A PARAMETRIC SURVIVAL MODEL WHICH GENERATES MONOTONIC AND NON-MONOTONIC HAZARD FUNCTIONS AND INCORPORATES TIME-DEPENDENT COVARIABLES

by

James Emmett Herndon II

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1851T

April 1988
A PARAMETRIC SURVIVAL MODEL
WHICH GENERATES MONOTONIC AND NON-MONOTONIC HAZARD FUNCTIONS
AND INCORPORATES TIME-DEPENDENT COVARIABLES

by
James Emmett Herndon II

A Dissertation submitted to the faculty of The
University of North Carolina at Chapel Hill in partial
fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Biostatistics

Chapel Hill
1988

Approved by:

[Signatures]

Adviser
Reader
Reader
Abstract

JAMES EMMETT HERDON II. A Parametric Survival Model Which Generates Monotonic and Non-Monotonic Hazard Functions and Incorporates Time-Dependent Covariates. (Under the direction of FRANK E. HARRELL JR.)

Cox's semiparametric proportional hazards model (1972) is commonly used to describe the relationship between survival time and fixed covariates. Estimation is efficient even though no assumptions are made about underlying distributional shape. However, the inclusion of time-dependent covariates is computationally expensive, especially when the sample size is large.

A parametric proportional hazards model where the baseline hazard is a cubic spline function with tails that are linearly restricted (Stone and Koo, 1985) is proposed as an alternative to the Cox model. Time-dependent covariates can be incorporated into this parametric model with little increase in computation time.

The properties of the proposed model with and without covariates are contrasted with those of the Kaplan-Meier and Cox models. The homogeneous model is shown to be able to describe a variety of hazard function shapes, including monotonic and non-monotonic hazards. The efficiency of estimation with the restricted cubic spline model is as good as, and sometimes better than, the Kaplan-Meier estimator for the underlying distributions studied. The restricted cubic spline hazard model with fixed covariates is usually more efficient than the Cox model at estimating survival probabilities. The efficiency of regression coefficient estimation with the spline model is comparable to the Cox model.

Additional properties of this parametric model include smooth survival curves, and confidence limits for survival and hazard estimates, even when time-dependent covariates are present.

Analyses previously obtained using the Cox model with a time-dependent covariate describing an intervening event are replicated using the spline model. Estimates of covariable
coefficients are virtually identical. Computation time for the spline model using an IBM 3081 computer was reduced by a factor of 213 over the Cox model.

The incorporation of repeated measurements into this model is demonstrated using coronary disease data. An approach to approximating continuous covariates by step functions is proposed. Such an approximation suggests an approach for testing the proportional hazards assumption.
## Table of Contents

List of Tables ........................................................................................................ viii

List of Figures ......................................................................................................... x

Chapter I - Literature Review

1.1 Basic Survival Model ...................................................................................... 1

1.2 Estimation in the Homogeneous Survival Model ........................................... 5
  1.2.1 Kaplan-Meier Estimation ........................................................................ 5
  1.2.2 Parametric Estimation in the Homogeneous Model ............................. 11
  1.2.3 Kaplan-Meier Estimation versus Parametric Estimation ....................... 13

1.3 Estimation in the Proportional Hazards Model .............................................. 15
  1.3.1 Incorporation of Covariates into the Homogeneous Model .................... 15
  1.3.2 Covariate Types .................................................................................. 16
  1.3.3 Estimation in Cox's Proportional Hazards Model ................................. 17
  1.3.4 Parametric Estimation in the Proportional Hazards Model ................. 24
  1.3.5 Cox's Model versus Parametric Estimation ........................................ 25

1.4 Numerical Optimization .................................................................................. 29
  1.4.1 Newton-Raphson Procedure and Related Methods ............................ 30
  1.4.2 Maximum Likelihood Estimation Using Weighted Least Squares ......... 34
Chapter I - Literature Review (Continued)

1.5 Parametric Distributions Applicable to Survival Analysis ........................................ 39
   1.5.1 Exponential Distribution ....................................................................................... 44
   1.5.2 Rayleigh and Generalized Rayleigh Distributions ........................................... 44
   1.5.3 Gompertz and Makeham-Gompertz Distributions ......................................... 49
   1.5.4 Weibull Distribution ......................................................................................... 50
   1.5.5 Gamma and Generalized Gamma Distributions .................................................. 54
   1.5.6 Lognormal and Log-logistic Distributions ........................................................ 55
   1.5.7 Generalized F Distribution .............................................................................. 59
   1.5.8 Homogeneous Models with Bathtub-Shaped Hazard Functions ...................... 59
   1.5.9 Regression Models ......................................................................................... 74

1.6 Splines .................................................................................................................. 78
   1.6.1 Introduction to Spline Functions ....................................................................... 78
   1.6.2 Spline Estimation of the Hazard Function ....................................................... 83

1.7 Research Topic ....................................................................................................... 84

Chapter II - Two Homogeneous Hazard Models

2.1 Introduction ............................................................................................................ 86

2.2 Generalized Weibull Hazard Function .................................................................... 87
   2.2.1 Motivation for the Generalized Weibull Model and Its Structure ......................... 87
   2.2.2 Parameter Estimation and Problems with Model Fitting ..................................... 91
   2.2.3 Preliminary Test of Weibullness ...................................................................... 93

2.3 Restricted Cubic Spline Hazard Function ............................................................... 97
   2.3.1 Derivation of Hazard Function ......................................................................... 98
   2.3.2 Maximum Likelihood Estimation of Unknown Parameters ............................. 101
   2.3.3 Confidence Intervals for the Hazard, Cumulative Hazard, and Survival Functions .................................................................................................................. 102

2.4 Modeling with the Restricted Cubic Spline Hazard Function ................................. 104

2.5 Efficiency of Estimating the Survival Function with the Spline Model .................... 117

2.6 Spline Models Applied to Cardiovascular Data ....................................................... 126
   2.6.1 Medical Patient Data ....................................................................................... 126
   2.6.2 Surgical Patient Data ..................................................................................... 132
### Chapter III - The Spline Function as Baseline Hazard in the Proportional Hazards Model with Fixed Covariates

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Introduction</td>
<td>137</td>
</tr>
<tr>
<td>3.2 Incorporation of Covariates into the Model</td>
<td>138</td>
</tr>
<tr>
<td>3.2.1 The Model and Log-Likelihood Function</td>
<td>138</td>
</tr>
<tr>
<td>3.2.2 Approach to Modeling</td>
<td>140</td>
</tr>
<tr>
<td>3.3 Efficiency of Estimation of Covariate Coefficient</td>
<td>140</td>
</tr>
<tr>
<td>3.4 Efficiency of Survival Function Estimation</td>
<td>147</td>
</tr>
<tr>
<td>3.5 Application to Cardiovascular Data</td>
<td>148</td>
</tr>
<tr>
<td>3.5.1 Medical Patient Data</td>
<td>148</td>
</tr>
<tr>
<td>3.5.2 Surgical Patient Data</td>
<td>151</td>
</tr>
</tbody>
</table>

### Chapter IV - The Spline Function as Baseline Hazard in the Proportional Hazards Model with Time-Dependent Covariates Expressed as Step Functions

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Introduction</td>
<td>156</td>
</tr>
<tr>
<td>4.2 Structure of the Model with Time-Dependent Covariates</td>
<td>156</td>
</tr>
<tr>
<td>4.2.1 Schematic and Notation Describing Time-Dependent Covariates</td>
<td>156</td>
</tr>
<tr>
<td>4.2.2 The Hazard Function and Log-Likelihood Function</td>
<td>159</td>
</tr>
<tr>
<td>4.2.3 Confidence Intervals</td>
<td>162</td>
</tr>
<tr>
<td>4.2.4 Maximum Likelihood Estimation</td>
<td>165</td>
</tr>
<tr>
<td>4.3 Covariate Describing an Intervening Event or Treatment Crossover</td>
<td>166</td>
</tr>
<tr>
<td>4.3.1 Comparison with the Cox Model</td>
<td>166</td>
</tr>
<tr>
<td>4.3.2 Modeling the Effect of an Intervening Event as a Function of the Time Since the Event's Occurrence</td>
<td>176</td>
</tr>
<tr>
<td>4.4 Repeated Measurements</td>
<td>183</td>
</tr>
</tbody>
</table>
Chapter IV - The Spline Function as Baseline Hazard in the Proportional Hazards Model with Time-Dependent Covariates Expressed as Step Functions (Continued)

4.5 Step Function Approximation of a Defined Covariate

4.5.1 Modeling Non-Proportionality of Hazards Using Indicator Functions

4.5.2 Modeling Non-Proportionality of Hazards Using the Interaction Between a Fixed Covariate and Log Time

Chapter V - Discussion and Future Research

5.1 Summary

5.2 Future Research Topics

Appendices

1 The Score Vector and Information Matrix Associated with the Generalized Weibull Hazard Function

2 The Score Vector and Information Matrix Associated with the Homogeneous Spline Hazard Function

3 The Score Vector and Information Matrix for the Spline Model with Fixed Covariates

4 The Score Vector and Information Matrix for the Spline Model with Time-Dependent Covariates Expressed as Step Functions

5 The PHSPLM Procedure

6 Examples of Use of the PHSPLM Procedure

Bibliography
List of Tables

Table                                      Page
1.1  Summary of Parametric Models Commonly Used in Survival Analysis ................................................................. 41
1.2  Additional Parametric Models Which Generate Bathtub-Shaped Hazard Functions ......................................................... 60
1.3  Parametric Models Generated by Slymen and Lachenbruch ................................................................. 71
2.1  Efficiency of Survival Function Estimation with the Restricted Cubic Spline Hazard Function Relative to the Kaplan-Meier Estimator for Data Generated from the Generalized Weibull Distribution with a=1, b=0.8, c=1, d=3 ......................................................... 123
2.2  Efficiency of Survival Function Estimation with the Restricted Cubic Spline Hazard Function Relative to the Kaplan-Meier Estimator for Data Generated from the Generalized Weibull Distribution with a=2, b=1, c=1, d=3 ......................................................... 124
3.1  Efficiency of Covariable Estimation with the Spline Model Relative to the Cox Model for Data Generated from the Generalized Weibull Proportional Hazards Model ................................................................. 142
3.2  Efficiency of Survival Function Estimation with the Spline Model Relative to the Cox Model for Data Generated from the Generalized Weibull Proportional Hazards Model with a=1, b=0.8, c=1, d=3, β=0.1 ......................................................... 143
3.3  Efficiency of Survival Function Estimation with the Spline Model Relative to the Cox Model for Data Generated from the Generalized Weibull Proportional Hazards Model with a=1, b=0.8, c=1, d=3, β=1.0 ......................................................... 144
3.4  Efficiency of Survival Function Estimation with the Spline Model Relative to the Cox Model for Data Generated from the Generalized Weibull Proportional Hazards Model with a=2, b=1, c=1, d=3, β=0.1 ......................................................... 145
3.5  Efficiency of Survival Function Estimation with the Spline Model Relative to the Cox Model for Data Generated from the Generalized Weibull Proportional Hazards Model with a=2, b=1, c=1, d=3, β=1.0 ......................................................... 146
4.1  Effect of an Intervening Nonfatal Myocardial Infarction Using the Cox and Spline Models ................................................................. 168
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2</td>
<td>Effect of an Intervening Nonfatal Myocardial Infarction Adjusted for HSM Using the Cox and Spline Models</td>
</tr>
<tr>
<td>4.3</td>
<td>Effect of an Intervening Nonfatal Myocardial Infarction Adjusted for HSM and Waiting Time Until an MI Using the Cox and Spline Models</td>
</tr>
<tr>
<td>4.4</td>
<td>Effect of an Intervening Nonfatal Myocardial Infarction as a Function of Time Since Its Occurrence Using Indicator Variables in the Spline Hazard Model</td>
</tr>
<tr>
<td>4.5</td>
<td>Effect of an Intervening Nonfatal Myocardial Infarction as a Function of Time Since Its Occurrence Using a Spline Function Covariable in the Spline Hazard Model</td>
</tr>
<tr>
<td>4.6</td>
<td>Effect of Congestive Heart Failure Status on Patient Survival Using a Spline Hazard Model</td>
</tr>
<tr>
<td>4.7</td>
<td>Assessment of the Proportional Hazards Assumption for Pain Index Using Indicator Functions in the Spline Hazard Model</td>
</tr>
<tr>
<td>4.8</td>
<td>Assessment of the Proportional Hazards Assumption for Pain Index Using a Step Function Approximation of the Interaction Between Pain Index and Log Time in the Spline Hazard Model</td>
</tr>
</tbody>
</table>
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Second-Order Generalized Rayleigh Hazard Function</td>
<td>46</td>
</tr>
<tr>
<td>1.2</td>
<td>Third-Order Generalized Rayleigh Hazard Function</td>
<td>47</td>
</tr>
<tr>
<td>1.3</td>
<td>Gompertz Hazard Function</td>
<td>48</td>
</tr>
<tr>
<td>1.4</td>
<td>Weibull Hazard Function</td>
<td>51</td>
</tr>
<tr>
<td>1.5</td>
<td>Gamma Hazard Function</td>
<td>52</td>
</tr>
<tr>
<td>1.6</td>
<td>Generalized Gamma Hazard Function</td>
<td>53</td>
</tr>
<tr>
<td>1.7</td>
<td>Lognormal Hazard Function</td>
<td>57</td>
</tr>
<tr>
<td>1.8</td>
<td>Log-Logistic Hazard Function</td>
<td>58</td>
</tr>
<tr>
<td>1.9</td>
<td>Power-Exponential Hazard Function</td>
<td>62</td>
</tr>
<tr>
<td>1.10</td>
<td>Gaver and Acar's Hazard Function</td>
<td>66</td>
</tr>
<tr>
<td>1.11</td>
<td>Murthy's Hazard Function</td>
<td>68</td>
</tr>
<tr>
<td>1.12</td>
<td>Modified Weibull Hazard Function</td>
<td>71</td>
</tr>
<tr>
<td>2.1a</td>
<td>Generalized Weibull Hazard Function</td>
<td>89</td>
</tr>
<tr>
<td>2.1b</td>
<td>Generalized Weibull Hazard Function (Continued)</td>
<td>90</td>
</tr>
<tr>
<td>2.2a</td>
<td>Survival Function Estimates with 95% Confidence Limits for a Sample of Size 200 Generated from the Generalized Weibull Distribution with $a=1, b=0.8, c=1, d=3$</td>
<td>107</td>
</tr>
<tr>
<td>2.2b</td>
<td>Hazard Function Estimate with 95% Confidence Limits for a Sample of Size 200 Generated from the Generalized Weibull Distribution with $a=1, b=0.8, c=1, d=3$</td>
<td>107</td>
</tr>
<tr>
<td>2.3a</td>
<td>Survival Function Estimates with 95% Confidence Limits for a Sample of Size 200 Generated from the Generalized Weibull Distribution with $a=1, b=0.8, c=2, d=1$</td>
<td>108</td>
</tr>
<tr>
<td>Figure</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>2.3b</td>
<td>Hazard Function Estimate with 95% Confidence Limits for a Sample of Size 200 Generated from the Generalized Weibull Distribution with $a=1, b=0.8, c=2, d=1$</td>
<td>108</td>
</tr>
<tr>
<td>2.4a</td>
<td>Survival Function Estimates with 95% Confidence Limits for a Sample of Size 200 Generated from the Generalized Weibull Distribution with $a=1, b=3, c=1, d=1.5$</td>
<td>109</td>
</tr>
<tr>
<td>2.4b</td>
<td>Hazard Function Estimate with 95% Confidence Limits for a Sample of Size 200 Generated from the Generalized Weibull Distribution with $a=1, b=3, c=1, d=1.5$</td>
<td>109</td>
</tr>
<tr>
<td>2.5a</td>
<td>Survival Function Estimates with 95% Confidence Limits for a Sample of Size 200 Generated from the Gompertz Distribution with $a=0, b=1$</td>
<td>110</td>
</tr>
<tr>
<td>2.5b</td>
<td>Hazard Function Estimate with 95% Confidence Limits for a Sample of Size 200 Generated from the Gompertz Distribution with $a=0, b=1$</td>
<td>110</td>
</tr>
<tr>
<td>2.6a</td>
<td>Survival Function Estimates with 95% Confidence Limits for a Sample of Size 50 Generated from the Generalized Weibull Distribution with $a=1, b=0.8, c=1, d=3$</td>
<td>111</td>
</tr>
<tr>
<td>2.6b</td>
<td>Hazard Function Estimate with 95% Confidence Limits for a Sample of Size 50 Generated from the Generalized Weibull Distribution with $a=1, b=0.8, c=1, d=3$</td>
<td>111</td>
</tr>
<tr>
<td>2.7a</td>
<td>Survival Estimates for a Sample Generated from the Generalized Weibull Distribution with $a=1, b=0.25, c=1, d=3$ Using Knots at the 5th, 33rd, 67th, and 95th Percentiles</td>
<td>112</td>
</tr>
<tr>
<td>2.7b</td>
<td>Survival Estimates for a Sample Generated from the Generalized Weibull Distribution with $a=1, b=0.25, c=1, d=3$ Using Knots at the 5th, 20th, 40th, and 90th Percentiles</td>
<td>112</td>
</tr>
<tr>
<td>2.7c</td>
<td>Hazard Estimate for a Sample Generated from the Generalized Weibull Distribution with $a=1, b=0.25, c=1, d=3$ Using Knots at the 5th, 20th, 40th, and 90th Percentiles</td>
<td>113</td>
</tr>
<tr>
<td>2.8a</td>
<td>Spline Survival Function Estimates for 25 Samples Generated from the Generalized Weibull Distribution with $a=1, b=0.8, c=1, d=3$</td>
<td>119</td>
</tr>
<tr>
<td>2.8b</td>
<td>Kaplan-Meier Curve for 25 Samples Generated from the Generalized Weibull Distribution with $a=1, b=0.8, c=1, d=3$</td>
<td>119</td>
</tr>
<tr>
<td>2.8c</td>
<td>Spline Hazard Function Estimate for 25 Samples Generated from the Generalized Weibull Distribution with $a=1, b=0.8, c=1, d=3$</td>
<td>120</td>
</tr>
<tr>
<td>2.9a</td>
<td>Spline Survival Function Estimates for 25 Samples Generated from the Generalized Weibull Distribution with $a=2, b=1, c=1, d=3$</td>
<td>121</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>2.9b</td>
<td>Kaplan-Meier Curve for 25 Samples Generated from the Generalized Weibull Distribution with ( a=2, b=1, c=1, d=3 )</td>
<td>121</td>
</tr>
<tr>
<td>2.9c</td>
<td>Spline Hazard Function Estimate for 25 Samples Generated from the Generalized Weibull Distribution with ( a=2, b=1, c=1, d=3 )</td>
<td>122</td>
</tr>
<tr>
<td>2.10a</td>
<td>Survival Function Estimates for Medical Patients Using Knots Located at the 5th, 33rd, 67th, and 95th Percentiles</td>
<td>127</td>
</tr>
<tr>
<td>2.10b</td>
<td>Hazard Function Estimate for Medical Patients Using Knots Located at the 5th, 33rd, 67th, and 95th Percentiles</td>
<td>127</td>
</tr>
<tr>
<td>2.10c</td>
<td>Survival Function Estimates for Medical Patients Using Knots Located at the 5th, 20th, 33rd, 67th, and 95th Percentiles</td>
<td>128</td>
</tr>
<tr>
<td>2.10d</td>
<td>Hazard Function Estimate for Medical Patients Using Knots Located at the 5th, 20th, 33rd, 67th, and 95th Percentiles</td>
<td>126</td>
</tr>
<tr>
<td>2.11a</td>
<td>Survival Function Estimates for Surgical Patients Using Knots Located at the 5th, 33rd, 67th, and 95th Percentiles</td>
<td>133</td>
</tr>
<tr>
<td>2.11b</td>
<td>Hazard Function Estimate for Surgical Patients Using Knots Located at the 5th, 33rd, 67th, and 95th Percentiles</td>
<td>133</td>
</tr>
<tr>
<td>2.11c</td>
<td>Survival Function Estimates for Surgical Patients Using Knots Located at the 5th, 20th, 33rd, 67th, and 95th Percentiles</td>
<td>134</td>
</tr>
<tr>
<td>2.11d</td>
<td>Hazard Function Estimates for Surgical Patients Using Knots Located at the 5th, 20th, 33rd, 67th, and 95th Percentiles</td>
<td>134</td>
</tr>
<tr>
<td>3.1</td>
<td>Survival Estimates as a Function of Left Ventricular Ejection Fraction for Medical Patients</td>
<td>150</td>
</tr>
<tr>
<td>3.2</td>
<td>Survival Estimates as a Function of Left Ventricular Ejection Fraction for Surgical Patients</td>
<td>153</td>
</tr>
<tr>
<td>4.1</td>
<td>Schematic of Time-Dependent Covariable Measurements for the ( i )-th Subject</td>
<td>158</td>
</tr>
<tr>
<td>4.2a</td>
<td>Survival Estimate with 95% Confidence Limits for a Patient Who Does Not Experience a Myocardial Infarction</td>
<td>171</td>
</tr>
<tr>
<td>4.2b</td>
<td>Hazard Estimate with 95% Confidence Limits for a Patient Who Does Not Experience a Myocardial Infarction</td>
<td>171</td>
</tr>
<tr>
<td>4.3a</td>
<td>Survival Estimate with 95% Confidence Limits for a Patient Who Experiences a Myocardial Infarction Soon After Catheterization</td>
<td>172</td>
</tr>
<tr>
<td>4.3b</td>
<td>Hazard Estimate with 95% Confidence Limits for a Patient Who Experiences a Myocardial Infarction Soon After Catheterization</td>
<td>172</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>4.4a</td>
<td>Survival Estimate with 95% Confidence Limits for a Patient Who Experiences a Myocardial Infarction 2.25 Years After Catheterization</td>
<td>173</td>
</tr>
<tr>
<td>4.4b</td>
<td>Hazard Estimate with 95% Confidence Limits for a Patient Who Experiences a Myocardial Infarction 2.25 Years After Catheterization</td>
<td>173</td>
</tr>
<tr>
<td>4.5a</td>
<td>Survival Estimate with 95% Confidence Limits for a Patient Who Has Congestive Heart Failure Assessments of Level 0 at Baseline, Level 1 at 3 Years, and Level 3 at 5 Years</td>
<td>186</td>
</tr>
<tr>
<td>4.5b</td>
<td>Hazard Estimate with 95% Confidence Limits for a Patient Who Has Congestive Heart Failure Assessments of Level 0 at Baseline, Level 1 at 3 Years, and Level 3 at 5 Years</td>
<td>186</td>
</tr>
<tr>
<td>4.6a</td>
<td>Survival Estimate with 95% Confidence Limits for a Patient Who Has Congestive Heart Failure Assessments of Level 1 at Baseline and Level 3 at 5 Years</td>
<td>187</td>
</tr>
<tr>
<td>4.6b</td>
<td>Hazard Estimate with 95% Confidence Limits for a Patient Who Has Congestive Heart Failure Assessments of Level 1 at Baseline and Level 3 at 5 Years</td>
<td>187</td>
</tr>
<tr>
<td>4.7</td>
<td>Kaplan-Meier and Cox Model Estimates in Three Strata Determined by the Tertiles of Pain Index</td>
<td>191</td>
</tr>
<tr>
<td>4.8</td>
<td>Log Hazard Ratio with 95% Confidence Limits in 12 Time Intervals for Covariable Pain Index</td>
<td>192</td>
</tr>
<tr>
<td>4.9</td>
<td>Log Hazard Ratio with 95% Confidence Limits in 5 Time Intervals for Covariable Pain Index</td>
<td>193</td>
</tr>
<tr>
<td>4.10</td>
<td>Step Function Approximation of Log(T) for a Patient Whose Survival Time is 5 Years</td>
<td>197</td>
</tr>
<tr>
<td>4.11</td>
<td>Log Hazard Ratio for Covariable PI Based on the Cox Model in 12 Time Intervals and the Spline Model with Interaction Between PI and Log (T)</td>
<td>199</td>
</tr>
</tbody>
</table>
Chapter I
Literature Review

1.1 Basic Survival Problem

Biological and epidemiological studies often involve analysis of the time to a response, whether it be time to failure, death, cure, or development of a given disease. These measurements are subject to random variation and hence, form a distribution. Much of survival analysis is involved with estimating this distribution and examining the relationship between survival times and subject characteristics, often referred to as covariates or concomitant variables.

Let $T$ be a nonnegative continuous random variable representing the survival time of an individual. Initially, it will be assumed that $T$ is measured in a homogeneous population and therefore no covariates are necessary to describe the probability distribution associated with $T$. The survival and cumulative distribution functions are defined as $S(t) = P(T > t)$ and $F(t) = 1 - S(t)$, respectively. The probability density function, also known as the unconditional failure rate, is defined as

$$f(t) = -rac{dS(t)}{dt}.$$

The hazard function or conditional failure rate specifies the instantaneous failure rate
conditional upon survival to time $t$ and is defined as

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} = -\frac{d\log S(t)}{dt}.$$ 

In many applications, the cumulative hazard function is relevant and is defined as

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

Relationships among these functions which will later prove to be useful are the following:

$$S(t) = \exp (-\Lambda(t)) \quad \text{and} \quad f(t) = \lambda(t) \, S(t). \quad (1.1)$$

It should be noted that each of the survival and hazard functions discussed above is a function of one or more parameters.

Survival times are usually known exactly for only a portion of all experimental units under study and the rest are censored. An observation is said to be right censored at time $C$ if the exact survival time is known to be greater than $C$. Similarly, an observation is said to be left censored if the exact survival time is known to be less than $C$. In biological applications, right censoring is more common than left censoring. Therefore only right censoring will be considered here.

There are several ways in which right censoring can occur. Let $t_1 \leq t_2 \leq \ldots \leq t_n$ be the observed survival times of $n$ patients under study. Type II censoring occurs when the $r$ ($r < n$) smallest observed survival times are uncensored. The remaining $n-r$ items are censored at the survival time for the $r$-th smallest observed survival time, $t_r$. This type of censoring scheme is often used in quality control where a total of $n$ items are simultaneously tested and the test is discontinued when $r$ items have failed. Type I censoring occurs when each unit is subjected to a limited period of observation of prespecified or known length $C_i$, $i = 1, 2, \ldots n.$
A unit's failure time $T_i$ is observed only when $T_i \leq C_i$. Type I censoring is frequently used in clinical research where a decision is made to terminate a study after a prescribed period of time.

In most medical and epidemiological situations, random censoring occurs. Each individual is assumed to have a survival time $T$ and a censoring time $C$. $T$ and $C$ are independent continuous random variables with survivor functions $S(t)$ and $G(t)$, respectively. The density functions associated with these random variables are $f(t)$ and $g(t)$, respectively.

Let $(T_i, C_i), i = 1, 2, \ldots, n$ be independent random variables. For each subject the two following random variables are observed:

$$t_i = \min (T_i, C_i) \quad \text{and} \quad \delta_i = \begin{cases} 1 & \text{if } T_i \leq C_i \\ 0 & \text{if } T_i > C_i \end{cases},$$

where $t_i$ is the observed survival time and $\delta_i$ indicates whether $t_i$ is censored or not. The probability density function of $t_i$ and $\delta_i$ is easily obtained by using the assumed independence properties of $T_i$ and $C_i$.

$$P(t_i = t, \delta_i = 0) = P(C_i = t, T_i > t) = S(t) \cdot g(t).$$

$$P(t_i = t, \delta_i = 1) = P(T_i = t, C_i > t) = G(t) \cdot f(t).$$

(It should be noted that the notation used above is slightly misleading. The first term in each of the above equations represents the joint probability of the two events $t_i = t$ and $\delta_i = 0$.)

The likelihood for the $i$-th patient is

$$L_i = P(t_i = t, \delta_i) = \left( S(t) \cdot g(t) \right)^{1-\delta_i} \left( G(t) \cdot f(t) \right)^{\delta_i}.$$
This implies that the full likelihood function is

\[ L = \prod_{i=1}^{n} L_i = \prod_{i=1}^{n} \left( \left( S(t_i) g(t_i) \right)^{1-\delta_i} \left( G(t_i) f(t_i) \right)^{\delta_i} \right) \]

\[ = \left[ \prod_{i=1}^{n} G(t_i)^{\delta_i} g(t_i)^{1-\delta_i} \right] \left[ \prod_{i=1}^{n} S(t_i)^{1-\delta_i} f(t_i)^{\delta_i} \right]. \] (1.2)

If \( G(t) \) and \( g(t) \) do not depend on any parameters of interest, then the first term of (1.2) can be neglected. The resulting likelihood function, which contains the information on the parameters of interest, is

\[ L = \prod_{i=1}^{n} S(t_i)^{1-\delta_i} f(t_i)^{\delta_i}. \] (1.3)

The following alternative form of (1.3) can be derived using the relationships in equation (1.1):

\[ L = \prod_{i=1}^{n} \lambda(t_i)^{\delta_i} S(t_i). \] (1.4)

Upon further use of these relationships, equation (1.4) can easily be shown to be

\[ \mathcal{L} = \log L = \sum_{i=1}^{n} \left[ \delta_i \log \lambda(t_i) + \log S(t_i) \right] \]

\[ = \sum_{i=1}^{n} \left[ \delta_i \log \lambda(t_i) - \Lambda(t_i) \right]. \] (1.5)

If \( D \) is the set of all uncensored observations, then (1.5) can be written in the following form:

\[ \mathcal{L} = \sum_{i \in D} \log \lambda(t_i) - \sum_{i=1}^{n} \Lambda(t_i). \] (1.6)
The likelihood function for Type I censoring is a special case of the likelihood function for random censoring.

Although the random censorship model is often reasonable to assume, the censoring-death process is sometimes more complicated. The censoring mechanism and the death process may not be independent. Lawless (1982) shows that under the following fairly general conditions, likelihood function (1.4) can be used to make inferences. The assumptions are: (1) the censoring-death mechanism for different individuals are acting independently, and (2) the probability of an uncensored death in a given time interval is independent of the number of censored and uncensored observations in previous time intervals. Kalbfleisch and Prentice (1980) present an argument similar to that given by Lawless.

Even though only the homogeneous survival model has been discussed, all concepts are applicable to models which include covariables. These will be discussed later.

1.2 Estimation of the Homogeneous Survival Model

1.2.1 Kaplan-Meier Estimation

In a homogeneous population, the survival curve is often estimated by the nonparametric approach developed by Kaplan and Meier (1958). The survival function is estimated as

\[ \hat{S}(t) = \prod_{i | t_i < t} \left[ \frac{n_i - d_i}{n_i} \right]^{t_i}, \]  

(1.7)

where \( n_i \) is the number of individuals alive just prior to time \( t_i \) and \( d_i \) is the number of individuals who die at time \( t_i \). Kaplan and Meier present an argument to show that this estimate is the maximum likelihood estimate when no distributional assumptions are made on survival times. Let \( t_1 \leq t_2 \leq \ldots \leq t_u \) be the observed uncensored ordered failure times and let \( t_0 = 0 \) and \( t_{u+1} = \infty \). These observations partition time into \( u+1 \) intervals \( I_1 = (0, t_1] \),
I_2 = (t_1, t_2], \ldots, I_u = (t_{u-1}, t_u], I_{u+1} = (t_u, \infty). With n_i and d_i defined as above, let m_i denote the number of observations censored in interval I_i and let t_{ij}, j = 1, 2, \ldots, m_i denote their censoring times. It should be noted that the total sample size n is
\[ n = m_{u+1} + \sum_{i=1}^{u} (d_i + m_i). \]

The probability of failure at time t_i is
\[ S(t_i) - S(t_i + 0), \]
where \( S(t_i + 0) = \lim_{\Delta t \to 0^+} S(t_i + \Delta t) \). The contribution made to the likelihood function by an observation censored at t_{ij} is S(t_{ij}). Therefore the likelihood function is
\[ L = \prod_{i=1}^{u} \left\{ \left[ S(t_i) - S(t_i + 0) \right]^{d_i} \prod_{j=1}^{m_i} S(t_{ij}) \right\}^{m_{u+1}} \prod_{j=1}^{m_{u+1}} S(t_{u+1,j}). \] (1.8)

In order to maximize L with respect to S(t), it should be noticed that \( \hat{S}(t) \) must be discontinuous at the uncensored survival times t_i, i = 1, 2, \ldots, u or else L = 0. In addition, \( \hat{S}(t_{ij}) \) should be as large as possible. This implies that \( \hat{S}(t_1) = \hat{S}(t_{1j}) = 1 \) for j = 1, 2, \ldots, m_1; \( \hat{S}(t_i) = \hat{S}(t_{i-1} + 0) = \hat{S}(t_{ij}) \) for i = 2, \ldots, u+1 and j = 1, 2, \ldots, m_i; and \( \hat{S}(t_u + 0) = \hat{S}(t_{u+1} + 0) \). Let \( \hat{S}(t_i) = P_{i-1} \) for i = 2, \ldots, u+1 and let \( P_0 = 1 \). Upon substituting these values into equation (1.8), the following likelihood equation is obtained:
\[ L = \prod_{i=1}^{u} \left[ P_{i-1} - P_i \right]^{d_i} P_i^{m_i+1}. \]
By letting $p_i = p_i / p_{i-1}$, the likelihood becomes

$$L = \prod_{i=1}^{n} \left[ p_1 p_2 \cdots p_i (1-p_i) \right]^{d_i} \left( p_1 p_2 \cdots p_i \right)^{n_i-1}.$$  

Through combining similarly subscripted terms and simplification, the following likelihood will result:

$$L = \prod_{i=1}^{n} (1-p_i)^{d_i} p_i^{n_i-d_i}. \quad \text{(1.9)}$$

This likelihood is maximized when

$$\hat{p}_i = \frac{n_i - d_i}{n_i}.$$  

(1.10)

By using the recursive relationship $p_i = p_i p_{i-1}$ and induction, it can easily be shown that

$$\hat{p}_i = \prod_{k=1}^{i} \left( 1 - \hat{p}_k \right) = \prod_{k=1}^{i} \left[ \frac{n_i - d_i}{n_i} \right]. \quad \text{(1.11)}$$

for $i = 1, 2, \ldots, n$. This implies that (1.7) is the function which maximizes the likelihood and hence is the maximum likelihood estimator for $S(t)$.

Kalbfleisch and Prentice (1980) provide a heuristic derivation for the asymptotic variance formula for the Kaplan-Meier estimator. Their development relies upon standard asymptotic theory for the maximum likelihood estimates of $p_i$, shown in equation (1.10). The log-likelihood equation needed to obtain the information associated with $\hat{p}_i$ is a consequence of equation (1.9).

$$\mathcal{L} = \log L = \sum_{i=1}^{n} \left[ d_i \log (1-p_i) + (n_i - d_i) \log p_i \right].$$
The first partial derivative of the log-likelihood equation with respect to an arbitrary $p_k$, $k = 1, 2, \ldots, u$ is

$$\frac{\partial \ell}{\partial p_k} = \frac{-d_k}{1 - p_k} + \frac{n_k - d_k}{p_k^2}.$$  

The second partial derivative with respect to $p_k$ is therefore

$$\frac{\partial^2 \ell}{\partial p_k^2} = \frac{d_k}{(1 - p_k)^2} - \frac{n_k - d_k}{p_k^2}.$$  

The sample information associated with $p_k$ is estimated to be

$$I(\hat{p}_k) = -\frac{\partial^2 \ell}{\partial p_k^2} \bigg|_{p_k = \hat{p}_k} = \frac{-d_k}{(1 - \hat{p}_k)^2} + \frac{n_k - d_k}{\hat{p}_k^2},$$  

(1.12)

where $\hat{p}_k$ is defined in (1.10). Equation (1.12) simplifies to

$$I(\hat{p}_k) = \frac{n_k^3}{d_k (n_k - d_k)}.$$  

(1.13)

Therefore the estimated asymptotic variance of $\hat{p}_k$ is $n_k^3 d_k (n_k - d_k)$. From (1.7) and (1.11), it can be seen that

$$\log S(t) = \sum_{i|t_i < t} \log (1 - \hat{p}_i).$$  

(1.14)

It is assumed in this heuristic argument that the terms in the summation of equation (1.14) are independent. Therefore

$$\text{Var} [\log S(t)] = \sum_{i|t_i < t} \text{Var} [\log (1 - \hat{p}_i)].$$  

(1.15)
The variance of each individual term can be determined by the method of statistical differentials (Elandt-Johnson and Johnson, 1980), which is sometimes referred to as the delta method (Miller, 1981).

\[
\text{Var} [\log (1 - \hat{p}_i)] = \left[ \frac{d[\log (1 - \hat{p}_i)]}{d\hat{p}_i} \right]^2 \text{Var} (\hat{p}_i)
\]

\[= \frac{1}{(1 - \hat{p}_i)^2} \text{Var} (\hat{p}_i).\]

This latter formula for the variance simplifies to

\[
\text{Var} [\log (1 - \hat{p}_i)] = \frac{\frac{d_i}{n_i(n_i - d_i)}}{1 - \hat{p}_i).
\]

Upon substitution into (1.15), the estimated asymptotic variance of \(\log S(t)\) is found to be

\[
\text{Var} [\log S(t)] = \sum_{t_i < t} \left[ \frac{\frac{d_i}{n_i(n_i - d_i)}}{1 - \hat{p}_i} \right].
\]

Let \(A = \log \hat{S}(t)\). Another application of the method of statistical differentials will generate the asymptotic variance of \(\hat{S}(t) = \exp A\).

\[
\text{Var} [S(t)] = \text{Var} [\exp A] = \left[ \frac{d[\exp A]}{dA} \right]^2 \text{Var} A
\]

\[= [S(t)]^2 \sum_{t_i < t} \left[ \frac{\frac{d_i}{n_i(n_i - d_i)}}{1 - \hat{p}_i} \right]. \quad (1.16)
\]

This expression for the estimated asymptotic variance of the Kaplan-Meier estimator is usually referred to as Greenwood's Formula.
Previously, it was discussed that $t_1$ equaled min $(T_1, C_1)$ where $T_1$ was the actual survival time and $C_1$ was the censoring time. These random variables are assumed independent under random censorship with survivor functions $S(t)$ and $G(t)$, respectively. The $t_i$'s constitute a random sample with survival function $S(t)$ $G(t)$. In such a situation, Breslow and Crowley (1974) show that the Kaplan-Meier estimator is asymptotically distributed as a normal random variable with mean $S(t)$ and

$$\text{Var} [S(t)] = \frac{1}{n} \frac{S(t)^2}{S(t)} \int_0^t \frac{dS(u)}{S^2(u) G(u)} = \frac{1}{n} S(t)^2 \int_0^t \frac{f(u) du}{S^2(u) G(u)} .$$

(1.17)

This result was developed by first considering the classical life table and showing that the limiting distribution, as the sample size becomes large, of the vector of estimated survival probabilities is multivariate normal. In order to examine the limiting distribution of the Kaplan-Meier estimator, the number of cells in the life table is allowed to become large and hence the partition size is allowed to become infinitesimally small. They show that

$$\sqrt{n} \left[ \hat{S}(t) - S(t) \right], \quad 0 < t < T$$

where $\hat{S}(t)$ is the Kaplan-Meier estimate, converges weakly to a mean zero Gaussian process with covariance function

$$S(t_1) S(t_2) \int_0^{t_1} \frac{dS(u)}{S^2(u) G(u)} , \quad 0 < t_1 \leq t_2 < T .$$

(1.18)

This implies the indicated asymptotic distribution of the Kaplan-Meier estimator. Equation (1.17) can be used as a basis for obtaining an estimate of the asymptotic variance. In the denominator, let $S(u) G(u)$ be estimated by the observed proportion of individuals alive and at risk at time $u$ and the remaining $S(u)$ by $\hat{S}(u + 0)$. Elsewhere, $S(u)$ is replaced by the Kaplan-Meier estimate. Upon simplification, Greenwood's formula is obtained (Lawless, 1980).
Breslow and Crowley's result also implies that the distribution of the Kaplan-Meier estimator is asymptotically normal with mean $S(t)$ and variance as given by Greenwood's formula.

1.2.2 Parametric Estimation in the Homogeneous Model

Parametric models can be considered as an alternative to the nonparametric Kaplan-Meier estimator for describing the survival experience in a homogeneous population. In many situations, the parametric form of the survival curve is known from prior experience or from knowledge of biological mechanisms. When such information is not available, graphical procedures, residual analysis, and some goodness-of-fit tests are available for determining which parametric form is most appropriate (Bain, 1978; Elandt-Johnson and Johnson, 1980; Lawless, 1982; Miller, 1981).

When the parametric form of the survival distribution is specified under random censorship, the likelihood function used to find the maximum likelihood estimate of the vector of unknown parameters $\theta' = (\theta_1, \theta_2, \ldots, \theta_p)$ is equation (1.3). Many times numerical optimization techniques must be used in order to find the value which maximizes equation (1.3).

When $\theta$ is interior to a closed, compact parameter space, and the likelihood function is thrice differentiable, and certain boundedness conditions are satisfied, then the maximum likelihood estimator $\hat{\theta}$ is asymptotically distributed as a normal random variable with mean $\theta$ (Cox and Hinkley, 1974; and Lehmann, 1983). Since observations are independently and identically distributed, the asymptotic variance is $\left[ n I(\theta) \right]^{-1}$ where $I(\theta)$ is the information matrix associated with an individual observation. The log-likelihood for an individual observation is:

$$L = \log L = \log \left[ f(t)^\delta S(t)^{1-\delta} \right] = \begin{cases} 
\log f(t) & \text{if } \delta = 1 \text{ (or } T \leq C) \\
\log S(t) & \text{if } \delta = 0 \text{ (or } T > C) 
\end{cases}.$$
Let the vector of first derivatives of the log-likelihood function be denoted as

$$
\mathbf{U}'(\theta) = (U_1(\theta), U_2(\theta), \ldots, U_k(\theta)) \quad \text{where} \quad U_j(\theta) = \frac{\partial \log \mathbf{f}(t)}{\partial \theta_j} \quad \text{if} \quad T \leq C
$$

$$
\quad \frac{\partial \log \mathbf{S}(t)}{\partial \theta_j} \quad \text{if} \quad T > C
$$

for \( j = 1, 2, \ldots, p \). Fisher's p x p information matrix associated with a single observation of the random variable \( t \) is

$$
\mathbf{I}(\theta) = E[\mathbf{U}(\theta) \mathbf{U}'(\theta)].
$$

The \( jk \)-th entry of the information matrix can be derived as follows:

$$
I_{jk} = E\left[ \frac{\partial^2 \mathbf{L}}{\partial \theta_j \partial \theta_k} \right] = \iint\limits_{T < C} \left[ \frac{\partial \log \mathbf{f}(u)}{\partial \theta_j} \frac{\partial \log \mathbf{f}(u)}{\partial \theta_k} \right] f(u) g(v) \, dv \, du
$$

$$
\quad + \iint\limits_{C < T} \left[ \frac{\partial \log \mathbf{S}(u)}{\partial \theta_j} \frac{\partial \log \mathbf{S}(u)}{\partial \theta_k} \right] f(u) g(v) \, du \, dv
$$

$$
= \int_0^\infty \int_u^\infty \left[ \frac{\partial \log \mathbf{f}(u)}{\partial \theta_j} \frac{\partial \log \mathbf{f}(u)}{\partial \theta_k} \right] f(u) g(v) \, dv \, du
$$

$$
\quad + \int_0^\infty \int_u^\infty \left[ \frac{\partial \log \mathbf{S}(u)}{\partial \theta_j} \frac{\partial \log \mathbf{S}(u)}{\partial \theta_k} \right] f(v) g(u) \, dv \, du
$$

$$
= \int_0^\infty \left[ \frac{\partial \log \mathbf{f}(u)}{\partial \theta_j} \frac{\partial \log \mathbf{f}(u)}{\partial \theta_k} \right] f(u) G(u) \, du
$$

$$
\quad + \int_0^\infty \left[ \frac{\partial \log \mathbf{S}(u)}{\partial \theta_j} \frac{\partial \log \mathbf{S}(u)}{\partial \theta_k} \right] S(u) g(u) \, du. \tag{1.19}
$$

By the method of statistical differentials, it can be shown that \( \hat{S}(t) \) is asymptotically distributed as a normal random variable with mean \( S(t) \) and variance

$$
\frac{1}{n} \mathbf{S}' \mathbf{I}^{-1}(\theta) \mathbf{S}, \tag{1.20}
$$
where $\partial S$ denotes the vector of partial derivatives of $S(t)$, i.e.

\[ \partial S' = \left[ \frac{\partial S(t)}{\partial \theta_1}, \frac{\partial S(t)}{\partial \theta_2}, \ldots, \frac{\partial S(t)}{\partial \theta_p} \right]. \]

When $f$ is a function of only one unknown parameter, the information matrix reduces to the following scalar:

\[ I(\hat{\theta}) = \int_0^\infty \left[ \frac{\partial \log f(u)}{\partial \theta} \right]^2 f(u) G(u) \, du + \int_0^\infty \left[ \frac{\partial \log S(u)}{\partial \theta} \right]^2 g(u) S(u) \, du. \] (1.21)

The asymptotic variance of $\hat{\theta}$ is $\frac{1}{n} \Gamma^{-1}(\theta)$ and the asymptotic variance of $S(\hat{\theta})$ is

\[ \frac{1}{n} \left[ \frac{\partial S(t)}{\partial \theta} \right]^2 \Gamma^{-1}(\theta). \] (1.22)

### 1.2.3 Kaplan-Meier Estimation versus Parametric Estimation

Miller (1983) examined the asymptotic efficiency of the Kaplan-Meier estimator relative to the parametric maximum likelihood estimator of the survival function under exponential censorship. Underlying distributions considered were the exponential distribution, the Weibull distribution with known shape parameter (one-parameter Weibull), and the Weibull distribution with unknown shape parameter (two-parameter Weibull). For the exponential model and the one-parameter Weibull model, efficiency was calculated as the ratio of the asymptotic variance of the parametric estimate, (1.22), to the asymptotic variance of the Kaplan-Meier estimate, (1.16), whereas for the two-parameter Weibull model, it was calculated as the ratio of (1.20) to (1.16). Asymptotic efficiency, when the underlying model was exponential or one-parameter Weibull, peaked at 0.65 and dropped to zero in the tails of the distribution. Though slightly higher for the two-parameter Weibull model, since two
parameters, rather than one, were estimated, the relative asymptotic efficiency also showed a drop in the tails. The drop in efficiency near 0 is not of much concern since estimates are extremely accurate, i.e., variances for both Kaplan-Meier estimates and parametric estimates are small. The drop in efficiency as time increases is a real problem since variances for both parametric and nonparametric estimates are increasing. Miller concludes that “If interest is focused on estimation for the extreme high quantiles . . . then the Kaplan-Meier estimator is usually worthless . . .” An additional observation made by Miller was that as the proportion of censored data increased, the relative efficiency of both the exponential and Weibull model estimates to the Kaplan-Meier estimate decreased. The asymptotic relative efficiency of the Kaplan-Meier estimator when the underlying parametric distribution is different from exponential and Weibull has not been studied. In addition, the effect of different censoring distributions has not been examined. The applicability of the large-sample results to small sample efficiency needs examination.

Other differences between the Kaplan-Meier and parametric estimators exist. First, the Kaplan-Meier estimator is a step function, whereas the parametric estimator is a continuous function. The Kaplan-Meier estimator makes the assumption that the hazard between deaths is zero. Such is probably not the case in most biological and physical applications. In addition, most biological and physical mechanisms probably adhere to a continuous survival function. Second, the Kaplan-Meier estimator is more robust to outliers than the parametric estimator. Third, the Kaplan-Meier estimator can easily describe any survival function shape unlike a specific parametric form. The parametric estimator is subject to bias when an incorrect parametric form is chosen for inferences, unlike the Kaplan-Meier estimator, which is always asymptotically unbiased. The size of this bias needs examination. Fourth, parametric estimation is usually more computationally intensive than Kaplan-Meier estimation since many passes through the data are required during numerical optimization.
1.3 Estimation in the Proportional Hazards Model

1.3.1 Incorporation of Covariables into the Homogeneous Model

Many times the relationship between survival time and some other factor or subject characteristic is of interest. A variety of ways exist to incorporate such a concomitant factor or variable into the homogeneous survival model. Included among these approaches are defining the unknown homogeneous parameters as functions of covariables (see Bailey, 1977; Feigl and Zelen, 1965; Zippin and Armitage, 1966; and Byar et al., 1974 for examples and details) and the accelerated failure model (see Kalbfleisch and Prentice, 1980 for details). However, probably the most popular approach is the proportional hazards model where covariables are allowed to act multiplicatively on the hazard.

The basic form of the proportional hazards model is

$$\lambda(t|\mathbf{x}) = \lambda_0(t) \ g(\mathbf{x}),$$

where $\lambda_0(t)$ can be thought of as the baseline hazard or hazard when $g(\mathbf{x}) = 1$ and $\mathbf{x}$ is a vector of covariables which are possibly time-dependent. The function $g(\mathbf{x})$ is often specified to be $\exp(\mathbf{x} \beta)$. When the covariables are not a function of time $t$, the hazard ratio of two individuals is proportional and not dependent on time, that is,

$$\frac{\lambda(t|\mathbf{x}_1)}{\lambda(t|\mathbf{x}_2)} = \frac{g(\mathbf{x}_1)}{g(\mathbf{x}_2)}.$$

This proportionality characteristic motivated the nonparametric or semiparametric generalization proposed by Cox (1972). He specified the hazard function as $\lambda(t|\mathbf{x}) = \lambda_0(t) \ \exp(\mathbf{x} \beta)$ where $\lambda_0(t)$ was an arbitrary unspecified baseline hazard. Cox's model has been found to adequately describe many biological and epidemiological situations. The parametric and
nonparametric versions of the proportional hazards model with \( g(s) = \exp (s \beta) \) will be considered further in the following material.

\[ S(t|\mathbf{s}), F(t|\mathbf{s}), f(t|\mathbf{s}), \lambda(t|\mathbf{s}), \text{ and } A(t|\mathbf{s}) \] will denote the survival, cumulative density, density, hazard, and cumulative hazard functions, respectively, when each is a function of the covariable vector \( \mathbf{s} \). When the covariable vector is a function of \( t \), it will be denoted as \( \mathbf{s}(t) \).

### 1.3.2 Covariable Types

Covariables can be categorized into two broad classes—external and internal covariates (Kalbfleisch and Prentice, 1980). External covariates are not directly involved with the failure mechanisms. Internal covariates are measurements taken directly from the subject and require the survival of the subject for their existence.

Kalbfleisch and Prentice further subdivide the class of external covariates into fixed, defined, and ancillary covariates. A fixed covariate is measured in advanced and fixed for the duration of the study. These are sometimes referred to as time independent or baseline covariables. These have been denoted as \( \mathbf{s} \) in this text. A defined covariate has its covariate path determined in advance for each subject. Usually the value of this covariate varies with time. When it is not, the defined covariate is a fixed covariate. An ancillary covariate is the output of a stochastic process which is external to the subject under study. An example is the level of air pollution as a predictor of asthma attacks.

An internal covariate is the output of a stochastic process which is generated by the subject and observed only while the subject is alive. As a result, values of the covariable carry information about the subject’s survival time. Internal covariates will not be considered in much, if any detail, in the subsequent material.
1.3.3 Estimation in Cox's Proportional Hazards Model

Cox (1972) proposed the proportional hazards model in which the hazard function was

$$
\lambda(t|z) = \lambda_0(t) \exp(z \beta),
$$

where $\lambda_0(t)$ is the baseline hazard and $z$ is a $1 \times r$ vector of covariates. Cox suggested that $z$ can be a function of time dependent covariates. Therefore in the subsequent discussion, the following hazard model will be considered

$$
\lambda(t|z(t)) = \lambda_0(t) \exp[z(t)\beta],
$$

(1.23)

where $z(t)$ is a vector of time dependent covariates.

Cox (1972, 1975) proposed the partial likelihood function for estimating the vector of unknown parameters $\beta$ in the proportional hazards model. The development of this technique will be similar to that given by Efron (1977). The general likelihood function based on model (1.23) is

$$
L = \prod_{i=1}^{n} \left[ \lambda_0(t_i) \exp(z_i(t_i)\beta) \right]^{d_i} \exp \left[ - \int_{0}^{t_i} \lambda_0(v) \exp(z_i(v)\beta) \, dv \right].
$$

(1.24)

Efron rewrites this likelihood as

$$
L = \exp \left[ - \sum_{i=1}^{n} \left\{ \int_{0}^{t_i} \lambda_0(v) \exp(z_i(v)\beta) \, dv \right\} \right] \prod_{i \in D} \left[ \lambda_0(t_i) \exp(z_i(t_i)\beta) \right],
$$

where $D$ denotes the set of individuals for whom observed survival times are uncensored, i.e.
the individual fails. Upon rearrangement, the likelihood becomes

\[
L = \exp \left[ - \int_0^\infty \left( \sum_{i \in R(v)} \exp \left\{ \beta_i(v) \theta \right\} \right) \lambda_0(v) \, dv \right] \\
\times \prod_{i \in D} \left[ \lambda_0(t_i) \exp \left( \beta_i(t_i) \theta \right) \right].
\]

(1.25)

where \( R(v) \) denotes the set of individuals alive just prior to time \( v \). Efron defines \( H(t) \) as the average hazard rate if all \( n \) patients were on test at time \( t \):

\[
H(t) = \frac{1}{n} \sum_{i=1}^n \lambda_0(t) \exp \left( \beta_i(t) \theta \right).
\]

Efron also defines "the average number of patients remaining at risk at time \( t \)" as

\[
N(t) = n \left[ \frac{\sum_{i \in R(t)} \exp \left( \beta_i(t) \theta \right)}{\sum_{i=1}^n \exp \left( \beta_i(t) \theta \right)} \right].
\]

Each of the above definitions assumes that the covariates are external. Using the definition of \( H(t) \) and \( N(t) \), likelihood (1.25) is rewritten as

\[
L = \left[ \prod_{i \in D} \frac{\exp \left( \beta_i(t_i) \theta \right)}{\sum_{j \in R(t_i)} \exp \left( \beta_j(t_i) \theta \right)} \right] \\
\times \exp \left\{ - \int_0^\infty N(u) H(u) \, du \right\} \prod_{i \in D} N(t_i) H(t_i).
\]

(1.26)
The first term in the above formula is Cox's partial likelihood function,

\[
L = \prod_{i \in D} \left[ \frac{\exp \left( x_i(t_i) \beta \right)}{\sum_{j \in R(t_i)} \exp \left( x_j(t_i) \beta \right)} \right],
\]

(1.27)
in which the term in brackets can be viewed as the conditional probability that individual \( i \) fails at \( t_i \) given all subjects at risk at \( t_i \). The second term is a likelihood assuming all patients have the same hazard and at any time \( t \) an "average" number of patients are at risk of failure. Cox (1972) argues that this second term contains little information about \( \beta \). One of the primary criticisms of the use of Cox's partial likelihood equation is that it ignores a factor relating to the "nonfailure intervals", i.e. intervals between deaths during which no failures occur. Cox argues that this factor conveys no information about \( \beta \) since the baseline hazard could conceivably be near 0 in the interval.

When covariables are fixed, the partial likelihood (1.27) reduces to the following:

\[
L = \prod_{i \in D} \left[ \frac{\exp \left( x_i \beta \right)}{\sum_{j \in R(t_i)} \exp \left( x_j \beta \right)} \right].
\]

(1.28)

This likelihood function can also be obtained through marginal likelihood arguments, which are based on the ranks of observed survival times. These arguments do not hold for censoring schemes more general than Type II censoring nor for time dependent covariables.

The development of the partial likelihood to this point has assumed no ties present in the data. Several modifications of these results have been proposed when ties in the data occur. Suppose \( t_1 \leq t_2 \leq \ldots \leq t_u \) are the uncensored failure times and \( d_i, i = 1, 2, \ldots, u \) are the number of deaths at time \( t_i \). Cox (1972) suggested the following generalization of
(1.28) for tied data:

\[
L = \prod_{i=1}^{n} \left[ \frac{\exp(s_i \beta)}{\sum_{j \in R_{d_i}(t_i)} \exp(s_j \beta)} \right], \tag{1.29}
\]

where \(s_i\) is the sum of the covariates associated with deaths at time \(t_i\) and \(R_{d_i}(t_i)\) is the set of all subsets of \(d_i\) individuals chosen from the risk set \(R(t_i)\) without replacement. This procedure is computationally difficult. If the number of individuals failing at each failure point, \(d_i\), is small compared to the size of the risk set, (1.29) can be approximated by

\[
L = \prod_{i=1}^{n} \left[ \frac{\exp(s_i \beta)}{\left\{ \sum_{j \in R(t_i)} \exp(s_j \beta) \right\}^{d_i}} \right]. \tag{1.30}
\]

(Breslow, 1974). The partial likelihood when no ties are present is a special case of (1.30).

Also, partial likelihood (1.30) is clearly generalizable to time-dependent covariates as

\[
L = \prod_{i=1}^{n} \left[ \frac{\exp(s_i(t_i) \beta)}{\left\{ \sum_{j \in R(t_i)} \exp(s_j(t_i) \beta) \right\}^{d_i}} \right], \tag{1.31}
\]

where \(s_i(t_i)\) is the sum of the covariate values associated with deaths at \(t_i\). The log-likelihood function associated with (1.31) is

\[
L = \log L = \sum_{i \in D} \left[ s_i(t_i) \beta - d_i \log \left\{ \sum_{j \in R(t_i)} \exp(s_j(t_i) \beta) \right\} \right]. \tag{1.32}
\]

The maximum likelihood estimate \(\beta\) can be obtained as a solution to the following system of
\[
\frac{\partial^2}{\partial \beta_k \partial \beta_h} = \sum_{i=1}^{u} \left[ s_{ik}(t_i) - d_i \frac{\sum_{j \in R(t_i)} z_{jk}(t_i) \exp \left( z_j(t_i) \beta \right)}{\sum_{j \in R(t_i)} \exp \left( z_j(t_i) \beta \right)} \right]
\]

for \( k = 1, 2, \ldots, r \) and where \( s_{ik}(t_i) \) is the \( k \)-th element in the \( \mathbf{s}_i(t_i) \) vector and \( z_{jk}(t_i) \) is the \( k \)-th element in the \( \mathbf{z}_j(t_i) \) vector. These equations are seldom tractable and hence numerical optimization techniques are used.

The \( kh \)-th entry in the sample information matrix \( \mathbf{I}(\hat{\beta}) \) is

\[
I_{kh}(\hat{\beta}) = - \frac{\partial^2}{\partial \beta_k \partial \beta_h} = \sum_{i=1}^{u} d_i C_{ikh}, \quad (1.33)
\]

where

\[
C_{ikh} = \frac{\sum_{j \in R(t_i)} z_{jk}(t_i) z_{jh}(t_i) \exp \left( z_j(t_i) \beta \right)}{\sum_{j \in R(t_i)} \exp \left( z_j(t_i) \beta \right)}
\]

The \( kh \)-th entry in the information matrix (1.33) reduces drastically when covariates are fixed (Kalbfleisch & Prentice, 1980).

Cox (1972, 1975) heuristically shows that \( \hat{\beta} \), the maximum partial likelihood estimator (MPLE) is asymptotically distributed as a normal random variable with mean \( \beta \) and covariance matrix \( \mathbf{I}^{-1}(\hat{\beta}) \). Tsiatis (1981) gives a rigorous proof of asymptotic normality.
and consistency for fixed covariates and no data ties. Asymptotic normality holds in the presence of tied data, though there is some bias in the estimate of $\hat{\beta}$ and its covariance matrix.

Assuming all covariates are fixed, the estimation of the survival curve given the maximum partial likelihood estimate $\hat{\beta}$, involves finding an estimate for the baseline survival function $S_0(t)$ where

$$S_0(t) = \exp \left( - \int_0^t \lambda_0(v) \, dv \right).$$

One such estimate is

$$\hat{S}_0(t) = \prod_{i | t_i < t} \hat{\alpha}_i,$$

where $\hat{\alpha}_i$ is the solution to the following normal equations (Kalbfleisch & Prentice, 1980):

$$\sum_{j=1}^{d_i} \frac{\exp (x_{ij}^T \hat{\beta})}{\exp (x_{ij}^T \hat{\beta}) - \beta_j} \frac{\exp (x_{ij}^T \hat{\beta})}{\exp (x_{ij}^T \hat{\beta})} = \sum_{j \in R(t_i)} \exp (x_{ij}^T \hat{\beta}),$$

(1.34)

for $i = 1, 2, \ldots, u$. These equations simplify when $d_i = 1$ and explicitly define $\hat{\alpha}_i$ as

$$\hat{\alpha}_i = 1 - \frac{\exp (x_i \hat{\beta})}{\sum_{j \in R(t_i)} \exp (x_{ij} \hat{\beta})} + \exp \left( \exp (x_i \hat{\beta}) \right).$$

When ties are present, iterative calculations are required to find the solution to (1.34). Once $\hat{S}_0(t)$ has been determined, an estimate for $S(t)$ is obtained as

$$\hat{S}(t|\mathbf{x}) = \hat{S}_0(t)^{x^T \hat{\beta}}.$$

(1.35)

This equation reduces to the Kaplan-Meier estimator for homogeneous populations when $\hat{\beta}$
equals \( \Theta \). The iterative procedure described above is derived by Lawless (1982) in a manner similar to that of the Kaplan-Meier estimator.

Breslow (1974) presents the following alternative estimator which does not require iteration when ties are present in the data:

\[
\hat{S}_0(t) = \exp \left( - \Lambda_0(t) \right) = \exp \left\{ \sum_{j|t_j < t} \left[ - \frac{d_j}{\sum_{k \in R(t_j)} \exp (z_k \beta)} \right] \right\}.
\] (1.36)

Breslow claims that this estimator is not much different numerically from the previously described baseline survival function estimator. However, the survival function estimate, \( \hat{S}(t|x) \), does not simplify in this case to the Kaplan-Meier estimator when \( \beta = 0 \). Rather, it reduces to a function of a cumulative hazard estimator proposed by Nelson (1969). The survival function estimate is

\[
\hat{S}(t|x) \bigg|_{\beta=0} = \exp \left[ \hat{\Lambda}(t|x) \right] \bigg|_{\beta=0} = \exp \left[ - \sum_{j|t_j < t} \left\{ \frac{d_j}{\sum_{k \in R(t_j)} I} \right\} \right].
\]

Tsiatis (1981) develops the asymptotic theory associated with the survival function estimator. Tsiatis shows that Breslow's cumulative hazard estimator,

\[
\hat{\Lambda}_0(t) = \sum_{j|t_j < t} \left[ - \frac{d_j}{\sum_{k \in R(t_j)} \exp (z_k \beta)} \right],
\] (1.37)

is distributed asymptotically as a normal random variable with mean \( \Lambda_0(t) \). The asymptotic
variance can be estimated by

\[ \text{Var} \left( \hat{A}_0(t) \right) = \sum_{i=1}^{n} \left[ \sum_{j \in R(t_i)} \exp \left( x_j \hat{\beta} \right) \right]^{-2} + A' \Gamma^{-1}(\hat{\beta}) A, \]  

(1.38)

where

\[ A = \sum_{i \in D} \left[ \left\{ \sum_{j \in R(t_i)} \exp \left( x_j \hat{\beta} \right) \right\}^{-2} \sum_{j \in R(t_i)} x_j \exp \left( x_j \hat{\beta} \right) \right]. \]

By the method of statistical differentials, it can be shown that asymptotically \( \hat{S}(t|\mathbf{x}) \) is normally distributed with mean \( S(t|\mathbf{x}) \). The asymptotic variance is

\[ \text{Var}[S(t|\mathbf{x})] = \exp \left(-2 \hat{A}_0(t)\right) \times \text{Var} \left( \hat{A}_0(t) \right). \]  

(1.39)

Kalbfleisch and Prentice (1980) indicate that these results are also applicable to \( \hat{S}(t|\mathbf{x}) \) obtained via the nonparametric maximum likelihood approach.

### 1.3.4 Parametric Estimation in the Proportional Hazards Model

Estimation in the parametric regression model is similar to estimation in the parametric homogeneous model. Likelihood function (1.24) is fully specified with the exception for a vector of \( p \) unknown parameters. This vector is denoted as \( \theta = (\alpha, \beta) \) where \( \alpha \) is the \((p-r) \times 1\) vector of unknown parameters associated with the baseline hazard and \( \beta \) is the \( r \times 1 \) vector of unknown covariable coefficients. The associated log-likelihood equation is

\[ L = \log L \]

\[ = \sum_{i \in D} \log \left[ \lambda_0(t_i) \exp \left( x_i(t_i)\beta \right) \right] - \sum_{i=1}^{n} \left[ \int_0^{t_i} \lambda_0(u) \exp \left( x_i(u)\beta \right) \, du \right] \]  

(1.40)
(Efron, 1977; and Oakes, 1977). Numerical optimization techniques are often needed to find the $\hat{\beta}$ which maximizes this equation. Numerical integration techniques are useful when $\lambda_0(v) \exp \left( z_1(v) \beta \right)$ is not integrable. Standard maximum likelihood theory indicates that the asymptotic distribution of $\hat{\beta}$ is normal with mean $\theta$. The asymptotic variance can be calculated using information matrix (1.19) which is associated with the estimate $\hat{\theta}$. The method of statistical differentials can be used to show that the asymptotic variance for $S(\hat{\theta})$ is calculated using (1.20).

1.3.5 Cox Model versus Parametric Estimation

Due to its nonparametric basis, the Cox proportional hazards model can describe a variety of shapes without specification of the underlying hazard function. Estimation is always asymptotically unbiased, unlike parametric estimation, which is biased if an incorrect parametric form is chosen for inferences.

The nonparametric basis for the Cox model also makes the procedure fairly robust to outliers. Using cancer survival data, Byar (1982) examined the effect of outliers in the estimation of coefficients in the Cox and Weibull models. Regression coefficients in the Weibull model were more affected by the presence of outliers than in the Cox model. Byar concluded that “the Cox model provides greater flexibility in fitting the nuisance hazard and, because it is based on ranks of death times, it is robust with respect to outliers in survival time.”

The relative efficiency of the partial likelihood estimator of $\beta$ versus the parametric estimator has been examined by Kalbfleisch (1974), Kalbfleisch and McIntosh (1977), Efron (1977), Oakes (1977), and Kay (1979) in various situations. In the following discussion, relative efficiency is the ratio of the variance of the parametric estimate of $\beta$ to the variance of the Cox estimate of $\beta$.

Kalbfleisch (1974) examined the asymptotic relative efficiency assuming either an underlying exponential or Weibull distribution, no censorship, and one fixed covariable. Under
the assumption that \( \beta = 0 \), partial likelihood estimation is greater than 75% efficient in finite samples and asymptotically fully efficient. As \( \beta \) departs from \( \theta \), the relative efficiency decreases to zero. Kalbfleisch indicated that "unless \( \beta \) is indicated to be fairly large, very little is to be gained from making the more stringent" parametric assumption. Kalbfleisch showed that when the underlying distribution is exponential the asymptotic relative efficiency is approximately \( \exp(-\mu_2 \beta^2) \) where \( \mu_2 \) is the second central moment of the covariable.

Efficiency against the Weibull model is asymptotically the same as against the exponential model and is slightly better in finite samples.

By considering up to two fixed covariables and also censorship, Kay (1979) extended the efficiency results of Kalbfleisch (1974) for the exponential distribution. The primary conclusion made was as censorship increases, the efficiency of nonparametric estimation increases.

Efron (1977) and Oakes (1977) developed general efficiency expressions which are summarized in Kalbfleisch and Prentice (1980). In their developments, covariables are allowed to be time-dependent. Oakes shows that full asymptotic efficiency is achieved by the Cox model if

\[
E \left\{ \frac{\sum_{i \in R(t)} z_i(t) \exp \left( z_i(t) \beta \right)}{\sum_{i \in R(t)} \exp \left( z_i(t) \beta \right)} \right\} = \text{constant}.
\]

This condition is satisfied if \( \beta = 0 \) and censorship does not depend on \( z \). Censoring rates which depend on \( z \) give rise to distributions of \( z \) which vary with time. Kalbfleisch and Prentice try to explain the loss in efficiency by the fact that such variations with time introduce correlations between the estimator of \( \beta \) and \( \lambda_0(t) \). Parametric analyses, unlike nonparametric procedures, can exploit such relationships. The asymptotic relative efficiency declines as \( \beta \) becomes different from \( \theta \). Efron argues that when the family of parametric distributions is reasonably rich, inferences about \( \beta \) are asymptotically equivalent.
Kalbfleisch and McIntosh (1977) examined the relative efficiency of parametric estimation compared to partial likelihood estimation for the time-dependent covariable, $\gamma$, in the following Weibull regression model

$$\lambda_0(t) = \theta \ t^{\alpha-1} \ \exp (\beta z + \gamma z \ \log \ t) .$$

where $z = 0, 1$. A test of $H_0: \gamma = 0$ is a test for proportional hazards as proposed by Cox (1972). Using simulated sample sizes of 50 for each level of the covariable, Kalbfleisch and McIntosh showed that the relative efficiency of the estimation of $\gamma$ decreases as the absolute value of $\beta$ increases. Estimation when $\beta = 0$ is at least 83% efficient and varies as a function of the remaining parameters of the underlying distribution.

Peace and Flora (1978) compared through Monte Carlo methods the power of hypothesis tests concerning regression parameters arising from Cox's model with four covariables and the exponential, Weibull, and Gompertz proportional hazards models. All parametric tests had equal or slightly greater power than the equivalent test in the Cox model. Peace and Flora concluded that "in the analysis of real-world data . . . it may not be worth the effort to first ascertain whether the parametric models fit the data; rather use Cox's methods of analysis. An exception to this may be in the case of small samples."

Additional differences in parametric and nonparametric estimation of the survival function need to be noted. First, the estimate of the survival function using the Cox model is a step function whereas the parametric estimate is continuous. As previously indicated during the discussion of the homogeneous model, most biological and physical mechanisms probably adhere to a continuous survival function.

Second, the log-likelihood function for the Cox model cannot be written as the sum of contributions by individual subjects, as it can be with a parametric model. The basic components of Cox's log-likelihood function are summations over risk sets.
The final difference between the Cox and parametric model is the computational complexity of the estimation of the unknown model parameters. These complexities involve numerical integration, tied data, and time dependent covariables. Numerical integration techniques, which are unnecessary in Cox model estimation, are instituted in parametric estimation when the hazard function is not integrable. This difference in computational time is minor relative to that generated by tied data and time dependent covariables.

The form of the partial likelihood function proposed by Cox (1972) for tied data, (1.29), involves many computations when the number of ties is large. Consider the contribution to the partial likelihood function made by the $d_i$ deaths at uncensored failure time $t_i$. All possible subsets of size $d_i$ among the subjects at risk of dying at $t_i$ must be considered in the computations. If $d_i$ is large, the computations can be considerable. The approximation proposed by Breslow (1974) and presented as (1.30) eases the computations. However, Kalbfleisch and Prentice (1980) indicate that if the proportion of dead among those at risk at any time $t_i$ is large, then the bias using Breslow’s approximation may be severe. Howard (1972) and Gail, Lubin, and Rubinstein (1981) have proposed a recursive algorithm which permits rapid computation of Cox’s partial likelihood function for tied data. The applicability and feasibility of this algorithm in large data sets in unclear. The problems discussed above concerning tied data do not arise in parametric estimation.

The presence of time dependent covariables in the Cox model increases the number of computations considerably. A comparison of partial likelihood (1.27) which includes time dependent covariables and partial likelihood (1.28) which includes only fixed covariables shows that the covariate vector in (1.27) must be computed for each patient in a risk set unlike in (1.28). The number of mathematical operations required for one evaluation of the log-likelihood (1.32) which involves time dependent covariables is of order $nr^2$ where $n$ is the number of subjects in the study, $u$ is the number of uncensored failure times, and $r$ is the number of covariables. The number of operations is of order $nr^2$ when no time dependent covariables are present. The number of mathematical operations required in the parametric
model for one evaluation of the log-likelihood (1.40) is of order $n^2$, whether time dependent covariables are present or not. Hoffman's (1985) numerical examples indicated that in the presence of time dependent covariables BMDP requires 90 times more CPU time than when no time dependent covariables are present.

1.4 Numerical Optimization

Two of the most common methods of parameter estimation are least squares and maximum likelihood. The primary difference of the two approaches is the objective function being optimized. In weighted least squares estimation, the weighted sum of squares of deviations from the model is minimized in order to find parameter estimates. In maximum likelihood estimation, parameter estimates are taken as those values which maximize the likelihood function, whether it has been developed based on general likelihood or partial likelihood principles. Maximum likelihood estimators obtained from general likelihood functions are asymptotically unbiased and normally distributed. The asymptotic properties of maximum partial likelihood estimators have not been as rigorously developed, but asymptotic unbiasedness and normality are indicated. Maximum likelihood estimation is the estimation method of choice due to these properties.

Closed form solutions for the estimator of $\theta$, a $p \times 1$ vector of unknown parameters, seldom occur in survival problems. As a result, iterative numerical optimization techniques must be used to find the estimates. These iterative techniques begin with an initial guess, $\hat{\theta}_0$, of the optimal value. A sequence of iterates $\hat{\theta}_1, \hat{\theta}_2, \ldots$ is generated which hopefully converges to the desired solution, whether it is the maximum likelihood or weighted least squares estimate for $\theta$. After obtaining $\hat{\theta}_i$, the computations necessary for finding the next iterate $\hat{\theta}_{i+1}$ consist of determining the direction of the next step $d_i$ and the scalar length or size $\alpha_i$ of such a step. These are used to compute the following iterate:

$$\hat{\theta}_{i+1} = \hat{\theta}_i + \alpha_i d_i . \quad (1.41)$$
The primary difference among numerical optimization procedures is the approach to determining the step direction and step size.

1.4.1 Newton-Raphson Procedure and Related Methods

The most popular of the numerical optimization techniques for finding the maximum likelihood estimator is the Newton-Raphson procedure. This method uses the first and second derivative of the log-likelihood function in its iterative procedure and displays a quadratic rate of convergence. The algorithm may diverge or converge to saddle points or relative maxima if the initial starting point \( \hat{\theta}_0 \) is too far from the absolute maximum. The objective function is \( L(\theta) = \log L(\theta) \). The score vector or vector of first derivatives will be denoted as

\[
U'(\theta) = \left( \frac{\partial L(\theta)}{\partial \theta_1}, \frac{\partial L(\theta)}{\partial \theta_2}, \ldots, \frac{\partial L(\theta)}{\partial \theta_p} \right).
\]

The information matrix, which is a matrix of the negative of the second partial derivatives, will be denoted as

\[
I(\theta) = \begin{bmatrix}
I_{11} & I_{12} & \cdots & I_{1p} \\
I_{21} & I_{22} & \cdots & I_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
I_{p1} & I_{p2} & \cdots & I_{pp}
\end{bmatrix},
\]

where \( I_{jk} = -\frac{\partial^2 L(\theta)}{\partial \theta_j \partial \theta_k} \). The iterative procedure is obtained by considering the following second-order Taylor series expansion of the objective function about the point \( \hat{\theta}_i \):

\[
L(\theta) = L(\hat{\theta}_i) + (\theta - \hat{\theta}_i)'U(\hat{\theta}_i) - \frac{1}{2} (\theta - \hat{\theta}_i)' I(\hat{\theta}_i) (\theta - \hat{\theta}_i) + \text{error}.
\]

Since the error term can be considered to be negligible when \( \hat{\theta}_i \) is near to \( \theta \), a good
approximation of $\mathcal{L}(\theta)$ is

$$
\mathcal{L}(\theta) = \mathcal{L}(\hat{\theta}_i) + (\theta - \hat{\theta}_i)' \mathbf{U}(\hat{\theta}_i) - \frac{1}{2} (\theta - \hat{\theta}_i)' \mathbf{I}(\hat{\theta}_i) (\theta - \hat{\theta}_i).
$$

The maximum of this approximation occurs when

$$
\theta = \mathbf{U}(\hat{\theta}_i) - \mathbf{I}(\hat{\theta}_i) (\theta - \hat{\theta}_i).
$$

Assuming $\mathbf{I}(\hat{\theta}_i)$ is positive definite, this equation suggests the following iterative procedure

$$
\hat{\theta}_{i+1} = \hat{\theta}_i + \Gamma^{-1}(\hat{\theta}_i) \mathbf{U}(\hat{\theta}_i),
$$

which is of the form (1.41) where $\alpha_i = 1$ and $d_i = \Gamma^{-1}(\hat{\theta}_i) \mathbf{U}(\hat{\theta}_i)$.

Rao (1984) indicates that convergence problems for the Newton-Raphson procedure may be improved by using the iterative procedure:

$$
\hat{\theta}_{i+1} = \hat{\theta}_i + \alpha_i \Gamma^{-1}(\hat{\theta}_i) \mathbf{U}(\hat{\theta}_i),
$$

where $\alpha_i$ is the step length in the direction $d_i = \Gamma^{-1}(\hat{\theta}_i) \mathbf{U}(\hat{\theta}_i)$ which maximizes the objective function. This modification requires fewer iterations, finds the minimum more often than in the unmodified Newton-Raphson procedure, and avoids saddle points or relative maxima.

Various linear search procedures have been suggested for determining the step size, $\alpha_i$, in such an iterative procedure (Kennedy and Gentle, 1980). The basic problem is to find $\alpha_i$ which maximizes

$$
\rho(\alpha_i) = \mathcal{L}[\hat{\theta}_i + \alpha_i \Gamma^{-1}(\hat{\theta}_i) \mathbf{U}(\hat{\theta}_i)].
$$
The first of these approaches, sometimes referred to as bisection, begins with an interval which brackets \( \alpha_1 \). The length of this interval is repeatedly reduced until \( \alpha_1 \) is determined. Such reductions are made after observing the functional value of \( \rho \) at various \( \alpha \)-values in the interval. The second of the two basic approaches to determining step size involves finding a second- or third-degree polynomial approximation to \( \rho \) and then finding the value of \( \alpha_1 \) which maximizes the function. Each of these approaches to determining the step length becomes computationally expensive since the likelihood function must be reevaluated at many different parameter values.

A compromise to finding the absolute maximum of \( \rho(\alpha_1) \) is step-halving. Originally proposed by Jennrich and Sampson (1968) within the context of the modified Gauss-Newton iterative procedure for least squares (Hartley, 1961), step-halving ensures that the objective function strictly increases and parameter values stay within constraints. If after obtaining

\[
\hat{\theta}_{i+1} = \hat{\theta}_i + \Gamma^{-1}(\hat{\theta}_i) \cdot U(\hat{\theta}_i),
\]

it is determined that \( L(\hat{\theta}_{i+1}) < L(\hat{\theta}_i) \) or \( \hat{\theta}_{i+1} \) is not within constraints, \( \hat{\theta}_{i+1} \) is recalculated using step-halving. The iterate is adjusted to be

\[
\hat{\theta}_{i+1} = \hat{\theta}_i + \frac{1}{2} \, \Gamma^{-1}(\hat{\theta}_i) \cdot U(\hat{\theta}_i)
\]

with \( \alpha_1 = \frac{1}{2} \). Adjustments are repeatedly made to \( \hat{\theta}_i \) until the objective function increases over \( L(\hat{\theta}_i) \) and the parameter values are within constraints. Additional adjustments are made with \( \alpha_1 \) equal to \( 1/4, 1/8, 1/16, 1/32, \ldots \). Once a suitable iterate is determined, the original iteration scheme is resumed. Dennis and Schnabel (1983) discuss a similar approach to step-halving or "back-tracking."
Rao (1973) replaced $I(\hat{\theta}_i)$ by its expected value in the Newton-Raphson iteration and obtained

$$\hat{\theta}_{i+1} = \hat{\theta}_i + \left[ E(\hat{\theta}_i) \right]^{-1} U(\hat{\theta}_i). \quad (1.42)$$

This procedure, known as the method of scoring, converges more slowly than the Newton-Raphson algorithm (Mantel and Myers, 1971).

Let $U_h(\theta)$ and $I_h(\theta)$ denote the contribution by the $h$-th patient to the score vector and information matrix, respectively. Since $E I_h(\theta) = E\{U_h(\theta) U_h'(\theta)\}$,

$$E I(\theta) = \sum_{h=1}^{n} E\{U_h(\theta) U_h'(\theta)\}.$$ 
Therefore (1.42) can be rewritten as

$$\hat{\theta}_{i+1} = \hat{\theta}_i + \left[ \sum_{h=1}^{n} E\{U_h(\theta) U_h'(\theta)\} \right]^{-1} U(\hat{\theta}_i). \quad (1.43)$$

Anderson (1959) approximated $E\{U_h(\theta) U_h'(\theta)\}$ by $U_h(\theta) U_h'(\theta)$ and the resulting iterative procedure was

$$\hat{\theta}_{i+1} = \hat{\theta}_i + \left[ \sum_{h=1}^{n} (U_h(\theta) U_h'(\theta)) \right]^{-1} U(\hat{\theta}_i). \quad (1.44)$$

Petersen (1986) refers to this procedure as the modified scoring algorithm or outer product method and states that it is guaranteed to converge to a local or absolute maximum.

Alternative numerical optimization procedures are available. Kennedy and Gentle (1980) and Chambers (1977) discuss several other variations on the Newton-Raphson algorithm in addition to discussing optimization procedures which require only the score statistics and not the information matrix as in (1.44). In most cases, the convergence time of these algorithms is much longer than that for the Newton-Raphson procedure.
1.4.2 Maximum Likelihood Estimation Using Weighted Least Squares

One of the most commonly used approaches to least squares estimation is the Gauss-Newton algorithm. Consider the nonlinear regression equation:

\[ y_j = \mu(x_j, \theta) + \epsilon_j, \]

for \( j = 1, 2, \ldots, n \), where \( y_j \) is the dependent variable, \( x_j \) is the vector of covariates, \( \theta \) is the \( p \times 1 \) vector of unknown parameters, \( \mu(x_j, \theta) \) is a function of \( x_j \) and \( \theta \), and \( \epsilon_j \) is an error term with mean 0. \( E(y_j) = \mu(x_j, \theta) \) and \( \Sigma \) denotes the covariance matrix for the observations. The least squares objective function is

\[ \text{SSE}(\theta) = (y - \mu(\theta))' \Sigma^{-1} (y - \mu(\theta)) = \epsilon' \Sigma^{-1} \epsilon, \]

where \( y' = (y_1, y_2, \ldots, y_n) \), \( \mu'(\theta) = (\mu(x_1, \theta), \mu(x_2, \theta), \ldots, \mu(x_n, \theta)) \) and \( \epsilon' = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n) \). Since observations are often assumed to be independent and identically distributed, it will be assumed that \( \Sigma = I \). The objective function reduces to

\[ \text{SSE}(\hat{\theta}) = (y - \mu(\hat{\theta}))' (y - \mu(\hat{\theta})) = \phi(\hat{\theta})' \phi(\hat{\theta}), \quad (1.45) \]

where \( \phi(\theta) = y - \mu(\theta) \). The Gauss-Newton iterative procedure is obtained by considering a first-order Taylor series expansion of \( \phi(\theta) \) about \( \hat{\theta}_i \), the \( i \)-th iterate:

\[ \phi(\theta) = \phi(\hat{\theta}_i) + J(\hat{\theta}_i) (\theta - \hat{\theta}_i), \]
where \( J(\hat{\theta}_i) \) is of the form

\[
J(\hat{\theta}_i) = \begin{bmatrix}
\frac{\partial \phi_1}{\partial \theta_1} & \frac{\partial \phi_1}{\partial \theta_2} & \ldots & \frac{\partial \phi_1}{\partial \theta_p} \\
\frac{\partial \phi_2}{\partial \theta_1} & \frac{\partial \phi_2}{\partial \theta_2} & \ldots & \frac{\partial \phi_2}{\partial \theta_p} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial \phi_n}{\partial \theta_1} & \frac{\partial \phi_n}{\partial \theta_2} & \ldots & \frac{\partial \phi_n}{\partial \theta_p}
\end{bmatrix}
\] (1.46)

evaluated at \( \theta = \hat{\theta}_i \). The objective function becomes

\[
SSE(\theta) = \left( \phi(\hat{\theta}_i) + J(\hat{\theta}_i) (\theta - \hat{\theta}_i) \right)' \left( \phi(\hat{\theta}_i) + J(\hat{\theta}_i) (\theta - \hat{\theta}_i) \right).
\]

The gradient vector of \( SSE(\theta) \), denoted as \( \frac{\partial SSE}{\partial \theta} \), is

\[
\frac{\partial SSE}{\partial \theta} = J(\hat{\theta}_i)' \left( \phi(\hat{\theta}_i) + J(\hat{\theta}_i) (\theta - \hat{\theta}_i) \right).
\]

After setting the gradient vector to 0 and rearrangement,

\[
J(\hat{\theta}_i)' \phi(\hat{\theta}_i) = - J(\hat{\theta}_i)' J(\hat{\theta}_i) (\theta - \hat{\theta}_i).
\]

This suggests the following iteration, which is known as the Gauss-Newton iterative procedure for least squares,

\[
\hat{\theta}_{i+1} = \hat{\theta}_i - \left[ J(\hat{\theta}_i)' J(\hat{\theta}_i) \right]^{-1} J(\hat{\theta}_i)' \phi(\hat{\theta}_i).
\] (1.47)
Just as in the Newton-Raphson procedure, the Gauss-Newton iterative procedure can be modified to improve convergence properties. The modified iterative procedure is

$$\hat{\theta}_{i+1} = \hat{\theta}_i - \alpha_i \left[ J(\hat{\theta}_i)' J(\hat{\theta}_i) \right]^{-1} J(\hat{\theta}_i)' \phi(\hat{\theta}_i),$$

where $\alpha_i$ is the maximizing step length (Hartley, 1961). Step-halving, as discussed in the previous section, is the most commonly used approach to determining $\alpha_i$ (Jennrich and Sampson, 1968).

Bradley (1973) showed the equivalence of maximum likelihood and weighted least squares estimation when independent observations are from a one-parameter exponential family. Jennrich and Moore (1975) extended this result to include multi-parameter exponential families and observations which are not independent.

Jennrich and Moore (1975) considered a likelihood function of the exponential form

$$L(t, \theta) = \exp \left( \nu(\theta) + \eta'(\theta) t \right),$$

where $t = (t_1, t_2, \ldots, t_n)$, $\nu$ is a scalar function and $\eta' = (\eta_1, \eta_2, \ldots, \eta_n)$ is a $1 \times n$ vector function. $\nu$ and $\eta$ are both functions of $\theta = (\theta_1, \theta_2, \ldots, \theta_p)$. $E_{\theta}(t) = \mu(\theta) = (\mu_1, \mu_2, \ldots, \mu_n)'$ and $\Sigma = \text{Var}_{\theta}(t)$ denote the mean vector and covariance matrix for the observed statistics. Note that $\Sigma$ is not a function of $\theta$. It was shown that the score vector can be written as

$$U(\theta) = \left( \frac{\partial \mu(\theta)}{\partial \theta} \right)' \Sigma^{-1} \left( t - \mu(\theta) \right),$$

(1.48)

where $\frac{\partial \mu(\theta)}{\partial \theta}$ denotes the matrix whose $ij$-th entry is $\frac{\partial \mu_i}{\partial \theta_j}$ for $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, n$. Maximum likelihood estimates are obtained by setting the left hand side of (1.48) to 0, whereas least square estimates are obtained by setting the right hand side to 0. Jennrich and
Moore also showed that if the mean square error is set to 1 in the least squares estimation procedure, the standard error of the least squares and maximum likelihood estimates are identical. The convergence criteria in least squares computer packages must be modified if it is used for maximum likelihood estimation. Least squares computer programs compare sums of squared deviations at the current and previous iterations, whereas maximum likelihood programs usually compare the current and previous values of the log-likelihood function.

Jennrich and Moore (1975) argued that the relationship between maximum likelihood and least squares estimation could not be further generalized. If \( L(t, \theta) \) were an arbitrary likelihood for which (1.48) held then

\[
\frac{\partial L}{\partial \theta} = \alpha(\theta) + t_1 \beta_1(\theta) + \ldots + t_n \beta_n(\theta)
\]

for some scalar functions \( \alpha(\theta), \beta_1(\theta), \ldots, \beta_n(\theta) \). Upon integration,

\[
\log L(t, \theta) = \nu(\theta) + t_1 \eta_1(\theta) + \ldots + t_n \eta_n(\theta),
\]

where \( \nu(\theta), \eta_1(\theta), \ldots, \eta_n(\theta) \) are the integrals of the associated scalars functions. This implies that the likelihood function is an exponential model.

BMDP Biomedical Computer Programs suggests that use of the Gauss-Newton procedure is appropriate for maximum likelihood estimation based on data from any distribution (Dixon, 1977). Consider the log-likelihood function

\[
L(\theta) = \log L(\theta) = \sum_{j=1}^{n} \log L_j(\theta),
\]

where \( L_j(\theta) \), represents the likelihood function for one individual. The normal equations can
be written as

\[ \sum_{j=1}^{n} \frac{\partial \mu_j(\theta)}{\partial \theta_k} = 0 \quad \text{for} \quad k = 1, 2, \ldots, p, \]

where \( \mu_j(\theta) = \log L_j(\theta) \). If a synthetic dependent variable \( y_j \) is defined as \( \mu_j(\theta) + 1 \), then

\[ 0 = \sum_{j=1}^{n} \frac{\partial \mu_j(\theta)}{\partial \theta_k} (y_j - \mu_j(\theta)) = \sum_{j=1}^{n} \frac{\partial \phi_j(\theta)}{\partial \theta_k} \phi_j(\theta), \]

for \( k = 1, 2, \ldots, p \), where \( \phi_j(\theta) = y_j - \mu_j(\theta) \). The right hand side of this equation is clearly in the form of a derivative of (1.43). Petersen (1986) gives more justification by comparing the iterative procedure obtained from defining these new variables to maximum likelihood iterative procedures. Since

\[ \frac{\partial \phi_j(\theta)}{\partial \theta_k} = - \frac{\partial \mu_j(\theta)}{\partial \theta_k} = - \frac{\partial \log L_j(\theta)}{\partial \theta_k}, \]

for \( k = 1, 2, \ldots, p \), equation (1.44) becomes

\[ J(\hat{\theta}_i)' = \begin{bmatrix} U_1(\hat{\theta}_i) & U_2(\hat{\theta}_i) & \cdots & U_n(\hat{\theta}_i) \end{bmatrix}, \]

where \( U_j(\hat{\theta}_i), \ j = 1, 2, \ldots, n \) is the contribution by the j-th patient to the score vector. In addition,

\[ J(\hat{\theta}_i) \phi(\hat{\theta}_i) = -U(\hat{\theta}_i). \]

Therefore the Gauss-Newton iteration (1.45) is

\[ \hat{\theta}_{i+1} = \hat{\theta}_i + \left[ \sum_{h=1}^{n} \left( U_h(\hat{\theta}_i) U_h'(\hat{\theta}_i) \right)^{-1} U(\hat{\theta}_i) \right]. \]
Equation (1.49) is identical to the iterative procedure that Petersen referred to as the modified scoring algorithm for maximum likelihood estimation.

1.5 Parametric Distributions Applicable to Survival Analysis

A variety of homogeneous parametric models have been used or proposed for use with survival data. The shape of the hazard functions associated with most of these models can be classified into five categories — constant, monotonically increasing, monotonically decreasing, bathtub-shaped, and upside-down bathtub-shaped. In some cases, a hazard function may be able to take on a variety of shapes depending on the values of the unknown parameters.

A constant hazard function occurs when \( \frac{d\lambda(t)}{dt} = 0 \). Gross and Clark (1975) cite individuals whose only risks of death are accidents or rare illness as an example of a population where a constant hazard is applicable.

A hazard function is monotonically increasing when \( \frac{d\lambda(t)}{dt} > 0 \). Many of the parametric models used today have such a shape. This type of hazard function applies to situations where subjects or devices deteriorate with age. Lawless (1982) indicates that populations which display a bathtub-shaped hazard function are sometimes purged of weak individuals, as when manufacturers use a "burn-in" process in order to remove defective items from the population. In such a case, the resulting population will display an increasing hazard function.

A monotonically decreasing hazard function, which occurs when \( \frac{d\lambda(t)}{dt} < 0 \), is not commonly used in practice. Such hazard functions describe individuals which become more robust with age. Gross and Clark (1975) present as an example children undergoing an operative procedure to correct a congenital defect such as a heart defect. Such an example may be subject to criticism. If a mechanical heart valve had been implanted in the child, it may be subject to increased risk of deterioration after a period of several years. In such a case, a bathtub-shaped hazard function may be more applicable.
The bathtub-shaped hazard function, sometimes called a U-shaped or J-shaped hazard function, describes a situation which has three phases: an initial phase during which the hazard rate decreases, followed by a phase during which the hazard rate is constant, and concluded by an increasing hazard phase. If the general population is examined, it would be observed that deaths follow this pattern. Infants are at high risk of mortality after which there is a period of relatively constant low-level of mortality. In old age, the death rate increases.

Bathtub-shaped hazard functions can also occur in clinical trials based on invasive procedures which have an early high risk period. Mathematically, a hazard function is bathtub-shaped if there exists a $t_0$ such that $\frac{d\lambda(t)}{dt} < 0$ for $t < t_0$, $\frac{d\lambda(t)}{dt} \bigg|_{t=t_0} = 0$, and $\frac{d\lambda(t)}{dt} > 0$ for $t > t_0$ (Glaser, 1980).

The upside-down bathtub-shaped hazard function is the counterpart to the bathtub-shaped hazard. It describes a situation involving three phases: an initial phase during which the hazard rate increases, followed by a phase during which the hazard rate is constant, and concluded by a decreasing hazard phase. Gross and Clark (1975) suggest that time to recovery from injury or surgery may be described by an upside-down bathtub-shaped hazard function. Most recoveries are quick, but a few are prolonged. Bennett (1983) observes that an upside-down bathtub-shaped hazard function is appropriate in a study of the curability of breast cancer, where peak mortality occurred after about three years. Mathematically, a hazard function is upside-down bathtub-shaped if there exists a $t_0$ such that $\frac{d\lambda(t)}{dt} > 0$ for $t < t_0$, $\frac{d\lambda(t)}{dt} \bigg|_{t=t_0} = 0$, and $\frac{d\lambda(t)}{dt} < 0$ for $t > t_0$ (Glaser, 1980).

Glaser (1980) developed sufficient conditions under which a given density function $f(t)$ has a monotonically increasing, monotonically decreasing, bathtub-shaped, or upside-down bathtub-shaped hazard function associated with it.

A non-exhaustive summary of parametric models is given in Table 1.1. Many of the models proposed to describe bathtub-shaped hazard functions are not listed in Table 1.1 and will be discussed later. Conditions under which each hazard displays shapes discussed above
### Table 1.1
Summary of Parametric Models Commonly Used in Survival Analysis

<table>
<thead>
<tr>
<th>Model</th>
<th>Hazard</th>
<th>Shape Properties</th>
<th>Nested Distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>$\lambda$</td>
<td>Constant</td>
<td>Exponential ($\lambda_1 = 0$)</td>
</tr>
</tbody>
</table>
| Rayleigh            | $\lambda_0 + \lambda_1 t$ | Monotone Increasing ($\lambda_1 > 0$)  \
|                     |                 | Constant ($\lambda_1 = 0$)                           | Rayleigh ($m-1$)                |
| Generalized Rayleigh| $\sum_{i=0}^{m} \lambda_i t^i$ | Variety of shapes depending on degree of polynomial and values of coefficients | Exponential ($m-0$) |
| Gompertz            | $\exp(a + bt)$  | Monotone Increasing ($b > 0$)  \
|                     |                 | Monotone Decreasing ($b < 0$)  \
|                     |                 | Constant ($b = 0$)                            | Exponential ($b-0$)            |
| Makeham-Gompertz    | $A + \exp(a + bt)$ | Monotone Increasing ($b > 0$)  \
|                     |                 | Monotone Decreasing ($b < 0$)  \
|                     |                 | Constant ($b = 0$)                            | Exponential ($b-0$)            \
|                     |                 |                                                        | Gompertz ($A-0$)                |
| Weibull             | $a b t^b-1$      | Monotone Increasing ($b > 1$)  \
|                     |                 | Monotone Decreasing ($b < 1$)  \
|                     |                 | Constant ($b = 1$)                            | Exponential ($b-1$)            |
| Gamma               | \[ \frac{\lambda}{\Gamma(\gamma)} (\lambda t)^{\gamma-1} \exp(-\lambda t) }{1 - \frac{1}{\Gamma(\gamma)} \int_0^\lambda u^{\gamma-1} \exp(-u) \, du} | Monotone Increasing ($\gamma < 1$)  \
|                     |                 | Monotone Decreasing ($\gamma > 1$)  \
|                     |                 | Constant ($\gamma = 1$)                        | Exponential ($\gamma-1$)        \
|                     |                 |                                                        | Chi-square ($\lambda-1/2, \gamma=\nu/2$) where $\nu$ is an integer |
Table 1.1 (Continued)
Summary of Parametric Models Commonly Used in Survival Analysis

<table>
<thead>
<tr>
<th>Model</th>
<th>Hazard</th>
<th>Shape Properties</th>
<th>Nested Distributions</th>
</tr>
</thead>
</table>
| Generalized  | $\frac{1}{\Gamma(\gamma)} \frac{\lambda \rho}{\rho \gamma - 1} \exp \left[- (\lambda t)^\rho\right]$ | M. decreasing ($\rho \gamma < 1, \rho \leq 1$)  
or ($\rho \gamma - 1, \rho < 1$)  
M. increasing ($\rho \gamma > 1, \rho \geq 1$)  
or ($\rho \gamma - 1, \rho > 1$)  
Constant ($\gamma - \rho - 1$)  
Bathtub ($\rho \gamma < 1, \rho > 1$)  
Upside-down bathtub ($\rho \gamma > 1, \rho < 1$) | Exponential ($\rho = \gamma - 1$)  
Gamma ($\rho - 1$)  
Weibull ($\gamma - 1$)  
Lognormal ($\gamma \rightarrow \infty$) |
| Gamma        | $\frac{1}{1 - \frac{1}{\Gamma(\gamma)} \int_0^\infty \rho \ u^{\rho - 1} \exp \left[- u^{\rho}\right] \ du}$ |                                            |                                 |
| Lognormal    | $\frac{1}{t \sigma \sqrt{2\pi}} \exp \left[- \frac{(\log t - \mu)^2}{2\sigma^2}\right]$ | Upside-down bathtub                      |                                |
| Log-Logistic | $\frac{\lambda \rho (\lambda t)^{\rho - 1}}{1 + (\lambda t)^\rho}$ | Monotone decreasing ($\rho \leq 1$)  
Upside-down bathtub ($\rho > 1$) |                                |
Table 1.1 (Continued)
Summary of Parametric Models Commonly Used in Survival Analysis

<table>
<thead>
<tr>
<th>Model</th>
<th>Hazard</th>
<th>Shape Properties</th>
<th>Nested Distributions</th>
</tr>
</thead>
</table>
| Generalized F | \[
f(t) = \begin{cases} \frac{f(t)}{1 - \int_0^t f(u) \, du} & \text{where } f(t) \\ 1 - \int_0^t f(u) \, du \end{cases}
\]
is as defined below                                                                 | Variety of shapes, including bathtub and upside-down bathtub. Details can be determined from individual nested distributions. | Generalized Gamma \(s_2 \rightarrow \infty\) \(s_2 \rightarrow \infty\) \(s_1 \rightarrow \infty\) |
|                |                                                                        |                                                                                  | Lognormal \(s_2 \rightarrow \infty, s_1 \rightarrow \infty\) | Weibull \(s_2 \rightarrow \infty, s_1 \rightarrow 1\) |
|                |                                                                        |                                                                                  | Exponent \(s_1 \rightarrow \infty, s_1 \rightarrow 1, \sigma \rightarrow 1\) | Gamma \(s_2 \rightarrow \infty, \sigma \rightarrow 1\) |
|                |                                                                        |                                                                                  | F \(\sigma \rightarrow 1\) | Log-Logistic \(s_1 \rightarrow s_2 \rightarrow 1, \sigma \rightarrow 1\) |

\[
f(t) = \frac{\exp \left( -\frac{s_1}{\sigma} \right) t^{s_1/\sigma} - 1}{(s_1/s_2)^{s_1} \left[ 1 + (s_1/s_2)^{s_1} \right]^{1/\sigma}} \left( s_1 + s_2 \right)
\]

\[
f(t) = \frac{\Gamma(s_1) \Gamma(s_2)}{\Gamma(s_1 + s_2)}
\]
are given. Special models nested within the overall framework of the particular parametric model are indicated.

1.5.1 Exponential Distribution

The exponential distribution is the only distribution with a constant hazard. Defined for \( t > 0 \), the hazard function, survival function, and density function are

\[
\lambda(t) = \lambda, \quad S(t) = \exp (-\lambda t), \quad f(t) = \lambda \exp (-\lambda t),
\]

respectively. Since the hazard function is independent of \( t \), the exponential distribution exhibits the unique "memoryless property." A closed form solution for the maximum likelihood estimate of \( \lambda \) under random censorship exists and is:

\[
\hat{\lambda} = \frac{\sum_{i=1}^{n} t_i}{\sum_{i=1}^{n} \delta_i}.
\]

The exponential distribution arises as the limiting distribution of a minimum of samples obtained from densities with support on \((0, \infty)\). This property can sometimes be used as the justification for its use in survival studies for which a complex mechanism fails if any one of its components fails. Due to this property and the ease of estimation the exponential distribution has been widely used in reliability and biomedical applications.

1.5.2 Rayleigh and Generalized Rayleigh Distributions

The Rayleigh hazard function is sometimes referred to as the linear hazard function. The general form is:

\[
\lambda(t) = \lambda_0 + \lambda_1 t.
\]

Lee (1982) indicates that \( \lambda_0 \) and \( \lambda_1 \) need to be values such that \( \lambda(t) \geq 0 \). This includes values
of $\lambda_1$ which are negative. Lawless (1980) argues that since $\Lambda(t)$ must be an increasing function
with $\Lambda(0) = 0$ and $\lim_{t \to \infty} \Lambda(t) = \infty$, the following more stringent constraints are necessary: $\lambda_0 \geq 0$ and $\lambda_1 \geq 0$. Lawless' constraints on the unknown parameters will be assumed. The Rayleigh hazard is monotonically increasing if $\lambda_1 > 0$ and constant if $\lambda_1 = 0$. When the hazard is constant, the distribution is exponential. Figure 1.1 provides a graphical presentation of the variety of shapes generated by the Rayleigh hazard function. The density and survival function for the Rayleigh or linear hazard function are

$$f(t) = (\lambda_0 + \lambda_1 t) \exp \left[ - \left( \lambda_0 t + \frac{\lambda_1 t^2}{2} \right) \right]$$

and

$$S(t) = \exp \left[ - \left( \lambda_0 t + \frac{\lambda_1 t^2}{2} \right) \right],$$

respectively.

Gross and Clark (1975) indicate that "the patterns of middle age groups demonstrate" the Rayleigh hazard. Lee (1982) indicates that the Rayleigh distribution has been used to describe the survival pattern of patients with plasmacytic myeloma and the circulation of milk bottles that are filled in a dairy.

The Rayleigh hazard can be generalized to polynomial hazards of the form:

$$\lambda(t) = \sum_{i=0}^{m} \lambda_i t^i,$$

where the degree of the polynomial $m$ is less than (hopefully much less than) the sample size $n$. Lawless (1982) adds the following constraints to the unknown parameters: $\lambda_0 \geq 0$, $\lambda_m \geq 0$, and $\lambda(t) \geq 0$ for $t \geq 0$. Gross and Clark (1974) indicate that in many ordinary applications $m \leq 3$. Figures 1.1 and 1.2 display shapes generated by the generalized Rayleigh hazard function when $m=2$ and $m=3$, respectively. When $m \geq 2$, this hazard can describe a variety
FIGURE 1.1
SECOND-ORDER GENERALIZED RAYLEIGH HAZARD FUNCTION

\[ \lambda(t) = \lambda_0 + \lambda_1 t + \lambda_2 t^2 \]
THIRD-ORDER GENERALIZED RAYLEIGH HAZARD FUNCTION

\[ \lambda(t) = \lambda_0 + \lambda_1 t + \lambda_2 t^2 + \lambda_3 t^3 \]

---

\[ \lambda_3 = 0.33 \lambda_2 - 1.5 \lambda_1 = 2 \lambda_0 - 1 \]

\[ \lambda_3 = 0.87 \lambda_2 - 8 \lambda_1 = 8 \lambda_0 - 8 \]

---

TIME

\[ \alpha < N < \beta \]
FIGURE 1.3
GOMPertz HAZARD FUNCTION

\[ \lambda(t) = \exp(a + bt) \]
of different shapes, including bathtub-shaped, and many complex non-monotonic shapes.

Lawless suggests that one of the uses of the generalized Rayleigh distribution is in providing adequate fits to data not readily handled by one of the more common models.

Closed form solutions for the maximum likelihood estimates are not available for either the Rayleigh or generalized Rayleigh models. Numerical optimization techniques need to be used.

### 1.5.3 Gompertz and Makeham-Gompertz Distributions

Both the Gompertz and Makeham-Gompertz distributions are widely used in actuarial work. The Gompertz distribution is described by the following hazard function:

\[ \lambda(t) = \exp(a + bt). \]

The density and survival functions are

\[ f(t) = \exp\left[ (a + bt) - b \left( \exp(a + bt) - e^a \right) \right] \]

and

\[ S(t) = \exp\left[ -\frac{e^a}{b} \left( \exp(bt) - 1 \right) \right], \]

respectively. The Gompertz hazard function is monotone increasing when \( b > 0 \), monotone decreasing when \( b < 0 \), and constant when \( b = 0 \). When \( b = 0 \), the Gompertz distribution reduces to the exponential distribution. Shapes of these hazard functions are displayed in Figure 1.3.

The Makeham-Gompertz distribution is a generalization of the Gompertz distribution. The hazard is:

\[ \lambda(t) = A + \exp(a + bt). \]
Elandt-Johnson and Johnson (1980) indicate that the parameter $A$ is added to the hazard "to allow for accidental deaths in addition to deaths from natural causes." The shapes that the Makeham-Gompertz hazard is capable of generating are the same as the Gompertz hazard with the exception that the plot is translated upward $A$ units.

1.5.4 Weibull Distribution

The Weibull distribution is probably the most widely used survival model. It is a generalization of the exponential distribution and is characterized by a shape parameter $b$ and a scale parameter $a$. The hazard function, defined for $t > 0$, is of the following form:

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

where $a, b > 0$. The density and survival functions are

$$f(t) = a \, b \, t^{b-1} \exp\left(-a \, t^b\right)$$

and

$$S(t) = \exp\left(-a \, t^b\right),$$

respectively. The hazard function is monotone increasing, monotone decreasing, and constant when $b > 1$, $b < 1$, and $b = 1$, respectively. When $b = 1$, the exponential distribution is generated. Examples of the shapes that the Weibull hazard function can generate are graphed in Figure 1.4.

Numerical optimization methods are used to find the maximum likelihood estimates since a closed form solution does not exist. Extensive work with the Weibull distribution has been done since there are no two-dimensional sufficient statistics for $a$ and $b$ (Lawless, 1982). In addition, distributions of most estimators are mathematically intractable.
Figure 1.4

Weibull Hazard Function

\[ \lambda(t) = \alpha t^{\beta-1} \]

TIME

0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2

\( \alpha = 1 \) \( \beta = 1 \) 
\( \alpha = 1 \) \( \beta = 2 \) 
\( \alpha = 1 \) \( \beta = 3 \)
FIGURE 1.5
GAMMA HAZARD FUNCTION

\[ \lambda(t) = \frac{\frac{\lambda}{\Gamma(\gamma)} (\lambda t)^{\gamma-1} \exp(-\lambda t)}{1 - \frac{\lambda t}{\Gamma(\gamma)} \int_0^{\lambda t} u^{\gamma-1} \exp(-u) \, du} \]
FIGURE 1.6

GENERALIZED GAMMA HAZARD FUNCTION

\[
\lambda(t) = \frac{\lambda^\rho (\lambda t)^{\rho - 1} \exp \left( - (\lambda t)^\rho \right)}{1 - \frac{1}{\Gamma(\gamma)} \int_0^{\lambda t} \rho u^{\rho - 1} \exp \left( - u^\rho \right) du}
\]
A motivating factor for the use of the Weibull distribution is that it can be developed as the limiting distribution of the minimum of a sample obtained from a continuous distribution with support on \([0, u]\) where \(0 < u < \infty\).

### 1.5.5 Gamma and Generalized Gamma Distributions

The gamma distribution is also a generalization of the exponential distribution. The general form of the hazard function for \(t > 0\) is:

\[
\lambda(t) = \frac{\frac{1}{\Gamma(\gamma)} (\lambda t)^{\gamma - 1} \exp(-\lambda t)}{1 - \frac{1}{\Gamma(\gamma)} \int_0^{\lambda t} u^{\gamma - 1} \exp(-u) \, du}.
\]

where \(\lambda > 0\) and \(\gamma > 0\) are unknown scale and shape parameters, respectively. The hazard function is monotone decreasing from infinity to \(\lambda\) if \(\gamma < 1\), monotone increasing from 0 to \(\lambda\) if \(\gamma > 0\), and constant if \(\gamma = 1\). These properties are exhibited in Figure 1.5. When \(\gamma = 1\), the distribution reduces to the exponential distribution and when \(\lambda = \frac{1}{2}\) and \(\gamma = \nu\) where \(\nu\) is an integer, the distribution reduces to the chi-square distribution.

The presence of the incomplete gamma function in the hazard, density and survival functions makes estimation computationally difficult. Both numerical optimization and numerical integration methods must be used.

The gamma distribution with integer \(\gamma\) can be derived as the distribution of the time till the \(\gamma\)-th failure where each failure is described by an exponential distribution with unknown parameter \(\lambda\).

The gamma distribution can be generalized by incorporation of an additional parameter. The form of the generalized gamma hazard function is:

\[
\lambda(t) = \frac{\frac{1}{\Gamma(\gamma)} (\lambda t)^{\rho \gamma - 1} \exp(- (\lambda t)^\rho)}{1 - \frac{1}{\Gamma(\gamma)} \int_0^{\lambda t} \rho u^{\rho \gamma - 1} \exp(- u^\rho) \, du}.
\]
This distribution includes as special cases the exponential \((\frac{1}{\lambda} = \gamma = 1)\), the gamma \((\frac{1}{\lambda} = 1)\), the Weibull \((\gamma=1)\), and the lognormal \((\gamma \to \infty)\). The hazard function incorporates a variety of shapes as indicated by the variety of nested distributions. Glaser (1980) developed conditions under which the various hazard shapes appear. These are summarized in Table 1.1. The variety of hazard shapes is illustrated in Figure 1.6. Lawless (1982) indicates that the generalized gamma model is useful for discriminating among the various nested models. Two generalized gamma distributions with similar \(\rho\) and \(\lambda\) values, but different \(\gamma\) values can be very much alike. This property of the generalized gamma distribution has resulted in computational difficulties, in addition to those generated by the presence of the incomplete generalized gamma functions. Reparameterization of the problem, as proposed by Prentice (1974), improves estimation, but does not completely solve the computational problems.

### 1.5.6 Lognormal and Log-Logistic Distributions

The general form of the hazard function for the lognormal distribution, defined for \(t > 0\), is:

\[
\lambda(t) = \frac{\frac{1}{t \sigma \sqrt{2\pi}} \exp \left[ -\frac{(\log t - \mu)^2}{2\sigma^2} \right]}{1 - \Phi \left( \frac{\log t - \mu}{\sigma} \right)},
\]

where \(\Phi\) is the cumulative distribution function of a standard normal variable and \(\sigma > 0\). The shape of this hazard function is upside-down bathtub. The density function and survival function are

\[
f(t) = \frac{\frac{1}{t \sigma \sqrt{2\pi}} \exp \left[ -\frac{(\log t - \mu)^2}{2\sigma^2} \right]}{1 - \Phi \left( \frac{\log t - \mu}{\sigma} \right)}
\]

and

\[
S(t) = 1 - \Phi \left( \frac{\log t - \mu}{\sigma} \right),
\]
respectively. Figure 1.7 gives the hazard function for various values of $\mu$ and $\sigma$. Lee (1980) indicates that the lognormal distribution has been used to describe the survival times of several diseases such as Hodgkin's disease and chronic leukemia since they are skewed to the right and the logarithms of survival times are approximately normally distributed.

The lognormal distribution arises as the asymptotic distribution of the product of $n$ independent positive random variables. Lee (1980) indicates that such a property makes the lognormal distribution appropriate for modeling the size of an organism whose growth is subject to many small impulses for which the effect is proportional to the momentary size of the organism.

Kalbfleisch and Prentice (1980) indicate that the lognormal model is simple to use if no censoring occurs since the likelihood function does not involve the incomplete normal integral. With censoring, computations become "formidable." In such case, the log-logistic distribution provides a good approximation to the log-normal model.

The hazard function for the log-logistic distribution is:

$$\lambda(t) = \frac{\lambda \rho (\lambda t)^{\rho - 1}}{1 + (\lambda t)^\rho},$$

where $\lambda > 0$ and $\rho > 0$ are unknown scale and shape parameters, respectively. This hazard is monotone decreasing for $\rho \leq 1$ and upside-down bathtub for $\rho > 1$. Figure 1.8 gives examples of the shapes that the log-logistic distribution can generate. The density and survival functions are

$$f(t) = \lambda \rho (\lambda t)^{\rho - 1} \left[1 + (\lambda t)^\rho\right]^2$$

and

$$S(t) = \frac{1}{1 + (\lambda t)^\rho},$$

respectively.
FIGURE 1.7
LOGNORMAL HAZARD FUNCTION

\[ \lambda(t) = \frac{\exp \left( \frac{-(\log t - \mu)^2}{2\sigma^2} \right)}{t \phi \left( \frac{\log t - \mu}{\sigma} \right)} \]
FIGURE 1.8
LOG-LOGISTIC HAZARD FUNCTION

\[ \lambda(t) = \frac{\lambda \rho (\lambda t)^{\rho - 1}}{1 + (\lambda t)^\rho} \]
Computationally, the log-logistic distribution is much simpler to use than the lognormal distribution since it has closed forms for the density and survival functions. It provides a good approximation to the lognormal distribution except in the extreme tails.

1.5.7 Generalized F Distribution

The generalized F distribution incorporates almost all distributions discussed above. A random variable $T$ has the generalized F distribution if for some $\mu > 0$ and $\sigma > 0$ the transformed variable $(e^{-\mu T})^{1/\delta}$ has the F distribution. When the underlying F distribution has $2s_1$ and $2s_2$ degrees of freedom, the density function of $T$ is

$$f(t) = \frac{\exp\left(-\mu_{s_1}/\sigma\right) t^{(s_1/\sigma) - 1}(s_1/s_2)^{s_1/2} \left[1 + (s_1/s_2)^{s_1}(e^{-\mu t})^{1/\sigma}\right]^{-(s_1 + s_2)}}{\sigma^{s_1/2} \Gamma(s_1) \Gamma(s_2) \Gamma(s_1 + s_2)}.$$  

Among the distributions nested in the generalized F distribution are the generalized gamma, lognormal, Weibull, exponential, gamma, and log-logistic. Table 1.1 provides the details of these special cases. The hazard function of the generalized F generates a great variety of distributional shapes as indicated by the various nested distributions. This distribution has been suggested as a tool for discriminating among the many nested distributions.

The family of distributions generated by the generalized F distribution is sufficiently large that many members resemble each other closely. As a result, optimization is difficult and expensive.

1.5.8 Homogeneous Models with Bathtub-Shaped Hazard Functions

A variety of parametric approaches have been suggested to model a bathtub-shaped hazard function. Three parametric distributions previously discussed can generate such hazard
### Table 1.2

**Additional Parametric Models Which Generate Bathtub-Shaped Hazard Functions**

<table>
<thead>
<tr>
<th>Model (Source)</th>
<th>Hazard</th>
<th>Shape Properties</th>
<th>Nested Distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential-power (Smith &amp; Bain, 1975)</td>
<td>$b \ a^{-b} \ t^{b-1} \ exp \left(\frac{t}{a}\right)^b$</td>
<td>M. increasing ($b \geq 1$)</td>
<td>Bathub ($b &lt; 1$)</td>
</tr>
<tr>
<td>Exponential Family (Glaser, 1980)</td>
<td>$C(a,b,c) \ exp \left(-at - bt^2 + c \ log \ t\right)$</td>
<td>Bathub ($b &gt; 0, c &lt; 0$)</td>
<td></td>
</tr>
<tr>
<td>where $(a,b,c) \in { \ -\infty &lt; a &lt; \infty, b &gt; 0, c &gt; -1 \ } \cup { \ a &gt; 0, b = 0, c &gt; -1 \ }$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cubic Exponential (Glaser, 1980)</td>
<td>$C(a,b,c) \ exp \left(-at - bt^2 - ct^3\right)$</td>
<td>Bathub ($a &lt; 0, c &lt; 1$)</td>
<td></td>
</tr>
<tr>
<td>Gamma Mixture with Common Scale Parameter (Glaser, 1980)</td>
<td>$f(t) = p \ f_1(t) + (1-p) \ f_2(t)$</td>
<td>Bathub ($\gamma_1 &gt; 1, \gamma_2 &lt; 1$) or ($\gamma_1 &gt; 2, \gamma_2 = 1$)</td>
<td></td>
</tr>
<tr>
<td>where $f_i(t) = \frac{\alpha_i}{\Gamma(\gamma_i)} \ t^{\gamma_i-1} \ exp(-\alpha_i t)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gamma Mixture (Glaser, 1980)</td>
<td>$f(t) = p \ f_1(t) + (1-p) \ f_2(t)$</td>
<td>Bathub (Conditions Complicated)</td>
<td></td>
</tr>
<tr>
<td>where $f_i(t) = \frac{\alpha_i}{\Gamma(\gamma_i)} \ t^{\gamma_i-1} \ exp(-\alpha_i t)$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 1.2 (Continued)

**Additional Parametric Models Which Generate Bathtub-Shaped Hazard Functions**

<table>
<thead>
<tr>
<th>Model (Source)</th>
<th>Hazard</th>
<th>Shape Properties</th>
<th>Nested Distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Gaver and Acar, 1979)</td>
<td>( \frac{A}{t+a} + Bt + C )</td>
<td>Constant (A=B=0)</td>
<td>Exponential (A=B=0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M. Increasing (A=0)</td>
<td>Rayleigh (A=0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M. Decreasing (B=C=0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bathtub (A&gt;0, B&gt;0)</td>
<td></td>
</tr>
<tr>
<td>(Hjorth, 1980)</td>
<td>( ct + \frac{a}{1 + bt} )</td>
<td>Constant (b=c=0)</td>
<td>Exponential (b=c=0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M. Increasing (c\geq ab)</td>
<td>Rayleigh (b=0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M. Decreasing (c=0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bathtub (0&lt;c&lt;ab)</td>
<td></td>
</tr>
<tr>
<td>(Murthy, Swartz &amp; Yuen, 1973)</td>
<td>( at^{b-1} + \frac{c}{1 + dt} )</td>
<td>Constant (a=d=0)</td>
<td>Exponential (b-1, c=0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M. Decreasing (a=0, c&gt;0)</td>
<td>Rayleigh (b-2, d=0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M. Increasing (b&gt;1, c=0)</td>
<td>Weibull (c=0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bathtub (a&gt;0, b&gt;1, c&gt;0)</td>
<td></td>
</tr>
<tr>
<td>(Canfield &amp; Borgman, 1975)</td>
<td>( S(t) = \exp \left[ - \sum_{i=1}^{k} b_i t^{a_i} \right] )</td>
<td>Bathtub (k=3, a_1&lt;1, a_2=1, a_3&gt;1) or other conditions</td>
<td></td>
</tr>
<tr>
<td>(Canfield &amp; Borgman, 1975)</td>
<td>( S(t) = \left[ 1 + \sum_{i=1}^{k} \left( \frac{t}{\theta_i} \right)^{\alpha_i} \right]^m )</td>
<td>Bathtub (k=3, a_1&lt;1, a_2=1, a_3&gt;1) or other conditions</td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 1.9

POWER-EXPONENTIAL HAZARD FUNCTION

\[ \lambda(t) = b a^{-b} t^{b-1} \exp \left( \frac{t}{a} \right)^b \]
functions for appropriately chosen parameter values. These include the generalized gamma, generalized F, and generalized Rayleigh distributions. Additional parametric models are summarized in Tables 1.2 and 1.3.

Smith and Bain (1975) developed the exponential-power distribution which is based on the hazard function:

\[
\lambda(t) = b \ a^{-b} \ t^{b-1} \ \exp \left[ \left( \frac{t}{\bar{t}} \right)^b \right],
\]

(1.50)

where \( a, b > 0 \) are unknown parameters. The hazard function for \( 0 < b < 1 \) is Weibull-like for small \( t \), but exponentially grows as \( t \) increases (i.e. bathtub-shaped). It is monotonically increasing when \( b \geq 1 \). Figure 1.9 shows some of the shapes this hazard function can generate. Smith and Bain indicate that “this model appears to be one of very few natural 2-parameter models which allows for a U-shaped” or bathtub-shaped hazard function.

Glaser (1980) considered the exponential family of densities of the form:

\[
f(t) = C(a, b, c) \ \exp \left( -at - bt^2 + c \log t \right),
\]

(1.51)

where the parameter space is the union of \( \{ (a, b, c) \mid -\infty < a < \infty, b > 0, c > -1 \} \) and \( \{ (a, b, c) \mid a > 0, b=0, c > -1 \} \). He showed that this distribution displayed a bathtub-shaped hazard function when \( b > 0 \) and \( c < 0 \) simultaneously. Glaser also examined the cubic exponential density function, i.e.

\[
f(t) = C(a, b, c) \ \exp \left\{ -at - bt^2 - ct^3 \right\}
\]

and showed that the hazard function is bathtub-shaped when both \( a < 0 \) and \( c < 1 \).
Gamma mixtures of the following form were examined by Glaser (1980):

\[ f(t) = p f_1(t) + (1-p) f_2(t) , \]

where \( 0 < p < 1 \). Initially, it was assumed that the density functions comprising the mixture had a common scale parameter \( \alpha \) and varying shape parameters \( \gamma_j \) and were of the following form:

\[ f_j(t) = \frac{\gamma_j}{\Gamma(\gamma_j)} t^{\gamma_j-1} \exp(-\alpha t) , \]

for \( j = 1, 2 \). Glaser showed that the hazard associated with this mixture is bathtub-shaped when either \( \gamma_1 > 1 \) and \( \gamma_2 < 1 \) or \( \gamma_1 > 2 \) and \( \gamma_2 = 1 \). When both the scale and shape parameters are allowed to vary, determination of shapes generated becomes more complicated. Bathtub-shaped hazard functions are possible, but conditions for such generation are not given by Glaser.

Gaver and Acar (1979), Hjorth (1980), and Murthy (1974) considered models which can loosely be described as “mixture hazard” models. The hazard functions from two or more models are summed to create a new “mixture hazard” model.

Gaver and Acar (1979) suggested a hazard model of the form:

\[ \lambda(t) = g(t) + k(t) + C , \]

where \( g(t) \) is a decreasing function of \( t \) such that \( \lim_{t \to \infty} g(t) = 0 \) and \( k(t) \) is an increasing function of \( t \) will generate a bathtub-shaped hazard. In their example, \( g(t) \) and \( k(t) \) were further specified to create

\[ \lambda(t) = \frac{A}{t + a} + Bt + C , \quad (1.52) \]
where $A$, $B$, $a$, and $C$ are unknown positive parameters. The hazard function is constant when $A=B=0$, monotone increasing when $A=0$, and bathtub-shaped when $A, B > 0$. Shapes generated by this hazard function are given in Figure 1.10. This model has several nested distributions, including the exponential model when $A=0$ and $B=0$ and the Rayleigh model when $A=0$. The survival function associated with (1.52) is calculated to be

$$S(t) = \left[ \frac{8}{a + t} \right]^A \exp \left( - \frac{B}{2} t^2 \right) \exp (-Ct). \quad (1.53)$$

Each component of (1.53) could be considered to be the survival function of a random variable, say $T_1$, $T_2$, and $T_3$, respectively. The survival function could be reexpressed as

$$S(t) = P(T_1 > t) P(T_2 > t) P(T_3 > t) = P \left( \min (T_1, T_2, T_3) > t \right). \quad (1.54)$$

Formula (1.54) clearly indicates that hazard (1.53) has a competing risk interpretation. This "mixture hazard" model describes the time till the first failure of three components. Gaver and Acar also indicate that (1.54) suggests an easy procedure for simulation of $T$ by obtaining the smallest value from among the realizations of $T_1$, $T_2$, and $T_3$.

Hjorth (1980) examined a "mixture hazard" model of the form:

$$\lambda(t) = ct + \frac{a}{1 + bt}, \quad (1.55)$$

where $a$, $b$, $c > 0$. Nested distributions include the Rayleigh ($a=0$) and the exponential distribution ($b=c=0$). The hazard function is decreasing when $c=0$, increasing when $c \geq ab$, and bathtub-shaped when $0 < c < ab$. This hazard function is a special case of the hazard function proposed by Gaver and Acar (1979). Hjorth used two different approaches for the development of this hazard function. He showed that the hazard has a competing risk interpretation as did Gaver and Acar (1979) and also a compound distribution interpretation.
FIGURE 1.10
GAVER AND ACAR'S HAZARD FUNCTION

\[ \lambda(t) = \frac{A}{t + a} + Bt + C \]
The latter approach assumes that the hazard rate for a particular individual is

\[ \lambda_u(t) = u + ct, \]

where the initial hazard has a density \( k(u) \). Hjorth considered the gamma distribution to be a convenient choice for \( k(u) \):

\[ k(u) = \frac{u^{a-1} \exp \left( -\frac{u}{b} \right)}{b^a \Gamma(a)}, \]

where \( u \geq 0, \ a, b > 0 \). The survival distribution of \( t \) is

\[ S(t) = \int_0^\infty S_u(t) \ k(u) \ du, \tag{1.56} \]

where \( S_u(t) \) is the survival function associated with \( \lambda_u(t) \). Upon substituting the formulas for \( S_u(t) \) and \( k(u) \) into equation (1.56), it becomes

\[
\begin{align*}
S(t) &= \int_0^\infty \left[ \exp \left\{ - \int_0^t (u + ct) \ du \right\} \frac{u^{a-1} \exp \left( -\frac{u}{b} \right)}{b^a \Gamma(a)} \right] \ du \\
&= \int_0^\infty \left[ \exp \left( -ut - \frac{c^2}{2} t^2 \right) \frac{u^{a-1} \exp \left( -\frac{u}{b} \right)}{b^a \Gamma(a)} \right] \ du.
\end{align*}
\]

Upon further simplification, this formula becomes

\[
S(t) = \frac{\exp \left( -\frac{c^2}{2} t^2 \right)}{b^a \Gamma(a)} \int_0^\infty u^{a-1} \exp \left( -\frac{u (bt + 1)}{b} \right) \ du = \frac{\exp \left( -\frac{c^2}{2} t^2 \right)}{(1 + bt)^a}.
\]

The hazard function associated with this survival function is (1.55).
FIGURE 1.11
MURTHY'S HAZARD FUNCTION

\[ \lambda(t) = a \cdot b \cdot t^{-1} + \frac{c}{1 + dt} \]
Murthy (1974) considered the hazard function:

\[ \lambda(t) = a \, b \, t^{b-1} + \frac{c}{1 + dt}, \tag{1.57} \]

where \( a, b, c, d > 0 \) are unknown parameters. When \( t \) is small the second term of the hazard function dominates, whereas for large \( t \) the Weibull-like first term dominates. The hazard function can be constant (\( a=d=0 \)), monotonically increasing (\( a=0, c>0 \)), monotonically increasing (\( b>1, c=0 \)), and bathtub-shaped (\( a>0, b>1, c>0 \)). The shapes generated by the hazard function are shown in Figure 1.11.

Canfield and Borgman (1975) developed two survival distributions which can generate bathtub-shaped hazard functions. Their development assumed a competing risks framework in which there were \( k \) types of failure mechanisms. The distribution of the number of type \( i \) failures \( i = 1, 2, \ldots, k \) in a population conditional on \( N \) the total number of failures was multinomial. In the first model, \( N \) was assumed to be Poisson distributed with mean \( \lambda \), whereas in the second model, \( N \) was assumed to be distributed as a negative binomial random variable with parameters \( p \) and \( m \). Upon letting \( \lambda \to \infty \) in the first model, the survival function becomes

\[ S(t) = \exp \left[ - \sum_{i=1}^{k} \left\{ \frac{t}{\theta_i^a} \right\} \right] = \exp \left[ - \sum_{i=1}^{k} b_i \, t \, a_i \right], \]

where \( \theta_i \) and \( a_i \) are unknown parameters and \( b_i = \theta_i^{-a_i} \). If \( a_i \) were fixed integers, the distribution is reduced to the generalized Rayleigh. The second of the survival functions is obtained by allowing \( p \to 1 \):

\[ S(t) = \left[ 1 + \sum_{i=1}^{k} \left( \frac{t}{\theta_i^a} \right) \right]^{-m}. \]
Canfield and Borgman indicate that for \( k = 3, a_1 < 1, a_2 = 1, \) and \( a_3 > 1, \) the hazard functions are bathtub-shaped.

Slymen and Lachenbruch (1984) introduced a unique framework within which families of distributions, capable of displaying a wide variety of hazard shapes, were generated. Let \( g(S(t)) \) denote a composite function of \( g \) and the survival function \( S(t) \) where \( g(S(t)) \) satisfies the following properties: (i) \( \lim_{t \to 0} g(S(t)) = -\infty, \) (ii) \( \lim_{t \to \infty} g(S(t)) = \infty, \) and (iii) \( g \) is monotonically increasing. The general form of the investigated families was

\[
g(S(t)) = a + b \ w(t),
\]

where \( -\infty < a < \infty, \ b > 0, \) and \( w \) is a monotonically increasing function characterized by

\[
\lim_{t \to 0} w(t) = -\infty \ \text{and} \ \lim_{t \to \infty} w(t) = \infty.
\]

By specifying the form of \( g(S(t)) \), a family of distributions are defined. The extreme value type and logistic type families were generated using

\[
g_E(S(t)) = \log [-\log S(t)]
\]

and

\[
g_L(S(t)) = \log \left[ \frac{1 - S(t)}{S(t)} \right],
\]

respectively. Distributions within these families were obtained by specifying \( w(t) \) to be \( \log t, \ \log (t-\theta), \ t^{\frac{k}{2k}}, \) and \( \log \left( \exp (kt) - 1 \right). \) It should be noted that \( \lim_{t \to \infty} \frac{k - t^{-k}}{2k} = \log t. \)

Table 1.3 summarizes the distributions which Slymen and Lachenbruch generated. Slymen and Lachenbruch concentrated primarily on the properties of the modified Weibull distribution. The modified Weibull hazard function is

\[
\lambda(t) = 0.5b \left\{ t^{k-1} + t^{-(k+1)} \right\} \exp \left\{ a + b \frac{t^k - t^{-k}}{2k} \right\}.
\]

(1.59)
Table 1.3
Parametric Models Generated by Slymen and Lachenbruch

Extreme Value Type Models where \( g_E(S(t)) = \log \left[ -\log S(t) \right] = a + b \, w(t) \)

<table>
<thead>
<tr>
<th>Model</th>
<th>( w(t) )</th>
<th>Hazard Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weibull*</td>
<td>( \log t )</td>
<td>( b , e^a , t^{b-1} )</td>
</tr>
<tr>
<td>Weibull with threshold parameter</td>
<td>( \log (t-\theta) )</td>
<td>( b , e^a , (t-\theta)^{b-1} )</td>
</tr>
<tr>
<td>Modified Weibull</td>
<td>( \frac{t^k - t^{-k}}{2k} )</td>
<td>( b , \frac{t^{k-1} + t^{-k-1}}{2} \exp \left{ a + b , \frac{t^k - t^{-k}}{2k} \right} )</td>
</tr>
<tr>
<td>Modified Gompertz</td>
<td>( \log \left{ \exp (kt) - 1 \right} )</td>
<td>( b , \frac{k \exp (kt)}{\exp (kt) - 1} \exp \left{ a + b , \log \left{ \exp (kt) - 1 \right} \right} )</td>
</tr>
</tbody>
</table>

* The hazard function presented here can be reparameterized to agree with the form previously given in this text.
Table 1.3 (Continued)

Parametric Models Generated by Slymen and Lachenbruch

Logistic Type Models where $g_L(S(t)) = \log \left( \frac{1 - S(t)}{S(t)} \right) = a + b \ w(t)$

<table>
<thead>
<tr>
<th>Model</th>
<th>$w(t)$</th>
<th>Hazard Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-logistic</td>
<td>$\log t$</td>
<td>$\frac{e^a \ b \ t^{b-1}}{e^a \ t^b + 1}$</td>
</tr>
<tr>
<td>Log-logistic with threshold parameter</td>
<td>$\log (t-\theta)$</td>
<td>$\frac{e^a \ b \ (t-\theta)^{b-1}}{e^a \ (t-\theta)^b + 1}$</td>
</tr>
<tr>
<td>Modified Log-logistic</td>
<td>$\frac{t^k - t^{-k}}{2k}$</td>
<td>$0.5 \ b \left( t^{k-1} + t^{-k-1} \right) \exp \left[ a + b \frac{t^k - t^{-k}}{2k} \right]$ $\exp \left[ a + b \frac{t^k - t^{-k}}{2k} \right] + 1$</td>
</tr>
<tr>
<td>LG</td>
<td>$\log \left( \exp (kt) - 1 \right)$</td>
<td>$\left{ \frac{e^a \ b \ k \ e^{kt} \left[ e^{kt} - 1 \right]^{b-1}}{e^a \left[ e^{kt} - 1 \right]^b + 1} \right}$</td>
</tr>
</tbody>
</table>

* The hazard function presented here can be reparameterized to agree with the form previously given in this text.
FIGURE 1.12

MODIFIED WEIBULL HAZARD FUNCTION

\[ \lambda(t) = 0.5b \left\{ t^{k-1} + (k+1) \right\} \exp \left\{ a + b \frac{t^k}{2k} \right\} \]
This hazard increases sharply from 0 to a maximum and then decreases rapidly, followed by a gradual increase as t becomes large. As k increases, the initial increase reaches a higher maximum and the decrease is steeper. The modified Weibull hazard curves are somewhat bathtub-shaped with the exception of the initial spike. Figure 1.12 is a plot showing the variety of shapes generated by the modified Weibull model. The shape of the modified log-logistic hazard function for k<1 is similar to that of the log-logistic hazard. When k>1, the hazard function may be strictly increasing or initially increase from 0 to a maximum, decrease, and then increase without bound. The modified Gompertz yields bathtub-shaped hazard functions for b<1. The LG hazard function is also capable of generating bathtub-shaped curves. Slymen and Lachenbruch present graphs which display some of the hazard shapes generated by this framework.

Govindarajula (1977) defined a distribution which had a bathtub-shaped-hazard function using the following implicit definition:

\[ t = \theta + \sigma \left\{ (\delta+1) F(t)^\delta - \delta F(t)^{\delta+1} \right\}, \]

for \( \theta \leq t \leq \theta + \delta \), where \( \theta, \sigma, \) and \( \delta \) are unknown parameters. The usefulness of this distribution is questionable due to the limited range of \( t \). In addition, computational problems will ensue since the distribution is not explicitly defined.

### 1.5.10 Regression Models

Many of the previously discussed homogeneous parametric models have been used as bases for parametric regression models. The unknown parameters in the exponential and Weibull distributions have been defined as functions of covariables (see Lawless, 1982). Bennett (1983) has defined the unknown parameters in the log-logistic distribution to be linear functions of the covariables. Exponential and Weibull hazards have been used as baseline hazards in the proportional hazards model.
Regression models have also been developed from the generalized gamma and generalized F distributions since both distributions can generate a variety of shapes. Farewell and Prentice (1977) developed a log-linear regression model for $T$, a generalized gamma random variable. They defined the linear model for $Y = \log T$ to be

$$Y = \mathbf{z} \beta + \sigma W,$$

where $\sigma = 1/p \sqrt{\gamma}$ and $W$ has a log-gamma distribution with density function,

$$f(w) = \begin{cases} 
\frac{\gamma^{1/2}}{T(\gamma)} \exp \left( -\frac{\gamma}{\sqrt{\gamma}} w - \gamma \frac{w}{\sqrt{\gamma}} \right) & 0 < \gamma < \infty \\
(2\pi)^{-1/2} \exp \left( -\frac{w^2}{2} \right) & \gamma = \infty 
\end{cases}$$

Ciampi, Hogg, and Kates (1986) suggested the use of the generalized F distribution as an alternative to the proportional hazards model in the regression analysis of survival data. They considered a log-linear regression model where $\left\{ \exp \left( -\mathbf{z} \beta / T \right) \right\}^{1/\beta}$ had an F distribution. Computational problems as encountered for both the generalized gamma and generalized F regression models were similar to those contended within the homogeneous models.

Nevertheless, Ciampi et al indicated that the generalized F distribution is a good "framework for discriminating among well-known models, for developing goodness-of-fit tests and for studying the robustness of the conclusions drawn on the regression parameters under departure of the error term from a given form." Such a statement is also equally valid for the generalized gamma regression model.

Blackstone, Naftel, and Turner (1986), and in a sense Taulbee (1979), decompose the time-varying pattern of risk into phases which have physical interpretation. Blackstone et al consider there to be k overlapping phases of risk. Each phase $j = 1, 2, \ldots, k$ is shaped by a function of time, denoted as $G_j(t, \omega_j)$, and scaled by a function of concomitant information,
denoted as $\mu_j(x_j, \beta_j)$. The vectors $x_j$ and $\beta_j$, $j = 1, 2, \ldots, k$ denote the unknown shaping and scaling parameters associated with the $j$-th phase. The components of vector $x_j$ are comprised of components from the covariable vector $x$ which are relevant to describing the magnitude of risk during the $j$-th phase. The cumulative hazard function is assumed to be the sum of the cumulative hazards from each of the $k$ phases, i.e.

$$\Lambda(t|x) = \sum_{j=1}^{k} \mu_j(x_j, \beta_j) G_j(t, x_j). \quad (1.60)$$

Blackstone et al develop the model for $k=3$. Three phases were identified— the early phase of decreasing hazard, the constant hazard phase, and the late phase of increasing hazard. The sum of the hazards from the three phases should be bathtub-shaped. For this particular model, Blackstone et al allowed the shaping function for the early and late phases to be the generic shaping functions which had been developed by Hazlerig, Turner, and Blackstone (1982) and generalized by Turner, Hazlerig, and Blackstone (1982). The shaping and scaling function for the early phase were

$$G_1(t, \alpha_1) = \frac{|v| - v}{2|v|} + \frac{v}{|v|} \left[ 1 + \frac{-s}{|m|} \left( \frac{|m| - m}{2|m|} + \frac{\exp(\delta t) - 1}{\delta \rho} \right)^{-1/v} \right]^{1/m} \quad (1.61)$$

and

$$\mu_1(x_1, \beta_1) = \log \left[ 1 + \exp(x_1 \beta) \right],$$

respectively where $\alpha_1 = (v, m, \delta, \rho)$ and $\beta_1$ are unknown parameters. The constant phase was defined with shaping function,

$$G_2(t, \alpha_2) = t,$$

and scaling function,

$$\mu_2(x_2, \beta_2) = \exp(x_2 \beta_2),$$
For the late phase, the generic shaping function was

\[ G_3(t, \varpi_3) = \left\{ \left( 1 + \frac{t}{\tau} \right)^{\gamma} \right\}^{1/v - 1} \eta, \]

(1.62)

where \( \varpi_3 = (\tau, v, \eta) \) is unknown and the scaling function was

\[ \mu_3(\varpi_3, \beta_3) = \exp (\varpi_3 \beta_3). \]

The experience of Blackstone et al. indicates that this model is fairly robust in describing a variety of phase structures. The complete model with no covariables contains 11 parameters. With some foresight and knowledge of the mechanisms under study, the number of parameters can be reduced to make estimation tractable.

Taulbee (1979) proposed a parametric hazard function which was flexible enough so as not to restrict the shape of the function. The hazard function was defined as

\[ \lambda(t|\varpi) = \sum_{i=0}^{m} g_i(\varpi) t^i, \]

(1.63)

where \( m \) is some non-negative integer and \( g_i(\varpi), i = 1, 2, \ldots, m \) are possibly different functions of the covariable vector. The form of this model is somewhat similar to that of Blackstone et al. (1986). Using Blackstone's terminology, (1.63) is a hazard model describing \( m \) phases, each of which is identified with a shaping function, \( t^i \), and a scaling function, \( g_i(\varpi) \). If \( g_i(\varpi) = 1 \), (1.63) becomes the hazard function for the generalized Rayleigh distribution. Given this relationship, it is clear that if \( m \geq 2 \), the hazard function is capable of generating bathtub-shaped hazard functions. In order to facilitate the testing of the hypothesis of proportional hazards, Taulbee defined

\[ g_i(\varpi) = \lambda_i h(\varpi, \beta_i) \]
and substituted it into (1.63) to obtain

$$\lambda(t|{\mathbf{x}}) = \sum_{i=1}^{m} \lambda_i h({\mathbf{x}}, \beta_i) t^i.$$  \hfill (1.64)

In addition to its flexibility in describing the hazard function's shape, model (1.64) allows hazards of individuals with different covariate values to be in differing ratio over time.

Taulbee suggests assuming the form of \(h({\mathbf{x}}, \beta_i)\) to be \(\exp({\mathbf{x}}\beta), (1 + {\mathbf{x}}\beta),\) or \((1 + {\mathbf{x}}\beta)^{-1}\).

Regardless of the choice, estimation of the unknown parameters can be computationally difficult unless a step-up procedure for determining the order of the polynomial is instituted.

Slymen and Lachenbruch (1984) indicate that their method for generating new distributions can be extended to examine the effects of covariates. They suggest the general form:

$$g(S(t)) = \alpha_0 + \sum_{i=1}^{m} \alpha_i z_i + w(t) \left[ \beta_0 + \sum_{j=1}^{k} \beta_j z_j \right],$$  \hfill (1.65)

where \(m\) and \(k\) are fixed integers and \(\alpha_i, i = 0, 1, \ldots, m\) and \(\beta_j, j = 1, 2, \ldots, k\) are unknown parameters. Restrictions on the parameters include \(-\infty < \alpha_i < \infty, i = 0, 1, \ldots, m\) and \(\beta_0 > 0, \beta_0 + \beta_1 > 0, \ldots, \beta_0 + \beta_1 + \ldots + \beta_k > 0\). Slymen and Lachenbruch show that this parameterization when \(k=1\) and \(m=1\) will allow a test of the proportional hazards assumption. The null hypothesis for such a test is \(H_0: \beta_1 = 0\).

1.6 Splines

1.6.1 Introduction to Spline Functions

Polynomials have been shown to inadequately describe data since global properties are determined by what occurs in any small interval. De Boor (1978) states that if the function to be approximated by a polynomial is badly behaved anywhere, then the approximation will be
poor everywhere. High-order polynomials, which exhibit oscillatory behavior, are often used to
describe such "badly behaved" data. The behavior of these polynomials does not usually agree
with the smooth behavior of the underlying physical phenomenon. Polynomial regression also
cannot easily model phenomena which involve structural changes as a function of the
independent variable (Eubank, 1984 and Devlin and Weeks, 1986).

Spline functions have been proposed as an alternative to polynomials. These functions
are piecewise polynomials of degree k joined together at points called knots. Consider m knots
\( v_1 < v_2 < \ldots < v_m \) which define \( m+1 \) intervals \( (-\infty, v_1), [v_1, v_2), \ldots, [v_m, +\infty) \).
A polynomial of degree k is used to describe the data in each interval. Various continuity
conditions, which restrict coefficient values, can be assumed to be satisfied at each of the knots.
The piecewise nature of spline functions results in better description of the data locally without
affecting global properties of the approximation. In addition, spline functions are more
reasonable for approximating data with changing structure. A detailed discussion of spline
functions is given by de Boor (1978).

Spline functions are often defined using a truncated power representation (Eubank,
1984; Smith, 1979). Let

\[
(t - v_i)_+ = \begin{cases} 
  t - v_i & \text{if } t > v_i \\
  0 & \text{otherwise}
\end{cases}
\]

The truncated power representation of a spline of order k with no continuity restrictions is

\[
s_{m,k}(t) = \sum_{j=0}^{k} \gamma_{0j} t^j + \sum_{i=1}^{m} \sum_{j=0}^{k} \gamma_{ij} (t - v_i)_+^j,
\]

(1.66)

where \( \gamma_{ij} \) are unknown coefficients. The presence of \( \gamma_{ij} (t - v_i)_+^j \) in the spline definition
allows a discontinuity of the j-th derivative of the spline function at \( v_i \). Therefore omission of
any term will add a continuity restriction to either the spline function or one of its derivatives
(Smith, 1979). The smoothest spline function of order \( k \) is

\[
sp_{m,k}(t) = \sum_{j=0}^{k} \gamma_{0j} t^j + \sum_{i=0}^{m} \gamma_{ik} (t - v_i)^k .
\] (1.67)

In spline regression, use of the truncated power basis can result in a design matrix which is poorly conditioned. As a result, numerical inaccuracies can occur, especially when many knots are used. Computational difficulties increase as the polynomial degree increases and the number of continuity constraints decreases (de Boor, 1978). An additional disadvantage of truncated power splines is that the number of mathematical operations needed to evaluate the spline function increases as \( t \) increases.

An alternative parameterization of splines is through the use of B-splines. Let \([a,b]\) denote an interval which contains all data and knots. Define \( 2k \) "additional knots":

\[
v_{-(k-1)} = \ldots = v_0 = a \quad \text{and} \quad v_{m+1} = \ldots = v_{m+k} = b .
\]

The values for \( B_{i,k}(t) \), the B-spline of order \( k \) with knots at \( v_i, \ldots, v_{i+k} \), are defined using the following recurrence relation:

\[
B_{i,k}(t) = \frac{t - v_i}{v_{i+k-1} - v_i} B_{i,k-1}(t) + \frac{v_{i+k} - t}{v_{i+k} - v_{i+1}} B_{i+1,k-1}(t)
\]

with

\[
B_{i,1}(t) = \begin{cases} 
1 & \text{if } v_i \leq t < v_{i+1} \\
0 & \text{otherwise}
\end{cases}
\]

(Eubank, 1984). De Boor (1978) shows that the spline function (1.67) can be rewritten in terms of the B-splines as

\[
sp_m(t) = \sum_{j=-(k-1)}^{m} \xi_j B_{j,k}(t) ,
\]
for $a \leq t \leq b$. For the cubic spline, Wold (1974) shows that the B-spline representation is

$$
sp_m(t) = \sum_{i=-1}^{m+2} \xi_i \left[ \sum_{j=i-2}^{i+1} \frac{\prod_{l=i-2}^{j-1} (v_j - v_l)}{(t - v_j)^3} \right].
$$

In spline regression, especially when dealing with large data sets and/or many knots, B-splines are often preferred over truncated power splines due to their attractive computational properties. The evaluation of $sp(t)$ for any $t$ involves only $k$ of the B-splines (Eubank, 1984).

Roundoff errors in one interval are not carried over to subsequent intervals resulting in a mitigation of multicollinearity problems (de Boor, 1978). Among the disadvantages to B-splines are that individual parameters do not have any physical or statistical interpretation, although linear combinations of these parameters have been shown to be meaningful (Smith, 1982). A final disadvantage of B-splines are that inferences are limited to the interval $[a, b]$ since B-splines vanish outside of this interval. Extrapolation beyond the interval $[a, b]$ is only possible if the spline function estimate is reparameterized with the truncated power basis.

The aspects of the spline function under control by the analyst are: the degree of the spline function, $k$; the number, $m$, and position of knots. Wold (1974) indicates that the degree of the spline function depends on what is a realistic assessment of the number of derivatives available in the regression function. He restricts himself to cubic splines since they can describe a variety of relationship given a sufficient number of knots.

Knots may be (i) specified from a priori knowledge of the physical situation (ii) considered as unknown parameters, (iii) determined from the marginal distribution of the data, or (iv) determined from initial ad hoc analyses of the data. Knots associated with the first, third, and fourth situations are referred to as fixed knots, whereas knots which are to be estimated are called variable or free knots. The use of free knots in spline regression requires considerable computation time (Eubank, 1984) and its use "carries the practical danger of overfitting the data" (Smith, 1979). Eubank (1984) indicates that the precise location of knots
is irrelevant in many cases unless the model does not fit the data adequately. Bloxom (1985) indicates that, with hazard function estimation, knots do not need to be close together to give the spline sufficient flexibility to describe a wide variety of hazard function shapes.

Various strategies for selection of the number and location of fixed knots have been proposed in spline regression (Wold, 1974; Stone and Koo, 1986). Wold recommends that knots be added in intervals where residuals are inadmissibly large. Thereafter knot location is systematically varied until the sum of squared residuals is minimized. With small sample size, the analyst's "sound judgement" is the primary source of guidance. From his experience with cubic regression splines, Wold recommends:

1. Use as few knots as possible. Ensure that there are at least 4 or 5 points per interval.

2. Have no more than one extremum point or inflection point per interval.

3. Have extrema centered in intervals, and inflexion points near knots.

Stone and Koo (1986) recommend the use of five knots when fitting a cubic spline with linear tail restrictions to a sample sized "in the hundreds." They recommend the use as knots of the fifth smallest and largest data values when the empirical distribution is "regular". The additional three knots should be equally spaced between these knots. When the empirical distribution is "irregular", it is necessary to locate knots at points of structural change.

Stone and Koo (1986) indicate that the typical spline function is "too flexible in the tails in relation to the amount of noisy data available there". They suggest restricting the tails to be linear by making the second derivative of the spline function vanish at the smallest and largest knots. Devlin and Weeks (1986) present, without proof, the following form of this restricted cubic spline using a truncated power basis:

\[ sp_m(t) = \alpha + \beta t + \sum_{i=1}^{m-2} \gamma_i \left[ (t - v_i)^3 - \frac{(t - v_{m-1})^3}{(v_m - v_{m-1})} \right] + \frac{(t - v_m)^3}{(v_m - v_{m-1})} \left( \frac{v_m - v_{m-1}}{v_m - v_{m-1}} \right). \]
1.6.2 Spline Estimation of the Hazard Function

The introductory discussion on spline functions has emphasized their use in a general setting and in regression applications. However, spline functions have also been applied in other areas of statistics, including density estimation, time series, and hazard function estimation. Wegman and Wright (1983) give details about the areas of density estimation and times series; these areas of application will not be further discussed. Anderson and Senthilselvan (1980), Bloxom (1985), and Etesadi-Amoli and Ciampi (1987) propose the use of spline functions to estimate hazard functions.

Given an estimate of \( \beta \) in Cox's (1972) proportional hazards model, Anderson and Senthilselvan estimate the baseline hazard function using penalized maximum likelihood estimation. A roughness penalty function was subtracted from the log-likelihood and maximized. This penalty function was

\[
K_0 \int \left[ \frac{\partial \lambda(t)}{\partial t} \right]^2 dt,
\]

where \( K_0 \) is a positive constant determining the amount of smoothing to be applied to the data. Given that the baseline hazard is restricted to be continuous and the first derivative to be piecewise continuous, the solution to such an estimation procedure is a quadratic spline with knots at each survival time. A shortcoming of this method is that the hazard function can take on negative values for certain values of \( K_0 \). These negative values, however, do not occur at any failure times. Anderson and Senthilselvan suggest that the problem can be avoided by increasing the value of the smoothing parameter or by considering the hazard function to be zero where the hazard function is negative.

Bloxom (1985) imposes constraints on the coefficients of the quadratic spline when it is used to approximate the hazard function in a homogeneous setting. These constraints specify the shape, the number of peaks, and/or inflections a priori. Knots were placed at each decile of the sample's distribution. The results of a simulation study using uncensored data showed that
there is a positive bias in the middle of the distribution. Bias is also large when the underlying hazard is relatively large.

Etezadi-Amoli and Ciampi (1987) propose the extended hazard regression model which incorporates both the proportional hazards and accelerated failure time models. The basic structure of the model is

\[ \lambda(t; \mathbf{z}) = g_1(\mathbf{x} \cdot \mathbf{z}) \lambda_0( g_2( \mathbf{\beta} \cdot \mathbf{z} ) t ) , \]

where \( g_1 \) and \( g_2 \) are positive functions equal to 1 when the argument is zero, \( \lambda_0(t) \) is the baseline hazard function, and \( \mathbf{x} \) and \( \mathbf{\beta} \) are vectors of regression parameters. A quadratic spline approximation with variable knots is used to approximate the baseline hazard function, which is forced to remain nonnegative in the range of observed data. Their general approach to model fitting is to estimate the baseline hazard first and then introduce covariates. The variable knots are entered into the model stepwise until the value of the likelihood does not vary substantially. These researchers claim that no special numerical difficulties were encountered in the determination of the values of the variable knots.

### 1.7 Research Topic

Cox's proportional hazards model (1972) is the most popular approach for describing the relationship between survival time and fixed covariates. Estimation is efficient even though no assumptions are made about underlying distributional shape. However, the inclusion of time-dependent covariates is computationally expensive when the sample size is large.

An alternative modeling approach when time-dependent covariates are present is a parametric model, such as the proportional hazards model in which the form of the baseline hazard function is specified. The most commonly used parametric models, such as the exponential and Weibull proportional hazards models, have monotonic hazard functions (see
Section 1.5 and Table 1.1) and cannot adequately describe many hazard functions, such as bathtub-shaped hazards. Some of the more flexible models, such as the generalized F, generalized gamma, and phase component models, have previously been discussed. In most cases, there are many parameters to be estimated. Computational problems ensue, and often no unique solution to the maximum likelihood equations exists.

The purpose of the following research is to develop a parametric hazard model which (i) can describe a variety of hazard function shapes, including monotonic and non-monotonic hazards, (ii) has a minimal number of parameters to be estimated, and (iii) is robust, (iv) can easily incorporate time-dependent covariables. The model which will be considered is a proportional hazards model where the form of the baseline hazard is specified parametrically. This baseline hazard will be able to generate a variety of hazard function shapes. Initially, two models will be considered as candidates for the baseline hazard. The first is the generalized Weibull hazard model, which is a "mixture hazard" model with two Weibull hazard components. The second model is a cubic spline function with linearity restrictions in the tails.

In chapter II, these two models will be developed in a homogeneous setting, ignoring covariables. Due to computational problems with the generalized Weibull model, only the restricted cubic spline hazard model will be further developed in subsequent chapters. The proportional hazards model with cubic spline baseline hazard will be examined in Chapter III when fixed covariables are incorporated into the model. Time-dependent covariables will be incorporated into the model in Chapters IV. Results and proposals for future research will be summarized in Chapter V.
Chapter II
Two Homogeneous Hazard Models

2.1 Introduction

Two homogeneous hazard models, the generalized Weibull and restricted cubic spline hazard functions, will be presented and examined in this chapter. Each model will be fit to data generated from a variety of underlying distributions. These underlying hazard functions will include monotone increasing, monotone decreasing, and bathtub-shaped hazard functions. The fit of these models will be examined, but not rigorously, since the homogeneous model is not of principle interest in this research.

The structure of the generalized Weibull hazard function will be presented in Section 2.2.1, along with motivation for consideration of this particular model. Sections 2.2.2 and 2.2.3 will discuss the problems associated with estimation of the model's unknown parameters. It will be clear after these sections that further development of the generalized Weibull model is unwarranted.

The remaining portion of Chapter II will develop the cubic spline hazard function with linear tails. The structure of this model will be developed in Section 2.3.1. The log-likelihood function and the approach to parameter estimation will be briefly discussed in Section 2.3.2. Confidence intervals for the hazard, cumulative hazard, and survival functions will be presented in Section 2.3.3. Section 2.4 will discuss issues of modeling, including knot placement and goodness of fit. Efficiency of survival function estimation will be examined in Section 2.5. The use of the restricted cubic spline hazard function will be demonstrated with two cardiovascular disease data sets in Section 2.6.
2.2 Generalized Weibull Hazard Function

2.2.1 Motivation for the Generalized Weibull Hazard Model and Its Structure

The need to describe the survival experience of a group of patients who had undergone coronary artery bypass grafting with a model which could easily incorporate time-dependent covariables was a major motivation for the research which was conducted. The hazard function associated with these patients is initially high during and immediately after the surgical procedure. This high risk tapers off, but later it begins rising presumably due to the increasing atherosclerotic plaque in the new coronary arterial graft. Two competing death mechanisms are in play. Each of the “mixture hazard” models proposed by Gaver and Acar (1979), Murthy (1974), and Hjorth (1980) and many of the bathtub-shaped hazard functions discussed in Section 1.5.8 might be able to adequately describe this competing hazard situation. However, each individual death mechanism is probably best described by a Weibull hazard function. The hazard function which would be generated if each component is individually modeled using a Weibull hazard function is

\[ \lambda(t) = a b t^{b-1} + c d t^{d-1} \]

(2.1)

where \( b, d > 0 \) and \( a, c \geq 0 \).

This model, which will be referred to as the generalized Weibull model, is a special case of the first model described by Canfield and Borgman (1975). It can generate a variety of hazard function shapes, including constant, monotonically increasing, monotonically decreasing, and bathtub-shaped. The conditions under which these shapes occur are noted
below.

<table>
<thead>
<tr>
<th>Generated Shape</th>
<th>Parameter Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>a = 0, d = 1, c &gt; 0</td>
</tr>
<tr>
<td></td>
<td>a &gt; 0, b = 1, c = 0</td>
</tr>
<tr>
<td></td>
<td>a, c &gt; 0, b = d = 1 †</td>
</tr>
<tr>
<td>Increasing</td>
<td>a = 0, c &gt; 0, d &gt; 1</td>
</tr>
<tr>
<td></td>
<td>c = 0, a &gt; 0, b &gt; 1</td>
</tr>
<tr>
<td>Decreasing</td>
<td>c = 0, a &gt; 0, b &lt; 1</td>
</tr>
<tr>
<td></td>
<td>a = 0, c &gt; 0, d &lt; 1</td>
</tr>
<tr>
<td>Bathtub-Shaped</td>
<td>a, c &gt; 0, b &lt; 1, d &gt; 1</td>
</tr>
</tbody>
</table>

† Computationally, the two components of this model will be indistinguishable.

Figures 2.1a and 2.1b show some of the shapes that the generalized Weibull hazard function can generate.

Nested within the generalized Weibull distribution are the exponential, Rayleigh, and Weibull distributions. The exponential distribution arises when b=1 and c=0 or when d=1 and a=0, the Rayleigh distribution arises when b=2 and d=1, and the Weibull distribution arises when c=0 or a=0.

The “competing risk” interpretation of the generalized Weibull hazard function can be exploited to generate data from this distribution. Suppose $T_1$ and $T_2$ describe the survival time associated with the first and second component of (2.1), respectively. The survival time for $T$ can be shown to be

$$S(t) = P(T>t) = P(T_1>t) P(T_2>t) = P(\min(T_1, T_2)>t).$$

Therefore the minimum of two Weibull random variates is a generalized Weibull random
GENERALIZED WEIBULL HAZARD FUNCTION

\[ \lambda(t) = a \cdot b \cdot \frac{1}{c + d} \cdot t \]

FIGURE 2.1a
FIGURE 2.1b

GENERALIZED WEIBULL HAZARD FUNCTION (CONTINUED)

\[ \lambda(t) = a b t^{b-1} + c d t^{d-1} \]
variate. A uniform random variate, \( U \), is transformed by

\[
\left[ \frac{-\log U}{b} \right]^\frac{1}{\delta}
\]

to obtain a Weibull random variate (Hoffman, 1985).

### 2.2.2 Parameter Estimation and Problems with Model Fitting

Maximum likelihood estimation of the unknown parameters of the generalized Weibull hazard model involves maximizing the following log-likelihood function:

\[
L = \sum_{i=1}^{n} l_i = \sum_{i=1}^{n} \left[ \delta_i \log \left( a \, b \, t_i^{b-1} + c \, d \, t_i^{d-1} \right) - a \, t_i^b - c \, t_i^d \right], \tag{2.2}
\]

where \( t_i \) is the observed survival time and \( \delta_i \) is an indicator function denoting whether \( t_i \) is censored \( (\delta_i = 0) \) or not \( (\delta_i = 1) \). Numerical optimization procedures must be used to find the maximum likelihood estimate since closed form solutions do not exist. The score vector and information matrix necessary for many of the optimization procedures can be found in Appendix 1.

The first step in the research concerning the generalized Weibull model was to determine the "best" approach to parameter estimation. Such a procedure should consistently converge to the maximum likelihood estimate in a "small" number of iterations for a variety of hazard function shapes.

Data were generated from several underlying generalized Weibull distributions. The hazard functions from these distributions were monotonically increasing, monotonically decreasing, and bathtub-shaped. Several optimization algorithms were used to estimate the unknown parameters. These included the Newton-Raphson algorithm, the modified scoring algorithm, and a constrained Newton-Raphson algorithm described by Scales (1985). Step-halving as described in Section 1.4.1 and a variation on step-halving or back-tracking as
described by Dennis and Schnabel (1983) were used in conjunction with these procedures. Different initial estimates of the unknown parameters were also considered. These included estimates assuming an exponential model, least squares estimates, grid searches, and estimates assuming a Weibull model. This latter approach entails a two-stage optimization procedure—the first for the Weibull model and the second for the generalized Weibull model.

None of these iterative procedures converged for the majority of the simulated data sets. Convergence was especially difficult when the underlying distribution was Weibull-like or nearly Weibull-like, regardless of the initial estimate. In these situations, the maximum likelihood estimate was on the boundary of the parameter space or very near the boundary. As intermediate estimates approached this boundary, the size of the step generally became large. The boundary was often overstepped, and step-halving invoked in order to return the estimate to the parameter space. Sometimes the step put the next iteration in the feasible parameter space, but far away from the boundary and the maximum likelihood estimate.

The log-likelihood function within the parameter space is not “smooth” enough for the information matrix to be positive definite at all parameter points. Without this property of positive definiteness at a particular iteration, subsequent Newton-Raphson iterations do not increase the log-likelihood. The initial estimate needs to be carefully chosen to ensure that the information matrix associated with this estimate is positive definite. A greater problem is to ensure steps are small enough so that parameter estimates remain within the positive definite region. As indicated above, steps near the boundary are sometimes too “large”.

Since the two components of the generalized Weibull hazard function have the same structure, there are expected to be identifiability problems. However, when additional constraints on the parameters, such as $0 < b \leq 1$ and $1 < d < \infty$, were added, the convergence problems mentioned above were not resolved.

A final observation concerning this model is that as $c$ (a) approaches 0, parameter $d$ (b) becomes irrelevant. Therefore when the underlying model is a Weibull model, the set of maximum likelihood estimates is a hyperplane. Thus the second component in the generalized
Weibull model seems to be the source of many of the estimation problems. A method is needed to determine whether the second component of the generalized Weibull model is necessary for describing the data. Such a procedure would reduce the need for fitting the $c$ and $d$ parameters when a Weibull distribution fits the data.

### 2.2.3 Preliminary Test of Weibullness

Three approaches to the development of a test to determine whether the data are adequately described by a Weibull model will be presented. These approaches include the standard score test and a modified score test, both within the context of the generalized Weibull model, and a score test within the context of a piecewise Weibull model.

Standard score test theory is not applicable to testing

$$H_0: c=0 \quad \text{or} \quad H_0: c=0 \text{ and } d=d_0$$

since parameter values under the null hypothesis are on the boundary of the parameter space (Cox and Hinkley, 1974). However, it is thought that the score statistic might be adaptable for current use, though under the null hypothesis, it might have a distribution different from the chi-square distribution with 1 degree of freedom. Let $\hat{a}_w$ and $\hat{b}_w$ be the maximum likelihood estimates under the null hypothesis, i.e. from the Weibull model. The score vector under the null hypothesis can be shown to be

$$U'(\hat{a}_w, \hat{b}_w, c=0, d) = \begin{pmatrix} 0 & 0 & U_c & 0 \end{pmatrix},$$

where $U_c$ is a non-zero component. The information matrix has the following structure:

$$I\left(\hat{a}_w, \hat{b}_w, c=0, d\right) = \begin{bmatrix} I_{11} & I_{12} & I_{13} & 0 \\ I_{12} & I_{22} & I_{23} & 0 \\ I_{13} & I_{23} & I_{33} & 0 \\ 0 & 0 & 0 & I_{44} \end{bmatrix}.$$
where $I_{11}, I_{12}, \ldots, I_{33}, I_{44}$ are non-zero elements. The inverse of this information matrix is

$$I^{-1}(\hat{a}_w, \hat{b}_w, c=0, d) = \begin{bmatrix} I_{11} & I_{12} & 0 & I_{14} \\ I_{12} & I_{22} & 0 & I_{24} \\ 0 & 0 & 0 & I_{34} \\ I_{14} & I_{24} & I_{34} & I_{44} \end{bmatrix},$$

where $I_{11}, I_{12}, I_{22}, I_{14}, \ldots, I_{44}$ are non-zero elements. Therefore the score test statistic is

$$S = I^{-1}(\hat{a}_w, \hat{b}_w, c=0, d) \cdot U(\hat{a}_w, \hat{b}_w, c=0, d) = 0.$$

The standard score test statistic can be removed from consideration as a preliminary test of Weibullness since it is a constant.

The second approach to preliminary test development is motivated by the fact that parameter $d$ is causing much of the instability in the numerical optimization procedures. It seems reasonable to fix $d$ as a constant, say $d_0$, prior to testing the hypothesis $H_0: c=0$. Such a procedure would need to be repeated for a variety of $d_0$'s.

Davies (1977, 1987) develops such a procedure for hypothesis testing when a nuisance parameter, such as $d$, is present only under the alternative hypothesis. Suppose $\theta$ is such a nuisance parameter when testing $H_0: \xi=0$. Further, suppose that when $\theta$ is known, an appropriate test statistic for testing $H_0: \xi=0$ is $\mathcal{I}(\theta)$. Davies proposes the following test statistic:

$$\sup_{\theta} \mathcal{I}(\theta),$$

where $\theta \in (L, U)$ to test $H_0: \xi=0$ when $\theta$ is unknown. He assumes that for each $\theta$, $\mathcal{I}(\theta)$,
has a standard normal distribution under the null hypothesis, and is continuous with a continuous derivative except possibly for a finite number of jumps in the derivative. If \( M = \max_{\theta} \mathcal{L}(\theta) \) and

\[
V = \sum_{i=1}^{m} | \mathcal{L}(\theta_i) - \mathcal{L}(\theta_{i+1}) | ,
\]

where \( \theta_1, \theta_2, \ldots, \theta_m \) are "successive turning points of \( \mathcal{L}(\theta) \)" and \( \mathcal{L}(\theta_0) = \mathcal{L}(L) \) and \( \mathcal{L}(\theta_m) = \mathcal{L}(U) \), an estimate of the significance probability is obtained as

\[
\Phi(-M) + V \exp \left( -\frac{1}{2} \frac{M^2}{\sqrt{2\pi}} \right) .
\]

(2.3)

The form of Davies' test statistic for the current problem is

\[
\sup_{d_0 > 0} \sqrt{\chi^2(c=0 \mid d=d_0)} ,
\]

where \( \chi^2(c=0 \mid d=d_0) \) is a score test statistic for the hypothesis \( H_0: c=0 \) given the model

\[
\lambda(t) = a \cdot b \cdot t^{b-1} + c \cdot d_0 \cdot t^{d_0-1} .
\]

It can be shown that

\[
\lim_{d_0 \rightarrow 0} \chi^2(c=0 \mid d=d_0) = n ,
\]

where \( n \) is the sample size. Therefore

\[
M = \sup_{d_0 > 0} \sqrt{\chi^2(c=0 \mid d=d_0)} \geq \sqrt{n} .
\]

This implies that the significance level given in (2.3) is less than \( \Phi(-\sqrt{n}) \), regardless of
observed data values. Clearly additional research needs to be done to investigate how Davies’
approach could be properly adapted to the current problem.

The third approach to the development of a preliminary test is based on some of the
properties of the generalized Weibull hazard function. For small \( t \), one component of the
generalized Weibull hazard function dominates, whereas for large \( t \), the other component
dominates. This hazard can be approximated by a piecewise Weibull hazard function, i.e.

\[
\lambda(t) = a \ b \ t^{b-1} I(t \leq \tau) + c \ d \ t^{d-1} I(t > \tau),
\]

(2.4)

where \( \tau \) is some convenient choice such as median survival time. Continuity of the hazard
function at \( \tau \) implies

\[
a \ b \ \tau^{b-1} = c \ d \ \tau^{d-1}
\]

or

\[
c = \frac{a \ b}{d} \ \tau^{b-d}.
\]

Equation (2.4) can be rewritten as

\[
\lambda(t) = a \ b \ t^{b-1} I(t \leq \tau) + \frac{a \ b}{d} \ \tau^{b-d} \ d \ t^{d-1} I(t > \tau).
\]

A test of the hypothesis \( H_0: b=d \) determines whether the two Weibull components are the
same. Standard score test theory is applicable in this situation. Unfortunately, the results of
this test were not fully compatible with the computational problems discussed in the previous
section. There were data sets for which the score test indicated that the two components of the
piecewise Weibull model were not the same and for which maximum likelihood estimates of the
generalized Weibull model were not obtainable due to convergence problems. On the other
hand, there were data sets for which the score test indicated that the Weibull model was
adequate though a non-Weibull estimate was obtained from optimization of the generalized Weibull log-likelihood. Clearly, the current structure of this preliminary test is not adequate for its conceived purposes.

At the current moment, there is no “good” approach to estimation of the parameters of the generalized Weibull model. Additional research needs to be done to determine a good estimation approach. However, this model will be put aside, as the spline model, which will be discussed in the next section, has more promise in solving the original research problem.

2.3 Restricted Cubic Spline Hazard Function

Spline functions are an ideal parametric function for modeling hazards in that they can describe a variety of structure with a minimal number of parameters. Use of the quadratic spline with fixed knots as an approximation of the hazard function has previously been proposed and was discussed in Section 1.6.1. However, Wold (1974) advocates the use of cubic splines since they can describe a greater variety of shapes. Stone and Koo (1986) indicate that since ordinary cubic splines can have anomalous behavior in the extreme tails, these tails should be linearly restricted. Etezadi-Amoli and Ciampi (1987) propose an extended hazard regression model where the baseline hazard function is approximated by a quadratic spline with variable knots. As indicated in section 1.6.1, the use of variable knots requires considerable computation time and carries with it the practical danger of overfitting the data.

The research in this paper will involve examining the properties of the cubic spline function with linearly restricted tails and fixed knots as it is used to estimate the hazard function. The research mentioned above suggests that such a model should be able to describe a variety of hazard function shapes with a minimal number of parameters. Since knots will be fixed, computation time will be less than that associated with the model proposed by Etezadi-Amoli and Ciampi (1987). It seems reasonable that the restricted cubic spline model could be incorporated into the proportional hazards model with time-dependent covariables as the baseline hazard without encountering the computational expense associated with the Cox
model. In this chapter, the spline model’s application to survival data ignoring covariates will be examined. In future chapters, the spline function will be incorporated into the proportional hazards model as the baseline hazard function.

2.3.1 Derivation of Hazard Function

The linearity restrictions in the tails of the cubic spline function with m knots, \( sp_m(t) \), as proposed by Stone and Koo (1986) imply that

\[
sp_m''(t) = sp_m'''(t) = 0,
\]

for \( t \leq v_1 \) and \( t \geq v_m \), where \( v_1 \) and \( v_m \) are the first and last knots, as described in section 1.6.1. The equivalent parameter restrictions are determined and presented below. In addition, the effect of these restrictions on the equation describing the cubic spline is shown.

Equation (1.67) implies that the smoothest cubic spline function is given by

\[
sp_m(t) = a + bt + ct^2 + dt^3 + \sum_{j=0}^{m} \gamma_j (t - v_j)^3_+.
\]  
(2.5)

The second and third derivatives of this function are

\[
sp_m''(t) = 2c + 6dt + \sum_{j=0}^{m} 6 \gamma_j (t - v_j)_+,
\]

and

\[
sp_m'''(t) = \sum_{j=0}^{m} 6 \gamma_j \frac{(t - v_j)_+}{t - v_j}.
\]

The linearity restrictions for \( t \leq v_1 \) imply that \( sp_m''(t) = 2c + 6dt = 0 \) and \( sp_m'''(t) = 6d = 0 \).
Therefore
\[ c = d = 0 . \] (2.6)

For \( t \geq v_m \),
\[ s_p^{''}(t) = \sum_{j=0}^{m} 6 \gamma_j ( t - v_j )^+ = t \sum_{j=0}^{m} 6 \gamma_j - \sum_{j=0}^{m} 6 \gamma_j v_j = 0 \] (2.7)

and
\[ s_p^{'''}(t) = \sum_{j=0}^{m} 6 \gamma_j = 0 . \] (2.8)

Equation (2.8) can be rewritten as
\[ \gamma_m = - \gamma_1 - \gamma_2 - \cdots - \gamma_{m-1} . \] (2.9)

Equation (2.9) indicates that the first term in (2.7) is 0. Upon substituting (2.9) into the remaining portion of (2.7), it can be shown that
\[ \gamma_1 ( v_1 - v_m ) + \gamma_2 ( v_2 - v_m ) + \cdots + \gamma_{m-1} ( v_{m-1} - v_m ) = 0 . \] (2.10)

Equation (2.10) can be used to solve for \( \gamma_{m-1} \) as a function of \( \gamma_1, \gamma_2, \ldots, \gamma_{m-2} \) and the knot points:
\[ \gamma_{m-1} = \frac{v_m - v_{m-1}}{v_m - v_{m-1}} \left[ \gamma_1 ( v_1 - v_m ) + \gamma_2 ( v_2 - v_m ) + \cdots + \gamma_{m-2} ( v_{m-2} - v_m ) \right] . \] (2.11)

Substitution of (2.11) into (2.9) gives
\[ \gamma_m = \frac{1}{v_m - v_{m-1}} \left[ \gamma_1 ( v_1 - v_{m-1} ) + \gamma_2 ( v_2 - v_{m-1} ) + \cdots + \gamma_{m-2} ( v_{m-2} - v_{m-1} ) \right] \] (2.12)
Equations (2.6), (2.11), and (2.12) are the parameter restrictions imposed by linearity in the tails. These restrictions reduce the number of unknown parameters by 4. Substitution of these constraints into (2.5) yields

\[
sp_m(t) = a + b \, t + \sum_{j=1}^{m-2} \gamma_j \left( t - v_j \right)_+^3
+ \frac{1}{v_m - v_{m-1}} \left( v_m - v_{m-1} \right)_+^3 \sum_{j=1}^{m-2} \left[ \gamma_j \left( v_j - v_m \right) \right]
- \frac{1}{v_m - v_{m-1}} \left( t - v_m \right)_+^3 \sum_{j=1}^{m-2} \left[ \gamma_j \left( v_j - v_{m-1} \right) \right].
\] (2.13)

When all terms associated with a particular unknown parameter are combined, the following form of the restricted cubic spline function is generated:

\[
sp_m(t) = a + b \, t + \sum_{j=1}^{m-2} \gamma_j \, w_j(t),
\] (2.14)

where

\[
w_j(t) = \left( t - v_j \right)_+^3 - \frac{\left( t - v_{m-1} \right)_+^3 \left( v_m - v_j \right)}{\left( v_m - v_{m-1} \right)} + \frac{\left( t - v_m \right)_+^3 \left( v_{m-1} - v_j \right)}{\left( v_m - v_{m-1} \right)}. \] (2.15)

This result for the restricted cubic spline is presented by Devlin and Weeks (1986) without proof.

### 2.3.2 Maximum Likelihood Estimation of Unknown Parameters

The following log-likelihood function is maximized to obtain estimates of the unknown parameters in the restricted cubic spline hazard function (2.14) with m knots:

\[
L = \sum_{i=1}^{n} \mathcal{L}_i = \sum_{i=1}^{n} \left[ \delta_i \log \lambda(t_i) - \Lambda(t_i) \right].
\] (2.16)
The hazard function \( \lambda(t) \) is the spline function given by (2.14) and the cumulative hazard function is

\[
\Lambda(t) = at + bt^2 + \sum_{j=1}^{m-2} \gamma_j \mathcal{W}_j(t),
\]

(2.17)

where

\[
\mathcal{W}_j(t) = \frac{1}{4} \left[ (t - v_j)^4 - \frac{(t - v_{m-1})^4 (v_m - v_j)}{(v_m - v_{m-1})} + \frac{(t - v_m)^4 (v_{m-1} - v_j)}{(v_m - v_{m-1})} \right].
\]

(2.18)

Since closed form maximum likelihood estimates do not exist, the Newton-Raphson procedure with step-halving, as described in Section 1.4.1, is used to numerically optimize the log-likelihood. The initial estimate for the parameter vector \( (a, b, \gamma_1, \gamma_2, \ldots, \gamma_{m-2}) \), which assumes data are from an exponential distribution, is \( \left( \sum_{i=1}^{n} t_i / \sum_{i=1}^{n} \delta_i, 0, 0, \ldots, 0 \right) \).

Step-halving will be invoked whenever the log-likelihood at a new step has not increased over the previous step. In addition, step-halving will be invoked whenever the hazard function estimate at any survival time smaller than the largest uncensored survival time is negative or \( \hat{a} \) is less than zero. Convergence of the iterative procedure is attained when the change in the likelihood function is less than 0.0125 and the norm of the score vector is less than \( 10^{-6} \). The latter criteria for convergence is somewhat excessive and could probably be loosened in practice.

The score vector and information matrix used in the Newton-Raphson optimization procedure are the first and second derivatives of the log-likelihood function, respectively. These matrices are presented in Appendix 2. It should be noted that the \( m \times m \) information matrix is a function of only the uncensored survival times.
2.3.3 Confidence Intervals for the Hazard, Cumulative Hazard, and Survival Functions

Estimates and confidence intervals for the hazard, cumulative hazard, and survival functions can be calculated based upon estimates of the model's unknown parameters and the corresponding information matrix, \( I(\hat{\theta}) = \begin{bmatrix} \hat{a}, \hat{b}, \hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_{m-2} \end{bmatrix} \), denoted as \( I(\hat{\theta}) \).

The estimates of hazard and cumulative hazard are obtained by substituting parameter estimates into (2.14) and (2.17), respectively. Since \( \hat{\theta} = \begin{bmatrix} \hat{a}, \hat{b}, \hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_{m-2} \end{bmatrix} \) is asymptotically distributed as a normal random variable with mean \( \begin{bmatrix} a, b, \gamma_1, \gamma_2, \ldots, \gamma_{m-2} \end{bmatrix} \) and variance \( I(\theta)^{-1} \), it can be shown by the delta method (Miller, 1981) that the hazard and cumulative hazard estimates are asymptotically normally distributed. The variance of these estimates are

\[
\text{Var} \left( \hat{\lambda}(t) \right) = A \ I(\theta)^{-1} A' \\
\text{Var} \left( \hat{A}(t) \right) = B \ I(\theta)^{-1} B'
\]

where

\[
A = \begin{bmatrix} 1 & t & w_1(t) & w_2(t) & \ldots & w_{m-2}(t) \end{bmatrix}
\]

and

\[
B = \begin{bmatrix} t & \frac{1}{2} t^2 & w'_1(t) & w'_2(t) & \ldots & w'_{m-2}(t) \end{bmatrix}
\]

The \((1-\alpha)\) confidence intervals for the hazard and cumulative hazard, which can be obtained by standard confidence interval theory for normally distributed random variables, are

\[
\hat{\lambda}(t) - z_{1-\alpha/2} \sqrt{A \ I(\hat{\theta})^{-1} A'} < \lambda(t) < \hat{\lambda}(t) + z_{1-\alpha/2} \sqrt{A \ I(\hat{\theta})^{-1} A'}
\]

(2.19a)
and

$$\hat{\Lambda}(t) - z_{1-\alpha/2} \sqrt{\mathbf{B} \mathbf{I}(\hat{\theta})^{-1} \mathbf{B}'} < \Lambda(t) < \hat{\Lambda}(t) + z_{1-\alpha/2} \sqrt{\mathbf{B} \mathbf{I}(\hat{\theta})^{-1} \mathbf{B}'}$$  \hspace{1cm} (2.19b)$$

respectively, where $z_{1-\alpha/2}$ is the $(1-\alpha/2) \times 100$ percentile of the normal distribution.

The standard approach to confidence interval estimation for survival functions is a direct application of the delta method. This approach implies that $\hat{S}(t)$ is asymptotically distributed as a normal random variable with mean $S(t)$ and variance

$$\text{Var} \left( \hat{S}(t) \right) = \left[ \hat{S}(t) \right]^2 \mathbf{B} \mathbf{I}(\hat{\theta})^{-1} \mathbf{B}'$$

Therefore the $(1-\alpha)$ confidence interval is

$$\hat{S}(t) - z_{1-\alpha/2} \sqrt{\mathbf{B} \mathbf{I}(\hat{\theta})^{-1} \mathbf{B}'} < S(t) < \hat{S}(t) + z_{1-\alpha/2} \sqrt{\mathbf{B} \mathbf{I}(\hat{\theta})^{-1} \mathbf{B}'}$$  \hspace{1cm} (2.20)$$

This confidence interval will be used in the following research. However, the upper and lower confidence limits are not restricted between 0 and 1. An alternative approach to confidence interval estimation is to develop a confidence interval for $\log \Lambda(t)$ using the delta method and then transform the resulting inequality. It can be shown that the asymptotic variance of $\log \hat{\Lambda}(t)$ is

$$\text{Var} \left( \log \hat{\Lambda}(t) \right) = \left[ \hat{\Lambda}(t) \right]^2 \mathbf{B} \mathbf{I}(\hat{\theta})^{-1} \mathbf{B}'$$

Therefore a $(1-\alpha)$ confidence interval for $\log \hat{\Lambda}(t)$ is

$$\log \hat{\Lambda}(t) - z_{1-\alpha/2} \left[ \hat{\Lambda}(t) \right]^{-1} \sqrt{\mathbf{B} \mathbf{I}(\hat{\theta})^{-1} \mathbf{B}'} < \log \Lambda(t)$$

$$< \log \hat{\Lambda}(t) + z_{1-\alpha/2} \left[ \hat{\Lambda}(t) \right]^{-1} \sqrt{\mathbf{B} \mathbf{I}(\hat{\theta})^{-1} \mathbf{B}'}$$

Since $S(t) = \exp \left\{ - \exp \left( \log \Lambda(t) \right) \right\}$, this inequality can be transformed to obtain the
following confidence interval for $S(t)$.

$$\exp \left\{ - \exp \left( \log \hat{A}(t) + z_{1-\alpha/2} \left[ \hat{A}(t) \right]^{-1} \sqrt{B \mathbf{I}(\hat{\theta})^{-1} B'} \right) \right\} < S(t) < \exp \left\{ - \exp \left( \log \hat{A}(t) - z_{1-\alpha/2} \left[ \hat{A}(t) \right]^{-1} \sqrt{B \mathbf{I}(\hat{\theta})^{-1} B'} \right) \right\}$$

(2.21)

It can be shown that the lower and upper confidence limits are between 0 and 1.

2.4 Modeling with the Restricted Cubic Spline Hazard Function

The following investigation concerning the use of the restricted cubic spline as a model to describe survival ignoring covariates was conducted for two primary purposes. First, the flexibility of the model in adequately describing a variety of hazard function shapes was assessed. Second, the approach to the choice of number and location of knots was examined.

Data were generated from several distributions. The shapes of the associated underlying hazard functions can be categorized into four of five hazard function shape categories described by Glaser (1980). These distributions, with identifying number, are:

<table>
<thead>
<tr>
<th>ID #</th>
<th>Distribution</th>
<th>Parameters</th>
<th>Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>1*</td>
<td>Generalized</td>
<td>a=1       b=0.8  c=1  d=3</td>
<td>Upside-down Bathtub</td>
</tr>
<tr>
<td>2</td>
<td>Weibull</td>
<td>a=1       b=0.8  c=1  d=6</td>
<td>Upside-down Bathtub</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>a=1       b=0.5  c=2  d=3</td>
<td>Upside-down Bathtub</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>a=1       b=0.5  c=1  d=6</td>
<td>Upside-down Bathtub</td>
</tr>
<tr>
<td>5*</td>
<td></td>
<td>a=1       b=0.8  c=2  d=1</td>
<td>Decreasing</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>a=1       b=0.5  c=2  d=1</td>
<td>Decreasing</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>a=2       b=1    c=1   d=3</td>
<td>Increasing</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>a=2       b=1    c=1   d=6</td>
<td>Increasing</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>a=1       b=0.5  c=0  d=1</td>
<td>Decreasing</td>
</tr>
<tr>
<td>10*</td>
<td></td>
<td>a=1       b=3    c=1   d=1.5</td>
<td>Increasing</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>a=2       b=1    c=0   d=1</td>
<td>Constant</td>
</tr>
<tr>
<td>12*</td>
<td></td>
<td>a=1       b=3    c=1   d=0.25</td>
<td>Upside-down Bathtub</td>
</tr>
<tr>
<td>13*</td>
<td>Gompertz</td>
<td>a=0       b=1</td>
<td>Increasing</td>
</tr>
</tbody>
</table>

* The spline hazard and survival function estimates for one sample are presented in Figures 2.2a - 2.7c.
The first four distributions are upside-down bathtub-shaped with differing rates of initial decrease and final increase. Distributions 5 - 8 have horizontal asymptotes at 2. The hazard function for distributions 5 and 6 decreases from infinity to 2, while the hazard function for distribution 7 and 8 increases from 2. Distribution 9 is a Weibull distribution. Distribution 10 is the sum of two increasing Weibull hazards. Distribution 11 is an exponential distribution. Distribution 12 is an upside-down bathtub-shaped hazard which sharply decreases in the left tail. The hazard in distribution 13 exponentially increases from 0. The underlying hazard function for the first twelve distributions are presented in Figure 2.1.

The data generated from these distributions were censored by a uniform distribution which allowed for approximately 40% censoring. Sample sizes of 50 and 200 were considered. At least 3 samples were generated from each distribution-sample size combination.

Three, four, and five knots at varying locations determined by the data were used to define the spline model. The locations of the knots were based on the empirical distribution formed by the uncensored survival times. It is assumed that since only uncensored observations contribute to the information matrix, the remaining censored observations should not be useful in determining knot locations. When tails are not linearly restricted, the coefficient associated with any knot located where the remaining survival times are censored has no information.

For samples of size 50, tail knots were placed at the fifth smallest and fifth largest uncensored observation, in accordance with the recommendations of Stone and Koo (1985). However, their recommendations were not followed for larger samples. It is noted that as sample size increases, the 5th smallest observation approaches 0 and the 5th largest observation is likely to become an outlier. Since these knot locations did not seem reasonable, alternative locations were chosen further into the interior of the observed data. For samples of size 200, the tail knots were placed at the 5th and 95th percentile of the empirical distribution formed by the uncensored observations. Such an assignment of knot location also seems reasonable for
samples larger than 200. However, examination of such issues were not examined. The remaining knots were distributed between the tail knots.

The "goodness of fit" of these models was crudely assessed by visual comparison of the spline survival estimate and confidence interval with the Kaplan-Meier survival curve and confidence interval, and with the true population survival function. The true survival model was not used as the primary measure for fit since it did not take into account sampling error and censorship. Residuals or differences between the Kaplan-Meier and spline curves were deemed more serious when survival times were small, since standard errors are smaller (Miller, 1983). Residuals considered "too large" for a well-fitted model are indicators of where knots should be relocated or where additional knots should be incorporated into the model.

For both sample size 50 and 200, three knots did not adequately describe the data. Optimization with five knots almost always encountered singular information matrices in the iterative procedure. Four knots with proper placement appeared to generate a good-fitting model for the distributions and sample sizes considered.

Spline models with knots located at the 5th, 33rd, 67th, and 95th percentiles for samples of size 200, and at the fifth smallest observation, the fifth largest observation, and the 33rd and 67th percentiles for samples of size 50 adequately described the data sets generated from distributions 1 - 11 and 13. The estimated spline survival and hazard functions are graphically displayed for data sets of size 200 generated from distributions 1, 4, 10 and 13 in Figures 2.2a - 2.5b and for a data set of size 50 generated from distribution 1 in Figures 2.6a and b. In all cases, the estimated spline survival function agrees fairly well with the Kaplan-Meier curve. Survival function confidence intervals from the Kaplan-Meier and spline estimates are almost superimposable. The estimated spline hazard function, Kaplan-Meier curve, and spline survival function differ from the corresponding underlying curve due to sampling error and censorship. Additional sets of four knots were chosen and used to estimate the model. The resulting estimates were similar to those discussed above and indicate that for these data sets the fit of the spline model is fairly robust to reasonable knot location choices.
FIGURE 2.2a
SURVIVAL FUNCTION ESTIMATES WITH 95% CONFIDENCE LIMITS FOR A SAMPLE OF SIZE 200
GENERATED FROM THE GENERALIZED WEIBULL DISTRIBUTION WITH $a=1$, $b=0.8$, $c=1$, $d=3$

FIGURE 2.2b
HAZARD FUNCTION ESTIMATE WITH 95% CONFIDENCE LIMITS FOR A SAMPLE OF SIZE 200
GENERATED FROM THE GENERALIZED WEIBULL DISTRIBUTION WITH $a=1$, $b=0.8$, $c=1$, $d=3$
FIGURE 2.3a
SURVIVAL FUNCTION ESTIMATES WITH 95% CONFIDENCE LIMITS FOR A SAMPLE OF SIZE 200 GENERATED FROM THE GENERALIZED WEIBULL DISTRIBUTION WITH $a=1$, $b=0.8$, $c=2$, $d=1$

FIGURE 2.3b
HAZARD FUNCTION ESTIMATE WITH 95% CONFIDENCE LIMITS FOR A SAMPLE OF SIZE 200 GENERATED FROM THE GENERALIZED WEIBULL DISTRIBUTION WITH $a=1$, $b=0.8$, $c=2$, $d=1$
FIGURE 2.4a
SURVIVAL FUNCTION ESTIMATES WITH 95% CONFIDENCE LIMITS FOR A SAMPLE OF SIZE 200 GENERATED FROM THE GENERALIZED WEIBULL DISTRIBUTION WITH $a=1$, $b=3$, $c=1$, $d=1.5$

FIGURE 2.4b
HAZARD FUNCTION ESTIMATE WITH 95% CONFIDENCE LIMITS FOR A SAMPLE OF SIZE 200 GENERATED FROM THE GENERALIZED WEIBULL DISTRIBUTION WITH $a=1$, $b=3$, $c=1$, $d=1.5$
FIGURE 2.5a
SURVIVAL FUNCTION ESTIMATES WITH 95% CONFIDENCE LIMITS FOR A SAMPEL OF SIZE 200
GENERATED FROM THE Gompertz DISTRIBUTION WITH $a=0$, $b=1$

FIGURE 2.5b
HAZARD FUNCTION ESTIMATE WITH 95% CONFIDENCE LIMITS FOR A SAMPLE OF SIZE 200
GENERATED FROM THE Gompertz DISTRIBUTION WITH $a=0$, $b=1$
FIGURE 2.6a
SURVIVAL FUNCTION ESTIMATES WITH 95% CONFIDENCE LIMITS FOR A SAMPLE OF SIZE 50
GENERATED FROM THE GENERALIZED WEIBULL DISTRIBUTION WITH a=1, b=0.8, c=1, d=3

FIGURE 2.6b
HAZARD FUNCTION ESTIMATE WITH 95% CONFIDENCE LIMITS FOR A SAMPLE OF SIZE 50
GENERATED FROM THE GENERALIZED WEIBULL DISTRIBUTION WITH a=1, b=0.8, c=1, d=3
FIGURE 2.7a
SURVIVAL ESTIMATES FOR A SAMPLE GENERATED FROM THE GENERALIZED WEIBULL DISTRIBUTION
WITH $a=1$, $b=0.25$, $c=1$, $d=3$ USING KNOTS AT THE 5TH, 33RD, 67TH, AND 95TH PERCENTILES

FIGURE 2.7b
SURVIVAL ESTIMATES FOR A SAMPLE GENERATED FROM THE GENERALIZED WEIBULL DISTRIBUTION
WITH $a=1$, $b=0.25$, $c=1$, $d=3$ USING KNOTS AT THE 5TH, 20TH, 40TH, AND 90TH PERCENTILES

SOLID CURVE — UNDERLYING SURVIVAL FUNCTION
SMOOTH DASHED CURVES — SPLINE ESTIMATE AND CONFIDENCE INTERVAL
DASHED STEP FUNCTION — KAPLAN-MEIER ESTIMATE AND CONFIDENCE INTERVAL

SAMPLE SIZE = 200
PERCENTAGE CENSORED = 43.0 %
FIGURE 2.7c
HAZARD ESTIMATE FOR A SAMPLE GENERATED FROM THE GENERALIZED WEIBULL DISTRIBUTION WITH a=1, b=0.25, c=1, d=3 USING KNOTS AT THE 5TH, 20TH, 40TH, AND 90TH PERCENTILES

SOLID CURVE — UNDERLYING HAZARD FUNCTION
SMOOTHED CURVES — SPLINE ESTIMATE AND CONFIDENCE INTERVAL.
HAZARDS GREATER THAN 10 ARE NOT DISPLAYED IN THIS PLOT.

SAMPLE SIZE = 200
PERCENTAGE CENSORED = 43.0 %
With knots placed at the 5th, 33rd, 67th, and 95th percentiles, the estimated spline survival model for distribution 12 did not adequately describe the data (Figure 2.7a). Near the origin, where the Kaplan-Meier curve shows a dramatic drop, the difference between the Kaplan-Meier and spline survival estimates is "large", with this difference being approximately 0.1. (It is sometimes easier to see differences between the spline and Kaplan-Meier curves by examining the actual estimated values rather than graphs.) Adjustment of the knots towards the origin generated a better fitting model (Figure 2.7b). These knots were placed at the 5th, 20th, 40th, and 90th percentiles of the uncensored data. Additional adjustment of the knots did not improve the model's fit. The estimate of the hazard function is graphically shown in Figure 2.7c. The left tail of this estimate shows a sharp drop consistent with the true model. The seemingly lack of smoothness in this estimate is due to there being three knots (0.00000262, 0.00450, 0.00648) near 0.

Based on this limited experience with fitting the restricted cubic spline hazard function, the following guidelines to knot location are suggested.

1. Knots should be located at approximately equally spaced quantiles except when the underlying hazard function exhibits sharp changes or the Kaplan-Meier curve "dramatically" decreases. Similar to the recommendation of Stone and Koo (1985), the first and last knots should be placed at the 5th smallest and 5th largest uncensored survival times for smaller samples, e.g. \( N \leq 100 \). For larger samples, tail knots should be placed at the 5th and 95th percentile of the uncensored survival times.

2. When the underlying hazard function exhibits sharp changes or the Kaplan-Meier curve "dramatically" decreases, more knots should be located in and/or relocated into the region.
3. Knots should not be located too near each other. Otherwise, the coefficients are highly correlated and the optimization procedure “blows up” due to a singular information matrix. If it is necessary to locate knots “too near” each other, an alternative, though untried, approach would be to combine the knots and remove a continuity restriction at the combined knot.

4. Four knots are sufficient to handle a sample of size 200 with approximately 40% censoring. Three knots do not seem to adequately describe the data. With larger data sets, as will be seen in Section 2.6, five knots may be needed. Additional knots beyond five can be added, but Wold (1974) warns that the number of knots should be kept to a minimum. Stone and Koo (1986) claim that five knots “should be enough to model the overall shape . . . that are likely to occur in practice.” They indicate that additional knots increase the model’s flexibility, but it also increases the variance of estimation.

The number of iterations needed to obtain convergence with the Newton-Raphson procedure generally ranged from 3 to 6 for samples of size 200. Smaller samples required 2 or 3 more iterations for convergence. When the underlying hazard function exhibited sharp changes or the estimated survival curve “dramatically” decreased, as in distribution 12, the number of iterations needed for convergence increases to approximately 10 or 11.

In the preceding research, the optimization algorithm did not converge within 20 iterations, the maximum number, in some highly censored small samples. These data sets were generated primarily from distributions where the underlying hazard increased from zero at the origin. The estimate for parameter \( a \) in model (2.14) had difficulty converging to 0 when most, if not all, of the smallest survival times were censored. Step-halving was invoked many times. This difficulty seems to indicate that either the first knot location needs to be moved nearer zero, or the linearity assumption needs to be relaxed. A reparameterization of the model to guarantee that \( \hat{a} \) is non-negative may be beneficial. Further investigation into how to deal with this problem is warranted.
It should be restated that the optimization procedure used in this research only guarantees that hazards are positive for censored and uncensored observations up to the maximum uncensored observation. No guarantee is made concerning estimates at points between observations nor after the final uncensored observation.

In some small highly-censored data sets, where the underlying hazard converged to zero as \( t \) became large, the hazard estimates after the final uncensored observation were negative. This problem is minor in that survival estimates for smaller \( t \) seem to be minimally affected. A possible, though untried, solution to this problem may be relaxation of the linear tail restriction. This solution would increase the number of unknown parameters in the hazard model. Another alternative is the use of a more appropriate constrained optimization algorithm.

The recommended approach to fitting a restricted cubic spline hazard model is as follows:

1. Try fitting a spline model with four knots. Two of these knots should be in the tails. The remaining knots should be located at approximately equally spaced quantiles. A good initial set of knots is at the 5th, 33rd, 67th, and 95th percentiles of the empirical distribution of uncensored observations.

2. Examine the residuals or difference between the spline survival function and the Kaplan-Meier curve.

3. If the residuals are large and indicate a poorly fitting model, adjust knot location or add additional knots. Removal of continuity restrictions may need to be considered.

4. Repeat steps 2 and 3 until a good model is obtained.

The research discussed has not dealt with upside-down bathtub-shaped hazard functions. Data was generated from the log-logistic model and an attempt was made to fit the
restricted cubic spline hazard model to the data. Convergence of the optimization algorithm was not obtained. Plots of the spline survival function at intermediate steps were made and agreement with the Kaplan-Meier curve was fair. However, hazard estimates for small t and large t were negative. The log-logistic hazard function has two properties which have previously been mentioned to be sometimes associated with optimization problems and negative hazards. These properties of the log-logistic model are that the hazard function increases from 0, and it also converges to 0 as survival time becomes large. Additional investigation is needed to determine whether other spline models may better describe upside-down bathtub-shaped hazard models. Alternative constrained optimization techniques may need to be considered in lieu of the Newton-Raphson algorithm.

2.5 Efficiency of Estimating the Survival Function with the Spline Model

The efficiency of the restricted cubic spline survival estimator, relative to the Kaplan-Meier estimator, is studied in this section. A rigorous examination of this subject would consider, among other things, the effects of underlying distribution, sample size, censoring, and knot location. However, a more modest study will be conducted using two underlying distributions and three levels of censorship.

Data were generated from the following distributions uniformly censored with 0%, 25%, and 50% censorship.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Parameters</th>
<th>Hazard Function Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalized Weibull</td>
<td>a=1, b=0.8, c=1, d=3</td>
<td>Bathtub Shaped</td>
</tr>
<tr>
<td></td>
<td>a=2, b=1, c=1, d=3</td>
<td>Increasing</td>
</tr>
</tbody>
</table>

Fifty samples of size 200 were generated from each distribution and at each level of censorship.
A spline model with knots placed at the 5th, 33rd, 67th, and 95th percentiles of the empirical
distribution formed by the uncensored survival times was estimated.

The efficiency of estimating survival was determined at the 5th, 10th, 25th, 50th, 75th,
90th, and 95th percentiles (or equivalently, at the 95th, 90th, 75th, 50th, 25th, 10th, and 5th
survival fractiles, respectively) of the underlying distribution. Relative efficiency of the spline
estimator was calculated as the ratio of the mean square error for the Kaplan-Meier estimator
to the mean square error of the spline estimator. Mean square error was computed using the
following formula:

\[ \text{MSE}(p) = \frac{1}{50} \sum_{i=1}^{50} ( \hat{S}(t_i^p) - p )^2, \]

where \( t_i^p \) is the survival fractile satisfying the property \( \Pr(T > t_i^p) = p \), and \( \hat{S}(t_i^p) \) is the
estimated survival probability at \( t_i^p \). Whenever \( t_i^p \) was larger than the sample's maximum
uncensored survival time, estimates of survival were not obtained at that particular \( t_i^p \) and
hence not included in the calculation of MSE(p).

The estimated spline survival curves, the associated Kaplan-Meier curves, and the
estimated hazard functions for the first 25 samples generated with approximately 25%
censorship are displayed in Figures 2.8a, b, and c for the distribution with parameters \( a=1, b=0.8, c=1, d=3 \) and in Figures 2.9a, b, and c for the distribution with parameters \( a=2, b=1, c=1, d=3 \). The overall amount of variation with the spline survival estimator appears to
be comparable to that of the Kaplan-Meier estimator for both distributions.

Table 2.1 presents efficiency results when data are generated from the generalized
Weibull distribution with parameters \( a=1, b=0.8, c=1, d=3 \). The restricted cubic spline
survival estimator is more efficient than the Kaplan-Meier estimator at the 90th survival
fractile and also, at the 5th through 50th survival fractiles. Efficiency with the spline
estimator at the 75th survival fractile is approximately the same as that of the Kaplan-Meier
estimator, except when data is uncensored. In that case, the Kaplan-Meier estimator is slightly
FIGURE 2.8a
SPLINE SURVIVAL FUNCTION ESTIMATES FOR 25 SAMPLES GENERATED FROM THE
GENERALIZED WEIBULL DISTRIBUTION WITH \( a=1, b=0.8, c=1, d=3 \)

FIGURE 2.8b
KAPLAN-MEIER CURVE FOR 25 SAMPLES GENERATED FROM THE
GENERALIZED WEIBULL DISTRIBUTION WITH \( a=1, b=0.8, c=1, d=3 \)
FIGURE 2.3c

SPLINE HAZARD FUNCTION ESTIMATE FOR 25 SAMPLES GENERATED FROM THE GENERALIZED WEIBULL DISTRIBUTION WITH $a=1$, $b=0.8$, $c=1$, $d=3$.

SOLID CURVE IS UNDERLYING HAZARD FUNCTION.
THE REMAINING 24 CURVES ARE SPLINE ESTIMATES BASED ON SAMPLES OF SIZE 200 WITH APPROXIMATELY 20% CENSORED.
FIGURE 2.9a
SPLINE SURVIVAL FUNCTION ESTIMATES FOR 25 SAMPLES GENERATED FROM THE
GENERALIZED WEIBULL DISTRIBUTION WITH $a=2$, $b=1$, $c=1$, $d=3$

FIGURE 2.9b
KAPLAN-MEIER CURVE FOR 25 SAMPLES GENERATED FROM THE
GENERALIZED WEIBULL DISTRIBUTION WITH $a=2$, $b=1$, $c=1$, $d=3$

SOLID CURVE IS UNDERLYING SURVIVAL FUNCTION.
THE REMAINING 24 CURVES ARE SPLINE ESTIMATES BASED ON SAMPLES OF
SIZE 250 WITH APPROXIMATELY 250 OBSERVED.
FIGURE 2.3c

SPLINE HAZARD FUNCTION ESTIMATE FOR 25 SAMPLES GENERATED FROM THE GENERALIZED WEIBULL DISTRIBUTION WITH $a=2$, $b=1$, $c=1$, $d=3$

LEGEND: The solid line is the underlying hazard function. The dashed lines are spline estimates based on samples of 25 data with approximately 250 censored.
Table 2.1

EFFICIENCY OF SURVIVAL FUNCTION ESTIMATION WITH THE RESTRICTED CUBIC SPLINE HAZARD FUNCTION RELATIVE TO THE KAPLAN-MEIER ESTIMATOR FOR DATA GENERATED FROM THE GENERALIZED WEIBULL DISTRIBUTION WITH a=1, b=0.8, c=1, d=3

<table>
<thead>
<tr>
<th></th>
<th>Cens.</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prop.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spl MSE</td>
<td>K-M MSE</td>
<td>Efficiency</td>
<td>Spl MSE</td>
<td>K-M MSE</td>
<td>Efficiency</td>
<td>Spl MSE</td>
</tr>
<tr>
<td>0.00</td>
<td>0.95</td>
<td>0.000452</td>
<td>0.000235</td>
<td>0.00148</td>
<td>0.00130</td>
<td>0.00124</td>
<td>0.000366</td>
<td>0.000186</td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>0.000244</td>
<td>0.000439</td>
<td>0.00124</td>
<td>0.00168</td>
<td>0.00127</td>
<td>0.000545</td>
<td>0.000297</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.540</td>
<td>1.868</td>
<td>0.838</td>
<td>1.292</td>
<td>1.024</td>
<td>1.489</td>
<td>1.597</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.000441</td>
<td>0.000385</td>
<td>0.00146</td>
<td>0.00133</td>
<td>0.00121</td>
<td>0.000503</td>
<td>0.000372</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.000358</td>
<td>0.000474</td>
<td>0.00145</td>
<td>0.00141</td>
<td>0.00131</td>
<td>0.000691</td>
<td>0.000580</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.812</td>
<td>1.231</td>
<td>0.993</td>
<td>1.060</td>
<td>1.083</td>
<td>1.374</td>
<td>1.559</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.000321</td>
<td>0.000424</td>
<td>0.00133</td>
<td>0.00149</td>
<td>0.00182 &amp;</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.000223</td>
<td>0.000546</td>
<td>0.00129</td>
<td>0.00198</td>
<td>0.00214 &amp;</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.695</td>
<td>1.288</td>
<td>0.970</td>
<td>1.329</td>
<td>1.176 &amp;</td>
<td>#</td>
<td>#</td>
</tr>
</tbody>
</table>

1 The point t at which the true survival fractile p occurs is determined from the underlying distribution from which data is generated. This point satisfies P(T>t) = p.
2 Efficiency is the ratio of the mean square error of the Kaplan-Meier estimate (K-M MSE) over the mean square error of the spline estimate (Spl MSE). The estimates of mean square error and efficiency are based on 50 simulated data sets, except when indicated otherwise.
© Based on 48 samples.
& Based on 47 samples.
# Insufficient number of survival estimates to compute efficiency.
Table 2.2

EFFICIENCY OF SURVIVAL FUNCTION ESTIMATION WITH THE RESTRICTED CUBIC SPLINE HAZARD FUNCTION RELATIVE TO THE KAPLAN-MEIER ESTIMATOR FOR DATA GENERATED FROM THE GENERALIZED WEIBULL DISTRIBUTION WITH a=2, b=-1, c=-1, d=-3

<table>
<thead>
<tr>
<th>Prop. Cens.</th>
<th>0.95</th>
<th>0.90</th>
<th>0.75</th>
<th>0.50</th>
<th>0.25</th>
<th>0.10</th>
<th>0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spl MSE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td>0.000182</td>
<td>0.000363</td>
<td>0.000704</td>
<td>0.00128</td>
<td>0.000727</td>
<td>0.000276</td>
<td>0.000139</td>
</tr>
<tr>
<td>K-M MSE</td>
<td>0.000334</td>
<td>0.000457</td>
<td>0.000821</td>
<td>0.00153</td>
<td>0.000969</td>
<td>0.000319</td>
<td>0.000203</td>
</tr>
<tr>
<td>Efficiency</td>
<td>1.835</td>
<td>1.259</td>
<td>1.166</td>
<td>1.195</td>
<td>1.333</td>
<td>1.156</td>
<td>1.460</td>
</tr>
<tr>
<td>Spl MSE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>0.000242</td>
<td>0.000508</td>
<td>0.000995</td>
<td>0.00149</td>
<td>0.00124</td>
<td>0.000493</td>
<td>0.000281@</td>
</tr>
<tr>
<td>K-M MSE</td>
<td>0.000413</td>
<td>0.000463</td>
<td>0.00124</td>
<td>0.00162</td>
<td>0.00153</td>
<td>0.000748</td>
<td>0.000366@</td>
</tr>
<tr>
<td>Efficiency</td>
<td>1.707</td>
<td>0.911</td>
<td>1.246</td>
<td>1.087</td>
<td>1.234</td>
<td>1.517</td>
<td>1.302</td>
</tr>
<tr>
<td>Spl MSE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>0.000195</td>
<td>0.000404</td>
<td>0.00103</td>
<td>0.00244</td>
<td>0.00226&amp; #</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K-M MSE</td>
<td>0.000279</td>
<td>0.000411</td>
<td>0.00154</td>
<td>0.00274</td>
<td>0.00269&amp; #</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Efficiency</td>
<td>1.431</td>
<td>1.017</td>
<td>1.495</td>
<td>1.123</td>
<td>1.190&amp; #</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 The point t at which the true survival fractile p occurs is determined from the underlying distribution from which data is generated. This point satisfies P(T > t) = p.

2 Efficiency is the ratio of the mean square error of the Kaplan-Meier estimate (K-M MSE) over the mean square error of the spline estimate (Spl MSE). The estimates of mean square error and efficiency are based on 50 simulated data sets, except when indicated otherwise.

@ Based on 45 samples.
& Based on 37 samples.
# Insufficient number of survival estimates to compute efficiency.
more efficient. At the 95th survival fractile, the Kaplan-Meier estimator is more efficient. It is hypothesized that the fit of the spline model with the knot locations used is not good in some of the data sets for small t and hence contributes to a "large" bias and mean square error. The underlying hazard function shows some curvature in the left tail, indicating that more knots may be needed near the origin for some data sets. Regardless, since mean square error is small for both estimators, the decrease in efficiency of the spline estimator does not matter much.

Table 2.2 presents efficiency results when data are generated from the generalized Weibull distribution with parameters a=2, b=1, c=1, d=3. The spline estimator is more efficient than the Kaplan-Meier estimator except at the 90th survival fractile with 25% censorship. In general, efficiency is greatest in the tails, i.e. at the 95th, 10th, and 5th survival fractiles. At the 95th, 90th, 25th, and 5th survival fractile, the relative efficiency of the spline estimator decreases as censorship increases.

The primary difference between the two distributions from which data were generated is the shape of the hazard function. The first hazard function has curvature in the left tail, while the second is flat for small t. The results described above seem to indicate that efficiency of spline estimation decreases when there is sharp curvature in the underlying hazard function. However, it is thought that if knot locations were determined for each individual data set based on the fit of the model, this bias could be removed. Additional investigation of this issue is needed.

The efficiency of estimation with the restricted cubic spline model is at least as good as, if not better than, the Kaplan-Meier estimator for the underlying distributions studied. If knots had been allowed to be determined based on the fit of each particular data set, results might have been even more favorable towards the spline model.
2.6 Spline Models Applied to Cardiovascular Data

The approach, discussed in Section 2.3.2, to fitting a restricted cubic spline hazard model is applied to two cardiovascular disease data sets from the Cardiovascular Disease Data Bank of Duke University Medical Center. These data sets contain information on the survival experience of patients who have undergone cardiac catheterization for chest pain at Duke and were found to have significant narrowing of at least one major coronary artery (≥75% diameter narrowing). The endpoint in both data sets is cardiovascular death, with non-cardiovascular death or termination of follow-up considered as censoring events.

All patients are included in the “Medical Patient” data set, where survival time is the time from the date of catheterization. Patients who underwent coronary artery bypass grafting (CABG) are considered censored on the day of the operation. Only survival time under medical therapy is considered.

The “Surgical Patient” data set contains follow-up information on all patients who underwent CABG. Survival time is measured from the date of the operation. No censoring occurs for additional surgical intervention during follow-up.

2.6.1 Medical Patient Data

The “Medical Patient” data set contains the survival experiences of 5795 patients, of which 592 died of cardiovascular causes. A restricted cubic spline hazard function using the standard set of knot locations (5th, 33rd, 67th, and 95th percentiles) based on the empirical distribution of nonzero uncensored survival times was fit to the data. Figures 2.10a and b show the estimated survival and hazard functions. In Figure 2.10a, the spline survival function and the Kaplan-Meier curve appear to be identical. Even though differences between the two curves are less than 0.01 in the left tail, confidence intervals are not overlapping in this region. Considering the large sample size and resulting small standard errors, the two survival curves are in close enough agreement for most purposes to consider the spline model to be a good fit of the data.
FIGURE 2.10a

SURVIVAL FUNCTION ESTIMATES FOR MEDICAL PATIENTS USING KNOTS LOCATED AT THE 6TH, 33RD, 67TH, AND 96TH PERCENTILES

FIGURE 2.10b

HAZARD FUNCTION ESTIMATE FOR MEDICAL PATIENTS USING KNOTS LOCATED AT THE 6TH, 33RD, 67TH, AND 96TH PERCENTILES
FIGURE 2.10c

SURVIVAL FUNCTION ESTIMATES FOR MEDICAL PATIENTS USING KNOTS LOCATED AT THE 5TH, 20TH, 33RD, 67TH, AND 95TH PERCENTILES

FIGURE 2.10d

HAZARD FUNCTION ESTIMATE FOR MEDICAL PATIENTS USING KNOTS LOCATED AT THE 5TH, 20TH, 33RD, 67TH, AND 95TH PERCENTILES
Optimization of a model using knots based on all uncensored survival times (including zero) did not converge within 20 iterations. The nonzero restriction was added primarily because it was a necessary restriction in describing survival in the surgical patient data set, which will be discussed in the next section.

The structure of the spline model fitted with four knots is:

\[
\lambda(t) = a + b \cdot t + \\
\sum_{i=1}^{2} \gamma_i \left[ \left( t - v_1 \right)^3_+ - \frac{(t - v_3)^3_+ (v_4 - v_1)}{(v_4 - v_3)} + \frac{(t - v_4)^3_+ (v_3 - v_1)}{(v_4 - v_3)} \right]
\]

where \((v_1, v_2, v_3, v_4) = (0.0246, 1.0840, 3.417, 9.131)\) are knots (in years). Estimates of the unknown parameters, which were obtained after 19 iterations of the Newton-Raphson algorithm, are

\[
\hat{\theta} = \begin{bmatrix} \hat{a} & \hat{b} & \hat{\gamma}_1 & \hat{\gamma}_2 \end{bmatrix} = \begin{bmatrix} 0.0814 & -0.0318 & 0.00319 & -0.00468 \end{bmatrix}
\]

The associated covariance matrix is

\[
\text{Cov} (\hat{\theta}) = 10^{-5} \times \begin{bmatrix}
4.42 & -3.81 & 0.419 & -0.625 \\
4.38 & -0.550 & 0.835 \\
0.0747 & -0.115 \\
0.177
\end{bmatrix}
\]

The following estimate of the hazard function results when the parameter estimates given
above are substituted into (2.22):

\[ \dot{\lambda}(t) = 0.0814 - 0.0318 \ t + 0.00319 \ (t - 0.0246)^3 + 0.00468 \ (t - 1.0840)^3 + 0.00151 \ (t - 3.417)^3 + 0.0000169 \ (t - 9.131)^3 . \]

This implies that the survival function estimate is:

\[ \hat{S}(t) = \exp \left( - \int_0^t \dot{\lambda}(u) \, du \right) \]

\[ = \exp \left\{ - 0.0814 \ t + 0.0159 \ t^2 - 0.000798 \ (t - 0.0246)^4 + 0.00117 \ (t - 1.0840)^4 - 0.000378 \ (t - 3.417)^4 + 0.00000423 \ (t - 9.131)^4 \right\} . \]

Even though it has been indicated that for most purposes model (2.22) is a good fit of the data, a spline model with knots located at the 5th, 20th, 33rd, 67th, and 95th percentiles of the empirical distribution formed by nonzero uncensored survival times was also fit to the data. The estimates of this model were obtained after 5 iterations of the Newton-Raphson algorithm. The estimated survival and hazard functions are shown in Figures 2.10c and d, respectively.

A likelihood ratio test can be conducted to determine whether the coefficient of the truncated power term associated with the 20th percentile knot is necessary in describing the data. The log-likelihood of the reduced model, i.e. the model with knots at the 5th, 33rd, 67th, and 95th percentiles, is -2385.1, whereas the log-likelihood of the model with the additional knot at the 20th percentile is -2358.6. The likelihood ratio statistic is

\[ \text{LR} = -2 \ (L_F - L_R) \], where \( L_F \) and \( L_R \) are the log-likelihood associated with the full and reduced model, respectively. Therefore \( LR = -2 \ (-2385.1 - (-2358.6)) = 53.0 \). This
statistically significant likelihood ratio statistic with 1 degree of freedom indicates that the additional knot contributes significantly to the fit of the model.

The structure of the spline based on five knots is:

\[ \lambda(t) = a + b \cdot t + \sum_{i=1}^{3} \gamma_i \left[ \frac{(t - v_i)^3_+}{(v_5 - v_4)} + \frac{(t - v_5)^3_+}{(v_5 - v_4)} \right], \quad (2.23) \]

where \((v_1, v_2, v_3, v_4, v_5) = (0.0246, 0.372, 1.084, 3.417, 9.131)\) are knots (in years). Estimates of the unknown parameters are

\[ \hat{\theta} = \begin{bmatrix} \hat{a} & \hat{b} & \hat{\gamma}_1 & \hat{\gamma}_2 & \hat{\gamma}_3 \end{bmatrix} = \begin{bmatrix} 0.143 & -0.253 & 0.261 & -0.396 & 0.137 \end{bmatrix} \]

The associated covariance matrix is

\[ \text{Cov} (\hat{\theta}) = 10^{-3} \times \begin{bmatrix} 0.180 & -0.426 & 0.420 & -0.632 & 0.214 \\ 1.160 & -1.225 & 1.858 & -0.645 \\ 1.349 & -2.059 & 0.726 \\ 3.144 & -1.112 \\ 0.395 \end{bmatrix} \]

The following estimate of the hazard function results when the parameter estimates given above are substituted into (2.23):

\[ \hat{\lambda}(t) = 0.143 - 0.253 \cdot t + 0.261 \cdot (t - 0.0246)^3_+ \\
- 0.396 \cdot (t - 0.372)^3_+ + 0.137 \cdot (t - 1.084)^3_+ \\
- 0.00250 \cdot (t - 3.417)^3_+ + 0.0000951 \cdot (t - 9.131)^3_+. \]
The survival function estimate is

\[ \hat{S}(t) = \exp \left\{ -0.143 \ t + 0.127 \ t^2 - 0.653 \ (t - 0.0246)^{\frac{4}{7}} \\
+ 0.0990 \ (t - 0.372)^{\frac{4}{7}} - 0.0343 \ (t - 1.084)^{\frac{4}{7}} \\
+ 0.000625 \ (t - 3.417)^{\frac{4}{7}} - 0.0000238 \ (t - 9.131)^{\frac{4}{7}} \right\} \]

2.6.2 Surgical Patient Data

The survival experience of 2967 patients (86.5% censored) who had undergone coronary artery bypass grafting was initially fit with a spline model having knots at the 5th, 33rd, 67th, and 95th percentiles of the empirical distribution formed by the non-zero uncensored survival times. The non-zero restriction was placed on the 5th percentile knot since more than 5% of the uncensored failure times were at 0 and it is desirable to have a non-zero knot. Figures 2.11a and b show the estimated survival curve, the Kaplan-Meier curve, and hazard function estimate. Figure 2.11a shows that the spline survival function underestimates the Kaplan-Meier survival curve in the 0 - 1 year interval. Since this difference was considered "large", an additional knot was placed at the 20th percentile of the empirical distribution.

Figures 2.11c and d show the curves associated with the spline hazard model with knots at the 5th, 20th, 33rd, 67th, and 95th percentiles. Figure 2.11c indicates that the additional knot is needed since there is now good agreement between the Kaplan-Meier and spline survival curves. A likelihood ratio test can be conducted to determine whether the coefficient of the truncated power term associated with the 20th percentile knot is necessary in describing the data. The log-likelihood of the reduced model is -1728.7, whereas the log-likelihood of the model with the additional knot at the 20th percentile is -1542.4. The likelihood ratio statistic is therefore equal to 372.6. This statistic with one degree of freedom, being clearly significant, indicates that the additional knot contributes significantly to the fit of the model.
FIGURE 2.11a
SURVIVAL FUNCTION ESTIMATES FOR SURGICAL PATIENTS USING KNOTS LOCATED AT THE 5TH, 33RD, 67TH, AND 95TH PERCENTILES

FIGURE 2.11b
HAZARD FUNCTION ESTIMATE FOR SURGICAL PATIENTS USING KNOTS LOCATED AT THE 5TH, 33RD, 67TH, AND 95TH PERCENTILES
FIGURE 2.11c

SURVIVAL FUNCTION ESTIMATES FOR SURGICAL PATIENTS USING KNOTS LOCATED AT THE 5TH, 20TH, 33RD, 67TH, AND 95TH PERCENTILES

FIGURE 2.11d

HAZARD FUNCTION ESTIMATE FOR SURGICAL PATIENTS USING KNOTS LOCATED AT THE 5TH, 20TH, 33RD, 67TH, AND 95TH PERCENTILES
The structure of the model which fits the data adequately is:

\[
\lambda(t) = a + bt + \\
\sum_{i=1}^{3} \gamma_i \left[ (t - v_1)_{+}^3 - \frac{(t - v_4)_{+}^3 (v_5 - v_1)}{(v_5 - v_4)} + \frac{(t - v_5)_{+}^3 (v_4 - v_1)}{(v_5 - v_4)} \right],
\]

(2.24)

where \((v_1, v_2, v_3, v_4, v_5) = (0.00548, 0.0849, 0.498, 5.262, 10.059)\) are knots (in years). The estimates of the parameters, which were obtained after 10 iterations of the Newton-Raphson algorithm, are

\[
\hat{\theta} = \begin{bmatrix} \hat{a} & \hat{b} & \hat{\gamma}_1 & \hat{\gamma}_2 & \hat{\gamma}_3 \end{bmatrix} = \begin{bmatrix} 0.504 & -2.491 & 21.166 & -25.063 & 4.063 \end{bmatrix}
\]

and the associated covariance matrix is

\[
\text{Cov} (\hat{\theta}) = \begin{bmatrix}
0.00146 & -0.00752 & 0.0642 & -0.766 & 0.0123 \\
0.0391 & -0.334 & 0.398 & -0.642 & \\
2.859 & -3.408 & 0.549 & \\
4.063 & -0.655 & \\
0.106 & 
\end{bmatrix}
\]

Substitution of the parameter estimates into (2.24) generates the following estimate of the hazard function:

\[
\hat{\lambda}(t) = 0.504 - 2.491 \ t + 21.166 \ (t - 0.00548)_{+}^{3} \\
- 25.229 \ (t - 0.0849)_{+}^{3} + 4.063 \ (t - 0.498)_{+}^{3} \\
- 0.0104 \ (t - 5.262)_{+}^{3} + 0.0104 \ (t - 10.059)_{+}^{3}.
\]
Figure 2.11d displays this hazard function which decreases sharply in the left tail, levels out, and then begins slowly rising. The survival function estimate is

\[ \hat{S}(t) = \exp \left\{ -0.504 \, t + 1.245 \, t^2 - 5.292 \left( t - 0.00548 \right)^4 + 6.307 \left( t - 0.0849 \right)^4 - 1.016 \left( t - 0.498 \right)^4 + 0.00260 \left( t - 5.262 \right)^4 - 0.00260 \left( t - 10.059 \right)^4 \right\} . \]

The analyses with the "Medical Patient" and "Surgical Patient" data sets seem to indicate that five knots are needed to adequately describe survival in large data sets. The preference for five knots is clear in the analysis of the surgical patient data. However, this preference is not as clear with the medical patient data. In this latter case, though the likelihood ratio test indicates that the models based on four and five knots are significantly different, it is not clear whether this small difference is clinically significant. The primary difference between the two models is the number of iterations needed for convergence of the Newton-Raphson algorithm. The model with four knots needed 19 iterations, while the model with five knots needed only 5 iterations.
Chapter III
The Spline Function as Baseline Hazard in the
Proportional Hazards Model with Fixed Covariables

3.1 Introduction

It was shown in chapter II that the restricted cubic spline hazard function can
adequately describe a variety of hazard function shapes. This flexible and robust model will be
incorporated into the proportional hazards model as the baseline hazard function. The current
chapter will examine this model when fixed covariables are present.

The model, the likelihood function, and the approach to incorporating covariables into
the model are described in Section 3.2. In Section 3.3, the efficiency of estimating the
coefficient of continuous covariables with the spline proportional hazards model will be
compared with that of Cox's semiparametric proportional hazards model. The efficiency of
estimating the survival function will be examined in Section 3.4. The use of the spline model
in describing survival experience in the cardiovascular data sets described in the previous
chapter will be demonstrated in Section 3.5.
3.2 Incorporation of Covariables into the Model

3.2.1 The Model and Log-Likelihood Function

When the restricted cubic spline hazard function (2.14) is incorporated into the proportional hazards model as the baseline hazard, the resulting hazard function is

\[
\lambda(t|\mathbf{s}) = \left\{ a + b \ t + \sum_{j=1}^{m-2} \gamma_j \ w_j(t) \right\} \exp \left\{ \mathbf{s}^\top \ \mathbf{\beta} \right\},
\]  

(3.1)

where

\[
w_j(t) = \left( t - v_j \right)_+^{3} - \frac{(t - v_{m-1})_+^{3} (v_m - v_j)}{(v_m - v_{m-1})} + \frac{(t - v_m)_+^{3} (v_{m-1} - v_j)}{(v_m - v_{m-1})},
\]

\[\mathbf{s} = (z_1, z_2, \ldots, z_T)\] is a row vector of covariables, and \(\mathbf{\beta}\) is a column vector of associated coefficients.

The log-likelihood function, which is used to obtain estimates of the unknown parameters in (3.1), is

\[
\mathcal{L} = \sum_{i=1}^{n} \ell_i = \sum_{i=1}^{n} \left[ \delta_i \log \lambda(t_i|\mathbf{s}_i) - \Lambda(t_i|\mathbf{s}_i) \right],
\]  

(3.2)

where \(\lambda(t)\) is the hazard function defined in (3.1) and the cumulative hazard function is

\[
\Lambda(t|\mathbf{s}) = \left\{ a \ t + \frac{1}{2} b \ t^2 + \sum_{j=1}^{m-2} \gamma_j \ W_j(t) \right\} \exp \left\{ \mathbf{s}^\top \ \mathbf{\beta} \right\}
\]  

(3.3)

with

\[
W_j(t) = \frac{1}{4} \left[ \left( t - v_j \right)_+^{4} - \frac{(t - v_{m-1})_+^{4} (v_m - v_j)}{(v_m - v_{m-1})} + \frac{(t - v_m)_+^{4} (v_{m-1} - v_j)}{(v_m - v_{m-1})} \right].
\]
The Newton-Raphson procedure, as described in Section 2.3.2, is used to numerically optimize this log-likelihood function. The initial estimates of the covariable coefficients are 0, while the initial estimates for the baseline hazard parameters, as described in Section 2.3.2, will be \( \left( \frac{n}{\sum_{i=1}^{n} t_i} / \sum_{i=1}^{n} \delta_i, 0, 0, \ldots, 0 \right) \). The score vector and information matrix used in this optimization algorithm are presented in Appendix 3.

As with the homogeneous spline model, estimates and confidence intervals for the hazard, cumulative hazard, and survival functions can be calculated based upon the estimates of the model's unknown parameters and the corresponding information matrix \( I(\hat{\alpha}, \hat{\beta}, \hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_{m-2}, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_r) \), denoted as \( I(\hat{\theta}) \). Estimates of the hazard and cumulative hazard are calculated from (3.1) and (3.3), respectively. The survival estimate is obtained by transforming the cumulative hazard estimate using equation (1.1).

The confidence intervals for the hazard, cumulative hazard, and survival functions which were presented in Section 2.3.3 can be used within the context of the current model with fixed covariables if \( A \) and \( B \) are defined as

\[
A' = \exp \left\{ \mathbf{z} \beta \right\}
\]

\[
\begin{bmatrix}
1 \\
t \\
w_1(t) \\
\vdots \\
w_{m-2}(t) \\
z_1 \left( a + b t + \sum_{j=1}^{m-2} \gamma_j w_j(t) \right) \\
\vdots \\
z_r \left( a + b t + \sum_{j=1}^{m-2} \gamma_j w_j(t) \right)
\end{bmatrix}
\]
and

\[ B' = \exp \left\{ s \beta \right\} \]

\[
\begin{bmatrix}
    t \\
    \frac{1}{2} t^2 \\
    \mathcal{W}_1(t) \\
    \ldots \\
    \mathcal{W}_{m-2}(t) \\
    z_1 \left( a t + \frac{1}{2} b t^2 + \sum_{j=1}^{m-2} \gamma_j \mathcal{W}_j(t) \right) \\
    \ldots \\
    z_r \left( a t + \frac{1}{2} b t^2 + \sum_{j=1}^{m-2} \gamma_j \mathcal{W}_j(t) \right)
\end{bmatrix}
\]

3.2.2 Approach to Modeling

The proposed approach to modeling will involve two steps. First, with no covariates present in the model, the form of the baseline hazard, including the number and location of knots, will be determined. The second step involves the addition of covariates, either fixed or time-dependent, into the model. Initially, only fixed covariates will be considered for inclusion into the model.

When fixed covariates are not centered, the Newton-Raphson does not converge or converges after a "large" number of iterations. Therefore fixed covariates will be centered, by subtraction of overall mean values, prior to inclusion in the model.

3.3 Efficiency of Estimation of Covariable Coefficients

The purpose of the following investigation is to examine the efficiency of estimation with the spline model relative to Cox's proportional hazards model. The efficiency of
estimating $\beta$ will be examined in this section, and the efficiency of survival function estimation will be addressed in Section 3.4.

As indicated in Section 2.5, a complete examination of efficiency would consider issues such as the effect of underlying distribution, sample size, censoring, and knot location. In the case of regression survival models, such as the proportional hazards models, the effect of covariable distribution, magnitude of the covariable coefficient, and number of covariables, could also be examined. In this research, the examination of efficiency was limited to the effects of two underlying baseline hazard functions, three levels of censoring, two values for the covariable coefficient, and one covariable which is normally distributed. The two generalized Weibull hazard functions mentioned in section 2.5 were used as the underlying baseline hazard function. Data were uniformly censored with 0%, 25%, and 50% censorship. Beta was allowed to equal 0.1 and 1.0.

Two hundred observations, which were used in all simulations as the set of covariables, were generated from a normal distribution with mean 2 and variance 0.5. Based on these covariable observations, the baseline hazard function, and the value of $\beta$, survival times were generated. Fifty samples were generated for each combination of underlying baseline hazard, level of censorship, and coefficient value using this same set of covariables. A spline model with knots placed at the 5th, 33rd, 67th, and 95th percentiles of the empirical distribution formed by the uncensored survival times was fit to the data. Preliminary investigation had indicated that these four knot locations were satisfactory for describing the survival data ignoring covariables.

The relative efficiency of the covariable coefficient estimator was calculated as the ratio of the mean square error for the Cox model to the mean square error of the spline estimator. Mean square error was computed as

$$\text{MSE} = \frac{1}{50} \sum_{i=1}^{50} (\hat{\beta}_i - \beta)^2,$$
Table 3.1

EFFICIENCY OF COVARIABLE COEFFICIENT ESTIMATION WITH THE SPLINE MODEL RELATIVE TO THE COX MODEL FOR DATA GENERATED FROM THE GENERALIZED WEIBULL PROPORTIONAL HAZARDS MODEL

<table>
<thead>
<tr>
<th>Prop. Cens.</th>
<th>a=0.8, b=1, c=1, d=3</th>
<th>a=2, b=1, c=1, d=3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β=0.1</td>
<td>β=1.0</td>
</tr>
<tr>
<td>0.00 Spl MSE</td>
<td>0.0277</td>
<td>0.0252</td>
</tr>
<tr>
<td></td>
<td>0.0273</td>
<td>0.0238</td>
</tr>
<tr>
<td></td>
<td>0.986</td>
<td>0.944</td>
</tr>
<tr>
<td>0.25 Spl MSE</td>
<td>0.0314</td>
<td>0.0368</td>
</tr>
<tr>
<td></td>
<td>0.0321</td>
<td>0.0350</td>
</tr>
<tr>
<td></td>
<td>1.022</td>
<td>0.951</td>
</tr>
<tr>
<td>0.50 Spl MSE</td>
<td>0.0536</td>
<td>0.0461</td>
</tr>
<tr>
<td></td>
<td>0.0538</td>
<td>0.0450</td>
</tr>
<tr>
<td></td>
<td>1.004</td>
<td>0.976</td>
</tr>
</tbody>
</table>

1 Efficiency is the ratio of the mean square error of the Cox estimate (Cox MSE) over the mean square error of the spline estimate (Spl MSE). The estimates of mean square error and efficiency are based on 50 simulated data sets.
Table 3.2

EFFICIENCY OF SURVIVAL FUNCTION ESTIMATION WITH THE SPLINE MODEL RELATIVE TO
THE COX MODEL FOR DATA GENERATED FROM THE GENERALIZED WEIBULL
PROPORTIONAL HAZARDS MODEL WITH \( a=1, b=0.8, c=1, d=3, \beta=0.1 \)

<table>
<thead>
<tr>
<th>Prop. Cens.</th>
<th>Covariate Value</th>
<th>0.95</th>
<th>0.90</th>
<th>0.75</th>
<th>0.50</th>
<th>0.25</th>
<th>0.10</th>
<th>0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>1.67</td>
<td>3.442</td>
<td>1.884</td>
<td>1.100</td>
<td>0.943</td>
<td>1.030</td>
<td>1.256</td>
<td>1.346</td>
</tr>
<tr>
<td>2.00</td>
<td>4.077</td>
<td>1.924</td>
<td>1.019</td>
<td>1.176</td>
<td>1.027</td>
<td>1.261</td>
<td>1.562</td>
<td></td>
</tr>
<tr>
<td>2.33</td>
<td>4.666</td>
<td>1.523</td>
<td>1.126</td>
<td>1.082</td>
<td>0.959</td>
<td>1.028</td>
<td>1.047</td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>1.67</td>
<td>3.238</td>
<td>1.559</td>
<td>1.155</td>
<td>0.905</td>
<td>1.202</td>
<td>1.623</td>
<td>1.977</td>
</tr>
<tr>
<td>2.00</td>
<td>4.113</td>
<td>2.056</td>
<td>0.947</td>
<td>1.117</td>
<td>1.259</td>
<td>1.701</td>
<td>1.989</td>
<td></td>
</tr>
<tr>
<td>2.33</td>
<td>4.196</td>
<td>2.113</td>
<td>1.058</td>
<td>1.020</td>
<td>1.184</td>
<td>1.067</td>
<td>1.117</td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>1.67</td>
<td>2.417</td>
<td>1.062</td>
<td>1.161</td>
<td>0.969</td>
<td>1.513</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>2.00</td>
<td>2.298</td>
<td>1.252</td>
<td>1.068</td>
<td>1.178</td>
<td>0.921</td>
<td>#</td>
<td>#</td>
<td></td>
</tr>
<tr>
<td>2.33</td>
<td>1.861</td>
<td>1.355</td>
<td>1.104</td>
<td>1.197</td>
<td>1.039</td>
<td>#</td>
<td>#</td>
<td></td>
</tr>
</tbody>
</table>

1 The point \( t \) at which the true survival fractile \( p \) occurs is determined from the underlying distribution from which data is generated. This point satisfies \( P(T > t | z) = p \). Efficiency is the ratio of the mean square error of the Cox estimate over the mean square error of the spline estimate. Estimates of efficiency are based on 50 simulated data sets. Mean square error for both estimators is generally greater than 0.001 for the 75th, 50th, and 25th fractiles and less than 0.001 elsewhere.

# Insufficient number of survival estimates to compute efficiency.
Table 3.3

EFFICIENCY OF SURVIVAL FUNCTION ESTIMATION WITH THE SPLINE MODEL RELATIVE TO
THE COX MODEL FOR DATA GENERATED FROM THE GENERALIZED WEIBULL
PROPORTIONAL HAZARDS MODEL WITH a=1, b=0.8, c=1, d=3, β=1

<table>
<thead>
<tr>
<th>Prop. Cens.</th>
<th>Covariate Value</th>
<th>0.95</th>
<th>0.90</th>
<th>0.75</th>
<th>0.50</th>
<th>0.25</th>
<th>0.10</th>
<th>0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>1.67</td>
<td>0.703</td>
<td>1.058</td>
<td>0.975</td>
<td>1.262</td>
<td>1.063</td>
<td>0.941</td>
<td>1.643</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>2.028</td>
<td>1.388</td>
<td>0.969</td>
<td>1.642</td>
<td>0.972</td>
<td>0.854</td>
<td>1.229</td>
</tr>
<tr>
<td></td>
<td>2.33</td>
<td>1.260</td>
<td>1.669</td>
<td>1.196</td>
<td>0.994</td>
<td>1.370</td>
<td>1.191</td>
<td>0.923</td>
</tr>
<tr>
<td>0.25</td>
<td>1.67</td>
<td>1.023</td>
<td>0.841</td>
<td>1.114</td>
<td>1.017</td>
<td>1.142</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>2.993</td>
<td>1.324</td>
<td>1.284</td>
<td>1.123</td>
<td>1.022</td>
<td>1.568</td>
<td>1.516@</td>
</tr>
<tr>
<td></td>
<td>2.33</td>
<td>2.211</td>
<td>2.277</td>
<td>1.155</td>
<td>1.189</td>
<td>1.050</td>
<td>1.122</td>
<td>1.448</td>
</tr>
<tr>
<td>0.50</td>
<td>1.67</td>
<td>1.464</td>
<td>1.036</td>
<td>1.092</td>
<td>1.059</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>3.103</td>
<td>1.275</td>
<td>1.254</td>
<td>1.146</td>
<td>2.206&amp;</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>2.33</td>
<td>2.549</td>
<td>1.527</td>
<td>1.191</td>
<td>1.188</td>
<td>1.176</td>
<td>#</td>
<td>#</td>
</tr>
</tbody>
</table>

1 The point t at which the true survival fractile \( p \) occurs is determined from the underlying distribution from which data is generated. This point satisfies \( P( T > t | z ) = p \). Efficiency is the ratio of the mean square error of the Cox estimate over the mean square error of the spline estimate. Estimates of efficiency are based on 50 simulated data sets. Mean square error for both estimators is generally greater than 0.001 for the 75th, 50th, and 25th fractiles and less than 0.001 elsewhere.

@ Based on 48 samples.
& Based on 30 samples.
# Insufficient number of survival estimates to compute efficiency.
Table 3.4

EFFICIENCY OF SURVIVAL FUNCTION ESTIMATION WITH THE SPLINE MODEL RELATIVE TO
THE COX MODEL FOR DATA GENERATED FROM THE GENERALIZED WEIBULL
PROPORTIONAL HAZARDS MODEL WITH a=2, b=1, c=1, d=3, β=0.1

<table>
<thead>
<tr>
<th>Prop. Cens.</th>
<th>Covariate Value</th>
<th>0.95</th>
<th>0.90</th>
<th>0.75</th>
<th>0.50</th>
<th>0.25</th>
<th>0.10</th>
<th>0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>1.67</td>
<td>1.910</td>
<td>1.384</td>
<td>1.243</td>
<td>1.258</td>
<td>1.067</td>
<td>1.337</td>
<td>1.146</td>
</tr>
<tr>
<td>2.00</td>
<td>2.306</td>
<td>1.325</td>
<td>1.325</td>
<td>1.297</td>
<td>1.179</td>
<td>1.138</td>
<td>1.616</td>
<td></td>
</tr>
<tr>
<td>2.33</td>
<td>2.145</td>
<td>1.499</td>
<td>1.264</td>
<td>1.171</td>
<td>1.014</td>
<td>1.223</td>
<td>1.190</td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>1.67</td>
<td>1.280</td>
<td>1.229</td>
<td>1.273</td>
<td>0.925</td>
<td>1.117</td>
<td>1.289</td>
<td>1.063</td>
</tr>
<tr>
<td>2.00</td>
<td>1.603</td>
<td>1.344</td>
<td>1.262</td>
<td>0.928</td>
<td>1.332</td>
<td>1.147</td>
<td>1.222</td>
<td></td>
</tr>
<tr>
<td>2.33</td>
<td>2.755</td>
<td>1.547</td>
<td>1.203</td>
<td>1.125</td>
<td>1.451</td>
<td>0.951</td>
<td>1.146</td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>1.67</td>
<td>1.110</td>
<td>1.230</td>
<td>1.055</td>
<td>0.954</td>
<td>1.010</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>2.00</td>
<td>1.265</td>
<td>1.252</td>
<td>1.067</td>
<td>0.977</td>
<td>1.178</td>
<td>#</td>
<td>#</td>
<td></td>
</tr>
<tr>
<td>2.33</td>
<td>1.770</td>
<td>1.209</td>
<td>1.111</td>
<td>1.206</td>
<td>1.350</td>
<td>#</td>
<td>#</td>
<td></td>
</tr>
</tbody>
</table>

1 The point \( t \) at which the true survival fractile \( p \) occurs is determined from the underlying distribution from which data is generated. This point satisfies \( P(T > t | z) = p \). Efficiency is the ratio of the mean square error of the Cox estimate over the mean square error of the spline estimate. Estimates of efficiency are based on 50 simulated data sets. Mean square error for both estimators is generally greater than 0.001 for the 75th, 50th, and 25th fractiles and less than 0.001 elsewhere.

@ Based on 49 samples.

# Insufficient number of survival estimates to compute efficiency.
Table 3.5

EFFICIENCY OF SURVIVAL FUNCTION ESTIMATION WITH THE SPLINE MODEL RELATIVE TO THE COX MODEL FOR DATA GENERATED FROM THE GENERALIZED WEIBULL PROPORTIONAL HAZARDS MODEL WITH $a=-2$, $b=1$, $c=1$, $d=3$, $\beta=1.0$

<table>
<thead>
<tr>
<th>Prop. Cens.</th>
<th>Covariate Value</th>
<th>0.95</th>
<th>0.90</th>
<th>0.75</th>
<th>0.50</th>
<th>0.25</th>
<th>0.10</th>
<th>0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>1.67</td>
<td>1.734</td>
<td>1.234</td>
<td>1.167</td>
<td>1.097</td>
<td>1.232</td>
<td>1.388</td>
<td>1.115</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>1.941</td>
<td>1.636</td>
<td>0.985</td>
<td>1.531</td>
<td>1.470</td>
<td>0.972</td>
<td>1.258</td>
</tr>
<tr>
<td></td>
<td>2.33</td>
<td>1.727</td>
<td>1.275</td>
<td>1.322</td>
<td>1.304</td>
<td>1.356</td>
<td>1.226</td>
<td>0.883</td>
</tr>
<tr>
<td>0.25</td>
<td>1.67</td>
<td>1.195</td>
<td>0.965</td>
<td>1.186</td>
<td>1.000</td>
<td>1.174</td>
<td>1.264</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>2.327</td>
<td>1.121</td>
<td>1.500</td>
<td>1.072</td>
<td>1.069</td>
<td>1.422</td>
<td>2.057</td>
</tr>
<tr>
<td></td>
<td>2.33</td>
<td>2.357</td>
<td>1.664</td>
<td>1.219</td>
<td>1.193</td>
<td>1.149</td>
<td>1.115</td>
<td>1.363</td>
</tr>
<tr>
<td>0.50</td>
<td>1.67</td>
<td>1.168</td>
<td>0.994</td>
<td>1.046</td>
<td>1.089</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>1.649</td>
<td>0.985</td>
<td>1.195</td>
<td>1.103</td>
<td>1.052</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>2.33</td>
<td>1.644</td>
<td>1.275</td>
<td>1.314</td>
<td>1.127</td>
<td>1.360</td>
<td>#</td>
<td>#</td>
</tr>
</tbody>
</table>

1 The point $t$ at which the true survival fractile $p$ occurs is determined from the underlying distribution from which data is generated. This point satisfies $P(T > t | z) = p$. Efficiency is the ratio of the mean square error of the Cox estimate over the mean square error of the spline estimate. Estimates of efficiency are based on 50 simulated data sets. Mean square error for both estimators is generally greater than 0.001 for the 75th, 50th, and 25th fractiles and less than 0.001 elsewhere.

# Insufficient number of survival estimates to compute efficiency.
where $\beta$ is the value of the underlying coefficient and $\hat{\beta}_i$ is the estimate of $\beta$ from the $i$-th sample.

Table 3.1 summarizes the efficiency of estimating the covariable coefficient with the spline proportional hazards model. Efficiency ranges from 0.944 to 1.022 and averages 0.981 when the underlying distribution is generalized Weibull with $a=1$, $b=0.8$, $c=1$, and $d=3$. Similarly, when the data has been generated from the generalized distribution with $a=2$, $b=1$, $c=1$, and $d=3$, efficiency ranges from 0.965 to 1.022 and averages 0.995. Efficiency is approximately 1, regardless of underlying baseline hazard, $\beta$, or censoring proportion. These results indicate that the spline and Cox models are equally efficient at estimating $\beta$.

### 3.4 Efficiency of Survival Function Estimation

The efficiency of estimating the survival function when the covariable equals 1.67, 2.00, and 2.33 was examined. In each of the simulations described in section 3.3, survival function estimates were obtained at the 95th, 90th, 75th, 50th, 25th, 10th, and 5th survival fractiles of the underlying distribution. The underlying baseline hazard, beta value, and covariable value determined the underlying distribution, and hence the survival fractiles at which survival estimates were obtained. Whenever $t^*_p\mid z$, the survival fractiles satisfying the property $P(T > t^*_p\mid z) = p$, was larger than the sample's maximum uncensored survival time, estimates of survival were not obtained. Mean square error and efficiency were calculated as described in section 2.5.

Tables 3.2, 3.3, 3.4, and 3.5 summarize these efficiency results. Twenty six of the 223 (11.7%) statistics measuring efficiency and reported reported in these tables are less than 1.00, with only 3 being less than 0.90. Over half (116/223, 52.0%) are greater than 1.20, with 22 being greater then 2.00. Even though many of these statistics are based on the same set of samples, it is clear that the spline estimator of survival is more efficient than the Cox estimator. The effects of censoring and choice of $\beta$ are not clear. However, efficiency is greatest at the 95th survival fractile and approaches 1 as time increases.
3.5 Application to Cardiovascular Data

The cardiovascular disease data described in Chapter II are used to demonstrate the use of the spline proportional hazards model with fixed covariables. The relationship between left ventricular ejection fraction (×100) truncated at 60%, and referred to as LVEF, and survival will be examined.

The first step in any analysis is to determine the number and location of knots needed for analysis. This has been conducted in chapter II for both the “Medical Patient” and “Surgical Patient” data sets.

3.5.1 Medical Patient Data

Three models were fit to the “Medical Patient” data set. These included the spline model using 4 knots located at the 5th, 33rd, 67th, and 95th percentiles, the spline model using 5 knots located at the 5th, 20th, 33rd, 67th, and 95th percentiles, and the Cox model. A summary of parameter estimates for β, standard errors, and the number of iterations necessary for convergence of the Newton-Raphson algorithm are given below.

<table>
<thead>
<tr>
<th>Model</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Number of Iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spline (4 knots)</td>
<td>-0.0516</td>
<td>0.00262</td>
<td>7</td>
</tr>
<tr>
<td>Spline (5 knots)</td>
<td>-0.0515</td>
<td>0.00261</td>
<td>6</td>
</tr>
<tr>
<td>Cox</td>
<td>-0.0520</td>
<td>0.00262</td>
<td>5</td>
</tr>
</tbody>
</table>

The primary observation that should be made is that the estimates of the coefficient of the covariable and their standard errors are identical.

The spline model with both four and five knots was fit to the data since it was not clear in chapter II which is a “better model.” In chapter II, the model with five knots was significantly better than the model with four knots statistically. However, since the size of the change in survival estimates was small, it was not clear whether this difference was clinically significant.
A likelihood ratio test can be conducted to compare the spline regression models with four and five knots. The log-likelihood associated with the model with four knots is -2214.36, and the log-likelihood associated with the model with five knots is -2187.77. The likelihood ratio test statistic is $LR = -2 (-2214.36 - (-2187.77)) = 53.18$, which is statistically significant. Therefore, as shown in chapter II, the addition of the knot at the 20th percentile significantly improves the fit of the model. The model with five knots will be considered in the remainder of this section.

One of the advantages of the spline model over the Cox model is that an explicit expression of the hazard function is given. With estimates of the unknown parameters of the hazard function and their covariance matrix, estimates and confidence intervals for hazard, cumulative hazard, and survival can be made.

The parameter estimate $\hat{\theta} = (a, b, \gamma_1, \gamma_2, \gamma_3, \beta)$ and associated covariance matrix for the spline model with 5 knots are:

$$\hat{\theta} = \begin{bmatrix} 0.110 & -0.195 & 0.202 & -0.306 & 0.106 & -0.0515 \end{bmatrix}$$

and

$$\text{Cov}(\hat{\theta}) = 10^{-4} \times \begin{bmatrix} 1.16 & -2.69 & 2.66 & 4.01 & 1.36 & 0.0811 \\ 7.17 & -7.58 & 11.5 & 4.00 & -0.144 \\ 8.35 & -12.8 & 4.50 & 0.149 \\ & & & & & \\ 19.5 & -6.89 & -0.225 & 2.46 & 0.00781 \\ & & & & & 0.00677 \end{bmatrix}$$
FIGURE 3.1

SURVIVAL ESTIMATES AS A FUNCTION OF LEFT VENTRICULAR EJECTION FRACTION FOR MEDICAL PATIENTS

SOLID STEP FUNCTION — COX ESTIMATE
SMOOTH DASHED CURVE — SPLINE ESTIMATE USING KNOTS AT THE 5TH, 20TH, 33RD, 67TH, AND 80TH PERCENTILES

SAMPLE SIZE = 5789
PERCENTAGE CENSORED = 69.8 %
The estimate of the hazard function is

\[
\lambda(t|z) = \left\{ \begin{array}{c}
0.110 - 0.195 \ t + 0.202 \ (t - 0.0246)^3 \\
- 0.306 \ (t - 0.372)^3 + 0.106 \ (t - 1.084)^3 \\
- 0.00197 \ (t - 3.417)^3 - 0.0000802 \ (t - 9.131)^3 \\
\end{array} \right\} \\
x \ \exp \left\{ -0.0515 \ (z - 50.0393) \right\},
\]

and the estimate of the survival function is

\[
S(t|z) = \exp \left\{ \begin{array}{c} \\
- 0.110 + 0.0975 \ t^2 - 0.0505 \ (t - 0.0246)^4 \\
+ 0.0765 \ (t - 0.372)^4 - 0.0265 \ (t - 1.084)^4 \\
+ 0.000493 \ (t - 3.417)^4 + 0.0000401 \ (t - 9.131)^4 \\
\end{array} \right\} \ \exp \left\{ -0.0515 \ (z - 50.0393) \right\}
\]

Figure 3.1 graphically displays the survival curve predicted from both the spline model and the Cox model when left ventricular ejection fraction equals 30, 40, and 50. The predicted survival curves from the two models are almost identical, with the Cox estimate being a step function and the spline estimate being a smooth curve. Clearly, the spline model describes the data as well, if not better, than the Cox model.

### 3.5.2 Surgical Patient Data

The spline and Cox's proportional hazards models were fit to the "Surgical Patient" data. Knots in the spline model were located at the 5th, 20th, 33rd, 67th, and 95th percentiles of the empirical distribution formed by uncensored non-zero survival times. The number of iterations required for convergence of the Newton-Raphson algorithm and estimates of the
The coefficient of the covariable describing cardiac ejection fraction are listed below.

<table>
<thead>
<tr>
<th>Model</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Number of Iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spline (5 knots)</td>
<td>-0.0516</td>
<td>0.0038</td>
<td>9</td>
</tr>
<tr>
<td>Cox</td>
<td>-0.0493</td>
<td>0.0039</td>
<td>6</td>
</tr>
</tbody>
</table>

The estimates of the coefficient of the covariable are fairly close.

The estimate of the parameter vector \( \hat{\theta} = (a, b, \gamma_1, \gamma_2, \gamma_3, \beta) \) associated with the spline model is

\[
\hat{\theta} = \begin{bmatrix}
0.438 & -2.166 & 18.412 & -21.946 & 3.534 & -0.0516
\end{bmatrix}
\]

The associated covariance matrix is

\[
\text{Cov}(\theta) = \begin{bmatrix}
0.00123 & -0.00628 & 0.0536 & -0.0639 & 0.0103 & 0.0000417 \\
0.0325 & -0.278 & 0.331 & -0.0534 & 0.000206 \\
2.375 & -2.831 & 0.456 & 0.00175 \\
3.375 & -0.544 & -0.00208 \\
0.0877 & 0.000336 \\
0.0000144
\end{bmatrix}
\]

The hazard estimate associated with the spline model is

\[
\lambda(t|z) = \left\{ \begin{array}{l}
0.438 - 2.166 t + 18.412 \left(t - 0.00548\right)_+^3 \\
- 21.946 \left(t - 0.0849\right)_+^3 + 3.534 \left(t - 0.498\right)_+^3 \\
+ 0.000165 \left(t - 5.262\right)_+^3 - 0.000221 \left(t - 10.059\right)_+^3 \\
x \exp \left\{ -0.0516 (z - 51.458) \right\}
\end{array} \right. 
\]
FIGURE 3.2

SURVIVAL ESTIMATES AS A FUNCTION OF LEFT VENTRICULAR EJECTION FRACTION
FOR SURGICAL PATIENTS

SOLID STEP FUNCTION — COX ESTIMATE
SMOOTH DASHED CURVE — SPLINE ESTIMATE USING KNOTS AT THE 5TH, 20TH, 33RD,
67TH, AND 95TH PERCENTILES

SAMPLE SIZE = 2067
PERCENTAGE CENSORED = 86.8%
where \( z \) is the covariable describing LVEF. The survival function is

\[
S(t|z) = \exp \left\{ \left( -0.438 t + 1.083 t^2 \\
- 4.603 (t - 0.00548)^{4\frac{1}{4}} + 5.487 (t - 0.0849)^{4\frac{1}{4}} \\
- 0.884 (t - 0.498)^{4\frac{1}{4}} - 0.0000413 (t - 5.262)^{4\frac{1}{4}} \\
+ 0.0000553 (t - 10.059)^{4\frac{1}{4}} \right) \right\} \exp \left\{ -0.0516 (z - 51.458) \right\}.
\]

The survival curve predicted from the spline model and the Cox model when left ventricular ejection fraction equals 30, 40, and 50 is shown in Figure 3.2. The predicted survival curves when LVEF equals 50 are in good agreement. However, as the covariable gets further away from the overall mean (51.458), agreement between the Cox and spline curves lessens. As will be shown, this phenomenon is explainable by the difference, though small, in the estimates of \( \beta \).

In general, the survival curve is

\[
S(t|z) = \left\{ S_0(t) \right\}^{\exp \left( (z - \bar{z}) \beta \right)}
\]

where \( S_0(t) \) is the baseline survival function and \( \bar{z} \) is the overall mean. In the current example, the baseline survival curve occurs when LVEF equals the overall mean (51.458). The curves corresponding to "LVEF=50" in Figure 3.2 are approximately this estimated baseline survival function. The agreement between the Cox and spline survival estimates for this value of LVEF indicate that the baseline survival functions are approximately equal. Therefore differences between the Cox and spline models at other values of LVEF are clearly determined by differing estimates of \( \beta \). For example, if \( \hat{S}_0(t) = 0.80 \), then the Cox estimate of survival
when LVEF equals 30 is

\[ \hat{S}_c(t|z) = \{ 0.80 \} \exp \left( (30 - 51.458) \times (-0.0493) \right) = 0.525, \]

and the spline estimate is

\[ \hat{S}_s(t|z) = \{ 0.80 \} \exp \left( (30 - 51.458) \times (-0.0516) \right) = 0.509. \]

These estimates are consistent with those graphically shown in Figure 3.2.
Chapter IV
The Spline Function as Baseline Hazard in the Proportional Hazards Model with Time-Dependent Covariates Expressed as Step Functions

4.1 Introduction

Time-dependent covariates defined as step functions will be incorporated into the restricted cubic spline proportional hazards model which was introduced in Chapter III. The structure of these covariates, which will allow for the number and time of repeated measurements to vary across subjects, will be discussed in section 4.2. Three classes of time-dependent covariates which can be described within the proposed covariate structure will be examined. A covariate which describes an intervening event or treatment crossover will be developed in Section 4.3. The standard covariate representing measurements repeated over time will be examined in Section 4.4. A step function approximation of a defined covariate, such as the interaction of a fixed covariate and log(time), will be developed in Section 4.5.

4.2 Structure of the Model with Time-Dependent Covariates

4.2.1 Schematic and Notation Describing Time-Dependent Covariates

Survival models in which time-dependent covariates are assumed to be step functions have been developed by Petersen (1986) and Gaynor (1987). Petersen assumed that
measurements are obtained at the same unequal time intervals for all subjects. Gaynor assumes that measurements are equally spaced. The structure of the model now proposed will allow the number and time of repeated measurements to vary across subjects.

Let \( \mathbf{z}_i(t) = \left( z_{i1}(t), z_{i2}(t), \ldots, z_{ir}(t) \right) \) denote the \( r \times 1 \) vector of time-dependent covariables associated with the \( i \)-th subject, \( i = 1, 2, \ldots, n \). Though each component in \( \mathbf{z}_i(t) \) is a stochastic process, continuous observation of the covariable is not possible. Covariable measurements are obtained at distinct time points and used to approximate the covariable process. Since it is assumed that covariables remain constant between observations, \( \mathbf{z}_i(t) \) will be considered to be a step function.

Let \( 0 = t^{(i0)} < t^{(i1)} < t^{(i2)} < \ldots < t^{(iq_i)} \) denote time points at which components of \( \mathbf{z}_i(t) \) change. In addition, let \( t^{(iq_i+1)} = t_i \), the survival time for the \( i \)-th subject. It is assumed that \( t^{(iq_i)} \leq t^{(iq_i+1)} \). The steps or repeated measurements for the \( j \)-th component of \( \mathbf{z}_i(t) \) are denoted as

\[
z_{ij}(t) = \begin{cases} 
  z_{ijk} & \text{if } t^{(iq_i)} \leq t < t^{(i,k+1)} \text{ for } k = 0, 1, 2, \ldots, q_i-1 \\
  z_{ijq_i} & \text{if } t^{(iq_i)} \leq t \leq t^{(iq_i+1)} 
\end{cases} \tag{4.1}
\]

A schematic explaining this notation and showing the measurement schedule is given in Figure 4.1. Indicator functions of the form \( I(\text{expression}) \), which equal 1 if the expression is true and 0 otherwise, can be used to rewrite (4.1) as

\[
z_{ij}(t) = z_{ijq_i} I \left( t^{(iq_i)} \leq t < t^{(i,k+1)} \right) + \sum_{k=0}^{q_i-1} z_{ijk} I \left( t^{(iq_i)} \leq t < t^{(i,k+1)} \right) \tag{4.2}
\]

Let \( \mathbf{z}^{(ik)} \), \( k = 0, 1, \ldots, q_i \) denote the covariable vector associated with the \( k \)-th step for the \( i \)-th subject, i.e. \( \mathbf{z}^{(ik)} = (z_{i1k}, z_{i2k}, \ldots, z_{irk}) \). The vector of time-dependent covariables can
FIGURE 4.1

SCHEMATIC OF TIME-DEPENDENT COVARIABLE MEASUREMENTS FOR THE I-TH SUBJECT

The point at which the covariable vector changes or is remeasured is marked by X on the time line. The times at which these changes occur are \( t^{(ij)} \), where \( t^{(i0)} = 0, t^{(ij)} < t^{(i,j+1)} \) for \( j = 0, 1, 2, \ldots, q_i - 1 \), \( t^{(i,q_i)} \leq t^{(i,q_i+1)} \), and \( t^{(i,q_i)} = t_i, \) subject i's failure time. The value of the covariable measured at \( t^{(ij)} \) is \( z^{(ij)} \), \( j = 0, 1, 2, \ldots, q_i \). The total number of repeated measurements or the number of steps in the covariable step function is \( q_i + 1 \).
then be written as

\[ z_i(t) = z_{(i,q_j)} I\left(t \leq t \leq t_{(i,q_j+1)}\right) + \sum_{k=0}^{q_i-1} z_{(ik)} I\left(t_{(ik)} \leq t < t_{(i,k+1)}\right) \]  (4.3)

### 4.2.2 The Hazard Function and Log-Likelihood Function

The restricted cubic spline proportional hazards model with time-dependent covariates corresponding to the i-th subject is

\[ \lambda(t_i | z_i(t)) = \left\{ a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\} \exp \left\{ z_i(t) \beta \right\} \]  (4.4)

where

\[ w_j(t) = \left( t - v_j \right)^3 - \frac{(t - v_{m-1})^3 (v_m - v_j)}{(v_m - v_{m-1})} + \frac{(t - v_m)^3 (v_{m-1} - v_j)}{(v_m - v_{m-1})} \]

for \( j = 1, 2, \ldots, m-2 \). Since \( z_i(t) = z_{(i,q_i)} \), the final observed value of the covariate, the hazard in (4.4) is

\[ \lambda(t_i | z_i(t)) = \left\{ a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\} \exp \left\{ z_{(i,q_i)} \beta \right\} \]  (4.5)

The following cumulative hazard function can be obtained by integrating (4.4).

\[ \Lambda(t_i | z_i(t)) = \int_0^{t_i} \left\{ a + b u + \sum_{j=1}^{m-2} \gamma_j w_j(u) \right\} \exp \left\{ z_i(u) \beta \right\} \, du \]  (4.6)

Upon substitution of (4.3) into (4.4) the following formula for the cumulative hazard is
obtained.

\[ \Lambda( t_i | z_i(t) ) = \int_0^{t_i} \left\{ a + b u + \sum_{j=1}^{m-2} \gamma_j w_j(u) \right\} \]

\[ \times \exp \left\{ \sum_{k=0}^{q_i-1} z^{(ik)}(k) I \left( t^{(ik)} \leq u < t^{(i,k+1)} \right) \right\} \] \[ \int_{t^{(ik)}}^{t^{(i,k+1)}} \left\{ a + b u + \sum_{j=1}^{m-2} \gamma_j w_j(u) \right\} \] \[ du. \] \tag{4.7} \]

The integral in (4.7) can be considered to be the sum of integrals defined over intervals in which the time-dependent covariable is constant. Therefore

\[ \Lambda( t_i | z_i(t) ) = \sum_{k=0}^{q_i} \left( \exp \left\{ z^{(ik)}(k) \right\} \right) \]

\[ \times \int_{t^{(ik)}}^{t^{(i,k+1)}} \left\{ a + b u + \sum_{j=1}^{m-2} \gamma_j w_j(u) \right\} \] \[ du \right\). \tag{4.8} \]

Let

\[ \Lambda^*(t) = \int_0^t \left\{ a + b u + \sum_{j=1}^{m-2} \gamma_j w_j(u) \right\} \] \[ du \]

\[ = a t + \frac{1}{2} b t^2 + \sum_{j=1}^{m-2} \gamma_j w_j(t) \]

where

\[ w_j(t) = \frac{1}{4} \left[ (t - v_j)_+^4 - \frac{(t - v_{m-1})_+^4 (v_m - v_j)}{(v_m - v_{m-1})} + \frac{(t - v_m)_+^4 (v_{m-1} - v_j)}{(v_m - v_{m-1})} \right]. \tag{4.9} \]
The cumulative hazard function in (4.8) can be rewritten as

\[ \Lambda(t_i | \mathbf{s}_i(t)) = \sum_{k=0}^{q_i} \exp \left( \mathbf{s}^{(ik)} \beta \right) \left[ \Lambda^*(t^{(i,k+1)}) - \Lambda^*(t^{(ik)}) \right]. \]  

(4.10)

The log-likelihood function, which is used to obtain estimates of the unknown parameters in (4.4), is

\[ \mathcal{L} = \sum_{i=1}^{n} \mathcal{L}_i = \sum_{i=1}^{n} \left[ \delta_i \log \Lambda(t_i | \mathbf{s}_i(t)) - \Lambda(t_i | \mathbf{s}_i(t)) \right] \]

where \( \Lambda(t_i | \mathbf{s}_i(t)) \) and \( \Lambda(t_i | \mathbf{s}_i(t)) \) are defined in (4.5) and (4.10), respectively. Therefore

\[ \mathcal{L} = \sum_{i=1}^{n} \left( \delta_i \log \left\{ a + b \ t_i + \sum_{j=1}^{m-2} \gamma_j \ w_j(t_i) \right\} + \delta_i \left\{ \mathbf{s}^{(i,q_i)} \beta \right\} \right. \]

\[- \sum_{k=0}^{q_i} \exp \left( \mathbf{s}^{(ik)} \beta \right) \left[ \Lambda^*(t^{(i,k+1)}) - \Lambda^*(t^{(ik)}) \right] \]. \]  

(4.11)

The log-likelihood function can be further manipulated to obtain

\[ \mathcal{L} = \sum_{i=1}^{n} \left( \delta_i \log \left\{ a + b \ t_i + \sum_{j=1}^{m-2} \gamma_j \ w_j(t_i) \right\} + \delta_i \left\{ \mathbf{s}^{(i,q_i)} \beta \right\} \right. \]

\[- \sum_{i=1}^{n} \sum_{k=0}^{q_i} \exp \left( \mathbf{s}^{(ik)} \beta \right) \left[ \Lambda^*(t^{(i,k+1)}) - \Lambda^*(t^{(ik)}) \right] \]. \]  

(4.12)

Failures contribute only to the first component of the log-likelihood function (4.12).

This component is a function of the final value of the subject’s covariables and failure time.

The time at which covariable values change is necessary information for the second component in (4.12); duration is not sufficient information. This summation is a function of all survival times, all time points at which covariables change, and all steps of the covariables. Petersen
(1986) indicates that for analysis purposes the data file can consist of a separate record for each interval of time in which covariates remain constant.

Given a data file structured according to Petersen's recommendation, the computation of the log-likelihood would proceed as follows.

Set log-likelihood $L = 0$;

Do over all records:

- Compute $L = L + \exp \left( \mathbf{s}^T \tilde{\mathbf{\beta}} \right) \left[ \Lambda^*(t_{\text{end}}) - \Lambda^*(t_{\text{beg}}) \right]$
  where $\mathbf{s}$ is the vector of covariates on the current record, $t_{\text{beg}}$ is the time at which the covariates were measured, and $t_{\text{end}}$ is the time at which the covariable vector next changed.

When the current record is the subject's final record and and survival time is uncensored, compute

$$L = L + \log \left( a + b t_{\text{end}} + \sum_{j=1}^{m-2} \gamma_j w_j(t_{\text{end}}) \right) + \mathbf{s}^T \tilde{\mathbf{\beta}}.$$ 

The score vector and information matrix are presented in Appendix 4. The computation of these quantities is similar to that for the log-likelihood function.

### 4.2.3 Confidence Intervals

Confidence intervals for the hazard, cumulative hazard, and survival functions were presented in Section 2.3.3 assuming a homogeneous restricted cubic spline hazard model. These formulas can be adapted for use when the restricted cubic spline model has been
incorporated into the proportional hazards model with time-dependent covariables by redefining matrices $A$ and $B$. These matrices will be defined using the notation given in Section 4.2.2. However, the subscript denoting subject will be dropped in the definitions for $A$ and $B$. The confidence interval for hazard functions when time-dependent covariables are present is computed in terms of

$$A' = \exp \left\{ z^{(q)} \beta \right\} = \begin{bmatrix}
1 \\
t \\
w_1(t) \\
\vdots \\
w_{m-2}(t) \\
 z_{1q} \left( a + b t + \sum_{j=1}^{m-2} \gamma_j w_j(t) \right) \\
\vdots \\
z_{rq} \left( a + b t + \sum_{j=1}^{m-2} \gamma_j w_j(t) \right)
\end{bmatrix}$$

where $z^{(q)} = (z_{1q}, z_{2q}, \ldots, z_{rq})$ is the $q$-th and final set of covariable measurements and $t$ is
the survival time. Cumulative hazards and survival function confidence intervals are computed in terms of

\[ B' = \sum_{k=0}^{q} \left\{ \exp \{ z^{(k)} \beta \} \right\} B_k \]

where

\[
B_k = \begin{bmatrix}
(t^{(k+1)} - t^{(k)}) \\
\frac{1}{2} \left\{ (t^{(k+1)})^2 - (t^{(k)})^2 \right\} \\
\omega_1(t^{(k+1)}) - \omega_1(t^{(k)}) \\
\vdots \\
\omega_{m-2}(t^{(k+1)}) - \omega_{m-2}(t^{(k)}) \\
z_{kl} \left\{ \Lambda^*(t^{(k+1)}) - \Lambda^*(t^{(k)}) \right\} \\
\vdots \\
z_{kr} \left\{ \Lambda^*(t^{(k+1)}) - \Lambda^*(t^{(k)}) \right\}
\end{bmatrix}
\]
4.2.4 Maximum Likelihood Estimation

Optimization in chapters II and III was inefficiently performed using programs written in SAS/IML. Since computations have the potential of being expensive when time-dependent covariates are present, a new SAS procedure, based on the PHGLM procedure (Harrell, 1986), was written in order to exploit PHGLM's efficient Newton-Raphson optimization and programming. A description of the resulting procedure PHSPLM is given in Appendix 5. Examples of the procedure's use are presented in Appendix 6. This procedure has been written to estimate the unknown parameters in the homogeneous model, the spline model with fixed covariates, and also the model with time-dependent covariates. PHSPLM reparameterizes the hazard model as

\[
\lambda(t \mid z_1(t)) = \left\{ a + b \frac{k}{D} + \sum_{j=1}^{m-2} \gamma_j w_j(t) \right\} \exp \left\{ z_1(t) \beta \right\},
\]

where

\[
w_j(t) = \left( \frac{t - v_j}{D} \right)_+^3 - \left( \frac{t - v_{m-1}}{D} \right)_+^3 \frac{v_m - v_j}{v_m - v_{m-1}} + \left( \frac{t - v_m}{D} \right)_+^3 \frac{v_{m-1} - v_j}{v_m - v_{m-1}}
\]

and \( D = v_m - v_{m-1} \), in order to minimize computational difficulties. Formulas presented in the early part of this chapter and in previous chapters and appendices were adapted for this parameterization. Analytical results discussed in the remaining part of this chapter will assume this parameterization.

A data file consisting of a separate record for each interval of time in which covariates remain constant is used as input to the PHSPLM procedure. When time-dependent covariates are present, each record should contain (i) the value of the covariate at the beginning of the interval, (ii) time at which the interval begins, and (iii) survival times for the subject. When the record is the first record for the subject, it should also indicate whether or not the subject is censored.
4.3 Covariable Describing an Intervening Event or Treatment Crossover

A covariable which describes an intervening nonfatal event or treatment crossover is the simplest time-dependent covariable having the structure laid out in Section 4.2.1. This covariable is characterized by an indicator variable

\[ z_1(t) = I\{t \geq \tau_1\} = \begin{cases} 1 & \text{if } t \geq \tau_1 \\ 0 & \text{if } t < \tau_1 \end{cases} \quad (4.13) \]

where \( \tau_1 \geq 0 \) is the time of occurrence of an event or change of treatment or condition for the i-th subject. Using the notation of section 4.2, \( q_i = 0 \) if the event does not occur and \( q_i = 1 \) when the event occurs at \( \tau_1 \). A time line depicting the measurements involved with a subject who experiences an intervening nonfatal event or treatment crossover at \( \tau_1 > 0 \) is below.

\[ \begin{array}{c}
\uparrow \\
(10)
\end{array} \quad z_1^{(11)} \quad \begin{array}{c}
\downarrow \\
\uparrow
\end{array} \quad z_1^{(10)} \quad \begin{array}{c}
\downarrow \\
X
\end{array} \quad (10) \quad (11) = \tau_1 \quad (12) = t_1
\]

A data file used for analysis has one or two records for each subject, depending on when and if the intervening event occurs.

4.3.1 Comparison with the Cox Model

Lee (1985) uses a binary time-dependent covariable when he addresses the question of how survival is affected by the occurrence of an intervening nonfatal myocardial infarction (MI) in patients with coronary artery disease. The indicator function \( z_1(t) \) equals 0 if no MI has occurred before time \( t \) and 1 otherwise. Lee's analyses were performed using BMDP2L (Dixon, 1981), a statistical program which allows for time-dependent covariables in the Cox
model. The results of his analyses will be contrasted with the results from the restricted cubic
spline proportional hazards model.

Lee's data, which were collected between November 1969 and January 1985, consist of
the survival experience of 3085 consecutive patients who were catheterized for chest pain at
Duke University Medical Center and medically treated. Survival time is measured from the
date of catheterization and endpoint is cardiovascular death. A more complete description of
the data, including a discussion of the perils of treating MI as a non-time-dependent covariate,
can be found in Lee (1985).

Lee's first analysis examined the effect of the binary covariable MI without adjusting
for any other factors. The results of his analysis and an analysis using the spline model are
summarized in Table 4.1. Five knots \((0.0137, 0.2163, 0.7365, 3.1538, 8.7803)\) are used in the spline analysis. These are the 5th, 20th, 33rd, 67th, and 95th
percentiles of the empirical distributions formed by nonzero uncensored failure times. The
covariable coefficients and standard errors in the two models agree well. The relative risk of
death after an intervening nonfatal MI compared to the risk before its occurrence is 2.45 based
on the Cox model and 2.44 based on the spline model. The striking difference between the two
analyses is the amount of CPU time used for estimation on an IBM 3081 mainframe computer.
The Cox analysis used almost 26 minutes, whereas the spline estimation used 6 seconds. Thus
time-dependent covariable analyses are much more feasible with the spline model.

Another advantage of the spline model is that explicit expressions for the survival and
hazard function can be obtained. The hazard function for this example is:

\[
\lambda(t | MI(t)) = \left\{ \begin{array}{l}
0.276 - 0.762 t + 1.796 (t - 0.0137)^3_+ \\
- 2.504 (t - 0.216)^3_+ + 0.710 (t - 0.736)^3_+ \\
- 0.00207 (t - 3.154)^3_+ + 0.0000905 (t - 8.780)^3_+ \\
\end{array} \right. \\
x \exp \left\{ 0.893 (MI(t) - 0.0858) \right\}.
\]
Table 4.1

Effect of an Intervening Nonfatal Myocardial Infarction
Using the Cox and Spline Models

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Cox Model</th>
<th>Spline Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff-</td>
<td>Std.</td>
</tr>
<tr>
<td></td>
<td>icient</td>
<td>Error</td>
</tr>
<tr>
<td>a</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>b</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>71</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>72</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>73</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MI</td>
<td>0.8947</td>
<td>0.1201</td>
</tr>
</tbody>
</table>

Global likelihood ratio $\chi^2$: 59.09 45.80

CPU time (min:sec): 25:57 0:06
The survival function is \( S(t \mid \text{MI}(t)) = \exp \left( - \Lambda(t \mid \text{MI}(t)) \right) \) where for \( t \leq \tau_i \),

\[
\Lambda(t \mid \text{MI}(t)) = \left\{ \begin{array}{l}
0.276 \ t - 0.381 \ t^2 + 0.449 \ (t - 0.0137)_+^4 \\
- 0.624 \ (t - 0.216)_+^4 + 0.178 \ (t - 0.736)_+^4 \\
- 0.000518 \ (t - 3.154)_+^4 + 0.0000226 \ (t - 8.780)_+^4 \\
\end{array} \right\}
\times \exp \left\{ \begin{array}{l}
0.893 (-0.0858) \\
\end{array} \right\},
\]

and for \( t > \tau_i \),

\[
\Lambda(t \mid \text{MI}(t)) = 
\left\{ \begin{array}{l}
0.276 \ \tau_i - 0.381 \ \tau_i^2 + 0.449 \ (\tau_i - 0.0137)_+^4 \\
- 0.624 \ (\tau_i - 0.216)_+^4 + 0.178 \ (\tau_i - 0.736)_+^4 \\
- 0.000518 \ (\tau_i - 3.154)_+^4 + 0.0000226 \ (\tau_i - 8.780)_+^4 \\
\end{array} \right\}
\times \exp \left\{ \begin{array}{l}
0.893 (-0.0858) \\
\end{array} \right\}
- \left\{ \begin{array}{l}
0.276 \left[ t - \tau_i \right] - 0.381 \left[ t^2 - \tau_i^2 \right] \\
+ 0.449 \left[ (t - 0.0137)_+^4 - (\tau_i - 0.0137)_+^4 \right] \\
- 0.624 \left[ (t - 0.216)_+^4 - (\tau_i - 0.216)_+^4 \right] \\
+ 0.178 \left[ (t - 0.736)_+^4 - (\tau_i - 0.736)_+^4 \right] \\
- 0.000518 \left[ (t - 3.154)_+^4 - (\tau_i - 3.154)_+^4 \right] \\
+ 0.0000226 \left[ (t - 8.780)_+^4 - (\tau_i - 8.780)_+^4 \right] \\
\end{array} \right\}
\times \exp \left\{ \begin{array}{l}
0.893 (1 - 0.0858) \\
\end{array} \right\}.
In addition, confidence intervals based on the covariance matrix of the unknown parameters, 

\[ \text{Var}(\hat{\theta}) = \begin{bmatrix} 
0.000647 \\
-0.0123 & 0.251 \\
0.913 & -19.326 & 1508.024 \\
-1.272 & 26.989 & -2108.520 & 2948.44 \\
0.360 & -7.694 & 603.547 & -844.261 & 242.026 \\
0.000145 & -0.00214 & 0.132 & -0.182 & 0.0496 & 0.0144 
\end{bmatrix} \]

can be obtained for hazard and survival estimates. Figures 4.2a and b show the survival and hazard functions with confidence intervals for a patient who experiences no intervening nonfatal MI. Figures 4.3a and b show these estimates for a patient who experiences an intervening nonfatal MI during catheterization or soon after. Figures 4.4a and b show estimates for a patient who experiences an intervening nonfatal MI after 2.25 years. This patient's hazard function for the first 2.25 years is identical to Figure 4.2b, and for the remaining years, it is identical to the latter portions of Figure 4.3b. In a sense, this patient's hazard jumps from one hazard curve where proportionality holds to another (Gaynor, 1987).

Lee also examined the effect of an intervening MI after adjusting for 24 baseline prognostic factors with the index variable HSM. The results for the Cox analysis and the identical analysis using the spline model are presented in Table 4.2. The covariable coefficients are almost identical. The risk of dying after a nonfatal MI compared to the risk prior to the MI is estimated to be 2.1616 in the Cox model and 2.1613 in the spline model.

A variation on the indicator time-dependent covariable is the following

\[ z_2(t) = \mathbb{I}\{t \geq \tau_i\} = \begin{cases} 
\tau_i & \text{if } t \geq \tau_i \\
0 & \text{if } t < \tau_i
\end{cases} \]

This covariable is nonzero whenever the intervening event has occurred and equals the time at
FIGURE 4.2a
SURVIVAL ESTIMATE WITH 95% CONFIDENCE LIMITS FOR A PATIENT WHO DOES NOT EXPERIENCE A MYOCARDIAL INFARCTION

FIGURE 4.2b
HAZARD ESTIMATE WITH 95% CONFIDENCE LIMITS FOR A PATIENT WHO DOES NOT EXPERIENCE A MYOCARDIAL INFARCTION
FIGURE 4.3a
SURVIVAL ESTIMATE WITH 95% CONFIDENCE LIMITS FOR A PATIENT WHO EXPERIENCES
A MYOCARDIAL INFARCTION SOON AFTER CATHETERIZATION

FIGURE 4.3b
HAZARD ESTIMATE WITH 95% CONFIDENCE LIMITS FOR A PATIENT WHO EXPERIENCES
A MYOCARDIAL INFARCTION SOON AFTER CATHETERIZATION
FIGURE 4.4a
SURVIVAL ESTIMATE WITH 95% CONFIDENCE LIMITS FOR A PATIENT WHO EXPERIENCES
A MYOCARDIAL INFARCTION 2.25 YEARS AFTER CATHETERIZATION

FIGURE 4.4b
HAZARD ESTIMATE WITH 95% CONFIDENCE LIMITS FOR A PATIENT WHO EXPERIENCES
A MYOCARDIAL INFARCTION 2.25 YEARS AFTER CATHETERIZATION
Table 4.2

Effect of an Intervening Nonfatal Myocardial Infarction adjusted for HSM Using the Cox and Spline Models

| Parameter | Cox Model | | | | Spline Model | | | |
|-----------|-----------|---|---|---|---|---|---|
|           | Coefficient | Std. Error | \(e^{|\text{coeff}|}\) | Coefficient | Std. Error | \(e^{|\text{coeff}|}\) |
| a         | -          | -            | -            | 0.1385     | 0.0151     | -            |
| b         | -          | -            | -            | -2.1079    | 0.2887     | -            |
| \(\gamma_1\) | -          | -            | -            | 158.8621   | 22.5549    | -            |
| \(\gamma_2\) | -          | -            | -            | -221.5882  | 31.5682    | -            |
| \(\gamma_3\) | -          | -            | -            | 62.9331    | 9.0749     | -            |
| HSM       | 1.0037     | 0.0368       | 2.7285       | 1.0126     | 0.0369     | 2.7529       |
| MI        | 0.7709     | 0.1203       | 2.1616       | 0.7707     | 0.1202     | 2.1613       |

Global likelihood ratio \(\chi^2\): 879.56 777.05

CPU time (min:sec): 24:19 0:09
Table 4.3

Effect of an Intervening Nonfatal Myocardial Infarction
Adjusted for HSM and Waiting Time Until an MI
Using the Cox and Spline Models

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Cox Model</th>
<th></th>
<th></th>
<th>Spline Model</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Std. Error</td>
<td>e-coeff</td>
<td>Coefficient</td>
<td>Std. Error</td>
<td>e-coeff</td>
</tr>
<tr>
<td>a</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.1404</td>
<td>0.0153</td>
<td>-</td>
</tr>
<tr>
<td>b</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-2.1373</td>
<td>0.2931</td>
<td>-</td>
</tr>
<tr>
<td>71</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>161.1957</td>
<td>22.8971</td>
<td>-</td>
</tr>
<tr>
<td>72</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-224.8568</td>
<td>32.0461</td>
<td>-</td>
</tr>
<tr>
<td>73</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>63.8728</td>
<td>9.2111</td>
<td>-</td>
</tr>
<tr>
<td>HSM</td>
<td>1.0019</td>
<td>0.0368</td>
<td>2.7235</td>
<td>1.0108</td>
<td>0.0369</td>
<td>2.7478</td>
</tr>
<tr>
<td>MI</td>
<td>0.6144</td>
<td>0.1525</td>
<td>1.8485</td>
<td>0.6209</td>
<td>0.1517</td>
<td>1.8606</td>
</tr>
<tr>
<td>WTIME</td>
<td>0.0975</td>
<td>0.0522</td>
<td>-</td>
<td>0.0930</td>
<td>0.0515</td>
<td>-</td>
</tr>
</tbody>
</table>

Global likelihood ratio $\chi^2$: 889.11
CPU time (min:sec): 43:33

Global likelihood ratio $\chi^2$: 780.09
CPU time (min:sec): 0:12
which the event occurs. Lee uses such a covariable to test whether the waiting time until infarction (WTIME) affects prognosis in his third analysis. A summary of the Cox and spline analyses examining the effect of an intervening MI adjusted for HSM and WTIME is presented in Table 4.3. Covariable coefficients, standard errors, and relative risks in the two models are almost identical. CPU usage is dramatically different, with the Cox model using over 43 minutes and the spline model using 12 seconds.

4.3.2 Modeling the Effect of an Intervening Event as a Function of the Time Since the Event’s Occurrence

In general, the effect of an intervening event may not have a constant, permanent effect. Rather the effect may wane and in some cases, may intensify over time. The models presented by Lee assume that a myocardial infarction has a permanent constant effect, rather than a time-varying one. However, it is possible that patients undergo a "high risk" phase after an infarction, and later the risk of death lessens.

The restricted cubic spline hazard model readily allows for such complexities. Let $\lambda_0(t)$ be the restricted cubic spline baseline hazard function, and $\tau_i$ be the time of the intervening event. A model describing the hazard for the i-th subject which assumes that the effect of an intervening event may change as a function of time is

$$
\lambda(t|\tau_i) = \lambda_0(t) \exp \left\{ \beta \xi(t|\tau_i) \right\}.
$$

where $\xi(t|\tau_i)$ is 0 for $t < \tau_i$ and is a function of $t - \tau_i$ when $t \geq \tau_i$. Since both the functional form of $\xi(t|\tau_i)$ is unknown and only step functions can be incorporated into the spline hazard model, it is necessary to approximate $\xi(t|\tau_i)$. Two approaches will be considered.

The first approach involves the creation of indicator time-dependent covariables which describe the effect of an intervening event during defined intervals of time. Various
parameterizations can be used. However, the parameterization which will be used here will allow covariates to be interpreted as the incremental effect over and above the effect in previous time periods. For example, the time following an intervening event may be partitioned as \([0, 1), [1, 2), [2, 3), \text{ and } [3, \infty)\) years. Covariates are then defined as:

\[
\begin{align*}
\xi_0(t|\tau_i) &= I\{ t \geq \tau_i \} \\
\xi_1(t|\tau_i) &= I\{ t \geq \tau_i + 1 \} \\
\xi_2(t|\tau_i) &= I\{ t \geq \tau_i + 2 \} \\
\xi_3(t|\tau_i) &= I\{ t \geq \tau_i + 3 \} .
\end{align*}
\]

The resulting model which would be used to examine the effect of an intervening event is

\[
\lambda(t|\tau_i) = \lambda_0(t) \exp \left\{ \beta_0 \xi_0(t|\tau_i) + \beta_1 \xi_1(t|\tau_i) + \beta_2 \xi_2(t|\tau_i) + \beta_3 \xi_3(t|\tau_i) \right\} .
\]

Such an approach is used to examine the effect of an MI in Lee’s data without adjustment for other covariates. Let \(\tau_i\) denote the time of the infarction. Let MI be defined as \(\xi_0\), MI1 be defined as \(\xi_1\), etc. Table 4.4 summarizes the model’s parameter estimates. The following three test statistics are available for testing whether the coefficients of MI1, MI2, and MI3 are zero.

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score Test</td>
<td>4.10</td>
<td>0.251</td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>3.91</td>
<td>0.274</td>
</tr>
</tbody>
</table>

The likelihood ratio test statistic is the difference between 4373.23, the log-likelihood \((-2 \log L)\) for the model with only MI, and 4369.32, the log-likelihood for the full model. The
Table 4.4

Effect of an Intervening Nonfatal Myocardial Infarction as a Function of Time Since Its Occurrence Using Indicator Variables in the Spline Hazard Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>Chi-Square</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0.2981</td>
<td>0.0278</td>
<td>115.26</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>b</td>
<td>-4.6469</td>
<td>0.5445</td>
<td>72.84</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>γ₁</td>
<td>347.6185</td>
<td>42.1531</td>
<td>68.01</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>γ₂</td>
<td>-484.6481</td>
<td>58.9430</td>
<td>67.61</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>γ₃</td>
<td>137.9251</td>
<td>16.8894</td>
<td>66.21</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>MI</td>
<td>1.1746</td>
<td>0.1775</td>
<td>43.77</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>MI₁</td>
<td>-0.4430</td>
<td>0.2807</td>
<td>2.96</td>
<td>0.085</td>
</tr>
<tr>
<td>MI₂</td>
<td>0.0914</td>
<td>0.3581</td>
<td>0.07</td>
<td>0.799</td>
</tr>
<tr>
<td>MI₃</td>
<td>0.0199</td>
<td>0.3294</td>
<td>0.00</td>
<td>0.958</td>
</tr>
</tbody>
</table>
contribution of the second (MI1), third (MI2), and fourth (MI3) steps in explaining the effect of a myocardial infarction is not significant.

In order to obtain a good approximation of $\xi(t|\tau_1)$, a large number of time intervals need to be generated. Such an approximation can potentially become numerically unstable and/or unwieldly. As the number of steps increases, the degrees of freedom associated with a test to determine whether the effect of an intervening event has a permanent constant effect increases. As a result the test becomes less powerful.

An alternative approach is to consider a spline approximation of $\xi(t|\tau_1)$. It has been established in chapters I and II that the restricted cubic spline function can describe a variety of functional shapes with a minimal number of parameters. Consider the following spline function in $t - \tau_1$ with $m$ knots defined for $t \geq \tau_1$.

$$
sp(t|\tau_1) = \beta_1 + \beta_2 (t - \tau_1) + \sum_{j=1}^{m-2} \beta_{j+2} u_{j+2}(t - \tau_1)
$$

(4.14)

where

$$
u_j(t - \tau_1) = (t - \tau_1 - v_j^*)^3 + \frac{(t - \tau_1 - v_{m-1}^*)^3}{(v_m^* - v_{m-1}^*)} + \frac{(t - \tau_1 - v_m^*)^3}{(v_m^* - v_{m-1}^*)}$$

for $j = 3, \ldots, m; v_1^*, \ldots, v_m^*$ are knots determined from the distribution of the difference between failure time and the time the intervening event occurred; and $\beta_1, \ldots, \beta_m$ are unknown regression coefficients. For notational purposes, let

$$u_1(t - \tau_1) = 1 \quad \text{and} \quad u_2(t - \tau_1) = t - \tau_1$$
Therefore

\[ sp(t|\tau_i) = \sum_{j=1}^{m} \beta_j \ u_j(t - \tau_i) \]

As will be explained in section 4.5, the spline function \( sp(t|\tau_i) \) can be approximated by a step function. Let \( c \) be the length of a time interval which defines the steps of the approximating step function. Using the left endpoint as the value of the step function within an interval, one approximation of \( sp(t|\tau_i) \) is

\[
sp_A(t|\tau_i) = \begin{cases} 
sp(0) & \text{if } 0 \leq t - \tau_i < c \\
sp(c) & \text{if } c \leq t - \tau_i < 2c \\
sp(2c) & \text{if } 2c \leq t - \tau_i \leq 3c \\
\vdots & \text{ } \end{cases} \quad (4.15)
\]

Alternative ways of approximating the spline function by a step function are suggested in section 4.5 and were investigated for this particular example. However, the approach described here seems to be the most satisfactory.

The components of \( sp(t|\tau_i) \), which are \( u_1(t - \tau_i), \ldots, u_m(t - \tau_i) \), can be treated as covariates which are repeatedly measured at the beginning of each time interval. In order to estimate the unknown parameters in the model using PHSPLM, it is necessary to generate additional records corresponding to steps of \( sp_A(t|\tau_i) \). Given a subject with survival time \( t_i \), a record corresponding to each of the remeasurement points or steps in (4.15) for which \( t \leq t_i \) is created as input to PROC PHSPLM.

Lee's data will be used to demonstrate this approach to determining whether an intervening event has a permanent, constant effect. Four knots are included in the regression part of the following model

\[
\lambda(t) = \left( a + b \ t + \sum_{k=1}^{3} \gamma_k \ w_k(t) \right) \ \exp \left\{ \sum_{j=1}^{4} \beta_j \ u_j(t - \tau_i) \right\} \quad (4.16)
\]
where $w_k(t)$ is defined in (4.4) and $u_j(t - \tau_j)$ is defined in (4.14). The location of the regression knots are determined from the distribution formed by the difference between the patient’s survival time, $t_i$, and the time of the intervening event, $\tau_i$, for patients who experienced a myocardial infarction. The chosen knot locations are 0.049048, 0.36683, 5.45801, and 9.46123 years. Time intervals of length 0.10 years will be used in order to approximate the spline function in the regression component of (4.16).

Table 4.5 summarizes the estimated model. The following score and likelihood ratio test are conducted to determine whether the effect of an MI is best described by a constant or is a function of $(t - \tau_i)$. The null hypothesis is $H_0: \beta_2 = \beta_3 = \beta_4 = 0$.

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score Test</td>
<td>3.26</td>
<td>0.353</td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>3.18</td>
<td>0.365</td>
</tr>
</tbody>
</table>

The likelihood ratio test statistic is the difference between 4373.23, the log-likelihood (-2 log L) for the model with $\beta_2 = \beta_3 = \beta_4 = 0$, and 4370.05, the log-likelihood for the full model. These tests clearly indicate that any fluctuation over time in the effect of a myocardial infarction is not statistically significant. These analyses were repeated with five regression knots and also with time intervals of length 0.05. All results were comparable.

This example demonstrates the flexibility of the spline model. However, before making clinical inferences from this particular example, several issues regarding choice of steps, regression knots, etc. in approximating $\xi(t|\tau_j)$ with the spline model need to be addressed. Some of these problems will be discussed in chapter V.
Table 4.5

Effect of an Intervening Nonfatal Myocardial Infarction as a Function of Time Since Its Occurrence Using a Spline Function Covariable in the Spline Hazard Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>Chi-Square</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0.4723</td>
<td>0.0614</td>
<td>59.22</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>b</td>
<td>-7.3521</td>
<td>1.0894</td>
<td>45.55</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>γ₁</td>
<td>550.1458</td>
<td>83.1096</td>
<td>43.82</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>γ₂</td>
<td>-767.0597</td>
<td>116.0836</td>
<td>43.66</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>β₁</td>
<td>1.2448</td>
<td>0.2299</td>
<td>29.32</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>β₂</td>
<td>-0.4090</td>
<td>0.3024</td>
<td>1.83</td>
<td>0.176</td>
</tr>
<tr>
<td>β₃</td>
<td>0.1199</td>
<td>0.1112</td>
<td>1.16</td>
<td>0.281</td>
</tr>
<tr>
<td>β₄</td>
<td>-0.1300</td>
<td>0.1213</td>
<td>1.15</td>
<td>0.284</td>
</tr>
</tbody>
</table>
4.4 Repeated Measurements

The structure of the time-dependent covariable has been described in Sections 4.1 and 4.2 in terms of measurements which have been taken serially, or repeated measurements. It should again be noted that these measurements are generated from an unknown covariable function or stochastic process. The use of the restricted cubic spline proportional hazards model with a covariable which is remeasured repeatedly will be demonstrated with a data set from Duke University Medical Center. The New York Heart Association congestive heart failure (CHF) classification on patients who had undergone cardiac catheterization and found to have significant coronary disease was assessed at catheterization. Follow-up measurements were repeated at 6 months, 1 year, and yearly thereafter. Classes 1–4 of CHF status represent progressively more severe symptoms, with 0 meaning that the patient never had CHF. Previous research has indicated that for prognostic purposes, CHF status should be categorized as classes 0, 1-3, and 4 and scored as levels 0, 1, 3, respectively. Gaynor (1983) presents a detailed description and analysis of an earlier version of these data.

The primary question of interest is whether repeated measurement of CHF status adds any additional prognostic information above that obtained from a baseline measurement. The following hazard model was fit to the data:

\[
\lambda(t | z_i(t)) = \left\{ a + b \cdot t + \sum_{j=1}^{m-2} \gamma_j \cdot w_j(t) \right\} \exp \left\{ \text{CHF}_0 \cdot \beta_0 + \text{CHF}_1(t) \cdot \beta_1 \right\}
\]

where \( \text{CHF}_0 \) is the baseline CHF assessment for the \( i \)-th patient and \( \text{CHF}_1(t) \) is the assessment at time \( t \).

Table 4.6 summarizes the results of an analysis using 4 knots in the baseline hazard function. These knots, located at 0.6899, 2.4110, 5.1113, and 9.6509 years, were the 5th, 33rd, 67th, and 95th percentiles of the empirical distribution formed by uncensored survival times. Comparisons between the homogeneous survival estimate based
Table 4.6

Effect of Congestive Heart Failure Status on Patient Survival Using a Spline Hazard Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>Chi-Square</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0.0046</td>
<td>0.0046</td>
<td>1.00</td>
<td>0.317</td>
</tr>
<tr>
<td>b</td>
<td>0.0855</td>
<td>0.0155</td>
<td>30.54</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-0.1284</td>
<td>0.0262</td>
<td>24.05</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.2384</td>
<td>0.0503</td>
<td>22.50</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\beta_0$ (CHF0)</td>
<td>0.6597</td>
<td>0.0712</td>
<td>85.87</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\beta_1$ (CHF)</td>
<td>0.4637</td>
<td>0.0725</td>
<td>40.96</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
on the spline model and the Kaplan-Meier estimate showed that the model with 4 knots
adequately describes the data. The homogeneous model does not converge with 5 knots in
the model.

The following table summarizes tests of $H_0: \beta_1=0$.

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wald Test</td>
<td>40.96</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>36.00</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

The significance of these tests indicates that repeated measurement of congestive heart failure
status improves prognostic capabilities. The following table summarizes tests of $H_0: \beta_0=0$.

Such a test determines whether the baseline measurement has any prognostic effect above
that obtained from the most recent CHF measurement.

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wald Test</td>
<td>85.87</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>61.26</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Clearly, both the baseline and most recent measurement of CHF status are significant
prognostic indicators of survival. As can be seen in Figures 4.5 and 4.6, the baseline CHF
status affects the magnitude of hazard estimates. Figures 4.5a and b show the survival and
hazard function for a patient who has a baseline CHF measurement of level 0, and has
subsequent measurements of 1 at 3 years and 3 at 5 years. Figures 4.6a and b show a similar
patient. The primary difference is that at baseline the CHF measurement was at level 1.
The measurements at 3 years and 5 years are identical to those given for the patient in
Figures 4.5a and b.

The importance of the baseline and current CHF measurement in predicting
prognosis suggests that measurements taken between the two measurements may have
prognostic value. Gaynor (1983) addresses such an issue in his analyses and shows that the
FIGURE 4.5a
SURVIVAL ESTIMATE WITH 95% CONFIDENCE LIMITS FOR A PATIENT WHO HAS CONGESTIVE HEART FAILURE ASSESSMENTS OF LEVEL 0 AT BASELINE, LEVEL 1 AT 3 YEARS, AND LEVEL 3 AT 5 YEARS

FIGURE 4.5b
HAZARD ESTIMATE WITH 95% CONFIDENCE LIMITS FOR A PATIENT WHO HAS CONGESTIVE HEART FAILURE ASSESSMENTS OF LEVEL 0 AT BASELINE, LEVEL 1 AT 3 YEARS, AND LEVEL 3 AT 5 YEARS
FIGURE 4.6a
SURVIVAL ESTIMATE WITH 95% CONFIDENCE LIMITS FOR A PATIENT WHO HAS CONGESTIVE HEART FAILURE ASSESSMENTS OF LEVEL 1 AT BASELINE, AND LEVEL 3 AT 5 YEARS

FIGURE 4.6b
HAZARD ESTIMATE WITH 95% CONFIDENCE LIMITS FOR A PATIENT WHO HAS CONGESTIVE HEART FAILURE ASSESSMENTS OF LEVEL 1 AT BASELINE AND LEVEL 3 AT 5 YEARS
second most recent and most recent measurements have prognostic value. Such a model
could be developed using the restricted cubic spline proportional hazards model.

4.5 Step Function Approximation of a Defined Covariable

Often a particular functional form of a time-dependent covariable is known or
hyphthesized. As an example the interaction between a fixed covariable and log t is frequently
used to test proportionality of hazards. In section 4.3.2, the spline function was used to
examine the effect of an intervening myocardial infarction as a function of the time since its
occurrence. The incorporation of such covariables into the restricted cubic spline proportional
hazards model is theoretically feasible if the functional form of the covariable is specified.
However, their inclusion almost always results in extremely complex log-likelihood functions
and entails the use of numerical integration techniques.

Fortunately, the structure of the restricted cubic spline model allows defined
covariables to be approximated by step functions. Two approaches will be considered. The
first approach, which involves the creation of indicator time-dependent covariables, has been
introduced in another context in section 4.3.1. The time axis is subdivided into intervals and
indicator functions are defined within each interval.

The second approach involves approximating the covariable function $\xi(t)$ by a step
function, somewhat like that described in section 4.3.2. The function $\xi(t)$ can be
approximated by

$$
\xi_A(t) = \begin{cases} 
c_1 & \text{if } t < a_1 \\
c_2 & \text{if } a_1 \leq t < a_2 \\
\cdots & \\
c_k & \text{if } a_{k-1} \leq t < a_k 
\end{cases}
$$

(4.17)

where $c_j$, $j = 1, 2, \ldots, k$ are constants. The primary question is how to subdivide the time
axis. Naturally, there are a variety of approaches. One approach was demonstrated in
Section 4.3.2. Another approach which individualizes the steps for each subject will be used in this section. Additional research is needed to determine which approach is the better approach for approximating a covariable function.

In general, when a subject's survival time is greater than 0, it will be partitioned into $k$ intervals of equal length. The approximating function $\xi_A(t)$ has the following form:

$$
\xi_A(t) = \begin{cases} 
  c_1 & \text{if } t < t_i/k \\
  c_2 & \text{if } t_i/k \leq t < 2t_i/k \\
  \vdots & \\
  c_k & \text{if } (k-1)t_i/k \leq t < t_i
\end{cases}
$$

(4.18)

The value of $\xi_A(t)$ within an interval can be based on any $t$ in the interval. When $\xi(t)$ is strictly increasing or decreasing, it is usually preferable to use the interval's midpoint in order to minimise approximation error. Otherwise, the left endpoint will be used to simplify computational formulas.

The step function $\xi_A(t)$ can be treated as a covariable which is remeasured at the beginning of each time interval. In order to estimate the unknown parameters in the model with the PHSPLM procedure, it is necessary to generate $k$ records corresponding to the steps of $\xi_A(t)$. This treatment is analogous to generating repeated measurements from an unknown stochastic process (see Section 4.4). Let $t_i$ be the $i$-th patient's survival time. Records corresponding to intervals beginning with $0$, $t_i/k$, $2t_i/k$, $\ldots$, $(k-1)t_i/k$ will be generated. An example can be found in Appendix 6.

These approaches to incorporating a covariable function into the hazard model will be demonstrated by assessing the proportional hazards assumption associated with a pain/ischemia index (PI) in a data set from the Cardiovascular Disease Data Bank of Duke University Medical Center. A primary component of this index is the presence of unstable angina which clinically has most of its effect for a 6 month duration (Califf et al, 1988). Harrell and Lee (1986) have used this data to demonstrate techniques for verifying
proportionality assumptions of the Cox model. Survival time is time from catheterization to death or myocardial infarction. Half of the data set was randomly selected for analysis. In Section 4.5.1, indicator variables will be used to assess proportionality. The interaction between PI and log t will be examined in Section 4.5.2.

4.5.1 Modeling Non-Proportionality of Hazards Using Indicator Functions

Graphical methods are available for showing that the proportional hazards assumption does not hold for PI (Harrell and Lee, 1986). Kaplan-Meier estimates in groups defined by the tertiles of PI are plotted against predictions based on the Cox model in Figure 4.7. The lack of agreement between the predicted (Cox) and observed (Kaplan-Meier) curves in the highest pain group indicates that the proportional hazards assumption is not valid. Log-log survival curves are not very informative for this particular example.

A graph of the log hazard ratio per unit increase in PI in various time intervals gives a visual impression of how well the proportionality assumption holds. The Cox model is used to estimate the log hazard ratio within an interval. All observations for events which occur in a previous time interval are excluded from the analysis and all events which occur after the end of the interval are censored. Figure 4.8 shows this plot given 12 subintervals. Since the log hazard ratio is not constant over time, it is clear that proportionality does not hold. This plot confirms conclusions based on the graphs of predicted versus observed survival estimates.

An analytical approach to assessing proportionality is available in the PHGLM procedure. This test is based on the linear correlation between the "partial residuals" and the rank of the survival time (Harrell, 1986). For the PI data set, the normal deviate is -4.37, indicating non-proportionality.

As indicated in the previous section, proportionality can be assessed by a step function. In a sense such a method is modeling the log hazard ratio. The log hazard plot
Figure 4.7

Kaplan-Meier and Cox Estimates in Three Strata Determined by the Tertiles of Pain Index

---

Survival Probability

\[ t \]

\[
\begin{array}{cccc}
_\text{S} & \text{a-Cox} & \text{a-K.M.} & \text{b-Cox} \\
\text{b-K.M.} & \text{c-Cox} & \text{c-K.M.}
\end{array}
\]
Figure 4.8

Log Hazard Ratio with 95% Confidence Limits in 12 Time Intervals for Covariable Pain Index
Figure 4.9

Log Hazard Ratio with 95% Confidence Limits in 5 Time Intervals for Covariable Pain Index

Dashed Lines Indicate 95% Confidence Limits
Table 4.7
Assessment of the Proportional Hazards Assumption for Pain Index Using Indicator Functions in the Spline Hazard Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>Chi-Square</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0.2656</td>
<td>0.0288</td>
<td>85.20</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>b</td>
<td>-3.1377</td>
<td>0.4762</td>
<td>43.41</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>γ₁</td>
<td>165.1425</td>
<td>27.7911</td>
<td>35.31</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>γ₂</td>
<td>-258.4765</td>
<td>44.1343</td>
<td>34.30</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>γ₃</td>
<td>93.3503</td>
<td>16.5122</td>
<td>31.96</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>β₁</td>
<td>0.1224</td>
<td>0.0107</td>
<td>130.08</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>β₂</td>
<td>0.0800</td>
<td>0.0172</td>
<td>21.53</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>β₃</td>
<td>0.0327</td>
<td>0.0295</td>
<td>1.23</td>
<td>0.268</td>
</tr>
<tr>
<td>β₄</td>
<td>0.1013</td>
<td>0.0391</td>
<td>6.72</td>
<td>0.010</td>
</tr>
<tr>
<td>β₅</td>
<td>0.2585</td>
<td>0.1243</td>
<td>4.33</td>
<td>0.038</td>
</tr>
</tbody>
</table>
using five intervals (0, 2), (2, 5), (5, 8), (8, 11), (11, ∞) is shown in Figure 4.9. The results of an analysis using indicator functions to describe each of these intervals is summarized in Table 4.7. The coefficient β₁ is associated with Π x I{t < 2}, β₂ is associated with Π x I{2 ≤ t < 5}, etc. The log-likelihood associated with this model is 2930.91. The log-likelihood associated with the model with a constant Π is 2943.31.

Therefore the likelihood ratio statistic with 4 degrees of freedom for testing H₀: β₁ = β₂ = β₃ = β₄ = β₅ is 12.40 (p < 0.05). This result showing that proportionality does not hold is consistent with the graphical analysis and the analytical test of proportionality provided by PROC PHGLM. It is interesting to note that the magnitude of the β coefficients are consistent with that shown in the log hazard plot of Figure 4.9.

A large number of time intervals need to be considered to generate a good approximation of ξ(t). As mentioned in section 4.3.2, there are problems inherent with such a model. The primary one is that the test of proportionality becomes less powerful as the number of degrees of freedom increases.

4.5.2 Modeling Non-Proportionality of Hazards Using the Interaction between a Fixed Covariable and Log Time

The proposed approach to approximating a covariable function involves subdividing a subject's survival time into k intervals. Within each interval, a time point is chosen and used to determine the functional value of the step. Since the logarithm of zero is undefined, the proposed approach needs modification when considering the interaction between a fixed covariable and log t.

The value of 0.001, being less than the smallest nonzero survival time (1/365 years), will be used in place of 0 in these approximations. If a patient's survival time tᵢ equals 0 years, the value of 0.001 will be used as tᵢ. If a patient's survival time tᵢ is greater than 0.001 years, the interval [0.001, tᵢ] will be divided into K subintervals. The length of each subinterval will be L = (tᵢ - 0.001) / K. The midpoint of an interval will be used to
calculate the step function's value. Let $M = L / 2$. Therefore the interaction between PI and
log (t) will be approximated by

$$
\xi_A(t) = \begin{cases} 
\text{PI} \times \log (0.001 + M) & \text{if } 0 < t \leq 0.001 + L \\
\text{PI} \times \log (0.001 + M + L) & \text{if } 0.001 + L < t \leq 0.001 + 2 L \\
\text{PI} \times \log (0.001 + M + 2 L) & \text{if } 0.001 + 2 L < t \leq 0.001 + 3 L . \\
\vdots & \\
\text{PI} \times \log (0.001 + M + (K+1) L) & \text{if } t_1 - L < t \leq t_1
\end{cases}
$$

Figure 4.10 shows how this step function approximates log t for a patient whose survival time is 5 years. This covariable (4.19) will be treated, as described in section 4.5, as a covariable which is repeatedly remeasured.

Using 10 subintervals, a restricted cubic spline hazard model was fit to the data in order to examine the proportionality of PI. The estimated parameters are summarized in Table 4.8. PILOGT is the coefficient of the step function approximation of PI \times log t. The score, Wald, and likelihood ratio test for $H_0$: PILOGT=0 are given below.

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score Test</td>
<td>170.61</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>113.04</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Wald Test</td>
<td>155.49</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

The likelihood ratio test statistic is the difference between 2943.31, the log-likelihood for the model with only PI, and 2830.27, the log-likelihood associated with the above model. All three of these test statistics with 1 degree of freedom are extremely significant. Conclusions based on these results are consistent with previous findings of non-proportionality. This test of proportionality produces a much larger $\chi^2$ statistic than produced by PHGLM's test of
FIGURE 4.10
STEP FUNCTION APPROXIMATION OF LOG(T) FOR A PATIENT WHOSE SURVIVAL TIME IS 5 YEARS

\[ \mu \]
Table 4.8

Assessment of the Proportional Hazards Assumption for Pain Index
Using a Step Function Approximation of the Interaction Between
Pain Index and Log Time in the Spline Hazard Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>Chi-Square</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0.2277</td>
<td>0.0293</td>
<td>60.61</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>b</td>
<td>-2.0848</td>
<td>0.5656</td>
<td>13.59</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>114.4712</td>
<td>34.7126</td>
<td>10.87</td>
<td>0.001</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-180.1105</td>
<td>55.4013</td>
<td>10.57</td>
<td>0.001</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>65.8625</td>
<td>20.9639</td>
<td>9.87</td>
<td>0.002</td>
</tr>
<tr>
<td>PI</td>
<td>0.0759</td>
<td>0.0100</td>
<td>57.15</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>PILOGT</td>
<td>-0.0344</td>
<td>0.0028</td>
<td>155.49</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
FIGURE 4.11

LOG HAZARD RATIO FOR COVARIABLE PI BASED ON THE COX MODEL IN 12 TIME INTERVALS AND THE SPLINE MODEL WITH AN INTERACTION BETWEEN PI AND LOG(T)
non-proportionality ($\chi^2_1 = z^2 = 19.10$) and suggests that this test may be more sensitive to departures from proportionality.

The log hazard ratio per unit increase in PI, described by

$$LHR(t) = 0.0759 - 0.0344 \log t,$$

is graphically shown in Figure 4.11. Superimposed on this graph are the log hazard estimates obtained from the Cox model in 12 time intervals (see Figure 4.8). The log hazard ratio for the spline model seems to approximate the Cox estimates reasonably well. However, the structure of the logarithmic function does not allow for non-monotonicity in the log hazard ratio. In the last two intervals, the Cox estimates are large. However, it should be noted that less than 10 failures occurred after $t=8$ years.
Chapter V
Discussion and Future Research

5.1 Summary

A parametric proportional hazards model where the baseline hazard is a cubic spline function with tails that are linearly restricted has been proposed as an alternative to the Cox model. The primary focus of this research has been to examine the properties of this model with and without covariates and contrast it with the Kaplan-Meier and Cox models. This parametric model has been shown to have several advantages, including flexibility of the hazard, efficiency of estimation, smooth survival curves, confidence limits for the survival and hazard estimates when time-dependent covariates are present, and computational efficiency.

Flexibility of Hazard Function Shape

One of the advantages of nonparametric approaches, such as the Kaplan-Meier estimator and the Cox model, is that it is unnecessary to specify an underlying hazard function. Estimation is robust and always asymptotically unbiased. With parametric estimation, a parametric form must be specified. Often these models are not robust and cannot adequately describe the data.

The restricted cubic spline model has been shown to be similar to the nonparametric approaches in that it can adequately describe survival data generated from a variety of distributions. These include distributions having both monotonic and non-monotonic hazard functions. The structure of the restricted cubic spline hazard function is extremely flexible in that the location and number of knots can be adjusted when necessary.
Efficiency of Estimation

The efficiency of estimation with the homogeneous restricted cubic spline model is as good as, and sometimes better than, the Kaplan-Meier estimator for the underlying distributions studied. For one distribution, estimation was more efficient for the Kaplan-Meier estimator at the 95th survival fractile. However the mean square error associated with both models was small, making the decrease in efficiency an unimportant issue. The restricted cubic spline hazard model with fixed covariables has also been shown to usually be more efficient that the Cox model at estimating survival probabilities. The efficiency of coefficient estimation using the spline model is the same as the Cox model.

Smoothness of Survival Curve

One of the minor criticisms of the Cox and Kaplan-Meier models is that survival curve estimates are step functions, even though underlying biological mechanisms may indicate that they should be continuous. Survival functions associated with spline models are continuous without sacrificing efficiency and robustness. Associated with this parametric model are formulas describing hazard, cumulative hazard, and survival, unlike the Kaplan-Meier and Cox models.

Confidence Limits

Confidence limits for survival and hazard estimates have been developed for the restricted cubic spline proportional hazards model based on standard asymptotic normal theory. These formulas are applicable when fixed and time-dependent covariables are present. Currently, confidence limits are not available with the Cox model when time-dependent covariables are present.
Computational Efficiency

Time-dependent covariates can be incorporated into the restricted cubic spline proportional hazards model without dramatically increasing computation time. Analyses previously obtained using the Cox model with a time-dependent covariate describing an intervening event were replicated using the spline model. Estimates of covariate coefficients were virtually identical. Computation time for the spline model was reduced by a factor of 213 over the Cox model.

In addition to covariates describing an intervening nonfatal event, two other classes of time-dependent covariates were shown to be easily incorporated into the spline model. These included repeated measurement covariates and step function approximations of continuous covariate functions. Neither of these covariates can be incorporated into the Cox model without outrageous computation time. Approximating covariate functions by step functions has suggested several approaches to testing the proportionality assumption of a fixed covariate.

5.2 Future Research Topics

Properties of the spline hazard model, especially flexibility and robustness will be fully known only after there is more experience using the model. Some of the general issues which need to be addressed include the following.

- Only samples of size 50 and 200 were examined prior to making suggestions concerning the choice of number and location of knots. With the cardiovascular data sets which had samples sizes in the thousands, an additional knot at the 20th percentile was necessary in order to describe the data adequately. What should the recommendations be concerning knot placement for samples of intermediate and large size? Are there distinct properties of data sets which will indicate prior to modeling that knots are needed at particular locations?
• When data were generated from the log-logistic distribution, estimates of the unknown parameters in the homogeneous restricted cubic spline model were not obtained due to lack of convergence of the Newton-Raphson algorithm. Was this lack of convergence a result of the model not being able to describe upside-down bathtub-shaped hazard functions? Or is it a result of an inadequate optimization procedure? Constrained optimization procedures need to be considered as alternatives to the Newton-Raphson algorithm. If the model specification is at fault, can the restricted cubic spline model be made more general so as to encompass upside-down bathtub-shaped hazard functions? Or is the log-logistic model unrealistic anyway?

• When data were generated from hazard functions which increased from 0 or converged to 0 as survival time became large, the Newton-Raphson algorithm had convergence problems. Step-halving was invoked since intermediate parameter estimates were outside the region of feasible estimates. The log-logistic model is a situation where the hazard function increases from 0 and also converges to 0 as survival time becomes large. Is there a better constrained optimization procedure which would allow such models to be adequately estimated? Is a model with fewer continuity restrictions more appropriate?

• Efficiency of estimation when fixed covariates are present in the model was examined under a limited number of conditions. What is the efficiency of estimation when data are generated from distributions other than those examined in this research? Additional investigation concerning the effect of covariate distribution, magnitude of the covariate coefficient, number of covariates, and censoring distribution is needed.

Step function approximation of covariate functions was introduced in chapter IV. Additional research concerning some of the aspects of this type of covariate are needed.
Two approaches for determining the steps in a covariable function approximation were proposed. In Section 4.3.2, the time axis was subdivided into intervals of equal length. In Section 4.5, a subject's survival time was subdivided into intervals of equal length. How many steps are needed for a good approximation with this latter approach? If the former approach is used, how small should time axis subdivisions be in order to generate a good approximation? Which approach is the most accurate and preferred approach for estimating a covariable function? A couple of the analyses in section 4.5.2 and 4.5.3 were conducted using 4 records per subject as opposed to the reported 10 records per subject. In the example dealing with the interaction between PI and log t, the model estimated based on 4 records per subject had a smaller log-likelihood (-2 log L) than the model based on 10 records. Exactly what is happening? Is this approach to determining steps inappropriate? The analysis based on 10 records was reported since it was felt that it should give a better approximation of the covariable function.

Knots in the baseline hazard function were determined based on the distribution of uncensored survival times. This approach seems inappropriate when choosing knots for spline functions in the regression component of the hazard model. What is the best way of choosing these knots? Or is the choice of knots fairly robust?

When the spline function was incorporated into the baseline hazard model, examination of survival function residuals was suggested as an approach for determining whether knots were appropriately placed and the model adequately described the data. It is not clear what sort of diagnostic examination should be done to determine whether the regression spline is properly specified. Additional thought needs to be done.
An approach for examining the proportional hazards assumption using a step function approximation of the interaction of a fixed covariable and log t was proposed in chapter IV. What is the power and sensitivity of this test? Steps were individualized in this test for each patient based on their survival time. Would subdivision of the time axis into intervals, as proposed in Section 5.4, result in a better test of proportional hazards?

The interaction between a fixed covariable and the logarithm of time does not model the departure from proportionality. Such modeling could be done by incorporating a step function approximation of the interaction of a fixed covariable and a spline function into the model. It is expected that such terms should provide a more powerful and sensitive test of proportionality than an interaction involving log t.

There are additional features that should be considered as part of the PHSPLM procedure in order to make it a more valuable statistical tool. These include the following.

A better default algorithm for determining the number and location of knots needs to be determined. This algorithm should be a function of the number of censored and uncensored observations and survival times.

An approach to calculating median survival time, a statistic desired and understood by researchers, needs to be developed and incorporated into the PHSPLM procedure. In a nonparametric structure, an estimate with confidence interval is not easily obtainable. However, with a parametric structure such as the restricted cubic spline proportional hazards model, median survival should be easily calculable.

The restricted cubic spline hazard model has the following structure:

$$\lambda(t_i \mid z_i(t)) = \left\{ a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\} \exp \left\{ z_i(t_i) \beta \right\}$$
where

\[ w_j(t) = (t - v_j)^3 + \frac{(t - v_{m-1})^3 (v_m - v_j)}{(v_m - v_{m-1})} + \frac{(t - v_m)^3 (v_{m-1} - v_j)}{(v_m - v_{m-1})} \]

and is a function of m unknown parameters. For clinical presentation, a model with all similar truncated power terms collected together is desirable. Upon combination, the model is

\[
\lambda(t | \mathbf{s}_1(t)) = \begin{cases} 
  a + b t + \sum_{j=1}^{m-2} \gamma_j (t - v_j)^3 + \\
  + \frac{1}{v_m - v_{m-1}} (t - v_{m-1})^3 \sum_{j=1}^{m-2} [\gamma_j (v_j - v_m)] \\
  - \frac{1}{v_m - v_{m-1}} (t - v_m)^3 \sum_{j=1}^{m-2} [\gamma_j (v_j - v_{m-1})] \\
  \times \exp \left\{ \mathbf{s}_1(t) \beta \right\}
\end{cases}
\]

The covariance structure of the coefficients of the truncated power terms can be calculated from the information matrix. Though this covariance matrix is singular, it can be useful for determining whether certain knots are essential in the model or whether the model can be simplified.

The restricted cubic spline proportional hazards model has many of the desirable properties of the nonparametric approaches, while retaining the desirable properties of parametric models. The spline model is as powerful a tool at describing the relationship between covariables and survival time as the Cox model. However, additional research concerning some of the properties of the spline model need examination.
Appendix 1

The Score Vector and Information Matrix Associated with the Generalized Weibull Hazard Function

The log-likelihood function which was maximized in order to obtain estimates of the unknown parameters in the generalized Weibull hazard function is

\[ L = \sum_{i=1}^{n} L_i \]

\[ = \sum_{i=1}^{n} \left[ \delta_i \log \left\{ \frac{a \, b \, t_i^{b-1}}{a \, b \, t_i^{b-1} + c \, d \, t_i^{d-1}} \right\} - \left\{ a \, t_i^b + c \, t_i^d \right\} \right]. \]

The score vector associated with this log-likelihood function and used in the Newton-Raphson optimization algorithm is

\[ U(a, b, c, d) = \left( \sum_{i=1}^{n} \frac{\partial L_i}{\partial a}, \sum_{i=1}^{n} \frac{\partial L_i}{\partial b}, \sum_{i=1}^{n} \frac{\partial L_i}{\partial c}, \sum_{i=1}^{n} \frac{\partial L_i}{\partial d} \right), \]

where

\[ \frac{\partial L_i}{\partial a} = \delta_i \frac {b \, t_i^{b-1}}{a \, b \, t_i^{b-1} + c \, d \, t_i^{d-1}} - t_i^b, \]

\[ \frac{\partial L_i}{\partial b} = \delta_i \frac {a \, t_i^{b-1} \left( 1 + b \, \log t_i \right)}{a \, b \, t_i^{b-1} + c \, d \, t_i^{d-1}} - a \, t_i^b \log t_i, \]
\[
\frac{\partial L_i}{\partial c} = \delta_i \frac{d t_i^{d-1}}{a b t_i^{b-1} + c d t_i^{d-1}} - t_i^d,
\]

and

\[
\frac{\partial L_i}{\partial d} = \delta_i \frac{c t_i^{d-1} (1 + d \log t_i)}{a b t_i^{b-1} + c d t_i^{d-1}} - c t_i^d \log t_i.
\]

The information matrix is

\[
I(a, b, c, d) = \sum_{i=1}^{n} I_i(a, b, c, d),
\]

where \( I_i \) is a symmetric matrix describing the contribution from the \( i \)-th observation to the total information matrix. The \( jh \)-th entry in \( I_i \), denoted as \( I_{ijk} \), is

\[
I_{i11} = -\frac{\partial^2 L_i}{\partial a^2} = \delta_i \frac{\left(b t_i^{b-1}\right)^2}{\left\{a b t_i^{b-1} + c d t_i^{d-1}\right\}^2},
\]

\[
I_{i12} = -\frac{\partial^2 L_i}{\partial a \partial b} = \delta_i \frac{-c d t_i^{d-1} t_i^{b-1} (1 + b \log t_i)}{\left\{a b t_i^{b-1} + c d t_i^{d-1}\right\}^2} + t_i^b \log t_i,
\]

\[
I_{i13} = -\frac{\partial^2 L_i}{\partial a \partial c} = \delta_i \frac{b t_i^{b-1} d t_i^{d-1}}{\left\{a b t_i^{b-1} + c d t_i^{d-1}\right\}^2},
\]

\[
I_{i14} = -\frac{\partial^2 L_i}{\partial a \partial d} = \delta_i \frac{b t_i^{b-1} c t_i^{d-1} (1 + d \log t_i)}{\left\{a b t_i^{b-1} + c d t_i^{d-1}\right\}^2},
\]

\[
I_{i15} = -\frac{\partial^2 L_i}{\partial a \partial t_i} = \delta_i \frac{b t_i^{b-1}}{a b t_i^{b-1} + c d t_i^{d-1}},
\]

\[
I_{i22} = -\frac{\partial^2 L_i}{\partial b^2} = \delta_i \frac{\left(c t_i^{d-1}\right)^2}{\left\{a b t_i^{b-1} + c d t_i^{d-1}\right\}^2},
\]

\[
I_{i23} = -\frac{\partial^2 L_i}{\partial b \partial c} = \delta_i \frac{-c d t_i^{d-1} t_i^{b-1}}{\left\{a b t_i^{b-1} + c d t_i^{d-1}\right\}^2},
\]

\[
I_{i24} = -\frac{\partial^2 L_i}{\partial b \partial d} = \delta_i \frac{c d t_i^{d-1}}{\left\{a b t_i^{b-1} + c d t_i^{d-1}\right\}^2},
\]

\[
I_{i25} = -\frac{\partial^2 L_i}{\partial b \partial t_i} = \delta_i \frac{c t_i^{d-1}}{a b t_i^{b-1} + c d t_i^{d-1}},
\]

\[
I_{i33} = -\frac{\partial^2 L_i}{\partial c^2} = \delta_i \frac{\left(c t_i^{d-1}\right)^2}{\left\{a b t_i^{b-1} + c d t_i^{d-1}\right\}^2},
\]

\[
I_{i34} = -\frac{\partial^2 L_i}{\partial c \partial d} = \delta_i \frac{c d t_i^{d-1}}{\left\{a b t_i^{b-1} + c d t_i^{d-1}\right\}^2},
\]

\[
I_{i35} = -\frac{\partial^2 L_i}{\partial c \partial t_i} = \delta_i \frac{d t_i^{d-1}}{a b t_i^{b-1} + c d t_i^{d-1}},
\]

\[
I_{i44} = -\frac{\partial^2 L_i}{\partial d^2} = \delta_i \frac{\left(c t_i^{d-1}\right)^2}{\left\{a b t_i^{b-1} + c d t_i^{d-1}\right\}^2},
\]

\[
I_{i45} = -\frac{\partial^2 L_i}{\partial d \partial t_i} = \delta_i \frac{c t_i^{d-1}}{a b t_i^{b-1} + c d t_i^{d-1}},
\]

\[
I_{i55} = -\frac{\partial^2 L_i}{\partial t_i^2} = \delta_i \frac{\left(c t_i^{d-1}\right)^2}{\left\{a b t_i^{b-1} + c d t_i^{d-1}\right\}^2}.
\]
\[ I_{122} = -\frac{\partial^2 L_i}{\partial b^2} = a t_i^b \left( \log t_i \right)^2 \]

\[ + \delta_i \left[ \frac{-a c d t_i^{d-1} t_i^{b-1} \log t_i \left( 2 + b \log t_i \right) + a^2 \left( t_i^{b-1} \right)^2}{\left\{ a b t_i^{b-1} + c d t_i^{d-1} \right\}^2} \right] \]

\[ I_{123} = -\frac{\partial^2 L_i}{\partial b \partial c} = \delta_i \frac{a t_i^{b-1} d t_i^{d-1} \left( 1 + b \log t_i \right)}{\left\{ a b t_i^{b-1} + c d t_i^{d-1} \right\}^2} \]

\[ I_{124} = -\frac{\partial^2 L_i}{\partial b \partial d} = \delta_i \frac{a c t_i^{b-1} t_i^{d-1} \left( 1 + b \log t_i \right) \left( 1 + d \log t_i \right)}{\left\{ a b t_i^{b-1} + c d t_i^{d-1} \right\}^2} \]

\[ I_{133} = -\frac{\partial^2 L_i}{\partial c^2} = \delta_i \left( d t_i^{d-1} \right)^2 \]

\[ \left\{ a b t_i^{b-1} + c d t_i^{d-1} \right\}^2 \]

\[ I_{134} = -\frac{\partial^2 L_i}{\partial c \partial d} = \delta_i \frac{-a b t_i^{b-1} t_i^{d-1} \left( 1 + d \log t_i \right)}{\left\{ a b t_i^{b-1} + c d t_i^{d-1} \right\}^2} + t_i^d \log t_i , \]

and

\[ I_{144} = -\frac{\partial^2 L_i}{\partial d^2} = c t_i^d \left( \log t_i \right)^2 \]

\[ + \delta_i \left[ \frac{-a b c t_i^{d-1} t_i^{b-1} \log t_i \left( 2 + \log t_i \right) + c^2 \left( t_i^{b-1} \right)^2}{\left\{ a b t_i^{b-1} + c d t_i^{d-1} \right\}^2} \right] . \]
Appendix 2
The Score Vector and Information Matrix Associated with the Homogeneous Spline Hazard Function

The log-likelihood function which was maximized in order to obtain estimates of the unknown parameters in the restricted cubic spline hazard function (2.9) was

\[ L = \sum_{i=1}^{n} L_i \]
\[ = \sum_{i=1}^{n} \left[ \delta_i \log \left\{ a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\} - \left\{ a t_i + \frac{1}{2} b t_i^2 + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\} \right], \]

where

\[ w_j(t) = \left( t - v_j \right)^3 \left( v_m - v_{m-1} \right)^3 \left( v_m - v_j \right) + \left( t - v_m \right)^3 \left( v_{m-1} - v_j \right) \left( v_m - v_{m-1} \right) \]
(A2.1)

and

\[ w_j(t) = \frac{1}{4} \left[ \left( t - v_j \right)^4 - \left( t - v_{m-1} \right)^4 \left( v_m - v_j \right) \left( v_m - v_{m-1} \right) + \left( t - v_m \right)^4 \left( v_{m-1} - v_j \right) \left( v_m - v_{m-1} \right) \right]. \]
(A2.2)

The score vector, which is the first derivative of the log-likelihood function, is

\[ U(a, b, \gamma_1, \ldots, \gamma_{m-2}) = \left( \sum_{i=1}^{n} \frac{\partial L_i}{\partial a}, \sum_{i=1}^{n} \frac{\partial L_i}{\partial b}, \sum_{i=1}^{n} \frac{\partial L_i}{\partial \gamma_1}, \sum_{i=1}^{n} \frac{\partial L_i}{\partial \gamma_2}, \ldots, \sum_{i=1}^{n} \frac{\partial L_i}{\partial \gamma_{m-2}} \right), \]
where
\[
\frac{\partial L_i}{\partial a} = \delta_i \left\{ \frac{1}{a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i)} \right\}^2 - t_i,
\]
and
\[
\frac{\partial L_i}{\partial b} = \delta_i \left\{ \frac{t_i}{a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i)} \right\}^2 - \frac{1}{2} t_i^2,
\]
for \( h = 1, 2, \ldots, m-2 \).

The information matrix is
\[
I(a, b, \gamma_1, \ldots, \gamma_{m-2}) = \sum_{i=1}^{n} I_i(a, b, \gamma_1, \ldots, \gamma_{m-2}),
\]
where \( I_i \) is a symmetric matrix describing the contribution from the \( i \)-th patient to the total information matrix. The components of \( I_i \) are:
\[
I_{i aa} = -\frac{\partial^2 L_i}{\partial a^2} = \delta_i \left\{ \frac{1}{a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i)} \right\}^2,
\]
\[
I_{i ab} = -\frac{\partial^2 L_i}{\partial a \partial b} = \delta_i \left\{ \frac{t_i}{a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i)} \right\}^2,
\]
\[
I_{i \gamma h} = -\frac{\partial^2 L_i}{\partial a \partial \gamma h} = \delta_i \left\{ \frac{w_h(t_i)}{a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i)} \right\}^2,
\]
\[ L_{ibb} = -\frac{\partial^2 L_i}{\partial b^2} = \delta_i \frac{t_i^2}{\left\{ a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\}^2} , \]

\[ L_{ib \gamma_h} = -\frac{\partial^2 L_i}{\partial b \partial \gamma_h} = \delta_i \frac{t_i w_h(t_i)}{\left\{ a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\}^2} , \]

and

\[ L_{i \gamma_h \gamma_h'} = -\frac{\partial^2 L_i}{\partial \gamma_h \partial \gamma_{h'}} = \delta_i \frac{w_h(t_i) w_{h'}(t_i)}{\left\{ a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\}^2} , \]

for \( h, h' = 1, 2, \ldots, m-2 \).
Appendix 3
The Score Vector and Information Matrix for the Spline Model with Fixed Covariates

In chapter III, the restricted cubic spline function was incorporated into the proportional hazards model as the baseline hazard. The log-likelihood function which was maximized in order to obtain estimates of the unknown parameters was:

\[
L = \sum_{i=1}^{n} L_i \\
= \sum_{i=1}^{n} \left\{ \delta_i \log \left( a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right) + \delta_i \mathbf{x}_i \mathbf{\beta} \\
- \left\{ a t_i + \frac{1}{2} b t_i^2 + \sum_{j=1}^{m-2} \gamma_j W_j(t_i) \right\} \exp \left( \mathbf{x}_i \mathbf{\beta} \right) \right\},
\]

where \( w(t) \) and \( W(t) \) are defined in (A2.1) and (A2.2), respectively, \( \mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ir}) \) is a vector of covariates associated with the i-th subject, and \( \mathbf{\beta} = (\beta_1, \ldots, \beta_r) \) is the vector of associated coefficients.

The score vector, which is the first derivative of the log-likelihood function, is

\[
U( a, b, \gamma_1, \ldots, \gamma_{m-2}, \beta_1, \ldots, \beta_r) = \\
\left( \sum_{i=1}^{n} \frac{\partial L_i}{\partial a}, \sum_{i=1}^{n} \frac{\partial L_i}{\partial b}, \sum_{i=1}^{n} \frac{\partial L_i}{\partial \gamma_1}, \ldots, \sum_{i=1}^{n} \frac{\partial L_i}{\partial \gamma_{m-2}}, \sum_{i=1}^{n} \frac{\partial L_i}{\partial \beta_1}, \ldots, \sum_{i=1}^{n} \frac{\partial L_i}{\partial \beta_r} \right),
\]
where

\[
\frac{\partial L_i}{\partial a} = \delta_i \frac{1}{a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i)} - t_i \exp \left( x_i \beta \right),
\]

\[
\frac{\partial L_i}{\partial b} = \delta_i \frac{t_i}{a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i)} - \frac{1}{2} t_i^2 \exp \left( x_i \beta \right),
\]

\[
\frac{\partial L_i}{\partial \gamma_h} = \delta_i \frac{w_h(t_i)}{a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i)} - w_h(t_i) \exp \left( x_i \beta \right),
\]

\[
\frac{\partial L_i}{\partial \beta_k} = \delta_i \frac{1}{a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i)} \left\{ a t_i + \frac{1}{2} b t_i^2 + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\} z_{ik} \exp \left( x_i \beta \right),
\]

for \( h = 1, 2, \ldots, m-2 \) and \( k = 1, 2, \ldots, r \).

The \((r+m) \times (r+m)\) information matrix is

\[
I( a, b, \gamma_1, \ldots, \gamma_{m-2}, \beta_1, \ldots, \beta_r ) = \sum_{i=1}^{n} I_i \left( b, \gamma_1, \ldots, \gamma_{m-2}, \beta_1, \ldots, \beta_r \right)
\]

where the components of \( I_i \) are:

\[
I_{aa} = \delta_i \frac{\partial^2 L_i}{\partial a^2} = \delta_i \frac{1}{\left\{ a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\}^2},
\]

\[
I_{ab} = \delta_i \frac{\partial^2 L_i}{\partial a \partial b} = \delta_i \frac{t_i}{\left\{ a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\}^2},
\]

\[
I_{a\gamma_h} = \delta_i \frac{\partial^2 L_i}{\partial a \partial \gamma_h} = \delta_i \frac{w_h(t_i)}{\left\{ a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\}^2},
\]

for \( h = 1, 2, \ldots, m-2 \).
\[ I_{ia\beta_k} = -\frac{\partial^2 L_i}{\partial a \partial \beta_k} = t_i z_{ik} \exp (z_i \beta), \]

\[ I_{ib\beta_k} = -\frac{\partial^2 L_i}{\partial b \partial \beta_k} = \delta_i \frac{t_i^2}{\left\{ a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\}^2}, \]

\[ I_{ib\gamma_h} = -\frac{\partial^2 L_i}{\partial b \partial \gamma_h} = \delta_i \frac{t_i w_h(t_i)}{\left\{ a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\}^2}, \]

\[ I_{ib\beta_k} = -\frac{\partial^2 L_i}{\partial b \partial \beta_k} = \frac{1}{2} t_i^2 z_{ik} \exp (z_i \beta), \]

\[ I_{i\gamma_h \gamma_h'} = -\frac{\partial^2 L_i}{\partial \gamma_h \partial \gamma_h'} = \delta_i \frac{w_h(t_i) w_{h'}(t_i)}{\left\{ a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\}^2}, \]

\[ I_{i\gamma_h \beta_k} = -\frac{\partial^2 L_i}{\partial \gamma_h \partial \beta_k} = w_h(t_i) z_{ik} \exp (z_i \beta), \]

and

\[ I_{i\beta_k \beta_{k'}} = -\frac{\partial^2 L_i}{\partial \beta_k \partial \beta_{k'}} = \left\{ t_i + \frac{1}{2} b t_i^2 + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\} z_{ik} z_{ik'} \exp (z_i \beta) \]

for \( k, k' = 1, 2, \ldots, r \) and \( h, h' = 1, 2, \ldots, m-2 \).
Appendix 4

The Score Vector and Information Matrix for the Spline Model with Time-Dependent Covariables Expressed as Step Functions

Time-dependent covariables expressed as step functions were incorporated into the restricted cubic spline proportional hazards model in chapter IV. The log-likelihood function which was maximized in order to obtain estimates of the unknown parameters was:

\[
L = \sum_{i=1}^{n} \left( \delta_i \log \left\{ a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\} + \delta_i \left\{ z^{(i,q_i)} \beta \right\} \right) \\
- \sum_{i=1}^{n} \sum_{k=0}^{q_i} \exp \left( z^{(ik)} \beta \right) \left[ \Lambda^*(t^{(i,k+1)}) - \Lambda^*(t^{(ik)}) \right],
\]

where

\[
\Lambda^*(t) = a t + \frac{1}{2} b t^2 + \sum_{j=1}^{m-2} \gamma_j w_j(t).
\]

The terms \( w_j(t) \) and \( w_j(t) \) are defined in (A2.1) and (A2.2), respectively, and \( z^{(ik)} \) and \( \beta \) are defined in section 4.2.1.

The score vector, which is the first derivative of the log-likelihood function, is

\[
U(a, b, \gamma_1, \ldots, \gamma_{m-2}, \beta_1, \ldots, \beta_r) =
\left( \sum_{i=1}^{n} \frac{\partial L_i}{\partial a}, \sum_{i=1}^{n} \frac{\partial L_i}{\partial b}, \sum_{i=1}^{n} \frac{\partial L_i}{\partial \gamma_1}, \ldots, \sum_{i=1}^{n} \frac{\partial L_i}{\partial \gamma_{m-2}}, \sum_{i=1}^{n} \frac{\partial L_i}{\partial \beta_1}, \ldots, \sum_{i=1}^{n} \frac{\partial L_i}{\partial \beta_r} \right).
\]
where

\[
\frac{\partial L_i}{\partial a} = \delta_i \frac{1}{a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i)} - \sum_{k=0}^{q_i} \exp \left( z^{(ik)} \beta \right) \left[ t^{(i,k+1)} - t^{(ik)} \right],
\]

\[
\frac{\partial L_i}{\partial b} = \delta_i \frac{t_i}{a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i)}
\]

\[-\frac{1}{2} \sum_{k=0}^{q_i} \exp \left( z^{(ik)} \beta \right) \left[ \left( t^{(i,k+1)} \right)^2 - \left( t^{(ik)} \right)^2 \right],
\]

\[
\frac{\partial L_i}{\partial \gamma_h} = \delta_i \frac{w_h(t_i)}{a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i)}
\]

\[-\sum_{k=0}^{q_i} \exp \left( z^{(ik)} \beta \right) \left[ \psi_h(t^{(i,k+1)}) - \psi_h(t^{(ik)}) \right],
\]

\[
\frac{\partial L_i}{\partial \beta_l} = \delta_i \sum_{k=0}^{q_i} \exp \left( z^{(ik)} \beta \right) \left[ \Lambda^* \left( t^{(i,k+1)} \right) - \Lambda^* \left( t^{(ik)} \right) \right]
\]

for h = 1, 2, \ldots, m-2 and l = 1, 2, \ldots r.

The (r + m) x (r + m) information matrix is

\[
I(\ a, \ b, \ \gamma_1, \ \ldots, \ \gamma_{m-2}, \ \beta_1, \ \ldots, \ \beta_r) = \sum_{i=1}^{n} I_i(\ a, \ b, \ \gamma_1, \ \ldots, \ \gamma_{m-2}, \ \beta_1, \ \ldots, \ \beta_r)
\]

where the components of I_i are:
\[ I_{\text{laa}} = -\frac{\partial^2 L_i}{\partial a^2} = \delta_i \frac{1}{\left\{ a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\}^2}, \]

\[ I_{\text{lab}} = -\frac{\partial^2 L_i}{\partial a \partial b} = \delta_i \frac{t_i}{\left\{ a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\}^2}, \]

\[ I_{\text{la} \gamma_h} = -\frac{\partial^2 L_i}{\partial a \partial \gamma_h} = \delta_i \frac{w_h(t_i)}{\left\{ a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\}^2}, \]

\[ I_{\text{la} \beta_l} = -\frac{\partial^2 L_i}{\partial a \partial \beta_l} = \frac{\sum_{k=0}^{q_i} z_{ik} \exp \left( s^{(ik)} \beta \right)}{t^{(i,k+1)} - t^{(ik)}}, \]

\[ I_{\text{lb} \beta_l} = -\frac{\partial^2 L_i}{\partial b \partial \beta_l} = \delta_i \frac{t_i^2}{\left\{ a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\}^2}, \]

\[ I_{\text{lb} \gamma_h} = -\frac{\partial^2 L_i}{\partial b \partial \gamma_h} = \delta_i \frac{t_i w_h(t_i)}{\left\{ a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\}^2}, \]

\[ I_{\text{lb} \beta_l} = -\frac{\partial^2 L_i}{\partial b \partial \beta_l} = \frac{1}{2} \sum_{k=0}^{q_i} z_{ik} \exp \left( s^{(ik)} \beta \right) \left[ \left( t^{(i,k+1)} \right)^2 - \left( t^{(ik)} \right)^2 \right], \]

\[ I_{\text{lb} \gamma_h' \gamma_h} = -\frac{\partial^2 L_i}{\partial \gamma_h \partial \gamma_h'} = \delta_i \frac{w_h(t_i) w_h(t_i)}{\left\{ a + b t_i + \sum_{j=1}^{m-2} \gamma_j w_j(t_i) \right\}^2}, \]

\[ I_{\text{lb} \gamma_h' \beta_l} = -\frac{\partial^2 L_i}{\partial \gamma_h' \partial \beta_l} = \sum_{k=0}^{q_i} z_{ik} \exp \left( s^{(ik)} \beta \right) \left[ W_h(t^{(i,k+1)}) - W_h(t^{(ik)}) \right]. \]
\[ I_{i \beta l \beta l'} = - \frac{\delta^2 L_i}{\partial \beta_i \partial \beta_{l'}} = \]

\[ \sum_{k=0}^{q_i} z_i z_{i'k} \exp \left( z^{(ik)} \beta \right) \left[ \Lambda^*(t^{(i,k+1)}) - \Lambda^*(t^{(ik)}) \right] \]

for \( l, l' = 1, 2, \ldots, r \) and \( h, h' = 1, 2, \ldots, m-2 \).
Appendix 5
The PHSPLM Procedure

The PHSPLM procedure fits the restricted cubic spline hazard model discussed in Chapters II — IV. The homogeneous model, the spline model with fixed covariates, and the spline model with step-function time-dependent covariates can be estimated using PHSPLM. This procedure provides an automatic test for the exponential model versus the spline model when the effect of regressors is assumed zero. In addition, it provides a global likelihood score and ratio test for the significance of the covariates relative to the homogeneous spline model.

The implementation of the PHSPLM procedure is similar to the PHGLM procedure (Harrell, 1986) with the following exceptions noted. Examples are presented in Appendix 6.

SPECIFICATIONS

The following statements are used with PROC PHSPLM:

PROC PHSPLM options;
  EVENT variable;
  COVTIME variable;
  MODEL dependent=independent/options;
  BY variables;
  KNOTS constants;
  INITIAL constants;
  ID variable;

PROC PHSPLM Statement
PROC PHSPLM options;

The following options on this statement differ from or are a modification of PROC PHGLM:
OUT=SASdataset
causes the procedure to output final parameter estimates and their covariance matrix as in PROC PHGLM. In addition, the knots and the value of the centering constant are output. Knots are labeled as _KNOT_1, _KNOT_2, etc. The centering constant $\mathbf{\bar{y}} \cdot \mathbf{\beta}$, where $\mathbf{\bar{y}}$ is the vector consisting of the covariable mean values, is labeled _CENTER_.

OUTP=SASdataset
causes PHSPLM to output predicted values into a SAS data set. Nuances of the PHSPLM output procedure are in the section entitled “Outputting Predicted Values.”

OUTH=SASdataset
causes PHSPLM to output Kaplan-Meier estimates with confidence limits and the corresponding homogeneous spline estimates and confidence limits. Hazard estimates and median survival estimates with confidence limits for the homogeneous model are also output. See “Outputting the Kaplan-Meier Curve.”

TINC=constant
is used in conjunction with the OUTP= option when the entire survival curve is output and also with the OUTH= option. Predictions are made at time points indicated by the dependent variable. When the interval between time points is greater than TINC, additional predictions are made increments of length TINC. (Default=maximum survival time/150).

MAXN=constant
specifies the maximum number of subjects (Default=2000).

KNOTS=constant
specifies the number of knots (Default=4). When KNOTS=1, an exponential model is used as the baseline hazard. When KNOTS=2, a Rayleigh hazard is used as the baseline hazard. When KNOTS is larger than 2, knots are placed at the following quantiles of the empirical distribution formed by uncensored failure times:

<table>
<thead>
<tr>
<th>KNOTS</th>
<th>Quantiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.1, 0.5, 0.9</td>
</tr>
<tr>
<td>4</td>
<td>0.05, 0.33, 0.67, 0.95</td>
</tr>
<tr>
<td>5</td>
<td>0.05, 0.20, 0.33, 0.67, 0.95</td>
</tr>
</tbody>
</table>

When the number of uncensored observations is less than or equal to 100, the smallest and largest knots are replaced by the 5th smallest and 5th largest uncensored observations. Whenever the first knot is zero, knots are determined from the empirical distribution formed by the nonzero uncensored observations.

ALPHA=constant
specifies type I error for confidence intervals (Default=0.05).

The following PROC PHGLM options are not applicable to PROC PHSPLM: BLOCK, NOPB, NT, and NTIED.
**COVTIME Statement**

**COVTIME** variable;

When time-dependent covariates are present, **COVTIME** must be specified. **COVTIME** is a variable which denotes the time points at which the steps of the time-dependent covariates begin. The input data set should be sorted within each subject by the **COVTIME** variable, and the first record for each subject should have **COVTIME** \( \leq 0 \). If the first record for a subject has any variable missing, all records for that subject are excluded from analysis. If a record after the initial record contains a missing variable value, that record is excluded and covariable values are assumed to be carried forward from the previous record.

**BY Statement**

**BY** variables;

A separate analysis is obtained on observations in groups defined by the **BY** variables except that the **OUT=** and **SW** (stepwise) options cannot simultaneously be specified when **BY** is present.

**KNOTS statement**

**KNOTS** constants;

The constants in the **KNOTS** list specify the values of the knots to be used. When the number of knots is known, but not the exact value, use the **KNOTS** option in the **PROC PHSPLOM** statement or the default value of 4 knots. The **KNOTS** statement and **KNOTS** option cannot be used simultaneously.

**INITIAL statement**

**INITIAL** constants;

The constants in the list specify initial parameter estimates. If there are \( m \) knots in the model, the first \( m \) constants in the **INITIAL** statement correspond to the parameters in the baseline
hazard. The remaining constants specify initial parameter estimates corresponding to the independent variables in the MODEL statement, in the same order. If the initial list contains more constants than there are independent variables, extra constants are ignored. If the list is shorter, starting values of zero are used for the right-most independent variables. When no INITIAL statement appears, PHSPLM uses an exponential estimate for the first parameter in the baseline hazard, with the remaining estimates being 0. When the regression model is estimated and no INITIAL statement appears, the maximum likelihood estimates from the homogeneous model are used with the remaining covariable coefficients being initially 0.

ID statement

ID variable;

The purpose of the ID variable is the same as in PROC PHGLM. However, some of the conventions for obtaining estimates when a COVTIME variable is specified are different and will be specified in "Outputting Predicted Values."

DETAILS

Input Data Set

The input data set used with PROC PHGLM is the same as that used with PROC PHSPLM when a homogeneous model or a regression hazard model with fixed covariables is being considered. When time-dependent covariables are present, each record in the input data set corresponds to an interval of time in which covariables are considered to remain constant. The time at which a step in the covariable vector begins is indicated in the COVTIME variable. Survival time or the time at which observation of a subject terminates and the subject becomes a censored or uncensored observation is included on each record as the dependent variable. The EVENT variable or censoring indicator can be defined on every record, but must be included on the subject's first record.
Computational Methods

The computational methods are as described in chapters II - IV and PROC PHGLM.

However, the model has been transformed to aid computational stability. After knots have been chosen, the difference between the largest and second largest knot is computed and denoted as D. The hazard model used is

\[ \lambda(t) = \left( a + b \frac{t - \gamma_j}{D} + \sum_{j=1}^{m-2} \gamma_j w_j^*(t) \right) \exp\left( \mathbf{z} \beta \right), \]

where

\[ w_j^*(t) = \left( \frac{t - \gamma_j}{D} \right)^3 - \left( \frac{t - \gamma_{m-1}}{D} \right)^3 \frac{(v_m - \gamma_j)}{(v_m - v_{m-1})} + \left( \frac{t - \gamma_m}{D} \right)^3 \frac{(v_{m-1} - \gamma_j)}{(v_m - v_{m-1})} \]

and \( \mathbf{z} \) has been centered by subtracting the overall mean of the covariates. All parameter and covariance estimates will be computed and presented assuming this structure.

Blocking and Conditional Logistic Regression

Not applicable in PROC PHSPLM.

Outputting Predicted Values

When the OUTP= option is specified in the PROC PHSPLM statement, estimated survival probabilities, hazard estimates, and their confidence intervals are output. Each observation in the output data set contains all the variables in the original input data set plus the following variables.

XBETA

is the linear combination of covariates in the final model weighted by the final regression coefficients and centered by subtracting covariable means. When there are no covariates, XBETA is missing.

SURVIVAL

is the estimated survival probability at the current time point (see below).
SURV_L
is the (1-ALPHA) lower confidence limit for the survival estimate. The upper and lower
confidence limits are computed using (2.21).

SURV_U
is the (1-ALPHA) upper confidence limit for the survival estimate.

HAZARD
is the hazard estimate at the current time point.

HAZARD_L
is the (1-ALPHA) lower confidence limit for the hazard estimate.

HAZARD_U
is the (1-ALPHA) upper confidence limit for the hazard estimate.

MEDIAN_S
is estimated median survival time.

MED_S_L
is the (1-ALPHA) lower confidence limit for median survival.

MED_S_U
is the (1-ALPHA) upper confidence limit for median survival.

The variable LOGLOG_S, as described in PHGLM, is not output.

The use of the ID variable when the model is homogeneous or fixed covariates are
present is the same as in PROC PHGLM. In all cases, the ID variable is coded as follows:

<table>
<thead>
<tr>
<th>Value of ID Variable</th>
<th>Observation Used for:</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID statement missing</td>
<td>Estimation and Prediction</td>
</tr>
<tr>
<td>Missing</td>
<td>Estimation only</td>
</tr>
<tr>
<td>0</td>
<td>Estimation and Prediction</td>
</tr>
<tr>
<td>1</td>
<td>Prediction only</td>
</tr>
</tbody>
</table>

When no time-dependent covariates are present, predictions correspond to the time indicated
by the dependent variable, and if the dependent variable value is missing, estimates are
generated with time varying from 0 to the maximum survival time by increments of TINC
(See above). When time-dependent covariates are present, predictions correspond to the time
indicated by the COVTIME variable. These predictions are a function of values of the covariable reported at or prior to COVTIME and the value of COVTIME at which these covariable values change. When COVTIME is missing, an error message will be printed and prediction will be terminated.

Outputting the Kaplan-Meier Curve

When the OUTH= option is specified in the PROC PHSPLM statement, estimates of survival based on the homogeneous spline model and the Kaplan-Meier estimator (1958) are output. Estimates corresponding to each uncensored failure time, time 0, and a midpoint between these failure times will be output. The data set consists of variables (SURVIVAL, SURV_L, SURV_U, HAZARD, HAZARD_U, MEDIAN_S, MED_S_L, and MED_S_U) which are based upon the spline model and described in "Outputting Predicted Values", and also the following variables which are associated with the Kaplan-Meier estimator:

SURV_KM
is the Kaplan-Meier survival estimate

SURV_KML
is the (1-ALPHA) lower confidence limit for the Kaplan-Meier estimate. It is based on Greenwood's formula, which is presented as (1.16).

SURV_KMU
is the (1-ALPHA) upper confidence limit for the Kaplan-Meier estimate.

Estimates of hazard and median survival are not available based on the Kaplan-Meier estimator.
Appendix 6
Examples of Use of the PHSPLM Procedure

Appendix 6 will illustrate the use of PROC PHSPLM. In each case the code and resulting output will be included.

**EXAMPLE 1:** Analysis of Fixed Covariables and Plotting Survival and Hazard Estimates

Data from a study of multiple myeloma taken from Krall, Uthoff, and Harley (1975), "A Step-up Procedure for Selecting Variables Associated with Survival," *Biometrics* 31: 49-57, are analyzed to study the relationship between survival time and patient characteristics. This data has been used in Harrell (1986) to demonstrate the use of the PHGLM procedure. Forward stepwise and fast backward variable selection methods are used in the first and second specification of PHSPLM. With the third PROC PHSPLM statement, Kaplan-Meier and spline survival estimates are computed and plotted for patients grouped by LOGBUN. Spline survival curves are estimated and plotted for various values of LOGBUN in the fourth specification of PHSPLM. The final PROC PHSPLM statement determines and plots 24-month spline survival probability estimates as a function of LOGBUN.

```
OPTIONS NODATE;
DATA MYELOMA;
TITLE 'SURVIVAL IN MULTIPLE MYELOMA';
INPUT T DEATH LOGBUN HG PLATE AGE LOGWBC FRAC LOGBM PROTEIN SCALT;
LABEL T='SURVIVAL TIME';
CARDS;
1.28 1 2.2175 9.4 1 67 3.6628 1 1.9542 12 10
1.28 1 1.9395 12 1 38 3.9868 1 1.9542 20 18
1.21 1 1.5185 9.8 1 81 3.8751 1 2 2 15
2 1 1.7482 11.3 0 75 3.8062 1 1.2553 0 12
2 1 1.301 5.1 0 57 3.7243 1 2 3 9
3 1 1.5441 6.7 1 46 4.4757 0 1.9345 12 10
```
PROC PHSPIM DATA=MYELOMA; EVENT DEATH;
MODEL T=LOGBUN HG PLATE AGE LOGMBC FRAC LOGBM PROTEIN SCALC /
   BACKWARD FAST SLE=0;
TITLE 'BACKWARD ELIMINATION EXAMPLE'; RUN; TITLE ;
*OBTAIN KAPLAN-MEIER AND HOMOGENEOUS SPLINE SURVIVAL ESTIMATES
   AND CONFIDENCE INTERVALS STRATIFIED BY
LOGBUN GROUPED INTO 2 INTERVALS CONTAINING NEARLY EQUAL NUMBERS
   OF OBSERVATIONS (<1.31,>=1.31). MEAN VALUES OF LOGBUN IN
   THESE INTERVALS ARE 1.143, 1.607. ALSO OBTAIN SPLINE HAZARD ESTIMATES:
DATA LB; SET MYELOMA;
LBUN=1*(LOGBUN<1.31)+2*(LOGBUN>=1.31);
PROC SORT; BY LBUN;
PROC PHSPIM OUT=SIM; BY LBUN; EVENT DEATH; MODEL T=;
PROC PLOT; BY LBUN;
   PLOT SURVIVAL*T="'" SURV_L*T = '..' SURV_U*T='.' / OVERLAY;
   PLOT SURV_KM*T = 'S' SURV_KM*T='.' / OVERLAY;
   PLOT HAZARD*T = 'K' HAZARD_L*T='.' HAZARD_U*T='.' / OVERLAY;
TITLE 'Estimates Stratified by LOGBUN Group';RUN;TITLE ;
*OBTAIN CORRESPONDING ESTIMATES FOR MEAN LOGBUN IN EACH GROUP BY MODELING;
DATA EST:I=1; DO LOGBUN=1.143,1.607;OUTPUT;END;
DATA BOTH; SET MYELOMA EST ;
PROC PHSPIM OUT=PRED;ID I; EVENT DEATH; MODEL T=LOGBUN;
PROC PLOT; BY logbun;
   PLOT SURVIVAL*T="'" SURV_L*T='.' SURV_U*T='.' / OVERLAY;
TITLE 'SURVIVAL PREDICTED FROM MODEL WITH LOGBUN AS CONTINUOUS VARIABLE';
RUN;TITLE ;
*PLOT ESTIMATED 24-MONTH SURVIVAL VERSUS CONTINUOUS LOGBUN;
DATA EST:I=1;T=24; DO LOGBUN=1 TO 2 BY .02;OUTPUT;END;
DATA BOTH; SET MYELOMA EST ;
PROC PHSPIM OUT=PRED;ID I; EVENT DEATH; MODEL T=LOGBUN;
PROC PLOT;
   PLOT SURVIVAL*LOGBUN="'" SURV_L*LOGBUN='.' SURV_U*LOGBUN='.' /OVERLAY;
TITLE 'PREDICTED 24-MONTH SURVIVAL VS. LOGBUN';
The output is given below.

**SURVIVAL IN MULTIPLE MYELOMA**

**STEPWISE SPLINE PROPORTIONAL HAZARD LINEAR MODEL PROCEDURE**

DEPENDENT VARIABLE: T  
SURVIVAL TIME

EVENT INDICATOR : DEATH

65 OBSERVATIONS
48 UNCENSORED OBSERVATIONS
0 OBSERVATIONS DELETED DUE TO MISSING VALUES

-2 LOG LIKELIHOOD FOR EXPONENTIAL MODEL | BETA=0 : 430.23

MODEL: RESTRICTED CUBIC SPLINE HAZARD WITH KNOTS AT DEPENDENT VARIABLE VALUES:
2 9 24 66

-2 LOG LIKELIHOOD FOR SPLINE MODEL | BETA=0 : 425.24

LIKELIHOOD RATIO TEST FOR H0: HAZARD(T) = CONSTANT | BETA=0:
CHI-SQUARE = 4.99 WITH 30.D.F.  
P=0.1722.

**STEP 0. HAZARD PARAMETERS ENTERED.**

MAX ABSOLUTE DERIVATIVE=10E-8  

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>BETA</th>
<th>STD. ERROR</th>
<th>CHI-SQUARE</th>
<th>P</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALPHA1</td>
<td>0.02057102</td>
<td>0.0121527</td>
<td>2.87</td>
<td>0.0905</td>
<td></td>
</tr>
<tr>
<td>ALPHA2</td>
<td>0.10466901</td>
<td>0.0788254</td>
<td>1.76</td>
<td>0.1842</td>
<td></td>
</tr>
<tr>
<td>ALPHA3</td>
<td>-0.76981875</td>
<td>0.4485026</td>
<td>2.95</td>
<td>0.0861</td>
<td></td>
</tr>
<tr>
<td>ALPHA4</td>
<td>1.23117188</td>
<td>0.6965840</td>
<td>3.12</td>
<td>0.0772</td>
<td></td>
</tr>
</tbody>
</table>

**TEST FOR ALL VARIABLES NOT IN THE MODEL:**

RESIDUAL CHI-SQUARE= 20.60 WITH 9 D.F.  P=0.0145

**SIMPLE UN-ADJUSTED CHI-SQUARE Q STATISTICS**

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>CHI-SQUARE</th>
<th>P</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOGBUN</td>
<td>9.71</td>
<td>0.0018</td>
<td>0.135</td>
</tr>
<tr>
<td>HG</td>
<td>5.47</td>
<td>0.0194</td>
<td>-0.090</td>
</tr>
<tr>
<td>PLATE</td>
<td>3.31</td>
<td>0.0688</td>
<td>-0.056</td>
</tr>
<tr>
<td>AGE</td>
<td>0.00</td>
<td>0.9633</td>
<td>0.000</td>
</tr>
<tr>
<td>LOGMSC</td>
<td>0.96</td>
<td>0.3261</td>
<td>0.000</td>
</tr>
<tr>
<td>FRAC</td>
<td>1.60</td>
<td>0.2059</td>
<td>0.000</td>
</tr>
<tr>
<td>LOGBM</td>
<td>0.68</td>
<td>0.4096</td>
<td>0.000</td>
</tr>
<tr>
<td>PROTEIN</td>
<td>0.02</td>
<td>0.8754</td>
<td>0.000</td>
</tr>
<tr>
<td>SCALC</td>
<td>1.33</td>
<td>0.2484</td>
<td>0.000</td>
</tr>
</tbody>
</table>
SURVIVAL IN MULTIPLE MYELOMA

STEPWISE SPLINE PROPORTIONAL HAZARD LINEAR MODEL PROCEDURE

DEPENDENT VARIABLE: T
EVENT INDICATOR : DEATH

STEP 1. VARIABLE LOGBUN ADDED WITH CHI-SQUARE= 9.71 P=0.0018 R= 0.135.

CONVERGENCE IN 6 ITERATIONS WITH 0 STEP HALVINGS R= 0.127.
MAX ABSOLUTE DERIVATIVE=41E-9 -2 LOG L= 416.37.
MODEL CHI-SQUARE= 8.87 WITH 1 D.F. (-2 LOG L.R.) P=0.0029.

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>BETA</th>
<th>STD. ERROR</th>
<th>CHI-SQUARE</th>
<th>P</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALPHA1</td>
<td>0.01455709</td>
<td>0.0112814</td>
<td>1.67</td>
<td>0.1969</td>
<td></td>
</tr>
<tr>
<td>ALPHA2</td>
<td>0.15723439</td>
<td>0.0816345</td>
<td>3.71</td>
<td>0.0541</td>
<td></td>
</tr>
<tr>
<td>ALPHA3</td>
<td>-1.03699896</td>
<td>0.4820903</td>
<td>4.63</td>
<td>0.0315</td>
<td></td>
</tr>
<tr>
<td>ALPHA4</td>
<td>1.64201078</td>
<td>0.7524969</td>
<td>4.76</td>
<td>0.0291</td>
<td></td>
</tr>
<tr>
<td>LOGBUN</td>
<td>1.88017818</td>
<td>0.6101961</td>
<td>9.49</td>
<td>0.0021 0.133</td>
<td></td>
</tr>
<tr>
<td>CENTER</td>
<td>2.61890019</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TEST FOR ALL VARIABLES NOT IN THE MODEL:
RESIDUAL CHI-SQUARE= 10.12 WITH 8 D.F. P=0.2564

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>CHI-SQUARE</th>
<th>P</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>HG</td>
<td>4.52</td>
<td>0.0335</td>
<td>-0.077</td>
</tr>
<tr>
<td>PLATE</td>
<td>2.03</td>
<td>0.1547</td>
<td>-0.008</td>
</tr>
<tr>
<td>AGE</td>
<td>0.70</td>
<td>0.4024</td>
<td>0.000</td>
</tr>
<tr>
<td>LOGWBC</td>
<td>0.10</td>
<td>0.7502</td>
<td>0.000</td>
</tr>
<tr>
<td>FRAC</td>
<td>1.53</td>
<td>0.2167</td>
<td>0.000</td>
</tr>
<tr>
<td>LOGBM</td>
<td>1.21</td>
<td>0.2722</td>
<td>0.000</td>
</tr>
<tr>
<td>PROTEIN</td>
<td>0.30</td>
<td>0.5859</td>
<td>0.000</td>
</tr>
<tr>
<td>SCALC</td>
<td>1.61</td>
<td>0.2050</td>
<td>0.000</td>
</tr>
</tbody>
</table>

RESIDUAL CHI-SQUARE IS NOT SIGNIFICANT AT THE 0.0500 LEVEL.

PARAMETER ESTIMATES AND COVARIANCE MATRIX

<table>
<thead>
<tr>
<th>OBS</th>
<th>_KNOT_1</th>
<th>_KNOT_2</th>
<th>_KNOT_3</th>
<th>_KNOT_4</th>
<th>CENTER</th>
<th>ALPHA1</th>
<th>ALPHA2</th>
<th>ALPHA3</th>
<th>ALPHA4</th>
<th>LOGBUN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>9</td>
<td>24</td>
<td>66</td>
<td>2.6189</td>
<td>0.0145571</td>
<td>0.157234</td>
<td>-1.0370</td>
<td>1.64201</td>
<td>1.88018</td>
</tr>
<tr>
<td>2</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>0.0001273</td>
<td>-0.000754</td>
<td>-0.0038</td>
<td>-0.00579</td>
<td>-0.00210</td>
</tr>
<tr>
<td>3</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>-0.0007541</td>
<td>0.006664</td>
<td>-0.0379</td>
<td>0.05853</td>
<td>0.01130</td>
</tr>
<tr>
<td>4</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>0.0037977</td>
<td>-0.037930</td>
<td>0.2324</td>
<td>-0.36243</td>
<td>-0.05286</td>
</tr>
<tr>
<td>5</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>-0.0057852</td>
<td>0.058527</td>
<td>-0.3624</td>
<td>0.56625</td>
<td>0.08075</td>
</tr>
<tr>
<td>6</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>-0.0020987</td>
<td>0.011296</td>
<td>-0.0529</td>
<td>0.08075</td>
<td>0.37234</td>
</tr>
</tbody>
</table>
BACKWARD ELIMINATION EXAMPLE

STEPWISE SPLINE PROPORTIONAL HAZARD LINEAR MODEL PROCEDURE

DEPENDENT VARIABLE: T
EVENT INDICATOR : DEATH

65 OBSERVATIONS
48 UNCESED OBSERVATIONS
0 OBSERVATIONS DELETED DUE TO MISSING VALUES

-2 LOG LIKELIHOOD FOR EXPONENTIAL MODEL | BETA=0 : 430.23
MODEL: RESTRICTED CUBIC SPLINE HAZARD WITH KNOTS AT DEPENDENT VARIABLE VALUES:
2 9 24 66

-2 LOG LIKELIHOOD FOR SPLINE MODEL | BETA=0 : 425.24
LIKELIHOOD RATIO TEST FOR HO: HAZARD(T) = CONSTANT | BETA=0:
CHI-SQUARE= 4.99 WITH 30.D.F. P=0.1722.

STEP 0. THE FOLLOWING VARIABLES ARE ENTERED:

ALPHA1 ALPHA2 ALPHA3 ALPHA4 LOGBUN HG PLATE AGE
LOGWBC FRAC LOGBM PROTEIN SCALC

MODEL CHI-SQUARE= 20.60 WITH 9 D.F. (SCORE STAT.) P=0.0145,
CONVERGENCE IN 7 ITERATIONS WITH 1 STEP HALVINGS R= 0.052.
MAX ABSOLUTE DERIVATIVE=59E-7
-2 LOG L= 406.10.
MODEL CHI-SQUARE= 19.14 WITH 9 D.F. (-2 LOG L.R.) P=0.0241.

FAST BACKWARD ELIMINATION

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>DELETED</th>
<th>CHI-SQUARE</th>
<th>P</th>
<th>RESIDUAL</th>
<th>CHI-SQUARE</th>
<th>D.F.</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROTEIN</td>
<td>0.16</td>
<td>0.6923</td>
<td>0.16</td>
<td>1</td>
<td>0.6923</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOGWBC</td>
<td>0.17</td>
<td>0.6774</td>
<td>0.33</td>
<td>2</td>
<td>0.8480</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLATE</td>
<td>0.18</td>
<td>0.6695</td>
<td>0.51</td>
<td>3</td>
<td>0.9163</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOGBM</td>
<td>0.69</td>
<td>0.4074</td>
<td>1.20</td>
<td>4</td>
<td>0.8784</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FRAC</td>
<td>1.02</td>
<td>0.3136</td>
<td>2.21</td>
<td>5</td>
<td>0.8189</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td>1.83</td>
<td>0.1766</td>
<td>4.04</td>
<td>6</td>
<td>0.6714</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCALC</td>
<td>1.97</td>
<td>0.1604</td>
<td>6.01</td>
<td>7</td>
<td>0.5386</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HG</td>
<td>3.65</td>
<td>0.0560</td>
<td>9.66</td>
<td>8</td>
<td>0.2896</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

STEP 1.

CONVERGENCE IN 4 ITERATIONS WITH 0 STEP HALVINGS R= 0.127.
MAX ABSOLUTE DERIVATIVE=28E-6
-2 LOG L= 416.37.
MODEL CHI-SQUARE= 8.87 WITH 1 D.F. (-2 LOG L.R.) P=0.0029.
BACKWARD ELIMINATION EXAMPLE

STEPWISE SPLINE PROPORTIONAL HAZARD LINEAR MODEL PROCEDURE

DEPENDENT VARIABLE: T
EVENT INDICATOR : DEATH
SURVIVAL TIME

TEST FOR ALL VARIABLES NOT IN THE MODEL:
RESIDUAL CHI-SQUARE = 10.12 WITH 8 D.F. P=0.2564

NO ADDITIONAL VARIABLES MET THE 0.0000 SIGNIFICANCE LEVEL FOR ENTRY.

FINAL PARAMETER ESTIMATES

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>BETA</th>
<th>STD. ERROR</th>
<th>CHI-SQUARE</th>
<th>P</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALPHA1</td>
<td>0.01455709</td>
<td>0.0112814</td>
<td>1.67</td>
<td>0.1969</td>
<td></td>
</tr>
<tr>
<td>ALPHA2</td>
<td>0.15723439</td>
<td>0.0816345</td>
<td>3.71</td>
<td>0.0541</td>
<td></td>
</tr>
<tr>
<td>ALPHA3</td>
<td>-1.036999895</td>
<td>0.4820903</td>
<td>4.63</td>
<td>0.0315</td>
<td></td>
</tr>
<tr>
<td>ALPHA4</td>
<td>1.64201076</td>
<td>0.7524969</td>
<td>4.76</td>
<td>0.0291</td>
<td></td>
</tr>
<tr>
<td>LOGBUN</td>
<td>1.88017820</td>
<td>0.6101961</td>
<td>9.49</td>
<td>0.0021</td>
<td>0.133</td>
</tr>
</tbody>
</table>
LBUN=1

SPLINE PROPORTIONAL HAZARD LINEAR MODEL PROCEDURE

DEPENDENT VARIABLE: \( T \) \hspace{1cm} \text{SURVIVAL TIME}
EVENT INDICATOR : \text{DEATH}

30 OBSERVATIONS
22 UNCENSORED OBSERVATIONS
0 OBSERVATIONS DELETED DUE TO MISSING VALUES

\(-2\) LOG LIKELIHOOD FOR EXPONENTIAL MODEL \( | \beta = 0 \): \( 204.05 \)

MODEL: RESTRICTED CUBIC SPLINE HAZARD WITH KNOTS AT DEPENDENT VARIABLE VALUES:
11 \hspace{1cm} 18.5 \hspace{1cm} 26 \hspace{1cm} 41

\(-2\) LOG LIKELIHOOD FOR SPLINE MODEL \( | \beta = 0 \): \( 200.88 \)

LIKELIHOOD RATIO TEST FOR H0: \text{HAZARD}(T) = \text{CONSTANT} \hspace{0.5cm} | \beta = 0:
CHI-SQUARE = 3.17 WITH 30.D.F. \hspace{1cm} P = 0.5656.

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>BETA</th>
<th>STD. ERROR</th>
<th>CHI-SQUARE</th>
<th>P</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALPHA1</td>
<td>0.000505879</td>
<td>0.0108977</td>
<td>0.22</td>
<td>0.6425</td>
<td></td>
</tr>
<tr>
<td>ALPHA2</td>
<td>0.03388436</td>
<td>0.0207506</td>
<td>2.99</td>
<td>0.0838</td>
<td></td>
</tr>
<tr>
<td>ALPHA3</td>
<td>-0.06094767</td>
<td>0.0388192</td>
<td>2.47</td>
<td>0.1164</td>
<td></td>
</tr>
<tr>
<td>ALPHA4</td>
<td>0.15176247</td>
<td>0.0984306</td>
<td>2.38</td>
<td>0.1231</td>
<td></td>
</tr>
</tbody>
</table>

LBUN=2

SPLINE PROPORTIONAL HAZARD LINEAR MODEL PROCEDURE

DEPENDENT VARIABLE: \( T \) \hspace{1cm} \text{SURVIVAL TIME}
EVENT INDICATOR : \text{DEATH}

35 OBSERVATIONS
26 UNCENSORED OBSERVATIONS
0 OBSERVATIONS DELETED DUE TO MISSING VALUES

\(-2\) LOG LIKELIHOOD FOR EXPONENTIAL MODEL \( | \beta = 0 \): \( 225.02 \)

MODEL: RESTRICTED CUBIC SPLINE HAZARD WITH KNOTS AT DEPENDENT VARIABLE VALUES:
3 \hspace{1cm} 6 \hspace{1cm} 16 \hspace{1cm} 51

\(-2\) LOG LIKELIHOOD FOR SPLINE MODEL \( | \beta = 0 \): \( 220.49 \)

LIKELIHOOD RATIO TEST FOR H0: \text{HAZARD}(T) = \text{CONSTANT} \hspace{0.5cm} | \beta = 0:
CHI-SQUARE = 4.54 WITH 30.D.F. \hspace{1cm} P = 0.2090.

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>BETA</th>
<th>STD. ERROR</th>
<th>CHI-SQUARE</th>
<th>P</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALPHA1</td>
<td>0.03362300</td>
<td>0.0236297</td>
<td>2.02</td>
<td>0.1548</td>
<td></td>
</tr>
<tr>
<td>ALPHA2</td>
<td>0.11867781</td>
<td>0.1591120</td>
<td>0.56</td>
<td>0.4357</td>
<td></td>
</tr>
<tr>
<td>ALPHA3</td>
<td>-2.69155043</td>
<td>2.2378975</td>
<td>1.45</td>
<td>0.2291</td>
<td></td>
</tr>
<tr>
<td>ALPHA4</td>
<td>3.65529281</td>
<td>2.9701204</td>
<td>1.51</td>
<td>0.2184</td>
<td></td>
</tr>
</tbody>
</table>
SPLINE PROPORTIONAL HAZARD LINEAR MODEL PROCEDURE

DEPENDENT VARIABLE: T \ (SURVIVAL TIME)  
EVENT INDICATOR : DEATH

65 OBSERVATIONS  
48 UNCENSORED OBSERVATIONS  
0 OBSERVATIONS DELETED DUE TO MISSING VALUES

-2 LOG LIKELIHOOD FOR EXPONENTIAL MODEL \ (BETA=0) \ :  \ 430.23

MODEL: RESTRICTED CUBIC SPLINE HAZARD WITH KNOTS AT DEPENDENT VARIABLE VALUES:  
\[2 \ 9 \ 24 \ 66\]

-2 LOG LIKELIHOOD FOR SPLINE MODEL \ (BETA=0) \ :  \ 425.24

LIKELIHOOD RATIO TEST FOR H0: HAZARD(T) = CONSTANT \ (BETA=0):  
CHI-SQUARE= 4.99 WITH 30 D.F. \ \ P=0.1722.

MODEL CHI-SQUARE= 9.71 WITH 1 D.F. \ (SCORE STAT.) \ P=0.0018.  
CONVERGENCE IN 6 ITERATIONS WITH 0 STEP HALVINGS \ R= 0.127.  
MAX ABSOLUTE DERIVATIVE=4.1E-9 \ -2 LOG L= 416.37.  
MODEL CHI-SQUARE= 8.87 WITH 1 D.F. \ (-2 LOG L.R.) \ P=0.0029.

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>BETA</th>
<th>STD. ERROR</th>
<th>CHI-SQUARE</th>
<th>P</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALPHA1</td>
<td>0.01455709</td>
<td>0.0112814</td>
<td>1.67</td>
<td>0.1969</td>
<td></td>
</tr>
<tr>
<td>ALPHA2</td>
<td>0.15723439</td>
<td>0.0816345</td>
<td>3.71</td>
<td>0.0541</td>
<td></td>
</tr>
<tr>
<td>ALPHA3</td>
<td>-1.03690896</td>
<td>0.4820903</td>
<td>4.63</td>
<td>0.0315</td>
<td></td>
</tr>
<tr>
<td>ALPHA4</td>
<td>1.64201078</td>
<td>0.7524969</td>
<td>4.76</td>
<td>0.0291</td>
<td></td>
</tr>
<tr>
<td>LOGBUN</td>
<td>1.88017818</td>
<td>0.6101961</td>
<td>9.49</td>
<td>0.0021</td>
<td></td>
</tr>
<tr>
<td>CENTER</td>
<td>2.61890019</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SPLINE PROPORTIONAL HAZARD LINEAR MODEL PROCEDURE

DEPENDENT VARIABLE: T
SURVIVAL TIME
EVENT INDICATOR: DEATH

65 OBSERVATIONS
48 UNCENSORED OBSERVATIONS
0 OBSERVATIONS DELETED DUE TO MISSING VALUES

-2 LOG LIKELIHOOD FOR EXPONENTIAL MODEL | BETA=0: 430.23

MODEL: RESTRICTED CUBIC SPLINE HAZARD WITH KNOTS AT DEPENDENT VARIABLE VALUES: 2 9 24 66

-2 LOG LIKELIHOOD FOR SPLINE MODEL | BETA=0: 425.24

LIKELIHOOD RATIO TEST FOR HO: HAZARD(T) = CONSTANT | BETA=0:
CHI-SQUARE = 4.99 WITH 30 D.F. | P=0.1722.

MODEL CHI-SQUARE = 9.71 WITH 1 D.F. (SCORE STAT.) P=0.0018.
CONVERGENCE IN 6 ITERATIONS WITH 0 STEP HALVINGS | R= 0.127.
MAX ABSOLUTE DERIVATIVE=4.1E-9
-2 LOG L = 416.37.
MODEL CHI-SQUARE = 8.87 WITH 1 D.F. (-2 LOG L.R.) P=0.0029.

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>BETA</th>
<th>STD. ERROR</th>
<th>CHI-SQUARE</th>
<th>P</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALPHA1</td>
<td>0.01455709</td>
<td>0.0112814</td>
<td>1.67</td>
<td>0.1969</td>
<td></td>
</tr>
<tr>
<td>ALPHA2</td>
<td>0.15723439</td>
<td>0.0816345</td>
<td>3.71</td>
<td>0.0541</td>
<td></td>
</tr>
<tr>
<td>ALPHA3</td>
<td>-1.03699896</td>
<td>0.4820903</td>
<td>4.63</td>
<td>0.0315</td>
<td></td>
</tr>
<tr>
<td>ALPHA4</td>
<td>1.64201078</td>
<td>0.7526969</td>
<td>4.76</td>
<td>0.0291</td>
<td></td>
</tr>
<tr>
<td>LOGBMN</td>
<td>1.86017818</td>
<td>0.6101961</td>
<td>9.49</td>
<td>0.0021</td>
<td>0.133</td>
</tr>
<tr>
<td>CENTER</td>
<td>2.61890019</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
EXAMPLE 2: Model with Binary Time-Dependent Covariate

The data analyzed in Lee (1986) and reanalyzed in Section 4.3 are used to demonstrate the use of PROC PHSPLM when a binary time-dependent covariable is present. Records corresponding to intervals of time in which covariables are constant are initially generated. Hazard and survival predictions are obtained for a patient who does not have a myocardial infarction, a patient who has a myocardial infarction at time 0, and a patient who has a myocardial infarction 2.25 years after catheterization. Only 25 of the 365 predictions will be printed in the output below. However, all were used when Figures 4.2a — 4.4c were created.

```
OPTIONS LS=60 NODATE;
TITLE 'EFFECT OF AN INTERVENING NON-FATAL MYOCARDIAL INFARCTION';
DATA ALL; SET DB.MEDGRP(KEEP=D.TIME CDEATH INFARC DM.TIME);
  * DB.MEDGRP IS THE SAS DATA SET WHICH CONTAINS LEE'S MYOCARDIAL INFARCTION DATA. VARIABLES DESCRIPTIONS ARE:
    D.TIME = TIME OF CARDIOVASCULAR DEATH OR CENSORING
    CDEATH = CENSORING INDICATOR (1 IF UNCENSORED, 0 OTHERWISE)
    INFARC = EQUALS 1 IF PATIENT EXPERIENCES AN MI, 0 OTHERWISE
    DM_TIME = TIME OF MYOCARDIAL INFARCTION

DATA ALL; SET ALL;
  PAT=_.;
  IF (INFARC NE 1) THEN DO:
    CT=0; MI=0; OUTPUT;
    * PATIENTS WHO DO NOT HAVE AN MI;
  END;
  ELSE DO:
    IF (DM_TIME EQ 0) THEN DO:
      CT=0; MI=1; OUTPUT;
      * PATIENTS EXPERIENCED AN MI AT CATHETERIZATION;
    END;
    ELSE DO:
      CT=0; MI=0; OUTPUT;
      * THIS RECORD CORRESPONDS TO TIME INTERVAL DURING WHICH INFARCTION HAS NOT OCCURRED;
    IF (D.TIME GE DM_TIME) THEN DO:
      MI=1; CT=DM_TIME; OUTPUT;
      * THIS RECORD CORRESPONDS TO TIME INTERVAL AFTER THE MI OCCURRED;
    END;
  END;
DATA P;
  *DATA SET CREATED WHICH CONTAINS INFORMATION FOR PREDICTIONS:
  I=1;
  MI=0; PAT=99997; DO CT=0 TO 12 BY 0.05; OUTPUT; END;
  MI=1; PAT=99996; DO CT=0 TO 12 BY 0.05; OUTPUT; END;
  PAT=99999; MI=0; DO CT=0 TO 2.2 BY 0.05; OUTPUT; END;
  CT=2.249; OUTPUT;
  MI=1; DO CT=2.25 TO 12 BY 0.05; OUTPUT; END;
DATA ALL; SET ALL P;
PROC PHSPLM DATA=ALL OUTP=PRED PCOV PCOR MAXN=4000 KNOTS=5;
  EVENT CDEATH;
  MODEL D_TIME-MI;
  COVTIME CT;
  ID I;
  OPTIONS OBS=25;
PROC PRINT DATA=PRED;
  TITLE2 'SURVIVAL AND HAZARD ESTIMATES';
```
Output is given below. PROC GPLOT is used to create Figures 2.2a – 2.4b based on the SAS data set PRED.

EFFECT OF AN INTERVENING NON-FATAL MYOCARDIAL INFARCTION

SPLINE PROPORTIONAL HAZARD LINEAR MODEL PROCEDURE

DEPENDENT VARIABLE: D_TIME Years
EVENT INDICATOR : CDEATH Cardiovascular Death
COVARIATE TIME : CT

3321 OBSERVATIONS
586 UNCENSORED OBSERVATIONS
3085 SUBJECTS
0 OBSERVATIONS DELETED DUE TO MISSING VALUES

-2 LOG LIKELIHOOD FOR EXPONENTIAL MODEL | BETA=0 : 4598.90

MODEL: RESTRICTED CUBIC SPLINE HAZARD WITH KNOTS AT DEPENDENT VARIABLE VALUES:
0.013689 .216276 .736481 3.153381 8.76207

-2 LOG LIKELIHOOD FOR SPLINE MODEL | BETA=0 : 4420.14

LIKELIHOOD RATIO TEST FOR HO: HAZARD(T) = CONSTANT | BETA=0:
CHI-SQUARE= 186.76 WITH 4 D.F. P=0.0000

MODEL CHI-SQUARE= 58.89 WITH 1 D.F. (SCORE STAT.) P=0.0000.
CONVERGENCE IN 8 ITERATIONS WITH 0 STEP HALVINGS R= 0.100.
MAX ABSOLUTE DERIVATIVE= 12-9
-2 LOG L= 4374.34.
MODEL CHI-SQUARE= 45.80 WITH 1 D.F. (-2 LOG L.R.) P=0.0000.

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>BETA</th>
<th>STD. ERROR</th>
<th>CHI-SQUARE</th>
<th>P</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALPHA1</td>
<td>0.276143285</td>
<td>0.0254372</td>
<td>117.85</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>ALPHA2</td>
<td>-4.28608873</td>
<td>0.5614017</td>
<td>73.07</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>ALPHA3</td>
<td>319.9718017</td>
<td>38.8332876</td>
<td>67.39</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>ALPHA4</td>
<td>-448.04855153</td>
<td>54.2925427</td>
<td>67.45</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>ALPHA5</td>
<td>126.42933367</td>
<td>15.5571997</td>
<td>66.04</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>MI</td>
<td>0.89224109</td>
<td>0.1200045</td>
<td>55.32</td>
<td>0.0000</td>
<td>0.110</td>
</tr>
</tbody>
</table>
EFFECT OF AN INTERVENING NON-FATAL MYOCARDIAL INFARCTION

SPLINE PROPORTIONAL HAZARD LINEAR MODEL PROCEDURE

DEPENDENT VARIABLE: D_TIME Years
EVENT INDICATOR: CDEATH Cardiovascular Death
COVARIATE TIME: CT

<table>
<thead>
<tr>
<th></th>
<th>ALPHA1</th>
<th>ALPHA2</th>
<th>ALPHA3</th>
<th>ALPHA4</th>
<th>ALPHA5</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALPHA1</td>
<td>0.0006470525</td>
<td>-0.0122542</td>
<td>0.9132667</td>
<td>-1.27231</td>
<td>0.3598189</td>
</tr>
<tr>
<td>ALPHA2</td>
<td>0.0001452727</td>
<td>-0.00213685</td>
<td>0.2514037</td>
<td>-19.3262</td>
<td>28.98871</td>
</tr>
<tr>
<td>ALPHA3</td>
<td>0.9132667</td>
<td>-19.3262</td>
<td>1508.024</td>
<td>-2108.52</td>
<td>603.5472</td>
</tr>
<tr>
<td>ALPHA4</td>
<td>0.1320432</td>
<td>26.98871</td>
<td>-2108.52</td>
<td>2948.44</td>
<td>-844.261</td>
</tr>
<tr>
<td>ALPHA5</td>
<td>-1.37321</td>
<td>7.69375</td>
<td>603.5472</td>
<td>-844.261</td>
<td>242.0264</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ALPHA1</th>
<th>ALPHA2</th>
<th>ALPHA3</th>
<th>ALPHA4</th>
<th>ALPHA5</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALPHA1</td>
<td>0.0001452727</td>
<td>-0.00213685</td>
<td>0.1320432</td>
<td>-0.181928</td>
<td>0.04956582</td>
</tr>
<tr>
<td>ALPHA2</td>
<td>0.01440109</td>
<td>0.046</td>
<td>-0.961</td>
<td>0.925</td>
<td>-0.921</td>
</tr>
<tr>
<td>ALPHA3</td>
<td>0.046</td>
<td>1.000</td>
<td>-0.993</td>
<td>0.991</td>
<td>-0.986</td>
</tr>
<tr>
<td>ALPHA4</td>
<td>-0.921</td>
<td>0.991</td>
<td>1.000</td>
<td>0.999</td>
<td></td>
</tr>
<tr>
<td>ALPHA5</td>
<td>0.028</td>
<td>-0.028</td>
<td>-0.999</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ALPHA1</th>
<th>ALPHA2</th>
<th>ALPHA3</th>
<th>ALPHA4</th>
<th>ALPHA5</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALPHA1</td>
<td>1.000</td>
<td>-0.961</td>
<td>0.925</td>
<td>-0.921</td>
<td>0.909</td>
</tr>
<tr>
<td>ALPHA2</td>
<td>0.046</td>
<td>1.000</td>
<td>-0.993</td>
<td>0.991</td>
<td>-0.986</td>
</tr>
<tr>
<td>ALPHA3</td>
<td>0.925</td>
<td>-0.993</td>
<td>1.000</td>
<td>-1.000</td>
<td>0.999</td>
</tr>
<tr>
<td>ALPHA4</td>
<td>0.028</td>
<td>0.991</td>
<td>-1.000</td>
<td>1.000</td>
<td>-0.999</td>
</tr>
<tr>
<td>ALPHA5</td>
<td>-0.921</td>
<td>-0.028</td>
<td>0.999</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>ALPHA1</td>
<td>0.046</td>
<td>-0.036</td>
<td>0.028</td>
<td>-0.028</td>
<td>0.027</td>
</tr>
<tr>
<td>ALPHA2</td>
<td>0.027</td>
<td>-0.986</td>
<td>0.999</td>
<td>-0.999</td>
<td>1.000</td>
</tr>
<tr>
<td>ALPHA3</td>
<td>1.000</td>
<td>-0.036</td>
<td>0.028</td>
<td>-0.028</td>
<td>0.027</td>
</tr>
<tr>
<td>ALPHA4</td>
<td>0.909</td>
<td>-0.986</td>
<td>0.999</td>
<td>-0.999</td>
<td>1.000</td>
</tr>
<tr>
<td>ALPHA5</td>
<td>0.027</td>
<td>0.028</td>
<td>-0.036</td>
<td>0.028</td>
<td>-0.028</td>
</tr>
<tr>
<td>ALPHA1</td>
<td>-0.921</td>
<td>-0.028</td>
<td>0.999</td>
<td>-0.999</td>
<td>1.000</td>
</tr>
<tr>
<td>ALPHA2</td>
<td>0.046</td>
<td>-0.036</td>
<td>0.028</td>
<td>-0.028</td>
<td>0.027</td>
</tr>
<tr>
<td>ALPHA3</td>
<td>0.027</td>
<td>-0.986</td>
<td>0.999</td>
<td>-0.999</td>
<td>1.000</td>
</tr>
<tr>
<td>ALPHA4</td>
<td>1.000</td>
<td>-0.036</td>
<td>0.028</td>
<td>-0.028</td>
<td>0.027</td>
</tr>
<tr>
<td>ALPHA5</td>
<td>0.909</td>
<td>-0.986</td>
<td>0.999</td>
<td>-0.999</td>
<td>1.000</td>
</tr>
</tbody>
</table>
**EFFECT OF AN INTERVENING NON-FATAL MYOCARDIAL INFARCTION SURVIVAL AND HAZARD ESTIMATES**

<table>
<thead>
<tr>
<th>OBS</th>
<th>CDEATH</th>
<th>DW_TIME</th>
<th>INFARC</th>
<th>PAT</th>
<th>CT</th>
<th>I_D_TIME</th>
<th>MI</th>
<th>SURVIVAL</th>
<th>SURV_L</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9997</td>
<td>0.00</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.05</td>
<td>1</td>
<td>0</td>
<td>1.00000</td>
<td>0.9997</td>
</tr>
<tr>
<td>2</td>
<td>0.9997</td>
<td>0.05</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.10</td>
<td>1</td>
<td>0</td>
<td>0.99816</td>
<td>0.99587</td>
</tr>
<tr>
<td>3</td>
<td>0.9997</td>
<td>0.10</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.15</td>
<td>1</td>
<td>0</td>
<td>0.99717</td>
<td>0.97405</td>
</tr>
<tr>
<td>4</td>
<td>0.9997</td>
<td>0.20</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.25</td>
<td>1</td>
<td>0</td>
<td>0.99689</td>
<td>0.96436</td>
</tr>
<tr>
<td>5</td>
<td>0.9997</td>
<td>0.30</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.35</td>
<td>1</td>
<td>0</td>
<td>0.99537</td>
<td>0.94546</td>
</tr>
<tr>
<td>6</td>
<td>0.9997</td>
<td>0.40</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.45</td>
<td>1</td>
<td>0</td>
<td>0.99469</td>
<td>0.94155</td>
</tr>
<tr>
<td>7</td>
<td>0.9997</td>
<td>0.50</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.55</td>
<td>1</td>
<td>0</td>
<td>0.99425</td>
<td>0.93355</td>
</tr>
<tr>
<td>8</td>
<td>0.9997</td>
<td>0.60</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.65</td>
<td>1</td>
<td>0</td>
<td>0.99389</td>
<td>0.92966</td>
</tr>
<tr>
<td>9</td>
<td>0.9997</td>
<td>0.70</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.75</td>
<td>1</td>
<td>0</td>
<td>0.99346</td>
<td>0.92597</td>
</tr>
<tr>
<td>10</td>
<td>0.9997</td>
<td>0.80</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.85</td>
<td>1</td>
<td>0</td>
<td>0.99308</td>
<td>0.92215</td>
</tr>
<tr>
<td>11</td>
<td>0.9997</td>
<td>0.90</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.95</td>
<td>1</td>
<td>0</td>
<td>0.99262</td>
<td>0.91930</td>
</tr>
<tr>
<td>12</td>
<td>0.9997</td>
<td>1.00</td>
<td>0.9997</td>
<td>0.9997</td>
<td>1.00</td>
<td>1</td>
<td>0</td>
<td>0.99218</td>
<td>0.91639</td>
</tr>
<tr>
<td>13</td>
<td>0.9997</td>
<td>1.05</td>
<td>0.9997</td>
<td>0.9997</td>
<td>1.10</td>
<td>1</td>
<td>0</td>
<td>0.99238</td>
<td>0.91208</td>
</tr>
<tr>
<td>14</td>
<td>0.9997</td>
<td>1.10</td>
<td>0.9997</td>
<td>0.9997</td>
<td>1.15</td>
<td>1</td>
<td>0</td>
<td>0.99204</td>
<td>0.90900</td>
</tr>
<tr>
<td>15</td>
<td>0.9997</td>
<td>1.20</td>
<td>0.9997</td>
<td>0.9997</td>
<td>1.20</td>
<td>1</td>
<td>0</td>
<td>0.99165</td>
<td>0.90579</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OBS</th>
<th>SURV_U</th>
<th>HAZARD</th>
<th>HAZARD_L</th>
<th>HAZARD_U</th>
<th>XBETA</th>
<th>MEDIAN_S</th>
<th>MED_S_L</th>
<th>MED_S_U</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00000</td>
<td>0.255784</td>
<td>0.209551</td>
<td>0.302017</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>2</td>
<td>0.99038</td>
<td>0.220584</td>
<td>0.182048</td>
<td>0.259119</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>3</td>
<td>0.98184</td>
<td>0.162294</td>
<td>0.158075</td>
<td>0.217812</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>4</td>
<td>0.97457</td>
<td>0.154158</td>
<td>0.129504</td>
<td>0.178612</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>5</td>
<td>0.96875</td>
<td>0.125424</td>
<td>0.106193</td>
<td>0.144656</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>6</td>
<td>0.96402</td>
<td>0.101252</td>
<td>0.085971</td>
<td>0.116533</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>7</td>
<td>0.96016</td>
<td>0.081795</td>
<td>0.068992</td>
<td>0.094597</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>8</td>
<td>0.95701</td>
<td>0.066570</td>
<td>0.055089</td>
<td>0.076072</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>9</td>
<td>0.95414</td>
<td>0.055087</td>
<td>0.044159</td>
<td>0.060018</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>10</td>
<td>0.95221</td>
<td>0.046854</td>
<td>0.036198</td>
<td>0.057613</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>11</td>
<td>0.95031</td>
<td>0.041377</td>
<td>0.030966</td>
<td>0.051792</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>12</td>
<td>0.94858</td>
<td>0.038159</td>
<td>0.028105</td>
<td>0.048234</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>13</td>
<td>0.94696</td>
<td>0.035735</td>
<td>0.027146</td>
<td>0.045634</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>14</td>
<td>0.94537</td>
<td>0.033583</td>
<td>0.027660</td>
<td>0.045605</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>15</td>
<td>0.94377</td>
<td>0.031722</td>
<td>0.028793</td>
<td>0.045600</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>16</td>
<td>0.94214</td>
<td>0.030161</td>
<td>0.030282</td>
<td>0.046041</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>17</td>
<td>0.94046</td>
<td>0.028704</td>
<td>0.031654</td>
<td>0.046494</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>18</td>
<td>0.93879</td>
<td>0.027992</td>
<td>0.032875</td>
<td>0.046983</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>19</td>
<td>0.93699</td>
<td>0.040727</td>
<td>0.033949</td>
<td>0.047506</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>20</td>
<td>0.93520</td>
<td>0.041471</td>
<td>0.034683</td>
<td>0.048058</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>21</td>
<td>0.93346</td>
<td>0.042161</td>
<td>0.035668</td>
<td>0.048634</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>22</td>
<td>0.93152</td>
<td>0.042796</td>
<td>0.036374</td>
<td>0.049223</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>23</td>
<td>0.92964</td>
<td>0.043385</td>
<td>0.036954</td>
<td>0.049816</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>24</td>
<td>0.92773</td>
<td>0.043922</td>
<td>0.037443</td>
<td>0.050402</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>25</td>
<td>0.92580</td>
<td>0.044411</td>
<td>0.037851</td>
<td>0.050972</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>
EXAMPLE 3: Modeling the Effect of an Intervening Effect

The following analyses were used in Section 4.3.2 to examine whether the effect of a myocardial infarction was a permanent constant effect or a time-varying effect. DASPLINE is a macro which generates formulas for dummy variables that allow fitting a restricted cubic spline function (Harrell et al., 1987). These dummy variables (MITIME1, MITIME2) are computed by the symbolic formula &_MITIME. A stepwise estimation procedure which allows all covariables to be entered into the model is used.

OPTIONS LS=60 NODATE;
TITLE 'MODELING THE EFFECT OF AN INTERVENING MYOCARDIAL INFARCTION';
DATA ALL; SET DB.MEDGRP(KEEP=D_TIME CDEATH INFARC DM_Time HSM);
   IF HSM=. THEN DELETE;
DATA MI; SET ALL;
   *DATA SET CREATED SO THAT KNOTS CAN BE ESTIMATED BY DASPLINE;
   IF INFARC EQ 1;
      MITIME=D_TIME-DM_TIME;
      KEEP MITIME;
DASPLINE MITIME NK=4 DATA=MI; * 4 KNOTS;
DATA ALL; SET ALL;
   PAT=MI;
   IF NOT( (INFARC EQ 1) AND (DM_TIME EQ 0) ) THEN DO;
      * A RECORD CORRESPONDING TO COVTIME=0 IS OUTPUT FOR ALL PATIENTS
      EXCEPT THOSE WHO EXPERIENCE AN MI AT TIME 0. THE FIRST RECORD
      FOR SUCH A PATIENT IS ASSOCIATED WITH MI=1 AND WILL BE OUTPUT
      BELOW;
      CT=0; MI=0; MITIME=0; MITIME1=0; MITIME2=0; OUTPUT;
   END;
   IF (INFARC EQ 1) THEN DO;
      * IF AN INFECTION OCCURS;
         MI=1;
         DIF=D_TIME-DM_TIME;
         DO MITIME=0 TO DIF BY 0.1;
         * STEP FUNCTION APPROXIMATION WITH RESOLUTION OF 0.1 YEARS;
         CT=CT+MITIME;
         W/MITIME;
         OUTPUT;
      END;
   END;
DROP DIF;
PROC PHSPLM DATA=ALL PCOV POOR MXN=4000 KNOTS=5;
   EVENT CDEATH;
   MODEL D_TIME-MI MITIME MITIME1 MITIME2/SW INCLUDE=1 SLE=1 SLS=1
   PRINTQ;
   COVTIME CT:
MODELING THE EFFECT OF AN INTERVENING MYOCARDIAL INFARCTION
STEPWISE SPLINE PROPORTIONAL HAZARD LINEAR MODEL PROCEDURE

DEPENDENT VARIABLE: D_TIME  Years
EVENT INDICATOR : CDEATH  Cardiovascular Death
COVARIATE TIME : CT

12059 OBSERVATIONS
586 UNCENSORED OBSERVATIONS
5081 SUBJECTS
0 OBSERVATIONS DELETED DUE TO MISSING VALUES

-2 LOG LIKELIHOOD FOR EXPONENTIAL MODEL  BETA=0 : 4598.24

MODEL: RESTRICTED CUBIC SPLINE HAZARD WITH KNOTS AT DEPENDENT VARIABLE VALUES:
.013669 .216278 .736981 3.15381 8.76527

-2 LOG LIKELIHOOD FOR SPLINE MODEL  BETA=0 : 4418.99
LIKELIHOOD RATIO TEST FOR EQ: HAZARD(T) = CONSTANT  BETA=0:
CHI-SQUARE= 179.25 WITH 4 D.F.  P=0.0000.

THE FOLLOWING VARIABLES WILL BE INCLUDED IN EVERY MODEL:
ALPHA1  ALPHA2  ALPHA3  ALPHA4  ALPHA5  MI

STEP 0. INCLUDED VARIABLES ENTERED.

MODEL CHI-SQUARE= 58.83 WITH 1 D.F.  (SCORE STAT.)  P=0.0000.
CONVERGENCE IN 6 ITERATIONS WITH 0 STEP HALVINGS  R= 0.100.
MAX ABSOLUTE DERIVATIVE=47E-8
MODEL CHI-SQUARE= 45.76 WITH 1 D.F.  (-2 LOG L.R.)  P=0.0000.

TEST FOR ALL VARIABLES NOT IN THE MODEL:
RESIDUAL CHI-SQUARE= 3.26 WITH 3 D.F.  P=0.3530

CHI-SQUARE Q STATISTICS ADJUSTED ONLY FOR VARIABLES IN THE MODEL
VARIABLE  CHI-SQUARE  P  R
MITIME  1.78  0.1825  0.000
MITIME1  0.86  0.3554  0.000
MITIME2  0.85  0.3617  0.000

STEP 1. VARIABLE MITIME ADDED WITH CHI-SQUARE= 1.78  P=0.1825  R= 0.000.
MODELING THE EFFECT OF AN INTERVENING MYOCARDIAL INFARCTION
STEPWISE SPLINE PROPORTIONAL HAZARD LINEAR MODEL PROCEDURE

DEPENDENT VARIABLE: D_TIME Years
EVENT INDICATOR: CDEATH Cardiovascular Death
COVARIATE TIME: CT

CONVERGENCE IN 4 ITERATIONS WITH 0 STEP HALVINGS R = 0.099.
MAX ABSOLUTE DERIVATIVE=74E-7 -2 LOG L= 4371.36.
MODEL CHI-SQUARE= 47.60 WITH 2 D.F. (-2 LOG L.R.) P=0.0000.

TEST FOR ALL VARIABLES NOT IN THE MODEL:
RESIDUAL CHI-SQUARE= 1.34 WITH 2 D.F. P=0.5118

CHI-SQUARE Q STATISTICS ADJUSTED ONLY FOR VARIABLES IN THE MODEL

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>CHI-SQUARE</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>MITIME1</td>
<td>0.18</td>
<td>0.6728</td>
</tr>
<tr>
<td>MITIME2</td>
<td>0.16</td>
<td>0.6847</td>
</tr>
</tbody>
</table>

STEP 2. VARIABLE MITIME1 ADDED WITH CHI-SQUARE= 0.18 P=0.6728 R= 0.000.

CONVERGENCE IN 4 ITERATIONS WITH 0 STEP HALVINGS R = 0.097.
MAX ABSOLUTE DERIVATIVE=45E-7 -2 LOG L= 4371.21.
MODEL CHI-SQUARE= 47.76 WITH 3 D.F. (-2 LOG L.R.) P=0.0000.

TEST FOR ALL VARIABLES NOT IN THE MODEL:
RESIDUAL CHI-SQUARE= 1.15 WITH 1 D.F. P=0.2829

CHI-SQUARE Q STATISTICS ADJUSTED ONLY FOR VARIABLES IN THE MODEL

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>CHI-SQUARE</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>MITIME2</td>
<td>1.15</td>
<td>0.2829</td>
</tr>
</tbody>
</table>

STEP 3. VARIABLE MITIME2 ADDED WITH CHI-SQUARE= 1.15 P=0.2829 R= 0.000.

CONVERGENCE IN 4 ITERATIONS WITH 0 STEP HALVINGS R = 0.096.
MAX ABSOLUTE DERIVATIVE=72E-6 -2 LOG L= 4370.03.
MODEL CHI-SQUARE= 46.93 WITH 4 D.F. (-2 LOG L.R.) P=0.0000.
MODELING THE EFFECT OF AN INTERVENING MYOCARDIAL INFARCTION
STEPWISE SPLINE PROPORTIONAL HAZARD LINEAR MODEL PROCEDURE

DEPENDENT VARIABLE: D_TIME  Years
EVENT INDICATOR : DETH  Cardiovascular Death
COVARIATE TIME  : CT

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>BETA</th>
<th>STD. ERROR</th>
<th>CHI-SQUARE</th>
<th>P</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALPHAL</td>
<td>0.47232559</td>
<td>0.0613767</td>
<td>59.22</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>ALPHAL2</td>
<td>-7.55801856</td>
<td>1.0895909</td>
<td>45.55</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>ALPHAL3</td>
<td>560.14584593</td>
<td>83.1098519</td>
<td>43.82</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>ALPHAL4</td>
<td>-767.05972246</td>
<td>116.0835768</td>
<td>43.66</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>ALPHAL5</td>
<td>217.55506023</td>
<td>33.1340020</td>
<td>43.11</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>MI</td>
<td>1.24484927</td>
<td>0.2299078</td>
<td>29.32</td>
<td>0.0000</td>
<td>0.079</td>
</tr>
<tr>
<td>MITIME1</td>
<td>-0.40896035</td>
<td>0.30236076</td>
<td>1.85</td>
<td>0.1762</td>
<td>0.000</td>
</tr>
<tr>
<td>MITIME2</td>
<td>0.11985030</td>
<td>0.1111601</td>
<td>1.16</td>
<td>0.2810</td>
<td>0.000</td>
</tr>
<tr>
<td>MITIME3</td>
<td>-0.12099619</td>
<td>0.1212890</td>
<td>1.15</td>
<td>0.2838</td>
<td>0.000</td>
</tr>
</tbody>
</table>

COVARIANCE MATRIX OF ESTIMATES

<table>
<thead>
<tr>
<th></th>
<th>ALPHAL</th>
<th>ALPHAL2</th>
<th>ALPHAL3</th>
<th>ALPHAL4</th>
<th>ALPHAL5</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI</td>
<td>0.000376105</td>
<td>-0.0649146</td>
<td>4.828925</td>
<td>-6.72857</td>
<td>1.90415</td>
</tr>
<tr>
<td>ALPHAL2</td>
<td>0.000947533</td>
<td>0.003828974</td>
<td>-0.000961513</td>
<td>0.00103635</td>
<td></td>
</tr>
<tr>
<td>ALPHAL3</td>
<td>0.0649146</td>
<td>-0.1267658</td>
<td>-90.0758</td>
<td>125.7019</td>
<td>-35.7538</td>
</tr>
<tr>
<td>ALPHAL4</td>
<td>4.828925</td>
<td>-90.0758</td>
<td>690.1286</td>
<td>-9647.198</td>
<td>2751.844</td>
</tr>
<tr>
<td>ALPHAL5</td>
<td>1.90415</td>
<td>-35.7538</td>
<td>2751.844</td>
<td>-3844.73</td>
<td>1097.862</td>
</tr>
<tr>
<td>MI</td>
<td>0.000947533</td>
<td>-0.0649146</td>
<td>4.828925</td>
<td>-6.72857</td>
<td>1.90415</td>
</tr>
<tr>
<td>MITIME1</td>
<td>0.05307976</td>
<td>-0.0555524</td>
<td>3.796828</td>
<td>-5.25294</td>
<td>1.451066</td>
</tr>
<tr>
<td>MITIME2</td>
<td>0.05307976</td>
<td>-0.0555524</td>
<td>3.796828</td>
<td>-5.25294</td>
<td>1.451066</td>
</tr>
</tbody>
</table>

CORRELATION MATRIX OF ESTIMATES

<table>
<thead>
<tr>
<th></th>
<th>ALPHAL</th>
<th>ALPHAL2</th>
<th>ALPHAL3</th>
<th>ALPHAL4</th>
<th>ALPHAL5</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI</td>
<td>1.000</td>
<td>-0.971</td>
<td>0.947</td>
<td>-0.944</td>
<td>0.936</td>
</tr>
<tr>
<td>ALPHAL2</td>
<td>0.067</td>
<td>0.206</td>
<td>-0.141</td>
<td>0.139</td>
<td>-0.991</td>
</tr>
<tr>
<td>ALPHAL3</td>
<td>-0.971</td>
<td>1.000</td>
<td>-0.995</td>
<td>0.994</td>
<td>-0.999</td>
</tr>
<tr>
<td>ALPHAL4</td>
<td>0.097</td>
<td>-0.169</td>
<td>0.110</td>
<td>-0.103</td>
<td>0.999</td>
</tr>
<tr>
<td>ALPHAL5</td>
<td>-0.944</td>
<td>-0.073</td>
<td>0.094</td>
<td>-0.089</td>
<td>1.000</td>
</tr>
<tr>
<td>MITIME</td>
<td>0.073</td>
<td>0.150</td>
<td>-0.085</td>
<td>0.084</td>
<td>1.000</td>
</tr>
<tr>
<td>MITIME1</td>
<td>0.936</td>
<td>-0.991</td>
<td>0.999</td>
<td>-1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>MITIME2</td>
<td>0.074</td>
<td>0.145</td>
<td>-0.085</td>
<td>0.084</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>HI</td>
<td>0.087</td>
<td>-0.087</td>
<td>0.073</td>
<td>-0.073</td>
<td>0.074</td>
</tr>
<tr>
<td>WITIME</td>
<td>1.000</td>
<td>-0.748</td>
<td>0.603</td>
<td>-0.597</td>
<td>0.145</td>
</tr>
<tr>
<td>WITIME1</td>
<td>-0.206</td>
<td>-0.169</td>
<td>0.131</td>
<td>-0.150</td>
<td>0.949</td>
</tr>
<tr>
<td>WITIME2</td>
<td>0.748</td>
<td>1.000</td>
<td>-0.953</td>
<td>0.091</td>
<td>-0.085</td>
</tr>
<tr>
<td></td>
<td>-0.141</td>
<td>0.110</td>
<td>-0.092</td>
<td>0.091</td>
<td>-0.085</td>
</tr>
<tr>
<td></td>
<td>0.603</td>
<td>-0.953</td>
<td>1.000</td>
<td>-1.000</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>0.139</td>
<td>-0.108</td>
<td>0.090</td>
<td>-0.089</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>-0.597</td>
<td>0.949</td>
<td>-1.000</td>
<td>1.000</td>
<td>0.084</td>
</tr>
</tbody>
</table>
EXAMPLE 4: Test of Proportionality using the interaction between a fixed
covariable and log t

The following code was used to model the interaction between pain index and log t. The
results of this analysis are reported in Section 4.5.2.

```
OPTIONS LS=80 NODATE;
TITLE 'TESTING THE PROPORTIONAL HAZARDS ASSUMPTION USING A STEP';
TITLES 'APPROXIMATION OF THE INTERACTION BETWEEN PI AND LOG T';
DATA ALL; SET DB.DMTIME(KEEP=DM_TIME CDEATHMI PI SAMPLE);
* DB.DMTIME IS A SAS DATA SET WHICH CONTAINS THE PAIN/ISCHEMIA
INDEX DATA. VARIABLE DESCRIPTIONS ARE:
DM_TIME = TIME OF MYOCARDIAL INFARCTION OR CARDIOVASCULAR DEATH
CDEATHMI = CENSORING INDICATOR (1 IF UNCESEORED, O OTHERWISE)
PI = PAIN/ISCHEMIA INDEX

IF SAMPLE=1; * OBTAINS RANDOM HALF OF DATA SET;
PAT=(_N_);
NINTERV=10; * NUMBER OF TIME INTERVALS;
IF DM_TIME=0 THEN DO;
* ONE RECORD IS OUTPUT WHEN FAILURE TIME IS 0;
PILOGT=PI*LOG(0.001);
CT=0;
OUTPUT;
END;
ELSE DO;
L=(DM_TIME-0.001)/(2*NINTERV); * HALF OF INTERVAL LENGTH;
DO I=0 TO (NINTERV-1); * DO OVER ALL INTERVALS;
IF I=0 THEN DO;
* RECORD CORRESPONDING TO FIRST INTERVAL IS OUTPUT;
CT=0; PILOGT=PI*LOG(0.001+L);
END;
ELSE DO;
* RECORDS CORRESPONDING TO OTHER INTERVALS ARE OUTPUT;
CT=0.001+2*I*L; PILOGT=PI*LOG(CT+L);
END;
OUTPUT;
END;
KEEP DM_TIME CT PI PILOGT PAT CDEATHMI;
PROC PHSLIP DATA=ALL PCOV PCOR MAIN=4000 KNOTS=5;
EVENT CDEATHMI;
MODEL DM_TIME=PI PILOGT/SW INCLUDE=1;
COVTIME CT;
```
The PHSPLM output follows.

TESTING THE PROPORTIONAL HAZARDS ASSUMPTION USING A STEP APPROXIMATION OF THE INTERACTION BETWEEN PI AND LOG T

STEPWISE SPLINE PROPORTIONAL HAZARD LINEAR MODEL PROCEDURE

DEPENDENT VARIABLE: DM_TIME  Years
EVENT INDICATOR : CDEATENI Cardiovascular Death or Infarction
COVARIATE TIME  : CT

29087 OBSERVATIONS
446 UNCENSORED OBSERVATIONS
2908 SUBJECTS
0 OBSERVATIONS DELETED DUE TO MISSING VALUES

-2 LOG LIKELIHOOD FOR EXPONENTIAL MODEL : BETA-0 : 3200.78

MODEL: RESTRICTED CUBIC SPLINE HAZARD WITH KNOTS AT DEPENDENT VARIABLE VALUES:
.043805 .298416 0.75 3.08842 8.33239

-2 LOG LIKELIHOOD FOR SPLINE MODEL : BETA-0 : 3056.48

LIKELIHOOD RATIO TEST FOR EQ: HAZARD(T) = CONSTANT : BETA-0:
CHI-SQUARE= 144.30 WITH 4 D.F.  P=0.0000.

THE FOLLOWING VARIABLES WILL BE INCLUDED IN EVERY MODEL:
ALPHA1  ALPHA2  ALPHA3  ALPHA4  ALPHA5  PI

STEP 0. INCLUDED VARIABLES ENTERED.

MODEL CHI-SQUARE= 141.85 WITH 1 D.F.  (SCORE STAT.) P=0.0000.
CONVERGENCE IN 6 ITERATIONS WITH 0 STEP HALVINGS  R= 0.191.
MAX ABSOLUTE DERIVATIVE=35E-9
-2 LOG L= 2943.31.
MODEL CHI-SQUARE= 113.17 WITH 1 D.F.  (-2 LOG L.R.) P=0.0000.

TEST FOR ALL VARIABLES NOT IN THE MODEL:
RESIDUAL CHI-SQUARE= 170.61 WITH 1 D.F.  P=0.0000

STEP 1. VARIABLE PILOT ADDED WITH CHI-SQUARE= 170.61 P=0.0000 R=-0.233.

CONVERGENCE IN 7 ITERATIONS WITH 1 STEP HALVINGS  R= 0.270.
MAX ABSOLUTE DERIVATIVE=10E-7
-2 LOG L= 2850.27.
MODEL CHI-SQUARE= 228.21 WITH 2 D.F.  (-2 LOG L.R.) P=0.0000.
TESTING THE PROPORTIONAL HAZARDS ASSUMPTION USING A STEP APPROXIMATION OF THE INTERACTION BETWEEN PI AND LOG T

STEPWISE SPLINE PROPORTIONAL HAZARD LINEAR MODEL PROCEDURE

DEPENDENT VARIABLE: DM_TIME Years
EVENT INDICATOR: CDEATHMI Cardiovascular Death or Infarction
COVARIATE TIME: CT

FINAL PARAMETER ESTIMATES

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>BETA</th>
<th>STD. ERROR</th>
<th>CHI-SQUARE</th>
<th>P</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALPHA1</td>
<td>0.22773606</td>
<td>0.0292528</td>
<td>60.61</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>ALPHA2</td>
<td>-2.0848375</td>
<td>0.5655783</td>
<td>13.59</td>
<td>0.0002</td>
<td></td>
</tr>
<tr>
<td>ALPHA3</td>
<td>114.4713216</td>
<td>34.7126002</td>
<td>10.87</td>
<td>0.0010</td>
<td></td>
</tr>
<tr>
<td>ALPHA4</td>
<td>-180.11049183</td>
<td>55.4013118</td>
<td>10.57</td>
<td>0.0011</td>
<td></td>
</tr>
<tr>
<td>ALPHA5</td>
<td>65.86249551</td>
<td>23.9639150</td>
<td>9.87</td>
<td>0.0017</td>
<td></td>
</tr>
<tr>
<td>PI</td>
<td>0.07592734</td>
<td>0.0100439</td>
<td>57.15</td>
<td>0.0000</td>
<td>0.134</td>
</tr>
<tr>
<td>PILOGT</td>
<td>-0.03438010</td>
<td>0.0027571</td>
<td>155.49</td>
<td>0.0000</td>
<td>-0.224</td>
</tr>
</tbody>
</table>

COVARIANCE MATRIX OF ESTIMATES

<table>
<thead>
<tr>
<th></th>
<th>ALPHA1</th>
<th>ALPHA2</th>
<th>ALPHA3</th>
<th>ALPHA4</th>
<th>ALPHA5</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI</td>
<td>0.0008557314</td>
<td>-0.0148507</td>
<td>0.8224678</td>
<td>-1.29617</td>
<td>0.4758269</td>
</tr>
<tr>
<td>PILOGT</td>
<td>-0.000555514</td>
<td>0.3198788</td>
<td>-19.2221</td>
<td>30.53546</td>
<td>-11.4138</td>
</tr>
<tr>
<td>ALPHA2</td>
<td>0.0004960894</td>
<td>-0.000610398</td>
<td>1204.965</td>
<td>-1922.66</td>
<td>3069.305</td>
</tr>
<tr>
<td>ALPHA3</td>
<td>0.8224678</td>
<td>-19.2221</td>
<td>1204.965</td>
<td>-1922.66</td>
<td>725.8613</td>
</tr>
<tr>
<td>ALPHA4</td>
<td>-0.0248277</td>
<td>0.03277442</td>
<td>-1922.66</td>
<td>3069.305</td>
<td>-1160.02</td>
</tr>
<tr>
<td>ALPHA5</td>
<td>0.4758269</td>
<td>-11.4138</td>
<td>725.8613</td>
<td>-1160.02</td>
<td>459.4857</td>
</tr>
<tr>
<td>PI</td>
<td>-0.000555514</td>
<td>0.01879676</td>
<td>-0.0248277</td>
<td>0.03877953</td>
<td>-0.0139434</td>
</tr>
<tr>
<td>PILOGT</td>
<td>0.000001008803</td>
<td>0.0001338137</td>
<td>0.03277442</td>
<td>-0.051513</td>
<td>0.01879676</td>
</tr>
</tbody>
</table>

CORRELATION MATRIX OF ESTIMATES

<table>
<thead>
<tr>
<th></th>
<th>ALPHA1</th>
<th>ALPHA2</th>
<th>ALPHA3</th>
<th>ALPHA4</th>
<th>ALPHA5</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI</td>
<td>1.000</td>
<td>-0.898</td>
<td>0.810</td>
<td>-0.800</td>
<td>0.776</td>
</tr>
<tr>
<td>ALPHA2</td>
<td>-0.898</td>
<td>1.000</td>
<td>-0.979</td>
<td>0.975</td>
<td>-0.963</td>
</tr>
<tr>
<td>ALPHA3</td>
<td>0.807</td>
<td>-0.979</td>
<td>1.000</td>
<td>-1.000</td>
<td>0.997</td>
</tr>
<tr>
<td>ALPHA4</td>
<td>0.810</td>
<td>-0.979</td>
<td>1.000</td>
<td>-1.000</td>
<td>0.999</td>
</tr>
<tr>
<td>ALPHA5</td>
<td>-0.070</td>
<td>0.342</td>
<td>-1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>PI</td>
<td>-0.066</td>
<td>0.325</td>
<td>-0.963</td>
<td>-0.999</td>
<td>-0.066</td>
</tr>
<tr>
<td>PILOGT</td>
<td>0.100</td>
<td>0.483</td>
<td>0.342</td>
<td>-0.337</td>
<td>0.325</td>
</tr>
</tbody>
</table>
Bibliography


