMETER REGRESSION

<table>
<thead>
<tr>
<th>which pm var</th>
<th>loglike</th>
<th>sic</th>
<th>npar</th>
</tr>
</thead>
<tbody>
<tr>
<td>both + treat</td>
<td>-1865.559</td>
<td>3759.117</td>
<td>14</td>
</tr>
<tr>
<td>b + treat</td>
<td>-1879.945</td>
<td>3779.891</td>
<td>10</td>
</tr>
<tr>
<td>b + met</td>
<td>-1887.398</td>
<td>3790.796</td>
<td>8</td>
</tr>
<tr>
<td>both + met</td>
<td>-1885.486</td>
<td>3790.972</td>
<td>10</td>
</tr>
<tr>
<td>b + alb</td>
<td>-1900.172</td>
<td>3816.344</td>
<td>8</td>
</tr>
<tr>
<td>a + treat</td>
<td>-1898.684</td>
<td>3817.368</td>
<td>10</td>
</tr>
<tr>
<td>both + alb</td>
<td>-1899.317</td>
<td>3818.635</td>
<td>10</td>
</tr>
<tr>
<td>a + met</td>
<td>-1905.212</td>
<td>3826.425</td>
<td>8</td>
</tr>
<tr>
<td>both + sod</td>
<td>-1907.762</td>
<td>3835.526</td>
<td>10</td>
</tr>
<tr>
<td>b + cell</td>
<td>-1909.641</td>
<td>3837.282</td>
<td>9</td>
</tr>
<tr>
<td>b + sod</td>
<td>-1911.239</td>
<td>3838.477</td>
<td>8</td>
</tr>
<tr>
<td>both + cell</td>
<td>-1908.339</td>
<td>3840.679</td>
<td>12</td>
</tr>
<tr>
<td>a + alb</td>
<td>-1915.492</td>
<td>3846.984</td>
<td>8</td>
</tr>
<tr>
<td>a + sod</td>
<td>-1916.780</td>
<td>3849.559</td>
<td>8</td>
</tr>
<tr>
<td>a + cell</td>
<td>-1917.797</td>
<td>3853.595</td>
<td>9</td>
</tr>
<tr>
<td>a + agegrp</td>
<td>-1918.841</td>
<td>3857.682</td>
<td>10</td>
</tr>
<tr>
<td>both + agegrp</td>
<td>-1915.578</td>
<td>3859.156</td>
<td>14</td>
</tr>
</tbody>
</table>

### Iter 2: both + treat

<table>
<thead>
<tr>
<th>which pm var</th>
<th>loglike</th>
<th>sic</th>
<th>npar</th>
</tr>
</thead>
<tbody>
<tr>
<td>both +/- [none]</td>
<td>-1923.637</td>
<td>3860.462</td>
<td>9</td>
</tr>
<tr>
<td>a + smoking</td>
<td>-1921.231</td>
<td>3860.462</td>
<td>9</td>
</tr>
<tr>
<td>b + sex</td>
<td>-1923.320</td>
<td>3860.640</td>
<td>7</td>
</tr>
<tr>
<td>a + sex</td>
<td>-1923.622</td>
<td>3861.246</td>
<td>7</td>
</tr>
<tr>
<td>both + sex</td>
<td>-1923.188</td>
<td>3862.377</td>
<td>8</td>
</tr>
<tr>
<td>b + smoking</td>
<td>-1922.243</td>
<td>3862.486</td>
<td>9</td>
</tr>
<tr>
<td>a + who</td>
<td>-1921.571</td>
<td>3863.143</td>
<td>10</td>
</tr>
<tr>
<td>both + smoking</td>
<td>-1921.032</td>
<td>3866.065</td>
<td>12</td>
</tr>
<tr>
<td>b + agegrp</td>
<td>-1923.437</td>
<td>3866.875</td>
<td>10</td>
</tr>
<tr>
<td>b - who</td>
<td>-2055.274</td>
<td>4114.549</td>
<td>2</td>
</tr>
</tbody>
</table>

### A: treat

### B: who + treat
### B.5. THE MPR CLASS

<table>
<thead>
<tr>
<th>Model</th>
<th>Variable</th>
<th>Loglike</th>
<th>AIC</th>
<th>Npar</th>
</tr>
</thead>
<tbody>
<tr>
<td>a + sex</td>
<td>-1865.491</td>
<td>3760.982</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>a + smoking</td>
<td>-1863.681</td>
<td>3761.362</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>both + sex</td>
<td>-1865.278</td>
<td>3762.556</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>b + smoking</td>
<td>-1864.557</td>
<td>3763.115</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>a + who</td>
<td>-1863.867</td>
<td>3763.735</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>b + agegrp</td>
<td>-1864.427</td>
<td>3764.855</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>both + smoking</td>
<td>-1863.513</td>
<td>3767.026</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>a - treat</td>
<td>-1879.945</td>
<td>3779.891</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>b - treat</td>
<td>-1898.684</td>
<td>3817.368</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>both - treat</td>
<td>-1923.637</td>
<td>3859.274</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>b - who</td>
<td>-1927.737</td>
<td>3875.474</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

---

*a: treat + met*

*b: who + treat + met*

*iter 3: both + met*

---

<table>
<thead>
<tr>
<th>Model</th>
<th>Variable</th>
<th>Loglike</th>
<th>AIC</th>
<th>Npar</th>
</tr>
</thead>
<tbody>
<tr>
<td>b + alb</td>
<td>-1822.355</td>
<td>3684.711</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>both + alb</td>
<td>-1821.196</td>
<td>3686.392</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>both + sod</td>
<td>-1822.175</td>
<td>3688.349</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>b + sod</td>
<td>-1826.148</td>
<td>3692.295</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>b + cell</td>
<td>-1826.946</td>
<td>3695.892</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>both + cell</td>
<td>-1824.991</td>
<td>3697.982</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>a + alb</td>
<td>-1832.208</td>
<td>3704.417</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>a + cell</td>
<td>-1831.219</td>
<td>3704.438</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>a + sod</td>
<td>-1832.223</td>
<td>3704.446</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>a + agegrp</td>
<td>-1831.626</td>
<td>3707.252</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>both +/- [none]</td>
<td>-1836.164</td>
<td>3708.328</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>b + sex</td>
<td>-1835.808</td>
<td>3709.617</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>both + agegrp</td>
<td>-1828.855</td>
<td>3709.710</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>a + sex</td>
<td>-1836.102</td>
<td>3710.203</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>a + smoking</td>
<td>-1834.431</td>
<td>3710.861</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>both + sex</td>
<td>-1835.768</td>
<td>3711.535</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>b + smoking</td>
<td>-1835.382</td>
<td>3712.764</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>a - met</td>
<td>-1840.422</td>
<td>3712.844</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>a + who</td>
<td>-1834.520</td>
<td>3713.040</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>b + agegrp</td>
<td>-1835.898</td>
<td>3715.796</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>both + smoking</td>
<td>-1834.404</td>
<td>3716.807</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>a - treat</td>
<td>-1850.630</td>
<td>3729.259</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>b - met</td>
<td>-1858.798</td>
<td>3749.595</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>both - met</td>
<td>-1865.559</td>
<td>3759.117</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>b - treat</td>
<td>-1868.549</td>
<td>3765.098</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>both - treat</td>
<td>-1885.486</td>
<td>3790.972</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>b - who</td>
<td>-1891.373</td>
<td>3810.746</td>
<td>14</td>
<td></td>
</tr>
</tbody>
</table>

---

*a: treat + met*

*b: who + treat + met + alb*

*iter 4: b + alb*
### Appendix B. Multi-Parameter Regression

<table>
<thead>
<tr>
<th>which pm</th>
<th>var</th>
<th>loglike</th>
<th>sic</th>
<th>npar</th>
</tr>
</thead>
<tbody>
<tr>
<td>b + cell</td>
<td>-1811.970</td>
<td>3669.940</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>both + cell</td>
<td>-1809.960</td>
<td>3671.920</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>both + sod</td>
<td>-1812.174</td>
<td>3672.349</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>b + sod</td>
<td>-1815.662</td>
<td>3675.324</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>a + cell</td>
<td>-1816.033</td>
<td>3678.066</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>a + sod</td>
<td>-1819.592</td>
<td>3683.184</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>a + agegrp</td>
<td>-1817.916</td>
<td>3683.832</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>both +/- [none]</td>
<td>-1822.355</td>
<td>3684.711</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>both + agegrp</td>
<td>-1814.938</td>
<td>3685.877</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>a + smoking</td>
<td>-1820.047</td>
<td>3686.094</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>a + alb</td>
<td>-1821.196</td>
<td>3686.392</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>b + sex</td>
<td>-1822.234</td>
<td>3686.469</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>a + sex</td>
<td>-1822.336</td>
<td>3686.672</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>both + sex</td>
<td>-1821.977</td>
<td>3687.954</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>b + smoking</td>
<td>-1820.979</td>
<td>3687.958</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>a - met</td>
<td>-1826.019</td>
<td>3688.039</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>a + who</td>
<td>-1820.822</td>
<td>3689.643</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>both + smoking</td>
<td>-1819.928</td>
<td>3691.855</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>b + agegrp</td>
<td>-1822.021</td>
<td>3692.043</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>a - treat</td>
<td>-1836.281</td>
<td>3704.562</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>b - alb</td>
<td>-1836.164</td>
<td>3708.328</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>b - met</td>
<td>-1843.888</td>
<td>3723.775</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>b - treat</td>
<td>-1850.260</td>
<td>3732.519</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>both - met</td>
<td>-1851.238</td>
<td>3734.476</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>both - treat</td>
<td>-1863.914</td>
<td>3751.829</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>b - who</td>
<td>-1868.918</td>
<td>3769.835</td>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>

---

**a:** treat + met  
**b:** who * treat + met + alb + cell  
iter 5 : b + cell

---

<table>
<thead>
<tr>
<th>which pm</th>
<th>var</th>
<th>loglike</th>
<th>sic</th>
<th>npar</th>
</tr>
</thead>
<tbody>
<tr>
<td>both + sod</td>
<td>-1802.835</td>
<td>3659.670</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>b + sod</td>
<td>-1805.969</td>
<td>3661.939</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>a + sod</td>
<td>-1809.358</td>
<td>3668.716</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>a + agegrp</td>
<td>-1807.705</td>
<td>3669.409</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>both + agegrp</td>
<td>-1803.729</td>
<td>3669.458</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>both +/- [none]</td>
<td>-1811.970</td>
<td>3669.940</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>b + sex</td>
<td>-1811.750</td>
<td>3671.500</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>a - met</td>
<td>-1814.801</td>
<td>3671.602</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>a + smoking</td>
<td>-1809.894</td>
<td>3671.788</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>a + cell</td>
<td>-1809.960</td>
<td>3671.920</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>a + sex</td>
<td>-1811.968</td>
<td>3671.935</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>a + alb</td>
<td>-1811.214</td>
<td>3672.428</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>both + sex</td>
<td>-1811.507</td>
<td>3673.013</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>b + smoking</td>
<td>-1811.069</td>
<td>3674.138</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>a + who</td>
<td>-1810.371</td>
<td>3674.741</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>b + agegrp</td>
<td>-1811.663</td>
<td>3677.326</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>both + smoking</td>
<td>-1809.858</td>
<td>3677.717</td>
<td>29</td>
<td></td>
</tr>
</tbody>
</table>
B.5. THE MPR CLASS

Inspection of the above output provides us with the extra insight of the order in which covariates were selected - this indicates their relative importance. Indeed we see agreement with the order of importance as determined previously in Example 5.7 (which was based on Table 5.9).
B.5.4 Piecewise Exponential

In this appendix we provide our implementation of the piecewise exponential MPR model (Section 5.2.6) which, as mentioned previously, has not been incorporated into the `mprfit` function.

Recall from Section 5.2.6 that the piecewise exponential model is defined using the vectors \( a(t) \) and \( d(t) \), respectively, which indicate the time-interval that \( t \) lies in and the time spent in each of these intervals. Furthermore, the likelihood function (given in Appendix B.2.6) requires \( a(t_i) \) and \( d(t_i) \) for each individual, \( i = 1, \ldots, n \). It is useful to store all of these vectors using matrices (\( A \) and \( D \) say) calculated using the `Afn` and `Dfn` functions:

### `Afn` function

```r
# Afn <- function(ti, tcut){
  nint <- length(tcut)-1
  A <- matrix(NA, length(ti), nint)
  for(i in 1:nint){
    A[,i] <- ifelse(ti >= tcut[i] & ti < tcut[i+1], 1, 0)
  }
  A
}
```

### `Dfn` function

```r
# Dfn <- function(ti, tcut, A){
  nint <- length(tcut)-1
  d <- diff(tcut)
  lenA <- dim(A)[1]
  D <- A
  D[D == 0] <- NA
  D <= D*t[i]
  D <- t(t(D) - tcut[-(nint+1)])
  fin <- rep(F, lenA)
  for(i in 1:nint){
    fin <- fin | (!is.na(D[i,i]))
    D[fin,i] <- d[i]
  }
  D[is.na(D)] <- 0
  D
}
```
The A and D matrices are used in the log-likelihood function below (compare with that of Appendix B.2.6).

```r
# ---------------------------------------------------------------------------------------- #
loglike <- function(param, surdat, A, D, nint, k.beta, nnon = 1) {
  p <- length(param)
  q <- dim(surdat)[2]
  nonint <- nnon * nint

  betamat <- matrix(NA, k.beta, nint)
  betamat[1:nonint,] <- matrix(param[1:nonint], nnon, nint)
  betamat[-(1:nonint),] <- param[-(1:nonint)]

  ti <- surdat[,1]
  deltai <- surdat[,2]
  xinon <- as(surdat[,3:(3+nnon-1)], "matrix")
  xiph <- as(surdat[-c(1,2,3:(3+nnon-1))], "matrix")
  xi <- cbind(xinon, xiph)

  xibetas <- xi %*% betamat
  lam <- exp(xibetas)

  dldalp <- (deltai*A)-(D*lam)
  dldnalp <- t(xi) %*% (dldalp)
  dldnon <- as.vector(dldalp[1:nonint,])
  if(nint > 1) {
    dldph <- ((dldalp[-(1:nonint),]) %*% rep(1, nint))
  } else {
    dldph <- dldalp[-(1:nonint),]
  }
  dldalp <- c(dldnon, dldph)

  loglike <- -sum(((deltai*A*xibetas) - (D*lam)) %*% rep(1, nint))
  attr(loglike, "gradient") <- -c(dldalp)
  loglike
}
# ---------------------------------------------------------------------------------------- #
```

In Example 5.3 we considered a single factor (treatment) piecewise exponential MPR model with four time intervals (approximately equal number of events in each interval). The code required to fit this model is given below.

```r
# ---------------------------------------------------------------------------------------- #
nint <- 4
enti <- LungC$surtim[LungC$cen==1]
ne <- length(enti)

tcut <- enti[seq(1, ne, by = ceiling(ne/nint))]
tcut[1] <- 0
```
APPENDIX B. MULTI-PARAMETER REGRESSION

A <- Afn(LungC$surtim, tcut)
D <- Dfn(LungC$surtim, tcut, A)
formnon <- formula(~treat)  # Note: formnon and formph cannot
formph <- formula(~1)  # have covariates in common.
beta.non <- model.matrix(formnon, data=LungC)
beta.ph <- model.matrix(formph, data=LungC)[,-1]
beta.xis <- cbind(beta.non, beta.ph)
surdat <- cbind(LungC$surtim, LungC$cen, beta.xis)
k.beta <- dim(beta.xis)[2]
nnon <- dim(beta.non)[2]
nonint <- nnon*nint
init <- rep(0.01, nonint + (k.beta - nnon))
loglikpwe.res <- nlm(loglikepwempr, init, surdat, A, D, nint, k.beta, nnon)

B.5.5 simfunc

In Chapter 3 we considered simulation by the method of inversion (of the survivor function). Recall that MacKenzie’s (1994) “J-function” is used to select an appropriate $\phi$ value, $C \sim \text{Exp}(\phi)$, ensuring that the censored proportion is equal to a pre-specified value, $p_{\text{fix}}$. In Section 3.5 we generalised this function to handle cure rate data where $p_{\text{fix}}$ is the censoring level for non-cured individuals (as cured individuals are censored by definition). Our generalised J-function, called the $J^*$-function in Section 3.5, is as follows:

\[
Jfunc <- function(phi, pcen, parvec, family="Weibull", tstr=Inf){
  func1<-function(t){
    mprsurv(parvec,t,family)*phi*(exp(-phi*t))
  }
  pstr <- mprsurv(parvec,tstr,family)
  pphi <- integrate(func1,lower=0,upper=tstr, stop.on.error=F)$value
  pphistr <- (pphi - pstr)/(1-pstr)
  (pcen-pphistr)^2
}
\]

The $J^*$-function must be minimised with respect to $\phi$ for a given survival distribution (i.e., family) and values of $pcen = p_{\text{fix}}$ and $parvec = \theta = (\log \gamma, \log \lambda)^T$ respectively. As $\phi$ is a scalar, the optimize function can be used:
Note that if \( \theta \) depends on covariates, we have a matrix of \( \theta \) vectors - one for each individual. In this case \( Jfunc \) must be minimised \( n \) times to produce a vector of \( \phi \) values.

Once the appropriate \( \phi \) value/vector has been calculated, as shown above, \( simfunc \) is used to simulate survival data with the correct censored proportion.

```r
# ---------------------------------------------------------------------------------------- #
phirange <- c(0, 50)
phi <- optimize(Jfunc,phirange,pcen,parvec,family,maximum=F)$minimum
# ---------------------------------------------------------------------------------------- #

# while loop ensures that number of individuals censored AND not cured
# is equal to "ncennotcure"
while ( sum(!cure) - sum(cen) != ncennotcure ){
  u1 <- runif(n)
  u2 <- runif(n)
  st <- mprsim(u1,parmat,family)
  ct <- weibsim(u2,cbind(0,log(phi)))
  cure <- is.na(st)
  st[cure] <- Inf
ti <- ifelse(st<ct, st, ct)
  cen <- ifelse(st<ct, 1, 0)
}
```

```r
# if "pcen" = 0.

u1 <- runif(n)
ct <- rep(Inf,n)
```
APPENDIX B. MULTI-PARAMETER REGRESSION

\[ \text{tbig} \leftarrow \text{mprsim} \left( \text{mprsurv(parmat,Inf,family)} + 1e-15, \text{parmat, family} \right) \]

\[ \text{ticure} \leftarrow \text{runif}(\text{sum(cure)}, 0, \text{tbig}) \]
\[ \text{ct[cure]} \leftarrow \text{ticure} \]

\[ \text{ti} \leftarrow \text{ifelse(st<ct, st, ct)} \]
\[ \text{cen} \leftarrow \text{ifelse(st<ct, 1, 0)} \]

\} \]

\} \]

\} \]

\} \]

# ---------------------------------------------------------------------------------------- #
Appendix C

Orthogonal Weibull

C.1 Introduction

In Section 5.4 we discussed the implications of orthogonality and noted that there exists an orthogonal parametrisation of the Weibull distribution. In this appendix we will derive the orthogonal Weibull distribution and show its extension to multi-parameter regression form which we then compare to the standard Weibull MPR model (Section 5.2.1).

Firstly, note the definition of orthogonality which is as follows. Upon partitioning the full parameter vector, $\theta$, into two sub-vectors, $\theta_1$ and $\theta_2$, of length $p_1$ and $p_2$ respectively, where $p_1 + p_2 = p = \dim(\theta)$, then we have that $\theta_1 \perp \theta_2$ if

$$E\left(-\frac{\partial^2 \ell}{\partial \theta_r \partial \theta_s}\right) = 0,$$

(C.1)

for $r = 1, \ldots, p_1$ and $s = p_1 + 1, \ldots, p$. Orthogonality guarantees that the maximum likelihood estimates, $\hat{\theta}_1$ and $\hat{\theta}_2$, are asymptotically independent (see Cox & Reid (1987)).

C.2 Gamma and Related Functions

In the Appendix C.3 we will calculate the expected information matrix (uncensored case) for the standard Weibull distribution; in doing so, integrals
related to the gamma function arise. Thus, we briefly review some properties of this function which will be necessary to make progress (for further information see Abramowitz & Stegun (1964, chap. 6)).

The *gamma* function is defined as

\[
\Gamma(x) = \int_0^\infty t^{x-1} \exp(-t) \, dt, \tag{C.2}
\]

where \(\Gamma(x) = (x-1)!\) when \(x\) is a natural number. Furthermore, the *digamma* function is given by

\[
\Psi(x) = \frac{d}{dx} \log \Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}, \tag{C.3}
\]

and the *trigamma* function is

\[
\Psi'(x) = \frac{d^2}{dx^2} \log \Gamma(x). \tag{C.4}
\]

The properties of the above three functions are well-known, thus enabling us to evaluate integrals which are functionally related to them. We will encounter integrals of the form

\[
\Gamma'(x) = \int_0^\infty t^{x-1} \exp(-t) \log t \, dt,
\]

and

\[
\Gamma''(x) = \int_0^\infty t^{x-1} \exp(-t)(\log t)^2 \, dt,
\]

respectively, which, using (C.3) and the product rule of differentiation, can be written as

\[
\Gamma'(x) = \Psi(x)\Gamma(x) \quad \Gamma''(x) = \Psi'(x)\Gamma(x) + \Psi(x)\Gamma'(x) \\
= \Psi'(x)\Gamma(x) + [\Psi(x)]^2\Gamma(x) \\
= \{\Psi'(x) + [\Psi(x)]^2\}\Gamma(x).
\]

Standard theory of the digamma and trigamma functions give the special values \(\Psi(2) = 1 - \gamma_E\), where \(\gamma_E \approx 0.57721\) is known as *Euler’s constant*, and
C.3. STANDARD WEIBULL EXPECTED INFORMATION

\( \Psi'(2) = \pi^2/6 - 1 \). We require these values as we will need to evaluate \( \Gamma'(2) \) and \( \Gamma''(2) \) in the next section. Thus, using the above, we arrive at

\[
\Gamma'(2) = \Psi(2)\Gamma(2)
\]

\[
= (1 - \gamma_E)(1!)
\]

\[
\Gamma''(2) = \left\{ \Psi'(2) + [\Psi(2)]^2 \right\} \Gamma(2)
\]

\[
= \left[ \frac{\pi^2}{6} - 1 + (1 - \gamma_E)^2 \right] (1!)
\]

\[
= 1 - \gamma_E
\]

\[
= \frac{\pi^2}{6} - 1 + 2\gamma_E + \gamma_E^2
\]

\[
= \frac{\pi^2}{6} - 2\gamma_E + \gamma_E^2.
\]

C.3 Standard Weibull Expected Information

In Appendix C.4, we will derive the orthogonal parametrisation of the Weibull distribution. This derivation requires the computation of the expected information matrix, \( \mathcal{I}(\theta) \), which we will carry out in this section. We will consider only the case of full information (i.e., the uncensored case) mirroring the work of Cox & Reid (1987); recall from Section 2.3.1 that the functional form of \( \mathcal{I}(\theta) \) is unknown in general (due to censoring).

For the standard Weibull distribution (Section 1.4.2), the hazard and survivor functions are given by

\[
\lambda(t) = \lambda \gamma t^{\gamma - 1} \quad S(t) = \exp(-\lambda t^\gamma).
\]

Thus the log-likelihood function, in the uncensored case, is given by

\[
\ell(\theta) = \sum_{i=1}^{n} \log \lambda + \log \gamma + (\gamma - 1) \log t_i - \lambda t_i^\gamma,
\]

where \( \theta = (\lambda, \gamma) \in \mathbb{R}_+^2 \). The first partial derivatives are therefore given by

\[
\frac{\partial \ell}{\partial \lambda} = \sum_{i=1}^{n} \frac{1}{\lambda} - t_i^\gamma
\]

\[
\frac{\partial \ell}{\partial \gamma} = \sum_{i=1}^{n} \frac{1}{\gamma} + \log t_i - \lambda t_i^\gamma \log t_i.
\]
The second partial derivatives are then given by
\[
\frac{\partial^2 \ell}{\partial \lambda^2} = \sum_{i=1}^{n} -\frac{1}{\lambda^2}
\]
\[
\frac{\partial^2 \ell}{\partial \gamma^2} = \sum_{i=1}^{n} -\frac{1}{\gamma^2} - \lambda t_i^\gamma (\log t_i)^2
\]
\[
\frac{\partial^2 \ell}{\partial \lambda \partial \gamma} = \sum_{i=1}^{n} -t_i^\gamma \log t_i,
\]
and these are the elements of the Hessian matrix, \( H(\theta) \), i.e., \( h_{11} = \partial^2 \ell / \partial \lambda^2 \), \( h_{22} = \partial^2 \ell / \partial \gamma^2 \), and \( h_{12} = h_{21} = \partial^2 \ell / \partial \lambda \partial \gamma \) respectively.

The observed information matrix is given by \( I_o(\theta) = -H(\theta) \) and, hence, the expected information matrix is
\[
I(\theta) = E[I_o(\theta)] = n \times \begin{pmatrix}
\frac{1}{\lambda^2} & E[t^\gamma \log t] \\
E[t^\gamma \log t] & \frac{1}{\gamma^2} + \lambda E[t^\gamma (\log t)^2]
\end{pmatrix}
\]
where expectation is with respect to the Weibull distribution. Therefore we must evaluate \( E[t^\gamma \log t] \) and \( E[t^\gamma (\log t)^2] \) respectively. Firstly,
\[
E[t^\gamma \log t] = \int_{0}^{\infty} (t^\gamma \log t) \lambda \gamma t^{\gamma-1} \exp(-\lambda t^\gamma) \, dt
\]
\[
= \lambda \gamma \int_{0}^{\infty} t^{2\gamma-1} \exp(-\lambda t^\gamma) \log t \, dt,
\]
which, after making the substitution \( u = \lambda t^\gamma \), becomes
\[
E[t^\gamma \log t] = \frac{1}{\lambda \gamma} \int_{0}^{\infty} u \exp(-u) \log \left( \frac{u}{\lambda} \right) \, du
\]
\[
= \frac{1}{\lambda \gamma} \int_{0}^{\infty} u \exp(-u)(\log u - \log \lambda) \, du
\]
\[
= \frac{1}{\lambda \gamma} \left( \int_{0}^{\infty} u \exp(-u) \log u \, du - \log \lambda \int_{0}^{\infty} \exp(-u) \, du \right)
\]
\[
= \frac{1}{\lambda \gamma} \left[ \Gamma'(2) - \Gamma(2) \log \lambda \right]
\]
\[
= \frac{1}{\lambda \gamma} \left[ \Gamma'(2) - \log \lambda \right], \quad \text{(C.5)}
\]
and, similarly,

\[ E[t^\gamma (\log t)^2] = \int_0^\infty [t^\gamma (\log t)^2] \lambda \gamma t^{\gamma-1} \exp(-\lambda t^\gamma) \, dt \]

\[ = \lambda \gamma \int_0^\infty t^{2\gamma-1} \exp(-\lambda t^\gamma) t^2 \, dt \]

\[ = \frac{1}{\lambda \gamma^2} \int_0^\infty u \exp(-u)(\log u - \log \lambda)^2 \, du \]

\[ = \frac{1}{\lambda \gamma^2} \int_0^\infty u \exp(-u)[(\log u^2) - 2 \log u \log \lambda + (\log \lambda)^2] \, du \]

\[ = \frac{1}{\lambda \gamma^2} [\Gamma''(2) - 2\Gamma'(2) \log \lambda + (\log \lambda)^2], \tag{C.6} \]

where, from Appendix C.2, we know the values of \( \Gamma'(2) \) and \( \Gamma''(2) \) respectively. Thus, the expected information is

\[ I(\theta) = n \times \begin{pmatrix} \frac{1}{\lambda} & \frac{1}{\lambda \gamma}[\Gamma'(2) - \log \lambda] \\ \frac{1}{\lambda \gamma}[\Gamma'(2) - \log \lambda] & \frac{1}{\lambda}[1 + \Gamma''(2) - 2\Gamma'(2) \log \lambda + (\log \lambda)^2] \end{pmatrix}. \]

### C.4 Derivation of the Orthogonal Weibull

While it is not possible to construct orthogonal parameters in general, Cox & Reid (1987) provide a set of differential equations whose solution yields a model where one parameter is orthogonal to the remaining parameters. Thus, the method can be used to transform the parameters of the Weibull model from \((\lambda, \gamma)\) to \((\lambda^*, g(\lambda, \gamma), \gamma)\), for some function \(g(\cdot)\), such that \(\lambda^* \perp \gamma\). In this particular case, Cox & Reid’s orthogonality equation is given by

\[ i \lambda \frac{\partial \lambda}{\partial \gamma} = -i \lambda \gamma \]

\[ \frac{n}{\lambda^2} \frac{\partial \lambda}{\partial \gamma} = \frac{n}{\lambda \gamma} [\log \lambda - \Gamma'(2)] \]

\[ \gamma \frac{\partial \lambda}{\lambda} = \log \lambda - \Gamma'(2), \tag{C.7} \]
which can be solved by making the substitution
\[ u = \log \lambda - \Gamma'(2) \]
\[ \Rightarrow \frac{\partial u}{\partial \gamma} = \frac{1}{\lambda} \frac{\partial \lambda}{\partial \gamma}. \]

Therefore, (C.7) becomes
\[ \gamma \frac{\partial u}{\partial \gamma} = u \]
\[ \Rightarrow u = c(\lambda^*) \gamma, \]
where \( c(\lambda^*) \) is an arbitrary function of \( \lambda^* \). As \( u = \log \lambda - \Gamma'(2) \) we have that
\[ \log \lambda - \Gamma'(2) = c(\lambda^*) \gamma \]
\[ \log \lambda = c(\lambda^*) \gamma + \Gamma'(2) \]
\[ \lambda = \exp[\Gamma'(2)] \exp[\lambda^* \gamma] \]
\[ \lambda = \exp[\Gamma'(2)] \{\exp[c(\lambda^*)]\} \gamma \]
\[ \lambda = \exp[\Gamma'(2)](\lambda^*)^\gamma, \quad (C.8) \]

where the last line follows by taking \( c(\lambda^*) = \log(\lambda^*) \) for mathematical convenience (but the solution is non-unique as it holds for any \( c(\lambda^*) \)).

Thus, the hazard and survivor functions for the orthogonal Weibull are given by
\[ \lambda(t) = \exp[\Gamma'(2)](\lambda^*)^\gamma^* t^{\gamma^* - 1} \quad S(t) = \exp\{-\exp[\Gamma'(2)](\lambda^* t)^\gamma\}, \]
where we have now superscripted both parameters to distinguish the model from the standard Weibull; clearly both are equivalent as \( \lambda = \exp[\Gamma'(2)](\lambda^*)^\gamma \) and \( \gamma = \gamma^* \). The log-likelihood function is
\[ \ell(\theta) = \sum_{i=1}^{n} \delta_i [\Gamma'(2) + \gamma^* \log \lambda^* + \log \gamma^* + (\gamma^* - 1) \log t_i] - \exp[\Gamma'(2)](\lambda^* t_i)^\gamma^*, \]
where \( \theta = (\lambda^*, \gamma^*) \in \mathbb{R}_+^2 \). It is easy to show that the mixed term in the expected information matrix, \( i_{\lambda^* \gamma^*} \), is now equal to zero and, hence, \( \lambda^* \perp \gamma^* \). However, we will not reproduce the calculation of \( i_{\lambda^* \gamma^*} \) here, as it amounts to evaluating integrals essentially identical to those of Appendix C.3.
C.5 Orthogonal Weibull MPR Model

We now extend the orthogonal Weibull to multi-parameter regression form by setting $\lambda^* = \exp(x^T \beta^*)$ and $\gamma^* = \exp(z^T \alpha^*)$, respectively, where $x = (1, x_1, \ldots, x_p)^T$, $z = (1, z_1, \ldots, z_q)^T$, $\beta^* = (\beta_0^*, \beta_1^*, \ldots, \beta_p^*)^T$ and $\alpha^* = (\alpha_0^*, \alpha_1^*, \ldots, \alpha_q^*)^T$ are the scale and shape covariate vectors and corresponding regression coefficients. Note that the scale and shape regression components are orthogonal in this MPR model (i.e., $\lambda^* \perp \gamma^* \Rightarrow \beta^* \perp \alpha^*$) which may be advantageous from an inferential point of view as the components can be considered separately (see Section 5.4 for further details); this only holds true when there is no censoring however.

As in Section 5.2 and Appendix B.3, we let $c$ be a binary covariate common to both the scale and shape regressions. Furthermore we assume that $x_1 = z_1 = c$ and write $x^T \beta^* = c\beta_1^* + \tilde{x}^T \beta^*$ and $z^T \alpha^* = \alpha_1^* + \tilde{z}^T \alpha^*$. Thus, the hazard function is

$$
\lambda(t \mid x, z) = \exp[\Gamma'(2)] [\exp(x^T \beta^*)]^{\exp(z^T \alpha^*)} \exp(z^T \alpha^*) t^{\exp(z^T \alpha^*) - 1} \\
= \exp[\Gamma'(2)] [\exp(c\beta_1^* + \tilde{x}^T \beta^*)]^{\exp(\alpha_1^* + \tilde{z}^T \alpha^*)} \exp(\alpha_1^* + \tilde{z}^T \alpha^*) t^{\exp(\alpha_1^* + \tilde{z}^T \alpha^*) - 1},
$$

and, the hazard ratio for $c$, at time $t$, is

$$
\psi(t) = \frac{\lambda(t \mid c = 1)}{\lambda(t \mid c = 0)} \\
= \frac{\exp[\Gamma'(2)] [\exp(c\beta_1^* + \tilde{x}^T \beta^*)]^{\exp(\alpha_1^* + \tilde{z}^T \alpha^*)} \exp(\alpha_1^* + \tilde{z}^T \alpha^*) t^{\exp(\alpha_1^* + \tilde{z}^T \alpha^*) - 1}}{\exp[\Gamma'(2)] [\exp(\tilde{x}^T \beta^*)]^{\exp(\tilde{z}^T \alpha^*)} \exp(\tilde{z}^T \alpha^*) t^{\exp(\tilde{z}^T \alpha^*) - 1}} \\
= \frac{[\exp(c\beta_1^* + \tilde{x}^T \beta^*)]^{\exp(\alpha_1^* \exp(\tilde{z}^T \alpha^*) \exp(\alpha_1^* \exp(\tilde{z}^T \alpha^*) - 1)}}{[\exp(\tilde{x}^T \beta^*)]^{\exp(\tilde{z}^T \alpha^*) \exp(\alpha_1^*) \exp(\tilde{z}^T \alpha^*) - 1)}} \\
= \left\{ \frac{\exp(c\beta_1^* + \tilde{x}^T \beta^*) \exp(\alpha_1^*)}{\exp(\tilde{x}^T \beta^*)} \right\}^{\exp(\tilde{z}^T \alpha^*)} \exp(\alpha_1^* \exp(\tilde{z}^T \alpha^*) - 1)}.
$$

(C.9)

Recall that in the standard Weibull MPR model, the hazard ratio is $\psi(t) = \exp(\beta_1) \exp(\alpha_1) \exp(\tilde{z}^T \alpha) \exp(\alpha_1 - 1)$ where $\alpha_1 = 0$ implies that the effect of $c$ is PH, i.e., the standard Weibull MPR model provides a test of proportionality.
It is clear from (C.9) that, in the orthogonal Weibull MPR model, the effect of \(c\) does not become proportional when \(\alpha_1^* = 0\).

Clearly the parameters in the standard and orthogonal MPR models have different interpretations. However, what may not be obvious at first sight is the fact that we now have entirely different models. Even though the two parametrisations are equivalent when there are no covariates (Appendix C.4), the generalisation to multi-parameter regression leads to models which are no longer equivalent. It is easy to see this when we look at both models in terms of the original \(\lambda\) parameter. In the standard Weibull MPR model we have

\[
\log(\lambda) = x^T \beta,
\]

which is linear in the covariates but, as \(\lambda = \exp[\Gamma'(2)](\lambda^*)^{\gamma^*}\), the orthogonal MPR model has

\[
\log(\lambda) = \Gamma'(2) + x^T \beta^* \exp(z^T \alpha^*),
\]

which is a non-linear regression model for the original scale parameter, \(\lambda\). In order for these two MPR models to be equivalent we require that the system of equations given by

\[
x^T \beta = \Gamma'(2) + x^T \beta^* \exp(z^T \alpha^*),
\]

(C.10)
can be solved for all \(\beta_r, r = 0, \ldots, p\), at all possible values of \(x\) and \(z\), i.e., each \(\beta\) coefficient must be expressible in terms of \(\beta^*\) and \(\alpha^*\) coefficients. This will not be true in general. However, two special cases where this occurs are: (i) when the shape is a constant and (ii) the single-factor model, i.e., a model containing a single binary covariate (or a categorical covariate coded as a binary design matrix) which is common to both the scale and shape.

In the case where the shape is constant (i.e., no covariates) (C.10) becomes

\[
\beta_0 + x_1 \beta_1 + \ldots + x_p \beta_p = \Gamma'(2) + \beta_0^* \exp(\alpha_0^*) + x_1 \beta_1^* \exp(\alpha_0^*) + \ldots + x_p \beta_p^* \exp(\alpha_0^*),
\]

and, trivially, we find that \(\beta_0 = \Gamma'(2) + \beta_0^* \exp(\alpha_0^*)\) and \(\beta_r = \beta_r^* \exp(\alpha_0^*),\) for \(r = 1, \ldots, p\) (obviously \(\gamma = \gamma^* \Rightarrow \alpha_0 = \alpha_0^*\) also). We now consider the single-factor model where both the scale and shape depend a categorical covariate. Assuming three levels \(A, B\) and \(C\), this covariate can be coded as two binary variables \(x_1\) and \(x_2\) such that \(A \Leftrightarrow x_1 = x_2 = 0, B \Leftrightarrow x_1 = 1, x_2 = 0\) and
C.5. ORTHOGONAL WEIBULL MPR MODEL

$C \iff x_1 = 0, x_2 = 1$. As the covariate appears in both regressions we have that $x = z = (1, x_1, x_2)^T$. Therefore (C.10) becomes

$$\beta_0 + x_1 \beta_1 + x_2 \beta_2 = \Gamma'(2) + (\beta^*_0 + x_1 \beta^*_1 + x_2 \beta^*_2) \exp(\alpha^*_0 + x_1 \alpha^*_1 + x_2 \alpha^*_2),$$

and this must be solvable for the three possible combinations of $x_1$ and $x_2$. One easily finds that

\begin{align*}
  x_1 = 0, x_2 = 0 : & \quad \beta_0 = \Gamma'(2) + \beta^*_0 \exp(\alpha^*_0) \\
  x_1 = 1, x_2 = 0 : & \quad \beta_1 = (\beta^*_0 + \beta^*_1) \exp(\alpha^*_0 + \alpha^*_1) - \beta^*_0 \exp(\alpha^*_0) \\
  x_1 = 0, x_2 = 1 : & \quad \beta_2 = (\beta^*_0 + \beta^*_2) \exp(\alpha^*_0 + \alpha^*_2) - \beta^*_0 \exp(\alpha^*_0).
\end{align*}

Furthermore, as $\gamma = \gamma^*$, we have that $\alpha_0 = \alpha^*_0$, $\alpha_1 = \alpha^*_1$ and $\alpha_2 = \alpha^*_2$ respectively. We reiterate the fact that (C.10) is not solvable in general. Thus, it is clear that the orthogonal Weibull MPR model is not just a reparametrisation of the standard Weibull MPR model - they are different models. Of course they are closely related being derived from two equivalent parametrisations of the Weibull distribution.

Example C.1. A Comparison of the Standard and Orthogonal Weibull MPR Models

We now show by example (using the lung cancer data) that the two Weibull MPR models are equivalent in the single-factor case but non-equivalent more generally. To this end we consider a single-factor model (treatment) and a two-factor model (treatment + sodium). The results are shown in Table C.1.

Firstly, we compare the two single-factor models and note that the log-likelihood values are the equal. Furthermore, the shape regression coefficients are also equal. From above, we know that the $\beta$ coefficients are be expressible in terms of the $\beta^*$ and $\alpha^*$ coefficients, i.e., $\beta_0 = \Gamma'(2) + \beta^*_0 \exp(\alpha^*_0)$ and $\beta_r = (\beta^*_0 + \beta^*_r) \exp(\alpha^*_0 + \alpha^*_r) - \beta^*_0 \exp(\alpha^*_0)$ for $r \neq 0$. We now check these relationships hold for the values in Table C.1. Noting that $\Gamma'(2) \approx 0.42$, we find $0.42 - 2.07 \exp(-0.19) \approx -1.28 = \hat{\beta}_0$. Furthermore, looking at the coefficient of surgery, we find that $(-2.07 - 1.70) \exp(-0.19 + 0.59) + 2.07 \exp(-0.19) \approx -3.91 = \hat{\beta}_1$. Similarly, we can show this for the other coefficients. Thus, the two models are clearly equivalent in this special case.
More generally, the models are not equivalent as can be seen by the inclusion of one additional covariate (sodium level). The log-likelihood values are not equal so the models are clearly non-equivalent. Furthermore, the shape coefficients are not equal and obviously we cannot express the $\beta$ coefficients in terms of the $\beta^*$ and $\alpha^*$ coefficients.

<table>
<thead>
<tr>
<th>Table C.1. Standard and Orthogonal Weibull MPR Fits</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scale</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>Treatment</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Sodium</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Shape</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>Treatment</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Sodium</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$\ell(\hat{\theta})$</td>
</tr>
</tbody>
</table>
Appendix D

Frailty

D.1 Introduction

In Chapter 7 we studied the topic of frailty modelling. This appendix provides all necessary supplementary material which was previously excluded in the interest of conciseness.

D.2 Marginal Hazard Function

In Section 7.2 we established that the marginal hazard function is given by

\[ \lambda_m(t) = E_U[\lambda(t \mid u) \mid T > t], \]

using the relationship \( \lambda_m(t) = f_m(t)/S_m(t) \) and previously derived expressions for \( f_m(t) \) and \( S_m(t) \). Here we offer an alternative derivation using probability statements which do not rely on having \( f_m(t) \) and \( S_m(t) \).

First note that the conditional hazard function is

\[ \lambda(t \mid u) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Pr(t \leq T < t + \Delta t \mid T \geq t, U = u), \]

using the relationship \( \lambda_m(t) = f_m(t)/S_m(t) \) and previously derived expressions for \( f_m(t) \) and \( S_m(t) \). Here we offer an alternative derivation using probability statements which do not rely on having \( f_m(t) \) and \( S_m(t) \).

First note that the conditional hazard function is

\[ \lambda(t \mid u) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Pr(t \leq T < t + \Delta t \mid T \geq t, U = u) \]

\[ = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \frac{\Pr(t \leq T < t + \Delta t, T \geq t, U = u)}{\Pr(T \geq t, U = u)}. \]
Now, multiplying both sides by \( \Pr(T \geq t, U = u) / \Pr(T \geq t) \) gives
\[
\lambda(t \mid u) \frac{\Pr(T \geq t, U = u)}{\Pr(T \geq t)} = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Pr(t \leq T < t + \Delta t, U = u \mid T \geq t).
\]
Hence, the marginal hazard function is
\[
\lambda_m(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Pr(t \leq T < t + \Delta t \mid T \geq t)
= \int \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Pr(t \leq T < t + \Delta t, U = u \mid T \geq t) \, du
= \int \lambda(t \mid u) \frac{\Pr(T \geq t, U = u)}{\Pr(T \geq t)} \, du,
= \int \lambda(t \mid u) \Pr(U = u \mid T \geq t) \, du,
= E_U[\lambda(t \mid u) \mid T \geq t],
\]
as required. Furthermore, for the multiplicative frailty model (Section 7.4), where \( \lambda(t \mid u) = u \lambda(t) \), the marginal hazard is given by
\[
\lambda_m(t) = \lambda(t) E(U \mid T \geq t).
\]
It is of interest to explore the behaviour of \( E(U \mid T \geq t) \) in this particular case of multiplicative frailty. Hence, we first write
\[
E(U \mid T \geq t) = \int_0^\infty u \Pr(U = u \mid T \geq t) \, du
= \int_0^\infty u \frac{\Pr(T \geq t \mid U = u) \Pr(U = u)}{\Pr(T \geq t)} \, du
= \int_0^\infty u \frac{S(t \mid u)g(u)}{S_m(t)} \, du,
\]
where \( S(t \mid u) \) and \( S_m(t) \) are the conditional and marginal survivor functions for \( T \), respectively, and \( g(u) \) is the density function for \( U \in (0, \infty) \) (see Section 7.2).
Now, using the fact that $S'(t) = -f(t) = -\lambda(t)S(t)$, we find

$$
\frac{d}{dt}E(U \mid T \geq t) = \int_0^\infty u \left[ -\frac{\lambda(t \mid u)S(t \mid u)g(u)}{S_m(t)} - \frac{-S(t \mid u)g(u)\lambda_m(t)S_m(t)}{S_m(t)^2} \right] du
$$

$$
= -\lambda(t) \int_0^\infty u^2 \frac{S(t \mid u)g(u)}{S_m(t)} du + \lambda_m(t) \int_0^\infty u \frac{S(t \mid u)g(u)}{S_m(t)} du
$$

$$
= -\lambda(t) E(U^2 \mid T \geq t) + \lambda_m(t) E(U \mid T \geq t)
$$

$$
= -\lambda(t) [E(U^2 \mid T \geq t) - E(U \mid T \geq t)^2]
$$

$$
= -\lambda(t) \text{var}(U \mid T \geq t),
$$

and, since $\lambda(t)$ and var$(U \mid T \geq t)$ are both non-negative,

$$
\frac{d}{dt}E(U \mid T \geq t) < 0,
$$

i.e., $E(U \mid T \geq t)$ is a decreasing function of time. Moreover, since $E(U) = 1$ (see Section 7.4), we have that $E(U \mid T \geq 0) = E(U) = 1$ and, therefore, $E(U \mid T \geq t) < 1$ for $t > 0$. Thus,

$$
\lambda_m(t) = \lambda(t) E(U \mid T \geq t)
$$

$$
< \lambda(t),
$$

for $t > 0$. In words, the marginal (or population) hazard function is less than the basic hazard.

## D.3 Laplace Transform

In Section 7.4.2 we showed that the marginal survivor and hazard functions (arising from multiplicative frailty) can be written in terms of the Laplace transform, $L(s)$, of the frailty distribution, $g(u)$. We now show how $L(s)$ is related to the moments of $g(u)$. Recall that

$$
L(s) = \int_0^\infty \exp(-us)g(u) \, du,
$$
and, therefore,
\[
L'(s) = -\int_0^{\infty} u \exp(-us) g(u) \, du
\]
\[
L''(s) = \int_0^{\infty} u^2 \exp(-us) g(u) \, du
\]
\[
\vdots
\]
\[
L^{(k)}(s) = (-1)^k \int_0^{\infty} u^k \exp(-us) g(u) \, du. \tag{D.1}
\]
Hence
\[
L^{(k)}(0) = (-1)^k \int_0^{\infty} u^k g(u) \, du.
\]
\[
= (-1)^k E(U^k),
\]
i.e., the Laplace transform of \(g(u)\) is simply its moment generating function multiplied by \((-1)^k\). Thus, in terms of \(L^{(k)}(s)\), the \(k\)th moment of \(g(u)\) is given by
\[
E(U^k) = (-1)^k L^{(k)}(0). \tag{D.2}
\]
In particular, we require that
\[
E(U) = -L'(0) = 1,
\]
for the multiplicative frailty model (see Section 7.4).

### D.3.1 Marginal Likelihood

In Section 7.4.2 we showed that the marginal log-likelihood for the multiplicative frailty model is given by

\[
\ell_m(\theta, \phi) = \sum_{i=1}^{n} \delta_i \log \lambda(t_i) + \log \{(-1)^{\delta_i} L^{(\delta_i)}[\Lambda(t)]\}.
\]

We now derive this result by different means. First recall from Section 7.3 that the marginal likelihood is given by

\[
L_m(\theta, \phi) = \prod_{i=1}^{n} \left[ \int \lambda(t_i \mid u_i, \theta)^{\delta_i} S(t_i \mid u_i, \theta) g(u_i \mid \phi) \, du_i \right],
\]
and, as we are considering multiplicative frailty, this becomes

\[ L_m(\theta, \phi) = \prod_{i=1}^{n} \left[ \lambda(t_i \mid \theta)^{\delta_i} \int u_i^{\delta_i} \exp[-u_i \Lambda(t_i \mid \theta)] g(u_i \mid \phi) \, du_i \right]. \]

Furthermore, from (D.1) above, we can see that

\[ \int u_i^{\delta_i} \exp[-u_i \Lambda(t_i \mid \theta)] g(u_i \mid \phi) \, du_i = (-1)^{\delta_i} \mathcal{L}^{(\delta_i)}[\Lambda(t_i \mid \theta)]. \]

Thus, the marginal likelihood is

\[ L_m(\theta, \phi) = \prod_{i=1}^{n} \lambda(t_i \mid \theta)^{\delta_i} (-1)^{\delta_i} \mathcal{L}^{(\delta_i)}[\Lambda(t_i \mid \theta)], \]

where it is understood that the \( \phi \) parameters appear on the right-hand side through the Laplace transform. Hence, the corresponding marginal log-likelihood is

\[ \ell_m(\theta, \phi) = \sum_{i=1}^{n} \delta_i \log \lambda(t_i \mid \theta) + \log \{(-1)^{\delta_i} \mathcal{L}^{(\delta_i)}[\Lambda(t_i \mid \theta)]\}; \]

as required.

### D.3.2 Gamma Frailty

In Section 7.5 we discussed the gamma frailty model. This is a multiplicative frailty model with frailty density

\[ g(u) = \frac{1}{\phi^{1/\phi} \Gamma(1/\phi)} u^{1/\phi - 1} \exp(-u/\phi). \]

In this case the Laplace transform is

\[
L(s) = \int_0^\infty \exp(-us) \frac{1}{\phi^{1/\phi} \Gamma(1/\phi)} u^{1/\phi - 1} \exp(-u/\phi) \, du \\
= \frac{1}{\phi^{1/\phi} \Gamma(1/\phi)} \int_0^\infty u^{1/\phi - 1} \exp[-u(1/\phi + s)] \, du,
\]
which, upon making the substitution \( y = u(1/\phi + s) \), becomes

\[
L(s) = \frac{1}{\phi^{1/\phi} \Gamma(1/\phi)} \int_{0}^{\infty} [y(1/\phi + s)^{-1}]^{1/\phi-1} \exp(-y) (1/\phi + s)^{-1} \, dy
\]

\[
= \frac{1}{\phi^{1/\phi} \Gamma(1/\phi)} (1/\phi + s)^{-1/\phi} \int_{0}^{\infty} y^{1/\phi-1} \exp(-y) \, dy
\]

\[
= \frac{1}{\phi^{1/\phi} \Gamma(1/\phi)} (1/\phi + s)^{-1/\phi} \Gamma(1/\phi)
\]

\[
= \phi^{-1/\phi} (1/\phi + s)^{-1/\phi}
\]

\[
= (1 + \phi s)^{-1/\phi}.
\]

Furthermore,

\[
L'(s) = (-1/\phi)(1 + \phi s)^{-1/\phi-1} \cdot \phi
\]

\[
= -(1 + \phi s)^{-(1/\phi + 1)},
\]

and

\[
L''(s) = (1/\phi + 1)(1 + \phi s)^{-(1/\phi + 1)-1} \cdot \phi
\]

\[
= (1 + \phi)(1 + \phi s)^{-(1/\phi + 2)},
\]

which can be used to calculate the first two moments of \( g(u) \) since \( E(U^k) = (-1)^k L^{(k)}(0) \) (as shown above). The expected value is

\[
E(U) = - [-(1 + \phi \cdot 0)^{-(1/\phi + 1)}]
\]

\[
= (1)^{-(1/\phi + 1)}
\]

\[
= 1.
\]

Similarly, we have that

\[
E(U^2) = (1 + \phi)(1 + \phi \cdot 0)^{-(1/\phi + 2)}
\]

\[
= 1 + \phi,
\]

and, therefore,

\[
\text{var}(U) = E(U^2) - E(U)^2
\]

\[
= \phi.
\]
Bibliography


individuals in censored survival data. *Biometrics*, 4, 1197–1205.

parametric, Semiparametric, and Parametric Families*. Springer.


and Hall, second edition.


Miller, A. J. (2002). *Subset Selection in Regression*. Chapman and Hall,
second edition.

63, 449–464.


Myers, L. E. (1981). Survival functions induced by stochastic covariate pro-


BIBLIOGRAPHY


BIBLIOGRAPHY


