

Confounding and misrounding in age-period-cohort models



Thomas Michael Higgins
University of Leeds
School of Mathematics

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

March 2019

Intellectual Property and Publication Statements

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement. The right of Thomas Michael Higgins to be identified as Author of this work has been asserted by Thomas Michael Higgins in accordance with the Copyright, Designs and Patents Act 1988.

Acknowledgements

My PhD was an amazing experience. A special thank you goes to my supervisors – John T. Kent and John Paul Gosling – for their continuing academic support and for nurturing my enthusiasm for mathematics. I will be forever grateful. I would like to thank the Engineering and Physical Sciences Research Council (EPSRC) as my PhD would not have been possible without their financial support. Thank you to my family for their continuing encouragement and for believing I could get this far in education. I am thankful for the support from my brother - James. Words cannot express how grateful I am to my mum and dad – Michelle and Jimmy.

Abstract

A natural starting point for an age-period-cohort analysis is to assess the suitability of an independence model. If necessary, modifications are then made to the independence model to account for the effects of period and cohort. Data are usually made available for analysis in a two-way contingency table categorised in terms of rounded age and rounded period. The linear relationship $\text{age} = \text{period} - \text{cohort}$ does not hold exactly under rounding and the observation in a cell rounded age-by-period is not necessarily the same as an observation in a cell rounded age-by-cohort. In practice, independence models are discretised incorrectly such that the age-by-period data are used for model fitting as if the data are rounded age-by-cohort. The independence model is often deemed to be unsuitable as a description of the data and modifications are made to the independence model according to a proportional hazards assumption. We investigate whether the need for modifications is only apparent due to the misrounded treatment of the data. The case of Bovine Spongiform Encephalopathy is used as an illustrative example.

Contents

List of Tables	7
List of Figures	9
1 Introduction	14
2 Time concepts in age-period-cohort analysis	24
2.1 Discrete time and continuous time	26
2.2 Rounded time	29
2.3 Discussion	36
3 Models for the survival distribution	38
3.1 Survival distribution	40
3.2 Parametric models	44
3.3 Covariate models with age-independent covariates	51
3.4 Covariate models with age-dependent covariates	58
3.5 Conclusion	60

4	Age-period-cohort modelling	63
4.1	Survival setting in continuous time	65
4.2	Survival setting in discrete time	71
4.3	Survival setting in the literature	77
4.4	Regression setting	81
4.5	Conclusion	85
5	Confounding in age-period-cohort models	88
5.1	Confounding	90
5.2	Confounding for polynomial functions	92
5.3	Confounding for factor variables	97
5.4	Conclusion	101
6	Misrounding effects for age-period-cohort modelling	103
6.1	Two-way rounding	105
6.2	Misrounding	109
6.3	Under-reporting of deaths	112
6.4	Longer life expectancy	118
6.5	Lengthening of life expectancy	127
6.6	Conclusion	136
7	Misrounding effects in the case of BSE incidence	139
7.1	Model for BSE incidence	141

7.2	Exact treatment of data	147
7.3	Misrounding effects for BSE incidence	154
7.4	Implications for BSE incidence	162
7.5	Conclusion	168
8	Conclusion	171
9	Appendix	178
9.1	Glossary	178
9.2	Estimation of the age-period-cohort model	180
9.3	Estimation of the independence model	187
9.4	Modifying the independence model	195
	Bibliography	200

List of Tables

3.1	Constants used to define the probability function and hazard function in rounded time.	44
3.2	A collection of parametric model families in continuous time.	46
3.3	A collection of parametric model families in discrete time. Note that $\eta(a, N) = \sum_{w=0}^N \pi_{6,w} a^w$. The models are all defined on $a \in \mathbb{N}_0$ unless stated otherwise. Some model families have an upper limit on the age-at-death so that $A \in \{0, 1, \dots, M\}$	50
3.4	The validity of covariate models in continuous time. A model is valid if the range of the base model and covariate model are the same.	54
3.5	The validity of functional covariate models in discrete time. A model is valid if the range of the base model and covariate model are the same.	56
6.1	There are three ways, labelled by s for “surrogate”, in which we can use $M_{i,j}$ to derive a surrogate for $N_{i,k}$. The surrogate $N_{i,k}^{(s)}$ has an expectation denoted as $\nu_{i,k}^{(s)}$. We defined $N_{i,k}$ and $M_{i,j}$ in (6.4) and (6.6), respectively.	110

7.1	Each cell is an observation of $N_{i,k}^{(1)}$ for the incidence of BSE, where $N_{i,k}^{(1)}$ is a random variable for a count in the region $R_{i,i+k}^{(AP)}$ from Figure 6.2.	143
7.2	Each cell is an observation of $\frac{N_{i+1,k}^{(1)}}{N_{i,k}^{(1)}}$ for the incidence of BSE, where $N_{i,k}^{(1)}$ and $N_{i+1,k}^{(1)}$ are random variables for a count in the regions $R_{i,i+k}^{(AP)}$ and $R_{i+1,i+1+k}^{(AP)}$ respectively. These Lexis regions are depicted in Figure 6.2.	145
7.3	Each cell is an estimate of the age-by-period discretised Poisson intensity, $\nu_{i,k}^{(1)}$, for the independence model described in (7.8) and (7.9). The calculations are rounded to the nearest integer.	152
7.4	Each cell is an estimate of the Pearson residual $e_{i,k} = \frac{n_{i,k}^{(1)} - \nu_{i,k}^{(1)}}{\sqrt{\nu_{i,k}^{(1)}}}$ to one decimal place. The quantities $n_{i,k}^{(1)}$ and $\hat{\nu}_{i,k}^{(1)}$ are displayed in Tables 7.1 and 7.3, respectively.	153
7.5	Each cell is an estimate of the age-by-cohort discretised Poisson intensity, $\nu_{i,k}$, for the independence model described in (7.8) and (7.9). The calculations are rounded to the nearest integer.	156
7.6	Each cell is a calculation for the ratio of age-by-cohort discretised Poisson intensities, $\frac{\hat{\nu}_{i+1,k}}{\hat{\nu}_{i,k}}$, for the independence model described in (7.8) and (7.9). Our calculations of $\hat{\nu}_{i,k}$ are displayed in Table 7.5. The calculations are presented to three decimal places.	159
7.7	Each cell is an estimate of the ratio of age-by-period rounded Poisson intensities, $\frac{\hat{\nu}_{i+1,k}^{(1)}}{\hat{\nu}_{i,k}^{(1)}}$, for the independence model described in (7.8) and (7.9). Our calculations of $\hat{\nu}_{i,k}^{(1)}$ are displayed in Table 7.3. The calculations are presented to three decimal places.	159

List of Figures

2.1	A Lexis diagram in age-cohort space. The triangle labels depicted with letter T are stated only for a rounding down style. The variables A , P and C are assumed to have fractional parts so they can not be defined at any point on the network of lines.	33
2.2	A square in the Lexis diagram separated into eight triangles. The near identity is considered for each triangle under the three rounding styles.	35
4.1	A histogram for the observed age at deaths for individuals $j = 1, 2, \dots, 1000$, which are drawn from an exponential distribution with parameter 0.5. The two curved lines are the fitted density functions for cohorts $c = 1$ and $c = 4$ for the model defined in equation (4.7).	70
4.2	A bar plot for the observed age at deaths for individuals $j = 1, 2, \dots, 1000$, which are drawn from a geometric distribution with parameter 0.4. The two curved lines are the fitted probability mass functions for cohorts $c = 1$ and $c = 4$ for the model described in equation (4.17).	76

5.1	A Lexis diagram in age-cohort space for discrete time. Age, period and cohort are only defined at the vertices. Period is calculated as $p = a + c - 2$. The six filled vertices represent cells of a hypothetical age-by-period contingency table for deaths. For example, cell two corresponds to $a = 1$, $p = 2$ and $c = 3$. This display of age-by-period cells in age-cohort space is similar to the data in Table 7.1.	92
6.1	A Lexis diagram in age-cohort space with some highlighted age-by-cohort regions.	106
6.2	A Lexis diagram in age-cohort space with some highlighted age-by-period regions. In Section 6.3, we compare the number of deaths occurring in regions $R_{i,k}^{(AC)}$ and $R_{i,i+k}^{(AP)}$. In Section 6.4, we compare the ratio of deaths in regions $R_{i+1,i+k+1}^{(AP)}$ and $R_{i,i+k}^{(AP)}$ with the ratio of deaths in regions $R_{i+1,k}^{(AC)}$ and $R_{i,k}^{(AC)}$. In Section 6.5, we investigate how the ratio of deaths in regions $R_{i+1,i+k+1}^{(AP)}$ and $R_{i,i+k}^{(AP)}$ changes for unit increases in k	107
6.3	The calculation of $\frac{\nu_{i,k}^{(1)}}{\nu_{i,k}}$ for an exponential survival distribution $A \sim \text{Exponential}(\kappa_1 = 0.5)$ with $a \geq 0$ and a cohort intensity $\kappa(c) = [c - 2000]^2$ with $c \geq 2000$. The exponential family was described in Table 3.2. The survival distribution is independent of cohort and the cohort intensity is strictly increasing.	114
6.4	The calculation of the ratio $\frac{\nu_{i,k}^{(1)}}{\nu_{i,k}}$ for a Gamma survival distribution $A \sim \text{Gamma}(\pi_7 = 3, \kappa_7 = 20)$ with $a \geq 0$ and a cohort intensity $\kappa(c) = [c - 2000]^{-2}$ with $c \geq 2000$. The gamma family was described in Table 3.2. The survival distribution is independent of cohort and the cohort intensity is strictly decreasing. A horizontal line is drawn at the point the y-axis equals one.	115

- 6.5 The calculation of $\frac{\nu_{i+1,k}}{\nu_{i,k}} \div \frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}}$ for an exponential survival distribution $A \sim \text{Exponential}(\kappa_1 = 0.5)$ with $a \geq 0$ and a cohort intensity $\kappa(c) = [c - 2000]^2$ with $c \geq 2000$. The exponential family was described in Table 3.2 and the calculation of $\frac{\nu_{i,k}^{(1)}}{\nu_{i,k}}$ was illustrated in Figure 6.3. The survival distribution is independent and log linear, while the cohort intensity is strictly increasing. 122
- 6.6 The calculation of $\frac{\nu_{i+1,k}}{\nu_{i,k}} \div \frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}}$ for a Gamma survival distribution $A \sim \text{Gamma}(\pi_7 = 3, \kappa_7 = 20)$ with $a \geq 0$ and a cohort intensity $\kappa(c) = [c - 2000]^{-2}$ with $c \geq 2000$. The gamma family was described in Table 3.2 and the calculation of $\frac{\nu_{i,k}^{(1)}}{\nu_{i,k}}$ was illustrated in Figure 6.4. The survival distribution is independent and strictly log concave, while the cohort intensity is strictly decreasing. 123
- 6.7 The calculation of $\frac{\nu_{i,k}^{(1)}}{\sum_i \nu_{i,k}^{(1)}}$ for a Gamma survival distribution $A \sim \text{Gamma}(\pi_7 = 3, \kappa_7 = 20)$ with $a \geq 0$ and a cohort intensity $\kappa(c) = [c - 2000]^{-2}$ with $c \geq 2000$. The quantity $\frac{\nu_{i,k}^{(1)}}{\sum_i \nu_{i,k}^{(1)}}$ is a discrete representation of the survival distribution that is apparent under misrounding for $s = 1$. The solid line is equal to f_i^* . The survival distribution is log concave and the cohort intensity is log convex. 130
- 7.1 An estimate of the intensity for cattle born at time c which are eventually deemed to have BSE. The estimate is denoted as $\hat{\kappa}(c)$ and the estimate was obtained by maximising an exact likelihood function. The specific parametric formula for $\hat{\kappa}(c)$ is stated in (7.8). The dashed vertical line indicates the point in time for the introduction of the feed ban. 149
- 7.2 An estimate of the probability density function for the age at onset of BSE, given that a cow is eventually deemed to have BSE. The estimate $\hat{f}(a)$ is plotted for $a \in [0, 12)$ and the estimate was obtained by maximising a likelihood function. The specific parametric formula for $\hat{f}(a)$ is stated in (7.9). 150

- 7.3 The curved lines are calculations for the estimated ratio of Poisson intensities $\frac{\hat{\nu}_{i,k}^{(1)}}{\hat{\nu}_{i,k}}$ for particular cohorts. The values of $\hat{\nu}_{i,k}^{(1)}$ and $\hat{\nu}_{i,k}$ are presented in Tables 7.3 and 7.5. Each curved line consists of nine points which have been joined together with straight lines. The horizontal line drawn at the point the y-axis equals one indicates an idealistic value for the ratio of intensities. 157
- 7.4 The curved lines are calculations for the relative difference between ratios for age-by-cohort and age-by-period Poisson intensities, $\frac{\hat{\nu}_{i+1,k}}{\hat{\nu}_{i,k}} \div \frac{\hat{\nu}_{i+1,k}^{(1)}}{\hat{\nu}_{i,k}^{(1)}}$, for particular cohorts. The values of $\hat{\nu}_{i+1,k}$ and $\frac{\hat{\nu}_{i+1,k}^{(1)}}{\hat{\nu}_{i,k}^{(1)}}$ are presented in Tables 7.6 and 7.7. Each curved line consists of eight points which have been joined together with straight lines. The horizontal line drawn at the point the y-axis equals 1.0 indicates an idealistic value of the relative difference between ratios. The cohort intensity is strictly log linear and the survival distribution is strictly log concave. 161
- 7.5 A simulation of the ratio of counts, $\frac{n_{i,k}^{(1)}}{n_{i,k}}$, for particular cohorts. The expected value of the simulated count ratio is $\frac{\hat{\nu}_{i,k}^{(1)}}{\hat{\nu}_{i,k}}$ and its values are presented in Figure 7.3. 164
- 7.6 A simulation of the relative ratio, $\frac{n_{i+1,k}}{n_{i,k}} \div \frac{n_{i+1,k}^{(1)}}{n_{i,k}^{(1)}}$, for cohorts 1980, 1987 and 1991. The expected value of the simulated relative ratio is $\frac{\hat{\nu}_{i+1,k}}{\hat{\nu}_{i,k}} \div \frac{\hat{\nu}_{i+1,k}^{(1)}}{\hat{\nu}_{i,k}^{(1)}}$. So this plot can be viewed as introducing noise to Figure 7.4. 165
- 7.7 A calculation of $\frac{\nu_{i+1,1987}}{\nu_{i,1987}} \div \frac{\nu_{i+1,1987}^{(1)}}{\nu_{i,1987}^{(1)}}$ under a dependence model with $\kappa(c)$ and $f(a | c)$ specified according to equations (7.1) and (7.16), respectively. The values of the cross ratio are evaluated at the parameters $\xi_1 = 70, 248.730$, $\xi_2 = 0.515$, $\mu_A = 5.825$, $\sigma_A^2 = 1.572$. The three curves differ in their value of β 167

Abbreviations

APC = age-period-cohort

BSE = Bovine Spongiform Encephalopathy

PH = proportional hazards

DL = discrete logistic

CLL = complementary log-log

PO = proportional odds

AA = accelerated age-at-death

LLS = log location scale

MAFF = Ministry of Agriculture, Fisheries and Farming

CVL = Central Veterinary Laboratory

Chapter 1

Introduction

The thesis is a contribution to the field of age-period-cohort (APC) modelling. In this chapter, we compare our research objectives to the research objectives of earlier work on APC modelling.

The existing literature

The modelling of mortality has been a topic of commercial and general scientific interest for hundreds of years (Dickson et al. 2009, page 1). Since the early eighteenth century, life insurance companies have employed actuaries to provide a scientific framework for managing the companies' assets and liabilities. The number of deaths amongst the insured lives determines the liabilities of the company each year. A policyholder agrees to pay a series of premiums to the company until death, and then receives a predetermined lump sum at the time of death. If the policyholder dies much sooner than expected, then the life insurance company makes a loss on that policy. A systematic understating of longevity for all policyholders could lead to bankruptcy for the company. It is important for insurance companies to accurately predict life expectancy to

ensure policies are profitable.

It is important to consider some basic characteristics that influence human longevity such as smoking status and nationality, but, in particular, year of birth. On average, human longevity has increased over time such that any person can expect to live a longer life than their biological parents. For the United Kingdom (UK), the Human Mortality Database (HMD) website shows that the duration of human life was expected to be 57 years for a person born in 1922 and 81 years for a person born in 2016 (Shkolnikov et al. 2018 (accessed January 27, 2018)). Not accounting for the increasing longevity would lead to substantial losses for life insurance companies due to the understating of longevity.

The risk of mortality faced by a person at a particular moment in time can be attributed to a combination of three related time variables: (i) their time of birth, “cohort”; (ii) the calendar time, “period”; and (iii) their age. An age effect captures the mortality risk at a particular age which people experience regardless of their cohort, while a period effect captures the risk that impacts everyone in a particular period regardless of his or her age. Typically, the age effect for mortality is U-shaped due to complications at birth, followed by a period of stability, and an increasing risk after age 30 (Lawless 1982, page 11). The period effect for mortality is typically decreasing due to improvements in healthcare and in the education of healthy living.

Ryder (1997, page 68) described the members of a cohort as having a unique location in the stream of history because people from different cohorts experience different ages in different periods. A person born at time 2000 would be aged five in period 2005. This would lead people not to experience mortality risk in the same way. It would be necessary to consider age-period interaction effects if the risk of mortality at a certain age depends on the current period. For example, the likelihood of contracting a fatal disease at birth is likely to be determined by the quality of healthcare, which will vary with period. A cohort effect captures the accumulation of age-period interaction effects and reflects

a constant level of mortality risk that is faced uniformly throughout life.

Modelling simultaneously the effects of age, period and cohort on mortality risk was the original motivation for the subject of APC modelling. In a classic APC study, Kermack et al. (1934, pages 446–448) reported that age and cohort have a significant effect on the mortality risk in England, Scotland and Wales. Their estimate of mortality risk increases strongly and is strictly increasing with age. Mortality risk fluctuates up and down for people born between the late eighteenth century and the mid nineteenth century, and then decreases significantly with cohort thereafter. This decreasing cohort effect is consistent with increasing longevity.

While APC models were originally used in actuarial science to study all-cause human mortality, they have since been used to answer substantial questions in other subject areas. APC models are frequently used in the area of medicine to study disease diagnosis. For example, in another classic APC study, Greenberg et al. (1950) explored how age, period and cohort simultaneously impact on the risk of syphilis diagnosis for black females in North Carolina. The age effect for syphilis diagnosis was reported to be unimodal and peaked at age 18. More recently, Murayama et al. (2006) studied how the risk of pleural malignant mesothelioma diagnosis varies with age and cohort. Overall, the most frequent application of APC models is in medicine to study cause-specific human mortality such as mortality due to lung cancer (Kupper et al. 1985; Clayton and Schifflers 1987a).

APC modelling is not only relevant to the study of human survival. APC models have been used in sociology to study attendance at religious services and belief in the afterlife (Schwadel 2011; Hayward and Krause 2015). In particular, Schwadel (2011, page 187) reported that age, period and cohort have a significant effect on the probability of church attendance in the United States. The estimated probability of attendance increases strongly with age from approximately 0.35 at age 20-24 to around 0.52 at age 70-74. The attendance probability is slowly decreasing on average with period and cohort to suggest

a gradual decline in religious participation over time. Schomerus et al. (2015) conducted an APC analysis to study the desire for social distance from a person with schizophrenia. It is not only humans that are the subjects of APC studies. Dealler and Kent (1995) conducted an APC study to investigate the effects of age, period and cohort on the diagnosis of a fatal neurodegenerative disease in cattle called Bovine Spongiform Encephalopathy (BSE).

In general, the purpose of APC modelling is to separate the effects of age, period and cohort on some phenomenon. However, it is impossible to separate the three effects into distinct contributions because of a fundamental confounding problem. Johnston et al. (2018, page 1958) define confounding as the situation in which the relationship between two variables is distorted because of a strong relationship between one or two of the variables and a third variable included in the analysis. Since age, period and cohort are linearly determined by the equation $\text{cohort} = \text{period} - \text{age}$, the relationship between the age-at-death distribution and period is distorted by the inclusion of cohort in a survival model. The confounding concept can be extended to a situation with four variables to say that the relationship between age, period and a response variable is distorted by the inclusion of cohort in a non-survival model. Thus, cohort is confounded with age and period (Rodgers 1982, page 775).

APC models suffer from the most severe case of confounding, the issue of identifiability, in which the relationship between variables cannot be identified uniquely (Mason et al. 1973; Smith and Wakefield 2016). Note that, a lack of identifiability can result from the exact linear dependency between age, period and cohort as well as from over-parameterisation (Blalock 1966). In APC literature, confounding is synonymous with a lack of identifiability.

Much of the APC literature consists of attempts to overcome confounding. One approach is to set additional constraints such as setting parameters equal to zero or minimising a penalty function (Mason et al. 1973; Osmond and Gardner 1982). Another approach is to remove one linear term for age, period or cohort from the model formulation, or to only interpret parameters for

nonlinear terms in age, period and cohort such as the coefficients for quadratic and cubic terms (Holford 1983). Many authors decide to parameterise models with only two of age, period and cohort in order to avoid the identifiability issue (Kermack et al. 1934; Sasieni and Adams 2001; Murayama et al. 2006). The remainder of the APC literature mostly consists of statisticians fitting APC models to data using the proposed solutions. A summary on the approaches to deal with confounding/identifiability issues was recently published in Smith and Wakefield (2016).

One reason that research on APC modelling is still being produced is to create a methodology that will help statisticians to gain a meaningful insight into the distinct effects of age, period and cohort. Some major new textbooks on APC modelling have recently been published such as Yang and Land (2013) and O'Brien (2015). The purpose of both textbooks is to introduce and illustrate a new method to overcome the confounding issues. In particular, Yang and Land (2013, Chapter 5) introduce an intrinsic estimator which has been used in applied APC studies to estimate simultaneously the parameters for age, period and cohort (Schwadel 2011; Kramer et al. 2015). Glenn (1976) is critical about the proposed solutions to the APC confounding problem and describes attempts to solve confounding as futile.

The thesis

APC models are typically fitted to data that take one particular format. For survival studies, the age-at-death, period-at-death and cohort are real numbers, but are interval censored (Collett 2015, page 3) in that for recording purposes they are rounded down to the nearest integer. A summary of the database is made available for analysis as a two-way contingency table categorised in terms of age and period. There are a large number of cases in which APC models are fitted to age-by-period data and this practice has occurred throughout the

entire history of APC modelling (Kermack et al. 1934; Clayton and Schifflers 1987a; Choi et al. 2016). The data are similarly made available in a table rounded age-by-period for non-survival APC studies (Schomerus et al. 2015, page 20).

When age, period and cohort are rounded down to the nearest integer, the exact relationship $\text{cohort} = \text{period} - \text{age}$ does not hold exactly. Given the value of age and period, the cohort variable can be determined from the age-by-period table as either $\text{cohort} = \text{period} - \text{age}$ or $\text{cohort} = \text{period} - \text{age} - 1$. For example, under rounding, a person born in the year 2000 could be either be aged four or five in the year 2005. This means that each cell of the age-by-period table contains two distinct cohort integers (Yang and Land 2013, page 16). In the formulation of an APC model to study how some phenomenon varies with age, period and cohort, the cohort variable is usually assumed by statisticians to be determined exactly as the difference between period and age (Clayton and Schifflers 1987a, page 459; Murayama et al. 2006, page 4). It is therefore assumed that each cell of the age-by-period table corresponds to one cohort value.

A natural starting point for an APC analysis is to consider the fit of an independence model, which assumes, for example, that the mortality risk of a person depends on age but is independent of period and cohort so that human longevity is not changing over time. The independence model may be deemed not to fit the data well, but it is still the natural starting point. If necessary, modifications are then made to the independence model to account for the effects of period and cohort on mortality risk. It is important to discretise the independence model correctly to reflect the fact that the data are rounded age-by-period. However, we explain that, in practice, independence models are discretised incorrectly, or equivalently are “misrounded”, such that the age-by-period data are used for model fitting as if the data are rounded age-by-cohort. There are many examples of misrounding, but a prime example that we consider in detail in this thesis is the APC study conducted by Dealler and Kent (1995) for the incidence of BSE.

The independence model is often deemed to be unsuitable as a description of age, period and cohort effects apparent in the age-by-period data, and modifications to the independence model are usually made according to a proportional hazards (PH) assumption (Holford 1983; Liu et al. 2001; Kramer et al. 2015). This thesis investigates the misleading conclusions that can arise under certain circumstances when assessing the fit an independence model for a misrounded treatment of the age-by-period data as rounded age-by-cohort. We then assess whether the need for modifications to the independence model is only apparent due to the misrounded treatment of the data.

One misleading conclusion is that, when the number of births is strictly increasing over time, there appears to be a strict under-reporting for the number of deaths in each cell of the age-by-period table. When instead the number of births is strictly decreasing over time, there is apparent over-reporting of the death counts. Another misleading conclusion is that life expectancy can falsely appear to be changing over time when in reality life expectancy is not changing under the independence model. The final misleading conclusion is that there can appear to be a systematic increase in life expectancy over time when in reality life expectancy is not changing. If it is found to be statistically significant, this final misleading conclusion could have important implications for the study of human mortality. However, we note that the potential for misleading conclusions is greatest when the number of births is changing substantially over time. If the number of births is changing slowly, then the misleading conclusions will not be apparent and a misrounded treatment of the age-by-period data is sufficient when fitting the independence model.

BSE is a fatal neurodegenerative disease in cattle, which is perhaps better known as “mad cow disease”. BSE was first observed in cattle in the United Kingdom in 1986 and its incidence grew rapidly over the following years to cause considerable havoc in the cattle industry. Dealler and Kent (1995) analysed BSE incidence data rounded in terms of age and period to assess the suitability of the independence model and to investigate various modifications. They found large departures from independence and concluded that the inde-

pendence model was not suitable as a description of BSE incidence. The case of BSE is used as an illustrative example to assess how a misrounded treatment of age-by-period data can affect the results of an APC analysis.

We aim to conclude whether an exact treatment of the age-by-period data is necessary to assess the suitability of an independence model in the case of BSE incidence. One possible limitation of our statistical analysis is that we assume that the number of BSE cases follows a Poisson model, so that cases occur independently. But if BSE was passed on by infected cattle due to the close proximity with other cattle, then it would be necessary to consider a spatial dependence for BSE cases. Infected cattle feed is believed to be the major source of BSE transmission and it is believed that infectivity can also be passed on from mother to calf at birth (Dealler and Kent 1995, page 6). BSE is an interesting case study for a consideration of misrounding issues because the potential for misleading conclusions is most severe when the number of cases is rapidly changing over time.

In summary, our research objective is to investigate the potential for misleading conclusions when assessing the fit of an independence model for a misrounded treatment of data, and to assess whether the need for modifications to the independence model is only apparent due to the misleading conclusions. This research is highly relevant due to the common publication of data as rounded age-by-period coupled with the common mistreatment of the data as rounded age-by-cohort. The research objectives of the previous APC literature can perhaps be summarised as the proposition and critique of approaches to overcome the APC confounding problem, as well as the use of such approaches to gain a meaningful insight into the effects of age, period and cohort in real case studies. So the objectives of this thesis and of past literature are very different.

If the potential for misleading conclusions were found to be statistically significant, then our research would have important implications for many previous APC studies. In particular, if an independence model is suitable for

human mortality, then the apparent increasing longevity, which has been reported on numerous occasions (Kermack et al. 1934; Lee and Hsieh 1996), is only apparent due to misrounding. In any case study, the potential for misleading conclusions can be overcome by following one of two exact treatments of data: (i) fitting a correctly discretised independence model to the age-by-period data, or (ii) fitting the mis-discretised independence model to age-by-cohort data. The second exact treatment may not be possible in many circumstances because age-by-cohort data is rarely made available for analysis. In this thesis, we demonstrate how to carry out the first exact treatment in the case of BSE incidence.

In actuarial mathematics, the hazard rate or “force of mortality” at a certain age is estimated as a ratio with the total number of deaths at that age taken as the numerator and the total time exposed to risk at that age taken as the denominator (CT4 2012, unit 9 page 4). Deaths and exposure time are initially recorded by age and period, and then the total deaths and exposure time are obtained for each age by summing over all periods (CT4 2012, unit 10 pages 4–7). The second exact treatment of data would require deaths and exposure time to be initially recorded by age and cohort rather than by age and period. On the other hand, the first exact treatment would involve no change in the recording of deaths and exposure time.

In Chapter 2, the APC model is defined for continuous, discrete and rounded time as a Poisson model for deaths. The independence model is defined as a special case in which the age-at-death distribution does not vary with cohort. In Chapter 3, we investigate how modifications should be made to the independence model to incorporate effects of period and/or cohort when formulating an APC model. The method of handling dependence is different for continuous and discrete time.

In Chapter 4, we establish a new method to fit some APC models to data and explain that the way of handling dependence in the APC literature is not consistent with our findings. We explain that statisticians usually fit a

discrete independence model to describe rounded data and that modifications are usually made according to a PH assumption. An important distinction we make is between survival and regression as two settings for APC modelling. In Chapter 5, we provide an overview for the concept of confounding in APC models. It is important to consider the confounding issue when modifications are made to the independence model.

In Chapter 6, we explain that data are usually provided in a contingency table rounded age-by-period, but that the data are used for model fitting as if they are rounded age-by-cohort. We explain that this misrounded treatment of data can be problematic when assessing the fit of an independence model due to the potential for misleading conclusions. The need for modifications to the independence model might only be apparent due to the misrounded treatment of data. In Chapter 7, we assess the suitability of an independence model for the case of BSE incidence. We investigate whether an independence model provides a good description of the age-by-period data, and, hence, whether a misrounded treatment of data led Dealler and Kent (1995) to find dramatic departures from the independence model which should not have been apparent in reality.

Chapter 2

Time concepts in age-period-cohort analysis

In the context of survival analysis, an individual enters a system at an initial event and exits the system later on at a final event. Consider “birth” and “death” as unifying terms for the entry and exit events, respectively. The time of birth for an individual is referred to as their cohort, and the time at which an individual is observed after birth is the period. The age of an individual is determined as the time-gap between cohort and period:

$$\text{age} = \text{period} - \text{cohort}. \quad (2.1)$$

Examples of exit events in the literature are discussed in Chapter 4 such as death due to lung cancer (Peto et al. 1995) and diagnosis of cervix cancer (Sasieni and Adams 2001). The entry event is birth for almost all applied APC studies.

In this chapter, we define three different time concepts for survival analysis: discrete time, continuous time and a discretisation of continuous time. The discretisation of continuous time is usually done by integer rounding so that age, period and cohort are grouped into one-year intervals. We distinguish between three integer rounding styles: down, up and nearest. This third

time concept is a convenient presentation of continuous data and other time groupings could be considered such as five-year groupings. The key distinction is between discrete time and rounded time. The three time concepts have been compared in the survival literature (Lawless 1982, Chapters 2 and 3), but the comparison has not been set out explicitly in the APC literature.

The time concept is a major theme throughout the thesis because data for APC modelling are typically made available in rounded time. While the identity (2.1) holds exactly in continuous and discrete time, the identity holds only approximately in rounded time. A confounding issue defined in Chapter 5 arises when the identity holds true. A misrounding issue defined in Chapters 6 and 7 can arise when the identity does not hold true. We explore the extent to which the linear identity holds true for different time concepts and rounding styles so that we can assess the extent to which there are issues of confounding and misrounding. While textbooks on survival analysis work mostly in continuous time (Cox and Oakes 1984; Collett 2015), we give much attention to all three time concepts and relatively more attention to rounded time. Our focus on rounded time is unique to the APC literature.

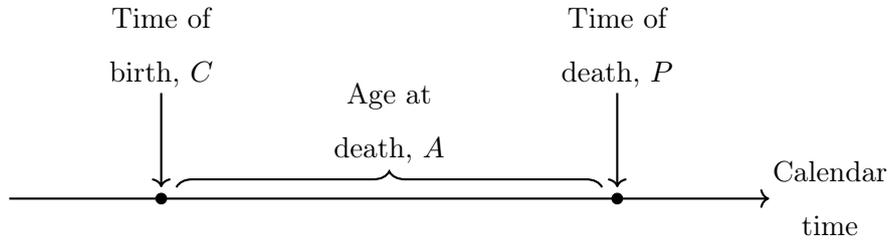
For each time concept, we define an APC model as a Poisson model for deaths indexed by age and cohort. The Poisson intensity is written as a product of a cohort intensity and a survival distribution, where the cohort intensity is the birth rate for a population and survival distribution is the probability distribution of the age-at-death for an individual. An independence model is defined as a special case of the APC model in which the survival distribution is assumed not to vary with period or cohort. In Chapter 3, we show how modifications should be made to the independence model to derive a formula for the APC model.

2.1 Discrete time and continuous time

In survival analysis, an individual has three main quantities of interest which we will treat as random: a cohort, C ; a period-at-death, P ; and an age-at-death, A . The three time variables are related by the linear identity

$$P = A + C. \quad (2.2)$$

In this thesis, we show that the linear identity makes APC modelling mathematically interesting.



Continuous time

In continuous time, the variables C and P are defined on the real numbers $\mathbb{R} = (-\infty, \infty)$ with $P > C$, so that $A \in (0, \infty) = \mathbb{R}^+$. Let $\lambda(a, c)$ denote the intensity of deaths for cohort $C = c$ at age a . The bivariate function, $\lambda(a, c) \in \mathbb{R}^+$, can be factorised into the product of a cohort intensity $\kappa(c) \in \mathbb{R}^+$ and a probability density function

$$f(a | c) = \lim_{\delta \rightarrow 0^+} \Pr(A \in [a, a + \delta) | C = c) \cdot \frac{1}{\delta} \in \mathbb{R}^+.$$

The cohort intensity is the intensity of individuals born in cohort c and can be written in terms of $\lambda(a, c)$ as $\kappa(c) = \int_0^\infty \lambda(a, c) da$. The probability density function can be written in terms of $\lambda(a, c)$ as

$$f(a | c) = \frac{\lambda(a, c)}{\kappa(c)} = \frac{\lambda(a, c)}{\int_0^\infty \lambda(a, c) da}. \quad (2.3)$$

Definition (2.3) implies that $\int_0^\infty f(a | c) da = 1$ for any $c \in \mathbb{R}$. The function $\lambda(a, c)$ characterises $f(a | c)$ because for a fixed c value, $f(a | c) \propto \lambda(a, c)$. We choose to omit the period label p in $\lambda(a, c)$ because the intensity of deaths indexed in terms of a , p and c can be equivalently indexed in terms of a and c due to equation (2.2).

We assume that deaths form a Poisson process so that the number of deaths in age interval (a_1, a_2) for a group of individuals born in cohort interval (c_1, c_2) is

$$N^{\text{death}}((a_1, a_2), (c_1, c_2)) \sim \text{Poisson} \left(\int_{a_1}^{a_2} \int_{c_1}^{c_2} \lambda(a, c) dc da \right) \quad (2.4)$$

where

$$\lambda(a, c) = \kappa(c) \cdot f(a | c). \quad (2.5)$$

The probability that there are exactly x deaths in the age-cohort space consisting of ages (a_1, a_2) and cohorts (c_1, c_2) can then be written as

$$\frac{1}{x!} \left(\int_{a_1}^{a_2} \int_{c_1}^{c_2} \lambda(a, c) dc da \right)^x \exp \left(- \int_{a_1}^{a_2} \int_{c_1}^{c_2} \lambda(a, c) dc da \right).$$

Alternatively, the probability of exactly x deaths occurring at ages (a_1, a_2) for cohorts (c_1, c_2) can be written under a Binomial model as $\binom{n}{x} p^x (1-p)^{n-x}$, where n is the number of births in cohort interval (c_1, c_2) and

$$p = \frac{1}{n} \int_{a_1}^{a_2} \int_{c_1}^{c_2} \lambda(a, c) dc da.$$

The Poisson model is a close approximation to a Binomial model for a small x and a small p (Lipschutz and Schiller 1998, page 191).

The Poisson model assumes that deaths occur independently so that the death of one individual does not influence the death of another individual. A Poisson model would not be appropriate if the death of one individual causes the death of another individual. In Chapter 7, we use a Poisson to describe the incidence of Bovine Spongiform Encephalopathy (BSE) in cattle. Infected cattle feed is believed to be the major source of BSE transmission. However, if BSE was passed on by infected cattle due to the close proximity with other

cattle, then it would be necessary to consider a spatial dependence for BSE cases and a Poisson model would not be suitable. A Poisson model is often used to describe deaths in the APC literature (Clayton and Schiffers 1987a; Lee and Hsieh 1996). Some other models for deaths used in applied APC studies include the Neyman Type A model (Barrett 1973) and the Negative Binomial model (Jean et al. 2013). In this thesis, we only use the Poisson model.

Discrete time

In discrete time, the variables C and P are defined on the set of integers $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ with $P \geq C$, so that $A \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$. Let $\lambda_{a,c}$ denote the expected number of deaths at age a for a group of individuals born in cohort c . The bivariate function, $\lambda_{a,c} \in \mathbb{R}^+$, can be factorised into the product of a cohort intensity $\kappa_c \in \mathbb{R}^+$ and a probability mass function

$$f_{a|c} = \Pr(A = a \mid C = c) \in (0, 1).$$

The cohort intensity is the expected number of individuals born in cohort c and can be written in terms of $\lambda_{a,c}$ as $\kappa_c = \sum_{a=0}^{\infty} \lambda_{a,c}$. The probability mass function can be written in terms of $\lambda_{a,c}$ as

$$f_{a|c} = \frac{\lambda_{a,c}}{\kappa_c} = \frac{\lambda_{a,c}}{\sum_{a=0}^{\infty} \lambda_{a,c}}. \quad (2.6)$$

Definition (2.6) implies that $\sum_{a=0}^{\infty} f_{a|c} = 1$ for any $c \in \mathbb{Z}$. The function $\lambda_{a,c}$ characterises $f_{a|c}$ because for a fixed c value, $f_{a|c} \propto \lambda_{a,c}$.

We assume that deaths form a Poisson process so that the number of deaths at age a for a group of individuals born in cohort c is

$$N_{a,c}^{\text{death}} \sim \text{Poisson}(\lambda_{a,c} = \kappa_c \cdot f_{a|c}). \quad (2.7)$$

The probability of exactly x deaths occurring at age a for cohort c is

$$\Pr(N_{a,c}^{\text{death}} = x) = \frac{(\lambda_{a,c})^x \exp(-\lambda_{a,c})}{x!}.$$

Alternatively, the probability of exactly x deaths occurring at age a for cohort c can be written under a Binomial model as $\binom{n}{x}p^x(1-p)^{n-x}$, where n is the number of births in cohort c and $p = \frac{\lambda_{a,c}}{n}$. Similar to continuous time, the Poisson model is a close approximation to a Binomial model for a small x and a small p .

When time is discrete, there is no concept of fractional time such as months or days. Deaths are counted at time instances rather than in time intervals. In discrete time, we allow for $A = 0$ so that an individual can die at birth. We explain in Chapter 4 that applied APC studies are mostly concerned with modelling $f_{a|c}$. The models (2.4) and (2.7) are APC models. In this thesis, we refer to $\lambda(a, c) = \kappa(c) \cdot f(a)$ and $\lambda_{a,c} = \kappa_c \cdot f_a$ as independence models because they do not allow the survival distribution to depend on period and cohort. In Chapter 3, we explore how modifications should be made to the independence models to allow for departures from independence due to period and cohort.

2.2 Rounded time

The underlying time concept can either be continuous or discrete. The time concept is typically continuous for applied APC studies, however, there is a third time concept that is created when the continuous variables A , P and C are discretised (Kermack et al. 1934; Clayton and Schiffers 1987a). A discretisation of time means that continuous time is treated as if it were discrete time. Time can be discretised into one-year intervals by integer rounding and we work exclusively with integer rounding in this thesis.

Rounding styles and near identities

Suppose time is continuous so that A , P and C all have a fractional part and let

$$I = [A], J = [P], K = [C]. \quad (2.8)$$

The square brackets $[\cdot]$ are a shorthand for either rounding down to the nearest integer, rounding up to the nearest integer, or rounding to the nearest integer. Taking the age-at-death as an example, the floor and ceiling functions can be defined as

$$\begin{aligned} \lfloor A \rfloor &= \max\{z \in \mathbb{Z} \mid a > z\} \\ \lceil A \rceil &= \min\{z \in \mathbb{Z} \mid a < z\}. \end{aligned}$$

In this thesis, we consider three styles of integer rounding:

1. Rounding down - The square brackets are replaced with the floor function $\lfloor \cdot \rfloor$;
2. Rounding up - The square brackets are replaced with the ceiling function $\lceil \cdot \rceil$;
3. Rounding nearest - Consider the A variable as an example and let

$$d(A) = A - \lfloor A \rfloor \in (0, 1)$$

denote the fractional part of A . We assume that $d(A)$ cannot take a value of zero. We round down to the nearest integer for $d(A) < 0.5$ and round up to the nearest integer for $d(A) \geq 0.5$. In summary, we replace $[A]$ with $\lfloor A + 0.5 \rfloor$ which is because $\lfloor A + 0.5 \rfloor = \lfloor A \rfloor$ for $d(A) < 0.5$ and $\lfloor A + 0.5 \rfloor = \lceil A \rceil$ for $d(A) \geq 0.5$. The same logic can be applied to the variables P and C . Note that, for nearest rounding we could not replace $[A]$ with $\lceil A - 0.5 \rceil$ because $\lceil A - 0.5 \rceil = \lfloor A \rfloor$ for $d(A) \leq 0.5$ and $\lceil A - 0.5 \rceil = \lceil A \rceil$ for $d(A) > 0.5$.

The variable I is defined on the set \mathbb{N}_0 for down and nearest rounding, and is defined on \mathbb{N} for rounding up. Note that, for rounding up, we can also write I in terms of the floor function as $\lfloor A \rfloor + 1$. The variables J and K can be written similarly for the three rounding styles.

Our work on integer rounding in this thesis can be extended for a general discretisation of A , P and C over θ -year intervals. For example, we can discretise A by using the following formula:

$$I = \theta \left(\left\lfloor \frac{A}{\theta} + \frac{z_1}{2} \right\rfloor + z_2 \right); \quad (2.9)$$

where $z_1 = z_2 = 0$ for a downward discretisation, $z_1 = 0$ and $z_2 = 1$ for an upward discretisation, and $z_1 = 1$ and $z_2 = 0$ for a nearest discretisation. For example, a value of $I = 50$ for a downward discretisation with $\theta = 5$ would correspond to a discretisation of $A \in (50, 55)$. The variable I would be defined on the set $\theta \cdot \mathbb{N}_0$ for a downward or nearest discretisation, and would be defined on $\theta \cdot \mathbb{N}$ for an upward discretisation. Substituting $\theta = 1$ into equation (2.9) would produce the definition of I for the three rounding styles.

In this section, we explain that the basic continuous identity $P = A + C$ from (2.2) transforms under a rounded discretisation into the “near identities”:

$$J = I + K, \quad (2.10)$$

$$J = I + K + 1, \quad (2.11)$$

$$J = I + K - 1. \quad (2.12)$$

It is widely known that (2.10) and (2.11) can hold for rounding down, however, these two near identities are rarely stated explicitly. We prove in Case 1 that these two near identities hold true and the proof requires us to consider the fractional parts of A and C . We do not let $d(A) + d(C)$ take values of zero or one because P cannot take an integer value. The rule of A , P and C not taking integer values is a consequence of time being defined on the real numbers.

The following rule for the near identities is true with respect to the three rounding styles:

- **Case 1** - When rounding down to the nearest integer,

$$J - I - K = \begin{cases} 0, & \text{if } d(A) + d(C) \in (0, 1) \\ 1, & \text{if } d(A) + d(C) \in (1, 2). \end{cases}$$

- **Case 2** - When rounding up to the nearest integer,

$$J - I - K = \begin{cases} -1, & \text{if } d(A) + d(C) \in (0, 1) \\ 0, & \text{if } d(A) + d(C) \in (1, 2). \end{cases}$$

- **Case 3** - When rounding to the nearest integer,

$$J - I - K = \begin{cases} +1, & \text{if } d(A), d(C) \in (0, 0.5) \text{ and } d(A) + d(C) \in [0.5, 1) \\ -1, & \text{if } d(A), d(C) \in [0.5, 1) \text{ and } d(A) + d(C) \in (1, 1.5) \\ 0, & \text{otherwise.} \end{cases}$$

Near identities in the Lexis diagram

A Lexis diagram is a two-dimensional graph with three axes, where each axis represents values of either age, period or cohort. There is a horizontal axis, a vertical axis and a diagonal axis. There are many possible ways to display the Lexis diagram and the direction of the diagonal axis depends on the choice of display. In this thesis, we only display the diagram in age-cohort space as in Figure 2.1, with age on the horizontal axis, cohort on the vertical axis, and period on a positive-diagonal axis. Positive diagonal means that period increases in a north-east direction.

The Lexis diagram is named after the German statistician Wilhelm Lexis for his work in Lexis (1875), which is written in the German language. Vandeschrick (2001, Chapter 2) provides a detailed description of the diagram in the English language. The name Lexis diagram might be a misnomer because Vandeschrick (2001, Chapters 3 and 4) explains that there are other researchers

including Gustav Zeuner and Otto Brasche who are more worthy than Wilhelm Lexis of being attributed as the inventor of the diagram.

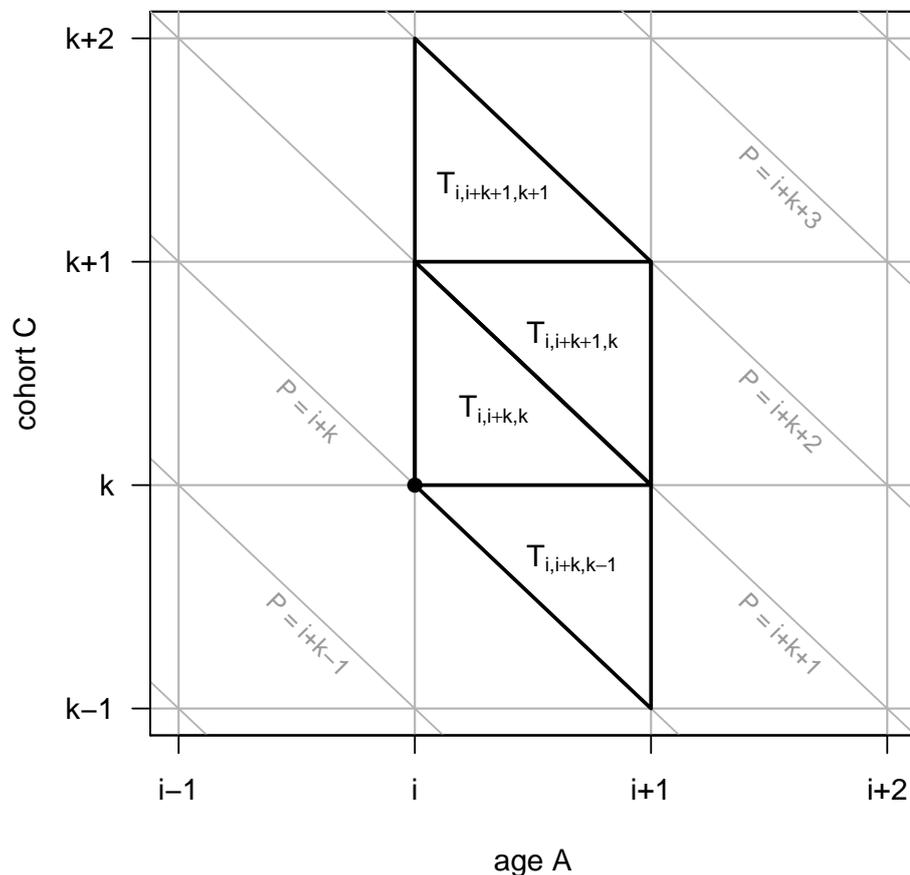


Figure 2.1: A Lexis diagram in age-cohort space. The triangle labels depicted with letter T are stated only for a rounding down style. The variables A , P and C are assumed to have fractional parts so they can not be defined at any point on the network of lines.

The network of lines drawn perpendicular to each axis produces a triangular grid with each triangle corresponding to a three-way rounding of age, period and cohort. Each triangle in the grid corresponds to one particular near identity when adopting the styles of rounding down or rounding up. For rounding nearest, however, each triangle in the grid does not correspond to one particular near identity and we must divide each triangle in the grid into

four smaller triangles to produce Lexis regions that each correspond to a near identity. Thus, we divide a square in the Lexis diagram into eight parts as in Figure 2.2.

Let T label a triangle in a Lexis diagram for a three-way rounding of age, period and cohort, but not necessarily outlined by the triangular grid, so that

$$T_{i,j,k} = \{(A, P, C) : I = i, J = j, K = k\}. \quad (2.13)$$

For any rounding style, the number of deaths in the region $T_{i,j,k}$ is derived as the following discretisation of the continuous model from (2.4):

$$N_{i,j,k}^{\text{death}} \sim \text{Poisson}(\lambda_{i,j,k}) \quad (2.14)$$

where

$$\lambda_{i,j,k} = \iint_{T_{i,j,k}} \lambda(a, c) da dc.$$

The triangle labels in Figure 2.1 are only true for a rounding down style. The triangle labelled by $T_{i,i+k,k}$ is enclosed vertically by ages $(i, i + 1)$, horizontally by cohorts $(k, k + 1)$ and diagonally by periods $(i + k, i + k + 1)$. If the floor function is applied to all values in these three time intervals, then this triangle is deduced to correspond to age $I = i$, period $J = i + k$ and cohort $K = k$. The Lexis diagram labels for a rounding up style are easy to deduce given the labels for a rounding down style. For rounding up, the triangle $T_{i,j,k}$ would be enclosed by ages $(i - 1, i)$, cohorts $(k - 1, k)$ and periods $(j - 1, j)$. Hence, a triangle labelled as $T_{i,j,k}$ in the Lexis diagram for a rounding down style would instead be labelled as $T_{i+1,j+1,k+1}$ for a rounding up style.

Suppose we classify each element of the triangular grid as either a lower triangle or an upper triangle:

$$T_{i,k}^{(\text{L})} = \{(A, P, C) : A \in (i, i + 1), P \in (i + k, i + k + 1), C \in (k, k + 1)\}, \quad (2.15)$$

$$T_{i,k}^{(\text{U})} = \{(A, P, C) : A \in (i, i + 1), P \in (i + k + 1, i + k + 2), C \in (k, k + 1)\}. \quad (2.16)$$

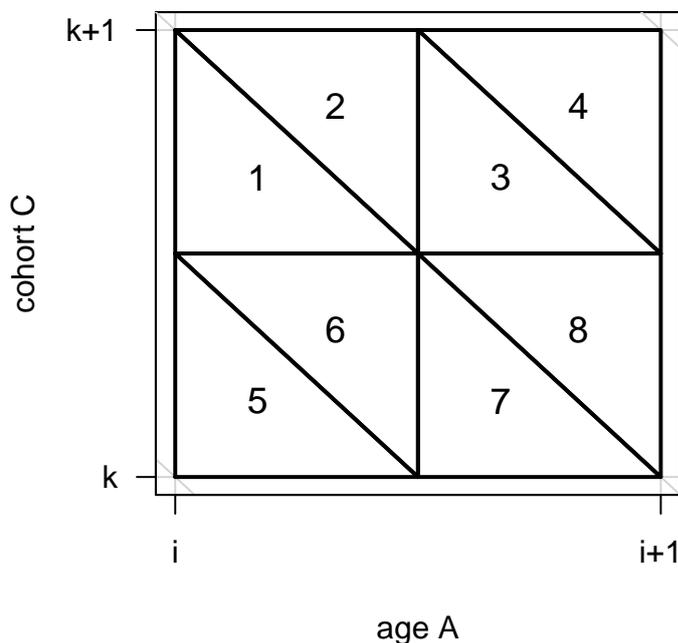


Figure 2.2: A square in the Lexis diagram separated into eight triangles. The near identity is considered for each triangle under the three rounding styles.

In Figure 2.2, a lower triangle is a collection of the triangles labelled numerically as 1, 5, 6 and 7. An upper triangle is a collection of the triangles labelled numerically as 2, 3, 4 and 8. This classification is mostly helpful when discussing near identities for the styles of rounding down and rounding up, and becomes helpful in Chapter 6 when distinguishing between Lexis squares and Lexis parallelograms. For rounding down, $T_{i,k}^{(L)} = T_{i,i+k,k}$ and $T_{i,k}^{(U)} = T_{i,i+k+1,k}$. But for rounding up, $T_{i,k}^{(L)} = T_{i+1,i+k+1,k+1}$ and $T_{i,k}^{(U)} = T_{i+1,i+k+2,k+1}$. When rounding down, lower triangles satisfy (2.10) and upper triangles satisfy (2.11). When rounding up, lower triangles satisfy (2.12) and upper triangles satisfy (2.10).

For nearest rounding, there are two near identities that hold true in each element of the triangular grid. Further, all three near identities can hold for nearest rounding compared to the two identities for rounding down and up. The near identities (2.10) and (2.11) hold for lower triangles, whereas the identities (2.10) and (2.12) hold for upper triangles. Specifically, in Figure

2.2, the first near identity (2.10), which is analogous to the continuous linear identity, holds for the triangles 1, 2, 4, 5, 7 and 8. To illustrate how to obtain a particular identity for a chosen triangle, consider the following example. The fractional parts of A , C and P in triangle 6 are $d(A) \in (0, 0.5)$, $d(C) \in (0, 0.5)$ and $d(P) = d(A) + d(C) \in [0.5, 1)$. When applying nearest rounding to triangle 6, we derive

$$I = \lfloor A \rfloor, K = \lfloor C \rfloor, J = \lceil P \rceil. \quad (2.17)$$

Using the fractional part expansion

$$A + C = (\lfloor A \rfloor + \lfloor C \rfloor) + d(A) + d(C),$$

the interval for P can be written as

$$A + C \in [0.5 + \lfloor A \rfloor + \lfloor C \rfloor, 1 + \lfloor A \rfloor + \lfloor C \rfloor). \quad (2.18)$$

Applying the ceiling function to each value in the interval (2.18) leads to

$$\lceil A + C \rceil = 1 + \lfloor A \rfloor + \lfloor C \rfloor. \quad (2.19)$$

For the variables stated in (2.17), the equation (2.19) implies that

$$J - I - K = 1 + \lfloor A \rfloor + \lfloor C \rfloor - \lfloor A \rfloor - \lfloor C \rfloor = 1.$$

2.3 Discussion

In this chapter, the APC model was defined for continuous, discrete and rounded time as a Poisson model for deaths. Other models for deaths can be considered if the deaths of individuals do not occur independently. The Poisson intensity of the APC model, which describes the intensity or expected quantity of deaths at a particular point in the age-cohort space, is written in continuous time as $\lambda(a, c) = \kappa(c) \cdot f(a | c)$ and in discrete time as $\lambda_{a,c} = \kappa_c \cdot f_{a|c}$. In rounded time, the Poisson intensity is a discretisation of $\lambda(a, c)$ over a region

in the Lexis diagram. In this thesis, the term APC model will be synonymous with the Poisson intensities $\lambda(a, c)$ and $\lambda_{a,c}$. When the survival distribution is independent of period and cohort, so that $\lambda(a, c) = \kappa(c) \cdot f(a)$ and $\lambda_{a,c} = \kappa_c \cdot f_a$, we refer to the Poisson intensities as independence models. In Chapter 3, we investigate how modifications should be made to $f(a)$ and f_a to handle dependence on period and cohort.

In Chapter 5, we show that there are confounding problems when the linear identity holds true. In this chapter, we showed that the linear identity $\text{age} = \text{period} - \text{cohort}$ holds true in continuous time and discrete time. The linear identity does not always hold true in rounded time, and, instead, three non-identities arise in rounded time. In particular, the value of $\text{period} - \text{age} - \text{cohort}$ can be equal to $+1$, -1 or 0 when age, period and cohort are rounded to the nearest integer. The distinction between discrete time and rounded time is very important in this thesis. In Chapter 6, we explain that data are typically made available for analysis in rounded time with age, period and cohort rounded down to the nearest integer. The Lexis diagram is helpful in illustrating the rounding of the data. We also show that there can be a substantial loss in accuracy for a statistical analysis when fitting discrete independence models to rounded data. The loss in accuracy is only explored for rounding down.

Chapter 3

Models for the survival distribution

In continuous time, the Poisson intensity of an APC model was described in equation (2.5) as equal to $\lambda(a, c) = \kappa(c) \cdot f(a | c)$. In discrete time, the Poisson intensity of an APC model was described in equation (2.7) as equal to $\lambda_{a,c} = \kappa_c \cdot f_{a|c}$. The Poisson intensity is a product of a cohort intensity and a survival distribution. The independence models $\lambda(a, c) = \kappa(c) \cdot f(a)$ and $\lambda_{a,c} = \kappa_c \cdot f_a$ are a special case of the APC model. In this chapter, we investigate how modifications should be made to $f(a)$ and f_a to derive a suitable parameterisation for $f(a | c)$ and $f_{a|c}$. In Section 3.1, we list five different ways to express the survival distribution and this list is standard in textbooks on survival analysis (Lawless 1982, Chapter 1). The survival distribution of an individual can depend on their covariates such as gender and smoking status, but, we focus particularly on period and cohort as covariates.

Covariate models take the survival distribution for a baseline group of individuals with the same covariates and describe a general survival distribution that is defined through departures from the baseline distribution due to difference in covariates. The functions $f(a)$ and f_a are viewed as the baseline survival distribution. A covariate is a modification of $f(a)$ and f_a to incor-

porate the effects of covariates. Cox and Oakes (1984, Chapter 5) present a list of covariate models such as PH models and accelerated failure time (AFT) models. In Section 3.3, we assess the suitability of various covariate models in continuous and discrete time for a situation in which covariates do not vary with age. The baseline survival distribution can be parametric so we review parametric survival models for continuous and discrete time in Section 3.2. Survival textbooks generally review parametric models in continuous time such as Kalbfleisch and Prentice (2002, Chapter 2), but, it is not standard to review parametric models in discrete time. We discuss parametric models in continuous, discrete and rounded time.

A covariate can be either a function of age or a constant value for each individual. Period $p = c + a$ is an example of an age-dependent covariate since period is a function of age, whereas cohort is age-independent since cohort stays constant throughout an individual's life. A distinction between age-independent and age-dependent covariates is made by Cox and Oakes (1984, Chapter 8). In Section 3.4, we explore the suitability of various covariate models in continuous and discrete time for a situation in which some covariates vary with age. We find that only covariate models for the hazard function are suitable under age-dependent covariates and that some of the covariate models that were suitable under age-independent covariates are no longer suitable. Further, only the PH model is suitable when modifying the independence model in continuous time, while only the complementary log-log (CLL) and discrete logistic (DL) models are suitable when modifying the independence model in discrete time. In Chapter 4, we explain that statisticians usually make modifications to the independence model by using a PH model.

3.1 Survival distribution

In Chapter 2, we defined A as the age-at-death for an individual and referred to the probability distribution of A as the survival distribution. The survival distribution was represented by a probability mass function in discrete time and a probability density function in continuous time. In this section, the survival distribution is defined as a function of age independently of covariates for all three time concepts: discrete, continuous and rounded. Covariates are introduced in Section 3.3 to allow the survival distribution to depend on covariates such as period and cohort.

Continuous time

In continuous time, the survival distribution has probability mass on the interval \mathbb{R}^+ and can be defined by a cumulative distribution function

$$F(a) = \Pr(A \leq a),$$

such that (i) $F(a)$ is continuous, (ii) $F(a)$ is strictly increasing, and (iii) $\lim_{a \rightarrow \infty} F(a) = 1$. Other ways to describe the survival distribution are:

- survivor function - $S(a) = \Pr(A > a) = \Pr(A \geq a) = 1 - F(a)$
- probability function - $f(a) = F'(a) = \lim_{\delta a \rightarrow 0^+} \frac{\Pr(a \leq A < a + \delta a)}{\delta a}$
- hazard function - $h(a) = \lim_{\delta a \rightarrow 0^+} \frac{\Pr(a \leq A < a + \delta a | A \geq a)}{\delta a}$
- cumulative hazard function - $H(a) = \int_0^a h(u) du$.

The hazard function and probability function always satisfy the inequality $0 < f(a) < h(a) < \infty$. The cumulative distribution function is strictly increasing because we assume that $f(a) > 0$ for all a .

We say probability function as a shorthand for both probability density and probability mass functions. In continuous time, $\Pr(A > a) = \Pr(A \geq a)$ because $\Pr(A = a) = 0$ for all $a \in \mathbb{R}^+$. It is standard for a textbook on survival modelling to mention all five of these expressions for the survival distribution, for example, see Lawless (1982, Chapter 1).

The hazard function captures the risk of death for an individual at a certain time in life conditional on the individual surviving to that particular time. The probability that an event, B , occurs given that another event, A , has already occurred is

$$\Pr(B | A) = \frac{\Pr(A \text{ and } B)}{\Pr(A)}. \quad (3.1)$$

We can use the conditional probability formula to write the hazard function in terms of the probability function and survivor function as

$$h(a) = \frac{1}{\Pr(A \geq a)} \cdot \lim_{\delta a \rightarrow 0^+} \frac{\Pr(a \leq A < a + \delta a)}{\delta a} = \frac{f(a)}{S(a)}.$$

Further, the hazard function can be written in terms of only the survivor function so that

$$h(a) = \frac{\frac{d}{da} F(a)}{S(a)} = -\frac{\frac{d}{da} S(a)}{S(a)} = -\frac{d}{da} \log S(a). \quad (3.2)$$

By rearranging (3.2), the survivor function is written as a function of the cumulative hazard function as

$$S(a) = \exp\left(-\int_0^a h(u) du\right) = \exp(-H(a)). \quad (3.3)$$

In continuous time, if we know the hazard function, then we can use (3.3) to obtain the probability function:

$$f(a) = S(a) \cdot h(a) = \exp\left(-\int_0^a h(u) du\right) \cdot h(a). \quad (3.4)$$

Notice that, since $\lim_{a \rightarrow \infty} S(a) = 1 - \lim_{a \rightarrow \infty} F(a) = 0$, the cumulative hazard function tends to infinity as $a \rightarrow \infty$.

Discrete time

In discrete time, the survival distribution has probability mass at \mathbb{N}_0 and can be defined by a cumulative distribution function

$$F_a = \Pr(A \leq a),$$

such that (i) F_a is strictly increasing and (ii) $\lim_{a \rightarrow \infty} F_a = 1$. Other ways to describe the survival distribution are:

- survivor function - $S_a = \Pr(A > a) = \Pr(A \geq a + 1) = 1 - F_a$
- probability function - $f_a = \Pr(A = a) = F_a - F_{a-1}$
- hazard function - $h_a = \Pr(A = a \mid A \geq a)$
- cumulative hazard function - H_a .

The hazard function and probability function satisfy the inequalities $0 < f_a < 1$ and $0 < h_a \leq 1$. The cumulative distribution function is strictly increasing because we assume that $f_a > 0$ for all a . The five expressions for the survival distribution in continuous and discrete time can be found in Kalbfleisch and Prentice (2002, Chapter 1).

The hazard function is the probability of death for an individual at a certain time in life conditional on their survival to that particular time. The quantity

$$1 - h_u = \Pr(A \neq u \mid A \geq u)$$

is the probability of survival for an individual at age u conditional on their survival to age u . Hence, the survivor function can be written in terms of the hazard function as

$$S_a = \Pr(A \geq a + 1) = \prod_{0 \leq u \leq a} \Pr(A \neq u \mid A \geq u) = \prod_{0 \leq u \leq a} (1 - h_u).$$

There are two ways in which we can define the cumulative hazard function in discrete time. One method is to assume that $H_a = -\log S_a$, as in equation (3.3), so that $H_a = -\sum_{u=0}^a \log(1 - h_u)$. The cumulative hazard function then tends to infinity as $a \rightarrow \infty$. The other method is to replace the integral in (3.3) with a sum so that $H_a = \sum_{u=0}^a h_u$. Both approaches are discussed by Collett (2015, pages 35–36) and both approaches are approximately equivalent for small h_u since $x \approx -\log(1 - x)$ for small x .

We can use the conditional probability formula in (3.1) to write the hazard function in terms of the probability function and survivor function as

$$h_a = \frac{\Pr(A = a)}{\Pr(A \geq a)} = \frac{f_a}{S_{a-1}}. \quad (3.5)$$

The probability function can then be written in terms of the hazard function as

$$f_a = \prod_{0 \leq u \leq a-1} (1 - h_u) \cdot h_a. \quad (3.6)$$

Note that, equation (3.6) holds only for ages greater than zero and that $f_0 = h_0 > 0$ is always true.

Rounded time

In Section 2.2, we discussed the discretisation of continuous time for three rounding styles. In Table 3.1, we use θ_1 to describe a lower bound and θ_2 to describe an upper bound for a time interval. For the three rounding styles, the probability function can be defined as

$$f_i^* = \Pr(I = i) = \int_{i+\theta_1}^{i+\theta_2} f(u) \, du = F(i + \theta_2) - F(i + \theta_1) \quad (3.7)$$

and the hazard function can be defined as

$$h_i^* = \Pr(I = i \mid I \geq i) = \frac{f_i^*}{\sum_{u \geq i} f_u^*} \neq \int_{i+\theta_1}^{i+\theta_2} h(u) \, du.$$

In particular, when rounding age down to the nearest integer, the probability function is defined by integrating $f(a)$ over the interval $(i, i + 1)$ so that

$$f_i^* = F(i + 1) - F(i). \quad (3.8)$$

The hazard function cannot be expressed as a discretisation of $h(a)$ over the interval $(i, i + 1)$, but it can be expressed in terms of a discretisation of $f(a)$. The quantity f_i^* is an important component of the independence model described in Chapter 6. The purpose of this thesis is to assess the misleading conclusions that can arise when assessing the suitability of an independence model for a mistreatment of data.

Table 3.1: Constants used to define the probability function and hazard function in rounded time.

Style of integer rounding	θ_1	θ_2
Rounding down	0	1
Rounding up	-1	0
Rounding nearest	$-\frac{1}{2}$	$\frac{1}{2}$

3.2 Parametric models

In this thesis, a model is called *parametric* if it contains a finite number of parameters. In this section, we present some parametric models for the survival distribution in continuous and discrete time; parametric models in rounded time can be obtained by equation (3.7). The models in this section can be considered as a base model to the covariate models of Section 3.3.

Continuous time

The five expressions for the survival distribution can be thought of as equivalent because one expression is enough to define the other four expressions. Since they are equivalent, we focus on just one expression when defining an APC survival model in Chapter 4 and we choose to focus on the hazard function because the hazard function allows for a convenient incorporation of age-dependent covariates; a conclusion we reach in Section 3.4.

The monotonicity properties of $h(a)$ can be split into five main categories:

- constant — $\frac{d}{da}h(a) = 0$ for all a (\rightarrow)
- increasing — $\frac{d}{da}h(a) > 0$ for all a (\nearrow)
- decreasing — $\frac{d}{da}h(a) < 0$ for all a (\searrow)
- unimodal — increasing, constant, then decreasing (\cap)
- U-shaped — decreasing, constant, then increasing. (\cup)

The hazard function for human mortality is typically U-shaped due to complications at birth, a low risk of death during midlife, and a high risk of death when reaching old age. Kermack et al. (1934) argued that the Gompertz–Makeham model with an increasing hazard function is a good representation of the hazard function for the mortality of Scottish and English males. As in discrete time, we assign symbols such as \nearrow to each of the five monotonicity categories for the hazard function. The five monotonicity symbols are used for parametric models in Table 3.2 and for covariate models in Table 3.4.

Let $F^{(\mathcal{D})}(a, \boldsymbol{\alpha})$ be the cumulative distribution function from family \mathcal{D} for $A \in \mathbb{R}^+$, where $\boldsymbol{\alpha}$ is defined in (3.9). A near exhaustive list of parametric models for the continuous survival distribution are presented in Table 3.2. Similar lists of continuous parametric models can be found in Cox and Oakes (1984,

Table 3.2: A collection of parametric model families in continuous time.

Model family \mathcal{D}	Parameter constraints	$S^{(\mathcal{D})}(a, \boldsymbol{\alpha})$	$h^{(\mathcal{D})}(a, \boldsymbol{\alpha})$	Shape of $h^{(\mathcal{D})}(a, \boldsymbol{\alpha})$
Exponential	$\kappa_1 > 0$	$\exp(-\kappa_1 a)$	κ_1	\rightarrow
Uniform	$a \leq M, M > 0$	$\frac{M-a}{M}$	$\frac{1}{M-a}$	\nearrow
Rayleigh	$\pi_2 > 0$	$\exp\left(-\frac{1}{2}\pi_2 a^2\right)$	$\pi_2 a$	\nearrow
Quadratic	$\kappa_3, \gamma_3 \geq 0, \pi_3 \geq -2\sqrt{\kappa_3 \gamma_3}$	$\exp\left(-\left[\kappa_3 a + \frac{1}{2}\pi_3 a^2 + \frac{1}{3}\gamma_3 a^3\right]\right)$	$\kappa_3 + \pi_3 a + \gamma_3 a^2$	$\rightarrow, \nearrow, \searrow, \cap, \cup$
Gompertz	$\kappa_4, \pi_4 > 0$	$\exp\left(-\frac{\kappa_4}{\pi_4}(\exp(\pi_4 a) - 1)\right)$	$\kappa_4 \exp(\pi_4 a)$	\rightarrow, \nearrow
Gompertz–Makeham	$\kappa_5, \pi_5, \gamma_5 > 0$	$\exp\left(-\gamma_5 a - \frac{\kappa_5}{\pi_5}(\exp(\pi_5 a) - 1)\right)$	$\gamma_5 + \kappa_5 \exp(\pi_5 a)$	\rightarrow, \nearrow
Weibull	$\kappa_6, \pi_6 > 0$	$\exp(-(\pi_6 a)^{\kappa_6})$	$\kappa_6 \pi_6 (\pi_6 a)^{\kappa_6 - 1}$	$\rightarrow, \nearrow, \searrow$
Gamma	$\kappa_7, \pi_7 > 0$	$1 - \frac{I_{\kappa_7}(\pi_7 a)}{\Gamma(\kappa_7)}$	$\frac{\pi_7 (\pi_7 a)^{\kappa_7 - 1} \exp(-\pi_7 a)}{\Gamma(\kappa_7) - I_{\kappa_7}(\pi_7 a)}$	$\rightarrow, \nearrow, \searrow$
Generalised Pareto	$\kappa_9 > 0, \gamma_9 \geq 0, \pi_9 \geq -\kappa_9 \gamma_9$	$\left(\frac{\kappa_9}{a + \kappa_9}\right)^{\pi_9} \exp(-\gamma_9 a)$	$\gamma_9 + \frac{\pi_9}{a + \kappa_9}$	\rightarrow, \searrow
Gamma–prime	$\kappa_{10}, \gamma_{10} > 0$	$\left(\frac{\kappa_{10}}{a + \kappa_{10}}\right)^{-\kappa_{10} \gamma_{10}} \exp(-\gamma_{10} a)$	$\frac{\gamma_{10} a}{a + \kappa_{10}}$	\rightarrow, \nearrow
Log–Logistic	$\kappa_{11}, \pi_{11} > 0$	$\frac{1}{1 + (\pi_{11} a)^{\kappa_{11}}}$	$\frac{\pi_{11} \kappa_{11} (\pi_{11} a)^{\kappa_{11} - 1}}{1 + (\pi_{11} a)^{\kappa_{11}}}$	\searrow, \cap
Log–Normal	$\kappa_{12}, \pi_{12} > 0$	$1 - \Phi\left(\frac{\log(a) - \kappa_{12}}{\pi_{12}}\right)$	$\frac{\phi\left(\frac{\log(a) - \kappa_{12}}{\pi_{12}}\right)}{\pi_{12} a \left(1 - \Phi\left(\frac{\log(a) - \kappa_{12}}{\pi_{12}}\right)\right)}$	\cap
IDB	$\kappa_{13}, \pi_{13}, \gamma_{13} > 0$	$\frac{\exp\left(-\kappa_{13} \frac{a^2}{2}\right)}{(1 + \gamma_{13} a)^{\frac{\pi_{13}}{\gamma_{13}}}}$	$\kappa_{13} a + \frac{\pi_{13}}{1 + \gamma_{13} a}$	$\rightarrow, \nearrow, \searrow, \cup$

Chapter 2) and Kalbfleisch and Prentice (2002, Chapter 2). In Table 3.2, we use $\phi(\cdot)$ and $\Phi(\cdot)$ to denote the probability function and cumulative distribution function of a standard normal distribution, respectively. The Gamma model is written in terms of an incomplete gamma function

$$I_{\kappa_7}(\pi_7 a) = \int_0^{\pi_7 a} u^{\kappa_7-1} e^{-u} du$$

as well as the gamma function $\Gamma(\kappa_7) = I_{\kappa_7}(\infty)$.

A model family can be a submodel of a more general model family, in particular, the exponential family is a submodel of the Weibull family if $\kappa_6 = 1$ and $\pi_6 = \kappa_1$. The hazard function of the Weibull family can take a variety of shapes: $h^{(\text{Weibull})}(a, \kappa_6, \pi_6)$ is constant for $\kappa_6 = 1$, increasing for $\kappa_6 > 1$ and decreasing for $\kappa_6 < 1$. The Gamma and Weibull models are special cases of a three-parameter generalised gamma model

$$f(a) = \frac{\rho(\pi)^\kappa a^{\kappa-1} e^{-(\pi a)^\rho}}{\Gamma\left(\frac{\kappa}{\rho}\right)}$$

for $\rho = 1$ and $\rho = \kappa$, respectively (Kalbfleisch and Prentice 2002, page 37). We explain in Chapter 4 that perhaps the first ever age-period-cohort model had a Gamma model to describe the survival distribution (Greenberg et al. 1950). A notable model family is increasing-decreasing-bathtub (IDB) which is U-shaped if $0 < \kappa_{13} < \pi_{13}\gamma_{13}$ (Hjorth 1980). The quadratic model is discussed by Elbatal and Butt (2014), and the generalised Pareto and Gamma-prime models are discussed by Davis and Feldstein (1979).

Each model family can be discretised to define a model family in rounded. We can calculate the probability function for a rounding down style as

$$f_{i,\boldsymbol{\alpha}}^{*(\mathcal{D})} = S^{(\mathcal{D})}(i, \boldsymbol{\alpha}) - S^{(\mathcal{D})}(i+1, \boldsymbol{\alpha}).$$

The probability function of the exponential family is $f_{i,\kappa_1}^{*(\text{Exp.})} = e^{-\kappa_1 i} \cdot (1 - e^{-\kappa_1})$, which is equivalent to the probability function of a Geometric family (Kalbfleisch and Prentice 2002, page 46). For the uniform family, the survival distribution is the same for continuous and rounded time:

$$f_{i,M}^{*(\text{Unif.})} = \frac{1}{M} = f^{(\text{Unif.})}(a, M) \quad 0 \leq i \leq M, \quad 0 < a < M.$$

However, it is difficult to write $f_{i,\boldsymbol{\alpha}}^{*(\mathcal{D})}$ in a simple and concise form for most model families. For instance, Gompertz is one of the simplest two-parameter families in our list and its discretised probability function is equal to

$$f_{i,\boldsymbol{\alpha}}^{*(\text{Gompertz})} = \exp\left(-\frac{\kappa_4}{\pi_4}e^{\pi_4 i}\right) - \exp\left(-\frac{\kappa_4}{\pi_4}e^{\pi_4 i}\right)^{e^{\pi_4}}.$$

Discrete time

The monotonicity properties of h_a can be split into five main categories:

- constant — $h_{a+1} - h_a = 0$ for all a (\rightarrow)
- increasing — $h_{a+1} - h_a > 0$ for all a (\nearrow)
- decreasing — $h_{a+1} - h_a < 0$ for all a (\searrow)
- unimodal — increasing, constant, then decreasing (\cap)
- U-shaped — decreasing, constant, then increasing. (\cup)

The list of five categories is not exhaustive and other monotonicity categories can exist. For product-reliability testing, the hazard function is typically U-shaped due to an initial burn-in period, followed by a period of stability, and followed later by a wear-and-tear period (Hjorth 1980). We have assigned a symbol such as \nearrow to each of the five monotonicity categories as shown in brackets and these symbols allow us to write concisely the shape of the hazard function for parametric models. The five monotonicity symbols are used to describe parametric models in Table 3.3 and covariate models in Table 3.5.

Given we know f_a , the function F_a is often difficult to write in a simple and concise form, which means h_a is also intractable to some extent. In particular, it is difficult to write a simple expression for h_a for the Poisson and Binomial models. It was proved by Gupta et al. (1997) that even if h_a is intractable to some extent, we can still determine its monotonicity by checking

the monotonicity of $\frac{f_{u+1}}{f_u}$. Specifically, letting $\eta_u = \frac{f_{u+1}}{f_u} - \frac{f_{u+2}}{f_{u+1}}$, Gupta et al. proved that:

- if $\eta_u = 0$ for all u , then h_a is constant;
- if $\eta_u > 0$ for all u , then h_a is increasing;
- if $\eta_u < 0$ for all u , then h_a is decreasing.

The key to the proof is that we can use equation (3.6) to write

$$\frac{1}{h_{a+1}} - \frac{1}{h_a} = \left(\frac{f_{a+2}}{f_{a+1}} - \frac{f_{a+1}}{f_a} \right) + \left(\frac{f_{a+3} f_{a+2}}{f_{a+2} f_{a+1}} - \frac{f_{a+2} f_{a+1}}{f_{a+1} f_a} \right) + \dots$$

One notable result is that a Poisson model has a hazard function that is increasing because if $A \sim \text{Poisson}(\lambda)$, then $f_a = \frac{\lambda^a e^{-\lambda}}{a!}$, $\frac{f_{a+1}}{f_a} = \frac{\lambda}{a+1}$ and $\eta_a = \frac{\lambda}{(a+1)(a+2)} > 0$.

Let \mathcal{D} denote a name for a model family such that $A \sim \mathcal{D}$. We use $F_{a,\boldsymbol{\alpha}}^{(\mathcal{D})}$ to denote the cumulative distribution function from family \mathcal{D} for $A \in \mathbb{N}_0$, where

$$\boldsymbol{\alpha}^T = (\alpha_1, \alpha_2, \dots, \alpha_m) \tag{3.9}$$

is a finite parameter vector. A comprehensive list of parametric models for the discrete survival distribution are presented in Table 3.3. This table presents some of the discrete parametric models reviewed by Johnson et al. (2005, Chapter 11). The last six model families in the table are introduced in Nakagawa and Osaki (1975), Salvia and Bollinger (1982), Stein and Dattero (1984), Xekalaki (1983), Adams and Watson (1989) and Lai and Wang (1995), respectively.

The simplest model family is Geometric because it assumes that the hazard function is constant such that, at each time step, an individual either dies with probability $h_{a,\pi_1}^{(\text{Geometric})} = \pi_1$ or survives with probability $1 - \pi_1$ independently of their current age.

A hazard function can take a variety of shapes depending on the parameter values. The discrete Weibull family proposed by Nakagawa and Osaki (1975) is named after the continuous Weibull family due to the variety of shapes the

Table 3.3: A collection of parametric model families in discrete time. Note that $\eta(a, N) = \sum_{w=0}^N \pi_{6,w} a^w$. The models are all defined on $a \in \mathbb{N}_0$ unless stated otherwise. Some model families have an upper limit on the age-at-death so that $A \in \{0, 1, \dots, M\}$.

\mathcal{D}	Parameter constraints	$f_{a,\alpha}^{(\mathcal{D})}$	$h_{a,\alpha}^{(\mathcal{D})}$	Shape of $h_{a,\alpha}^{(\mathcal{D})}$
Geometric	$\pi_1 \in \mathbb{R}^+$	$\pi_1 (1 - \pi_1)^a$	π_1	\rightarrow
Uniform	$M \in \{1, 2, \dots\}$	$\frac{1}{M}$	$\frac{1}{M-a}$	\nearrow
Discrete Weibull	$0 < \pi_2 < 1, \gamma_2 \in \mathbb{R}^+$	$(\pi_2)^{a\gamma_2} - (\pi_2)^{(a+1)\gamma_2}$	$1 - \frac{(\pi_2)^{(a+1)\gamma_2}}{(\pi_2)^{a\gamma_2}}$	$\rightarrow, \nearrow, \searrow$
Salvia–Bollinger	$0 < \pi_3 \leq 1$	$\pi_3 \prod_{u=1}^a \frac{u-\pi_3}{(a+1)!}$	$\frac{\pi_3}{a+1}$	\searrow
Stein–Dattero	$0 < \pi_4 \leq 1; \gamma_4 \in \mathbb{R};$ if $\gamma_4 > 1$, then $M = \lfloor 1 + \pi_4^{-1/(\gamma_4-1)} \rfloor$.	$\pi_4 a^{\gamma_4} \prod_{u=0}^{a-1} (1 - \pi_4 u^{\gamma_4})$	$\pi_4 a^{\gamma_4}$	\nearrow
Xekalaki	$\pi_5 \in \mathbb{R}^+; \gamma_5 \in \mathbb{R};$ if $\gamma_5 < 0$, then $M = \lfloor \frac{1-\pi_5}{\gamma_5} \rfloor$.	Depends on γ_5	$\frac{1}{\pi_5 + \gamma_5 a}$	$\rightarrow, \nearrow, \searrow$
Adams–Watson	$\pi_6, 0 \dots, \pi_{6,N} \in \mathbb{R}; N \in \mathbb{N}_0$	$\frac{1}{1 + \exp(-\eta(a, N))} \prod_{u=0}^{a-1} \frac{1}{1 + \exp(\eta(u, N))}$	$\frac{1}{1 + \exp(-\eta(a, N))}$	$\rightarrow, \nearrow, \searrow$
Lai–Wang	$\pi_7 \in \mathbb{R}, M \in \mathbb{N}$	$\frac{a^{\pi_7}}{\sum_{u=0}^M u^{\pi_7}}$	$\frac{a^{\pi_7}}{\sum_{u=0}^M u^{\pi_7}}$	\nearrow, \cup

hazard function can take. The discrete Weibull family has a hazard function that is increasing for $\gamma_2 > 1$, decreasing for $0 < \gamma_2 < 1$ and constant at $1 - \pi_2$ for $\gamma_2 = 1$. For other parametric families, the hazard function is strictly increasing when $\gamma_5 < 0$, $\pi_7 \geq 0$ or $\eta(a, N)$ is strictly increasing with a . The hazard function is strictly decreasing if $\gamma_5 > 0$ or $\eta(a, N)$ is strictly decreasing with a . The Lai–Wang family has a U-shaped hazard function if $\pi_7 < 0$. The Geometric family is a special case for a Xekalaki family with $\gamma_5 = 0$ or an Adams–Watson family with $\eta(a, N)$ constant in a . Johnson et al. (2005, page 517) explains that the probability function of a Xekalaki family comes from a Waring distribution for $\gamma_5 > 0$ and from a negative hypergeometric distribution for $\gamma_5 < 0$. The Adams–Watson family assumes that the logit of the hazard function is a polynomial.

3.3 Covariate models with age-independent covariates

The survival distribution of an individual can depend on their covariates such as gender or smoking status, and we focus on period and cohort as covariates. We use the term *covariate model* to refer to a model for the survival distribution that is conditional on covariates. There are many examples of covariate models in the survival literature such as the proportional hazards (PH) model and accelerated failure time (AFT) model (Cox 1972; Collett 2015). In this section, we assess the suitability of various covariate models in continuous and discrete time when covariates are age-independent.

Let $\mathbf{x}^T = (x_1, \dots, x_q)$ be a q -vector of covariates for an individual and let

$$\psi = \mathbf{x}^T \boldsymbol{\beta} \in \mathbb{R} \tag{3.10}$$

be a linear predictor associated with \mathbf{x} and a parameter vector

$$\boldsymbol{\beta}^T = (\beta_1, \dots, \beta_q). \quad (3.11)$$

We use the term *base model* to refer to the survival distribution for a reference individual or reference group of individuals. The individual or group of individuals represented by the base model has a particular set of covariate values. The base model can be viewed as the independence model before modifications are made to consider the effect of covariates such as period and cohort.

For example, the base model could refer to the survival distribution for males so that $\psi = \beta x$, where $x = 1$ if the individual is female and $x = 0$ if the individual is male. The parameter β defines the survival distribution for females through a departure from the male survival distribution. By considering a second covariate, the base model could refer to the survival distribution for males born at time $c = 2000$ so that $\psi = \beta_1 x + \beta_2 [c - 2000]$.

A covariate model takes the base model and describes the survival distribution for any individual by a departure from the base model through ψ . The base model corresponds to $\psi = 0$ and is obtained by substituting the covariates of the reference individual or individuals into ψ . The Cox model (Cox 1972), in which the departures are defined by shifting the hazard function by a constant of proportionality, is a prime example of a covariate model.

A covariate model takes a function of the base model and ψ . The base model is not necessarily parametric and can be nonparametric. We say that a model is *nonparametric* if the model cannot be expressed in terms of a finite number of parameters. Letting \mathcal{G} be a functional, a covariate model in continuous time is

$$F_\psi(a) = \mathcal{G}(F_0(a), \psi) \quad (3.12)$$

and a covariate model in discrete time is

$$F_{a,\psi} = \mathcal{G}(F_{a,0}, \psi). \quad (3.13)$$

We say that a covariate model is *valid* if each of the five expressions for the survival distribution has the same range before and after the functional transformation. For example, in continuous time, the survivor function should always have a range of $[0, 1)$ and the hazard function should have a range of \mathbb{R}^+ . It is not necessary to assess the range for every expression since $S(a) \in [0, 1]$ would imply that $F(a) \in [0, 1]$.

In Table 3.4, we present a list of five covariate models in continuous time. The first and third columns of the table state the name and form of the covariate model. In the case of a parametric base model, the final column states a family for which the covariate model is closed. A model is *closed* if the base model and covariate model both share the same family. If the range of the base model in column two is the same as the range of the covariate model in column three, then we say that the covariate model is valid. An important assumption we make is that the base model is valid.

A prime example of a valid functional in continuous time is the proportional hazards (PH) functional introduced by Cox (1972), which assumes that the hazard functions between individuals differ by a constant of proportionality. The PH functional has a hazard function $h_\psi(a) = h_0(a)e^\psi$ which is restricted to the interval \mathbb{R}^+ and a survivor function $S_\psi(a) = S_0(a)e^{-\psi a}$ which is restricted to $[0, 1]$. Parallel hazards is an example of an invalid functional because the hazard function $h_\psi(a) = h_0(a) + \psi$ takes values on \mathbb{R} . Parallel hazards is considered here as an illustrative example of an invalid model, but it is never used in practice. We present two covariate models for the quantile function of A :

$$Q(u) = F^{-1}(u) \in \mathbb{R}^+ \quad \text{for } u \in [0, 1].$$

The quantile function is a sixth expression for the survival distribution.

A similar list of covariate models was compiled by Cox and Oakes (1984, Chapter 5) which included parallel hazards, proportional hazards, proportional odds (PO) and accelerated age-at-death (AA). Note that, the AA functional is commonly known as accelerated failure time (AFT). The “logit” of a vari-

Table 3.4: The validity of covariate models in continuous time. A model is valid if the range of the base model and covariate model are the same.

Functional \mathcal{G}	Base model	Covariate model	Valid?	\mathcal{D}
Parallel hazards	$h_0(a) \in \mathbb{R}^+$	$h_\psi(a) = h_0(a) + \psi \in \mathbb{R}$	No	
Proportional hazards (PH)	$h_0(a) \in \mathbb{R}^+$ $S_0(a) \in [0, 1]$ $F_0(a) \in [0, 1]$	$h_\psi(a) = h_0(a)e^\psi \in \mathbb{R}^+$ $S_\psi(a) = S_0(a)e^\psi \in [0, 1]$ $F_\psi(a) = 1 - [1 - F_0(a)]e^\psi \in [0, 1]$	Yes	Gompertz, Weibull, Gompertz-Makeham, IDB
Proportional odds (PO)	$\frac{S_0(a)}{1-S_0(a)} \in \mathbb{R}^+$ $S_0(a) \in [0, 1]$	$\frac{S_\psi(a)}{1-S_\psi(a)} = \frac{S_0(a)}{1-S_0(a)}e^\psi \in \mathbb{R}^+$ $S_\psi(a) = \frac{S_0(a)e^\psi}{S_0(a)e^\psi + F_0(a)} \in [0, 1]$	Yes	Log-logistic
Log location scale (LLS)	$Q_0(u) \in \mathbb{R}^+$ $F_0(a)$	$Q_\psi(u) = Q_0(u)e^\psi e^\psi \in \mathbb{R}^+$ $F_\psi(a) = F_0 \left(\exp \left[\frac{\log a - \psi}{e^\psi} \right] \right) \in [0, 1]$	Yes	
Accelerated age-at-death (AA)	$Q_0(u) \in \mathbb{R}^+$ $F_0(a)$	$Q_\psi(u) = Q_0(u)e^{-\psi} \in \mathbb{R}^+$ $F_\psi(a) = F_0(e^\psi a) \in [0, 1]$	Yes	Weibull, Log-Normal, Log-logistic

able x means that $\text{logit}(x) = \log\left(\frac{x}{1-x}\right)$. The PO functional, which assumes that the logit of the survivor function differs between individuals by an additive constant, is valid because its survivor function $S_\psi(a) = \frac{S_0(a)e^\psi}{S_0(a)e^\psi + F_0(a)}$ is restricted to the interval $[0, 1]$. In Table 3.4, multiple entries for a particular functional indicate equivalent ways to write a covariate model. For example, proportional hazards implies that the cumulative distribution function can be written as $F_\psi(a) = 1 - [1 - F_0(a)]e^\psi$.

The AA functional, which assumes that the quantile function differs between individuals by a constant of proportionality, is another very widely used functional. Cox and Oakes (1984, pages 64–65) state that the AA functional implies that the cumulative distribution function and hazard function can be written as $F_0(e^\psi a)$ and $h_0(e^\psi a)e^\psi$, respectively. Textbooks on survival analysis tend to only mention the PH and AA functionals so it is rare to see a comparison of three or more functionals.

Letting X and Y be two variables, a location-scale model can be written as $Y = a + bX$ where a and b are real numbers. The log location scale (LLS) model

$$\log Q_\psi(u) = \psi + e^\psi \log Q_0(u)$$

adjusts the location and scale of the logarithm of the base quantile function. An AA functional $\log Q_\psi(u) = \log Q_0(u) - \psi$ adjusts only the location. The AA functional is valid because $Q_\psi(u) = e^{-\psi}Q_0(u) \in \mathbb{R}^+$. A location-scale functional $Q_\psi(u) = e^\psi Q_0(u) + \psi$ is invalid since the quantile function is allowed to be negative. Models for the quantile function such as the AA and LLS functionals differ from other functionals in that the linear predictor ψ acts on age itself. For example, in the expression for the cumulative distribution function in the AA functional, the effect of covariates is to multiply age a by e^ψ . In comparison, the effect of covariates in the PH model is to multiply the hazard function by e^ψ .

The final column of Table 3.4 states a family or multiple families for which the covariate model is closed. For example, the Gompertz family is closed under

Table 3.5: The validity of functional covariate models in discrete time. A model is valid if the range of the base model and covariate model are the same.

Functional \mathcal{G}	Base model	Covariate model	Valid?
Proportional hazards (PH)	$h_{a,0} \in (0, 1]$ $S_{a,0} \in [0, 1)$	$h_{a,\psi} = h_{a,0}e^\psi \in \mathbb{R}^+$ $S_{a,\psi} = \prod_{u \leq a} (1 - h_{u,0}e^\psi) \in \mathbb{R}$	No
Discrete logistic (DL)	$\frac{h_{a,0}}{1-h_{a,0}} \in \mathbb{R}^+$ $S_{a,0} \in [0, 1)$	$\frac{h_{a,\psi}}{1-h_{a,\psi}} = \frac{h_{a,0}}{1-h_{a,0}}e^\psi \in \mathbb{R}^+$ $S_{a,\psi} = \prod_{u \leq a} \left(\frac{1-h_{u,0}}{(1-h_{u,0})+h_{u,0}e^\psi} \right) \in [0, 1)$	Yes
Complementary log-log (CLL)	$h_{a,0} \in (0, 1]$ $S_{a,0} \in [0, 1)$	$h_{a,\psi} = 1 - (1 - h_{a,0})e^\psi \in (0, 1]$ $S_{a,\psi} = \prod_{u \leq a} (1 - h_{u,0})e^\psi = [S_{a,0}]^{e^\psi} \in [0, 1)$	Yes
Proportional odds (PO)	$\frac{S_{a,0}}{1-S_{a,0}} \in \mathbb{R}^+$ $S_{a,0} \in [0, 1)$	$\frac{S_{a,\psi}}{1-S_{a,\psi}} = \frac{S_{a,0}}{1-S_{a,0}}e^\psi \in \mathbb{R}^+$ $S_{a,\psi} = \frac{S_{a,0}}{1+S_{a,0}(e^\psi-1)}e^\psi \in [0, 1)$	Yes

the PH functional since

$$h_{\psi}^{(\text{Gomp.})}(a, \kappa_4, \pi_4) = h_0^{(\text{Gomp.})}(a, \kappa_4, \pi_4)e^{\psi} = h_0^{(\text{Gomp.})}(a, \kappa_4 e^{\psi}, \pi_4). \quad (3.14)$$

Collett (2015, Chapters 5 and 6) explains how to fit the majority of these closed models to data.

In Table 3.5, we present a list of covariate models in discrete time. It is necessary to use different functionals in continuous and discrete time because the probability function and hazard function have a range of \mathbb{R}^+ in continuous time, but have a range of $(0, 1)$ and $(0, 1]$ respectively in discrete time. The PH model in discrete time is invalid because it allows the hazard function to exceed one and allows the survivor function to be negative. A discrete logistic (DL) functional assumes that the logit of the hazard function differs between individuals by an additive constant, so that

$$\text{logit}(h_{a,\psi}) = \text{logit}(h_{a,0}) + \psi.$$

The DL functional was suggested by Cox (1972) to ensure that the hazard function takes values between zero and one. The survivor function of a DL functional can perhaps be written in its simplest form as

$$S_{a,\psi} = \prod_{u \leq a} (1 - h_{u,\psi}) = \prod_{u \leq a} \left(\frac{1 - h_{u,0}}{(1 - h_{u,0}) + h_{u,0}e^{\psi}} \right) \in [0, 1).$$

Another way to bound the hazard function between zero and one is to use the complementary log-log (CLL) functional suggested by Kalbfleisch and Prentice (2002, page 47). The CLL functional has a survivor function that can be written in the following simple form:

$$S_{a,\psi} = [S_{a,0}]^{e^{\psi}} \in [0, 1).$$

The CLL functional and the continuous time PH functional both assume that the survivor function differs between individuals by a power constant. The PO functional is valid in both continuous and discrete time.

3.4 Covariate models with age-dependent covariates

In this section, we consider age-dependent covariates and reassess the suitability of the covariate models from the previous section. We conclude that only the PH, CLL and DL functionals are suitable to describe the survival distribution of an APC survival model. This conclusion justifies the widespread adoption of the PH functional in APC modelling (Clayton and Schifflers 1987b; Carstensen 2007; Jean et al. 2013).

A covariate is age-independent if the covariate is fixed throughout the life of an individual. A covariate is age-dependent if the covariate varies either deterministically or randomly with age. An individual's blood pressure is a possible covariate that would change randomly with age, whereas period is a covariate that increases linearly with age according to the identity $p = c + a$. Cohort is a covariate that remains the same for all ages so is age-independent. Random covariates do not feature any further in this thesis. Modelling with age-dependent covariates is well established in survival modelling (Collett 2015, Chapter 8). Collett explains how to fit the Cox PH model to data under age-dependent covariates.

Letting $x_j(a)$ be the value of the j th covariate at age $a \in \mathbb{R}^+$, the linear predictor associated with $\boldsymbol{\beta}$ and a covariate vector $\mathbf{x}(a)^T = (x_1(a), \dots, x_q(a))$ is

$$\psi(a) = \mathbf{x}(a)^T \boldsymbol{\beta}. \quad (3.15)$$

Similarly, letting $x_{j,a}$ be the value of the j th covariate at age $a \in \mathbb{N}_0$, the linear predictor associated with covariate vector $\mathbf{x}_a^T = (x_{1,a}, \dots, x_{q,a})$ and a parameter vector $\boldsymbol{\beta}$ is

$$\psi_a = \mathbf{x}_a^T \boldsymbol{\beta}. \quad (3.16)$$

If we consider period as the only covariate, then an individual born at time 2000 will have a period covariate of $x_{1,a} = 2000 + a$. Age is implicit in the

systematic part of the model, but is not explicit because age cannot be a covariate. The presence of age in the systematic part of the model results in some covariate models becoming illogical.

Under age-dependent covariates, we can adjust our definitions from (3.12) and (3.13) to write the covariate model in continuous time as

$$F_{\psi(a)}(a) = \mathcal{G}(F_0(a), \psi(a)) \quad (3.17)$$

and in discrete time as

$$F_{a, \psi_a} = \mathcal{G}(F_{a,0}, \psi_a). \quad (3.18)$$

We cannot simply exchange ψ with $\psi(a)$ in Table 3.4, or exchange ψ with ψ_a in Table 3.5, to derive expressions for a covariate model. For instance, the PH model under age-dependent covariates is written as

$$h_{\psi(a)}(a) = h_0(a)e^{\psi(a)}. \quad (3.19)$$

The survivor function implied by this PH functional is not simply equal to $S_0(a)e^{\psi(a)}$, but is instead derived with the use of definition (3.3) as

$$S_{\psi(a)}(a) = \exp\left(-\int_0^a h_0(u)e^{\psi(u)}du\right). \quad (3.20)$$

Some of the covariate models in Section 3.3 do not make logical sense when at least one covariate is age-dependent. The quantile function of the AA model can be written as $Q_{\psi(a)}(u) = Q_0(u)e^{-\psi(a)}$ for $u \in (0, 1]$ under age-dependent covariates. The AA does not seem logical due to an assumption, for example, that the median age-at-death corresponding to $u = 0.5$ depends on age itself. The LLS model similarly assumes that the median age-at-death depends on age. So we argue that covariate models for the quantile function are also not suitable to describe the survival distribution under age-dependent covariates.

For age-independent covariates, we stated in Tables 3.4 and 3.5 that the PO functional is valid for continuous and discrete time. Under age-dependent covariates, the PO survivor function can be written in continuous time as

$$S_{\psi(a)}(a) = \frac{S_0(a)e^{\psi(a)}}{S_0(a)e^{\psi(a)} + F_0(a)} \in [0, 1]. \quad (3.21)$$

We argue that the PO functional does not make sense because it assumes that the probability of survival to age a only depends on the covariates of the individual at age a . Hence, the survival to age a for an individual is independent of their survival at ages $0 < u < a$. It is important to consider past survival so that the survivor function at age a depends on $\mathbf{x}(u)$ for all $u \in (0, a]$. There is an element of fortune telling by the PO functional because the unknown future values in the covariate vector, $\mathbf{x}(a)$, are treated as known at birth. This fortune telling does not apply to period since period is known for an individual at age a given the cohort of the individual is known.

It is appropriate, however, to assume that the hazard function depends only on $\mathbf{x}(a)$ since the hazard function conditions on $\mathbf{x}(a)$ being known at age a . Further, the PH, CLL and DL functionals are the only covariate models discussed in this chapter that are logical for age-dependent covariates. For the definition of the PH survivor function in equation (3.20), the integration of $h_{\psi(u)}(u)$ over the interval $u \in (0, a]$ takes into account the survival experience for an individual before age a . The CLL model for age-dependent covariates is written as

$$h_{a,\psi_a} = 1 - (1 - h_{a,0})^{e^{\psi_a}}. \quad (3.22)$$

The survivor function of a CLL model is then written as

$$S_{a,\psi_a} = \prod_{u \leq a} (1 - h_{u,0})^{e^{\psi u}} \in [0, 1). \quad (3.23)$$

3.5 Conclusion

The survival distribution for an individual in continuous and discrete time can be expressed in various ways including in terms of the survivor function, the probability function, the hazard function and the quantile function. Each expression for the survival distribution can be thought of as equivalent since, for

example, if we know the hazard function for an individual, then their probability function can be obtained by some integral or product formula. In Chapter 2, APC models were expressed in terms of the probability functions, $f(a | c)$ and $f_{a|c}$. The independence models $\lambda(a, c) = \kappa(c) \cdot f(a)$ and $\lambda_{a,c} = \kappa_c \cdot f_a$ were defined as a special case of the APC models.

In this chapter, we considered a variety of parametric models for $f(a)$ and f_a that consisted of at most three parameters. We then took $f(a)$ and f_a to define a baseline survival distribution and explored through a variety of functionals how a dependence on the covariates period and/or cohort should be introduced to the independence models. We concluded that the choice of functional is contingent on whether the APC model is to be parameterised with only cohort, only period, or both cohort and period.

We distinguished between cohort as an age-independent covariate and period as an age-dependent covariate. If the survival distribution is assumed to vary only with cohort, so that there are no age-dependent covariates, then there are many possible ways to handle dependence in the APC model. For instance, in continuous time, the dependence can be introduced by functionals such as PO, LLS and PH. PO is a functional for the survivor function, LLS is a functional for the quantile function, and PH is a functional for the hazard function.

If the survival distribution is assumed to vary only with period or with both cohort and period, then there are few ways to handle dependence because only functionals for the hazard function are appropriate under age-dependent covariates. The reason is that the hazard function at age a can be assumed to depend only on the covariates at age a . It is not appropriate however to assume that the survivor function at age a only depends on the covariates at age a . We conclude that $f(a)$ should be modified according to a PH functional, while f_a should be modified according to either a CLL functional or a DL functional.

In Chapter 4, we show how to fit the modified models to data and explain

that the way of handling dependence in the APC literature is not consistent with our findings. Statisticians usually fit the discrete model $\lambda_{a,c} = \kappa_c \cdot f_{a|c}$ to rounded data, where $f_{a|c}$ is specified according to a PH functional. In Chapter 6, we explain that as opposed to estimating a discrete Poisson intensity, statisticians should estimate a Poisson intensity that is discretised over an appropriate region in the Lexis diagram.

Chapter 4

Age-period-cohort modelling

In this chapter, we establish a method to fit APC models to data in the statistical package R. A simple parameterisation of the APC model is used as an illustrative example to explain how the method works. In Chapter 3, we explained that if the independence model indexed in continuous time is to be modified to incorporate the effects of both period and cohort, then $f(a | c)$ should be specified according to a PH functional. But if the independence model indexed in discrete time is to be modified to incorporate period and cohort, then $f_{a|c}$ should be specified according to either a CLL functional or DL functional. It is not necessary to consider both the CLL and DL functionals, and we choose to write discrete models in terms of the CLL functional. Our method for fitting APC models is different for continuous time and discrete time. Our method allows the user to specify a range of parameterisations for the baseline survival function and for the linear predictor.

In Section 4.3, we present a range of APC models studied in the literature. The exit event is not necessarily death and there are a range of exit events studied in the literature including lung cancer (Peto et al. 1995) and homicide arrest (Fu 2008). The entry event is birth for almost every APC study. We explain that our methodology for making modifications to the independence model is not consistent with the APC literature. This is because statisticians

usually formulate discrete models to describe rounded data and make modifications to the discrete independence model according to the discrete PH functional from Table 3.5 (Barrett 1973; Clayton and Schifflers 1987b; Jean et al. 2013). Models should be discretised correctly to match the rounding of the data. In Chapter 6, we investigate whether the need for modifications is only apparent due to misleading conclusions that arise when fitting discrete independence models to rounded data.

An important distinction we make in this thesis is that an APC model can either be a survival model or a regression model. This distinction between model settings has not been made explicit in the literature and is discussed more in Section 4.4. An APC regression model considers how the distribution of some response variable, that is unrelated to the death of an individual, varies with age, period and cohort. Chapters 2 and 3 are exclusive to the survival setting of APC modelling. The confounding issue discussed in Chapter 5 is relevant for both the survival and regression settings. The issue of misrounding discussed in Chapters 6 and 7 is only relevant to the survival setting.

APC survival modelling can be viewed from two perspectives: the population level and the individual level. At the individual level, the survival distribution for an individual is modelled conditional on the period and their cohort. Textbooks on survival modelling tend to focus on the individual perspective (Cox and Oakes 1984; Collett 2015). At the population level, both the survival distribution and cohort intensity are modelled. A population level perspective is adopted when describing Poisson models for deaths in Chapter 2 and an individual level perspective is adopted in Chapter 3. All APC survival models in this chapter adopt a population level perspective.

4.1 Survival setting in continuous time

Recall from (2.4) that the Poisson model indexed in continuous time by age and cohort has an intensity function $\lambda(a, c) = \kappa(c) \cdot f(a | c)$. In equation (3.15), we defined the linear predictor $\psi(a)$ as a product of a covariate vector $\mathbf{x}(a)^T$ and a parameter vector $\boldsymbol{\beta}$. If period and cohort are the only covariates that affect the survival distribution for an individual, so that $\psi(a)$ is only a function of p and c , the probability function defined under the covariate model in (3.17) can be written as

$$f_{\psi(a)}(a) \equiv f(a | c).$$

In Chapter 3, we showed that, in continuous time, dependence in the APC model should be handled by specifying the survival distribution according to a PH model. Recalling the survivor function in (3.20), the Poisson intensity for deaths in the APC model is then written as

$$\begin{aligned} \lambda(a, c) &= \kappa(c) \cdot f(a | c) \\ &= \kappa(c) \cdot S(a | c) \cdot h(a | c) \\ &= \kappa(c) \cdot \exp\left(-\int_0^a h_0(u)e^{\psi(u)} du\right) \cdot h_0(a)e^{\psi(a)}. \end{aligned} \quad (4.1)$$

Any valid APC model must ensure that $\kappa(c) > 0$ for all c .

Let $a_j \in \mathbb{R}^+$ and $c_j \in \mathbb{R}$ be the age-at-death and cohort respectively for individuals $j = 1, 2, \dots, n$. The data can be written in vectors as $\mathbf{a}^T = (a_1, a_2, \dots, a_n)$ and $\mathbf{c}^T = (c_1, c_2, \dots, c_n)$. While the linear predictor $\psi(a)$ consists of the parameter vector $\boldsymbol{\beta}$, the base hazard function $h_0(a)$ consists of the parameter vector $\boldsymbol{\alpha}$. The likelihood function for the data can be written as

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta} | \mathbf{a}, \mathbf{c}) = \prod_{j=1}^n f(a_j | c_j).$$

The logarithm of the likelihood function for the data is then

$$\sum_{j=1}^n \log f(a_j | c_j). \quad (4.2)$$

In this section, we establish a method to fit $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to data by maximising the log likelihood function. To illustrate this method, it is helpful to consider a simple dataset. In the following R code, we draw $n = 1000$ observations of the age-at-death from an exponential distribution with parameter $\kappa_1 = 0.5$ and draw 1000 observations of cohort from a uniform distribution on the interval $[1, 5]$:

```
n <- 1000; a <- rexp(n,0.5); c <- runif(n,1,5).
```

In Table 3.2, the exponential distribution was defined with a parameter κ_1 .

It is also helpful to consider a simple parametric form for $f(a_j | c_j)$. Suppose that the baseline hazard function follows a Weibull distribution so that

$$h_0(a_j) = \alpha_1 \alpha_2 (\alpha_2 a_j)^{\alpha_1 - 1}$$

where $\alpha_1 > 0$ and $\alpha_2 > 0$. The Weibull distribution was defined in Table 3.2. We write the baseline hazard function in R as a function of data vector \mathbf{a} and parameter vector $\boldsymbol{\alpha}^T = (\alpha_1, \alpha_2)$:

```
h0 <- function(a,alpha) {
alpha[1]*alpha[2]* ( (alpha[2]*a)^(alpha[1]-1) )
}
Nalpha <- 2 ##the length of alpha.
```

It is necessary to specify the length of the vector $\boldsymbol{\alpha}$ through `Nalpha`. The function `h0` creates a vector of base hazard functions evaluated at the observed ages at death,

$$\mathbf{h}_0^T = (h_0(a_1), h_0(a_2), \dots, h_0(a_n)). \quad (4.3)$$

Suppose that we specify the covariate vector at age a_j as $\mathbf{x}(a_j)^T = (c_j, p_j)$ with a corresponding parameter vector $\boldsymbol{\beta}^T = (\beta_1, \beta_2)$, where β_1 and β_2 can

take values on \mathbb{R} . The matrix of covariates at death is

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}(a_1)^T \\ \mathbf{x}(a_2)^T \\ \vdots \\ \mathbf{x}(a_n)^T \end{pmatrix} = \begin{pmatrix} c_1 & p_1 \\ c_2 & p_2 \\ \vdots & \vdots \\ c_n & p_n \end{pmatrix}.$$

The following R function, `Xmake`, takes the data vectors `a` and `c` as inputs and produces the matrix `X`:

```
Xmake <- function(a,c) { cbind(c,a+c) } ##covariate matrix
Nbeta <- 2 ##the length of  $\beta$ .
```

It is necessary to specify the number of columns in `X`, or equivalently, the length of `β` , through `Nbeta`. The function `Xmake` is written only in terms of `a` and `c` because the period covariates for $j = 1, 2, \dots, n$ can be determined as the sum `a+c` due to the linear identity $p = a + c$.

In this example, the linear predictor at age a_j is $\psi(a_j) = \mathbf{x}(a_j)^T \boldsymbol{\beta} = \beta_1 c_j + \beta_2 p_j$ and the hazard function at age a_j is

$$\begin{aligned} h(a_j | c_j) &= h_0(a_j) \cdot \exp(\psi(a_j)) \\ &= \alpha_1 \alpha_2 (\alpha_2 a_j)^{\alpha_1 - 1} \cdot \exp(\beta_1 c_j + \beta_2 p_j). \end{aligned} \quad (4.4)$$

The probability density function implied by our choice of $h_0(a_j)$ and $\psi(a_j)$ is obtained by the definition

$$f(a_j | c_j) = \exp\left(-\int_0^{a_j} h(u | c_j) du\right) \cdot h(a_j | c_j). \quad (4.5)$$

It is difficult to calculate the integral of $h(u | c_j)$ over the interval $u \in [0, a_j]$ so we write the probability density function as the following Riemann Sum:

$$\exp\left(-\frac{1}{m} \cdot \sum_{\ell=1}^m h\left(a_j \cdot \frac{\ell}{m} \mid c_j\right)\right) \cdot h(a_j | c_j). \quad (4.6)$$

The Riemann Sum approximation to an integral is discussed by Adams (2006, page 286).

The user must specify the value of m . The accuracy of the Riemann Sum approximation to $f(a_j | c_j)$ will increase as the value of m increases. In this example, we choose to set

```
m <- 30.
```

The user can also specify a variable `scalePar` to scale the vectors α and β , which is helpful because the value of $\exp(\psi(a_j))$ will be close to or equal to infinity when the observed cohorts are large and/or when the matrix of covariates is specified with quadratic or cubic terms in cohort or period. Similarly, the value of $h_0(a_j)$ could be close to infinity when the observed age at deaths are large. In this example, we consider linear terms for period and cohort as well as small values for a and c , so it is sufficient to set

```
scalePar <- 1.
```

We establish a function `apc.cont` in the statistical package R to calculate the log likelihood function for a particular value of α and β . Details on this function are given in Appendix 9.2. For our particular specification of `n`, `a`, `c`, `h0`, `Xmake`, `Nalpha`, `Nbeta`, `m` and `scalePar`, we find that the value of `apc.cont(rep(0.1,4))` is equal to -452.224 to three decimal places. That is, if the parameters α_1 , α_2 , β_1 and β_2 are all set equal to 0.1, then the value of the log likelihood function is equal to -452.224. If the four parameters are set equal to one, we find that the log likelihood function is equal to $-\infty$.

The initial values for an optimisation procedure, `init` must be chosen so that `apc.cont(init)` does not return $-\infty$, otherwise the procedure will not converge to set of optimal parameters. To find the set of parameters that maximise the log likelihood function, we run the following optimisation procedure:

```
init <- rep(0.1, Nalpha+Nbeta) ##initial values
fit <- optim( init, apc.cont,
control=list(fnscale=-1,maxit=5000) ) ##optimisation procedure
```

```
fit$par <- fit$par / scalePar ##scaling of parameters.
```

The argument `fnscale=-1` states that the function `apc.cont` is to be maximised rather than minimised. The argument `maxit=5000` states that the optimisation should be stopped after 5,000 iterations if an optimal solution to `param` has not been found beforehand. The command `fit$par / scalePar` rescales the vector `param`. It is helpful to try different values for `init` in the function `fit` to check that the set of parameters returned do maximise the log likelihood function.

After typing `fit` into R, we find that the maximum likelihood estimates of the parameters are $\hat{\alpha}_1 = 1.373$, $\hat{\alpha}_2 = 0.780$, $\hat{\beta}_1 = -0.147$ and $\hat{\beta}_2 = 0.135$ to three decimal places. The corresponding value of the log likelihood function is -756.585. Recalling (4.4) and (4.5), the fitted probability density function is

$$\hat{f}(a | c) = \exp\left(-\int_0^a \hat{h}(u | c) du\right) \cdot \hat{h}(a | c) \quad (4.7)$$

where

$$\hat{h}(a | c) = 1.373 \cdot 0.780 (0.780a)^{0.373} \cdot \exp(-0.147c + 0.135p) \quad (4.8)$$

$$= 0.976 \cdot a^{0.373} \cdot \exp(0.135a) \cdot \exp(-0.012c). \quad (4.9)$$

Since an exponential distribution with parameter $\kappa_1 = 0.5$ is equivalent to a Weibull distribution with parameters $\alpha_1 = 1$ and $\alpha_2 = 0.5$, our estimates $\hat{\alpha}_1$ and $\hat{\alpha}_2$ do not provide a very accurate description of the data. The observed age at deaths were simulated from an exponential distribution with no period or cohort dependence, so the parameters $\hat{\beta}_1$ and $\hat{\beta}_2$ should be close to zero. However, $\hat{\beta}_1$ and $\hat{\beta}_2$ do seem quite large. Perhaps the loss in accuracy is a result of age entering implicitly into the linear predictor through period. This is because, when substituting $p = a + c$ into (4.8) to derive (4.9), the contribution to the hazard function for cohort c , written as $\exp(-0.012c)$, is very close to one.

In Figure 4.1, a histogram is plotted for the $n = 1000$ observed age at deaths drawn from an exponential distribution with parameter 0.5. The Rie-

mann Sum approximations of $\hat{f}(a | 1)$ and $\hat{f}(a | 4)$ are also plotted. Relative frequency takes the number of deaths for a particular year of age and divides by the total number of observations, n . The two curves provide a reasonable description of the data since the fitted density and the bars both peak in the first year of age and are both decreasing on average thereafter. The two curves do not change very much with cohort which reflects the fact that the data are generated with no cohort dependence. So the fitted density function $\hat{f}(a | c)$ seems to describe the exponential data well.

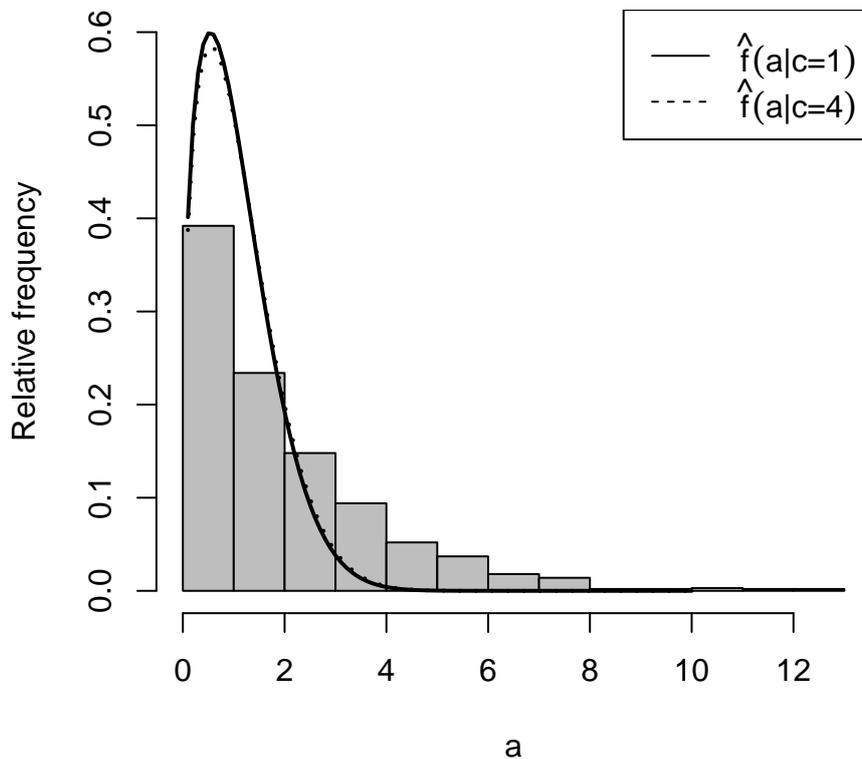


Figure 4.1: A histogram for the observed age at deaths for individuals $j = 1, 2, \dots, 1000$, which are drawn from an exponential distribution with parameter 0.5. The two curved lines are the fitted density functions for cohorts $c = 1$ and $c = 4$ for the model defined in equation (4.7).

The method described in this section can be applied to many other ex-

amples. We could fit models to real data or simulate some observations of the age at death with a dependence on cohort. The parameterisation of the APC model is determined by `h0` and `Xmake`. The parametric form for `h0` could be specified according to any of the models from Table 3.2. For example, the Exponential and Gompertz models would be coded respectively as

```
h0 <- function(a,alpha) { alpha[1] }
h0 <- function(a,alpha) { alpha[1] * exp(alpha[2]) }.
```

We could consider quadratic period and cohort terms if the cohort and period effects are nonlinear, so that

```
Xmake <- function(a,c) { cbind( c, c^2, a+c, (a+c)^2 ) }.
```

An existing R function `coxph` can also be used to estimate an APC model under a PH assumption. The APC model is estimated by maximising the Cox partial likelihood function described by Collett (2015, page 297). The partial likelihood function only depends on the rank order of the ages at death a_1, a_2, \dots, a_n (Collett 2015, page 66). A benefit of `apc.cont` is that it can be used to estimate both the base model and linear predictor simultaneously. Also, `apc.cont` accounts for the exact time-gaps between deaths by calculating the hazard function for all individuals at all intermediate times between birth and death.

4.2 Survival setting in discrete time

Recall from (2.7) that the Poisson model indexed in discrete time by age and cohort has an intensity function $\lambda_{a,c} = \kappa_c \cdot f_{a|c}$. In equation (3.16), we defined the linear predictor ψ_a as a product of a covariate vector \mathbf{x}_a^T and a parameter vector $\boldsymbol{\beta}$. If period and cohort are the only covariates that affect the survival distribution for an individual, so that ψ_a is only a function of p and c , the

probability function defined under the covariate model in (3.18) can be written as

$$f_{a,\psi_a} \equiv f_{a|c}.$$

In Chapter 3, we showed that, in discrete time, dependence in the APC model should be handled by specifying the survival distribution according to either a CLL functional or a DL functional. It is not necessary to consider both functionals and we choose to write discrete APC models in terms of the CLL functional. Recalling the hazard function and survivor function defined in equations (3.22) and (3.23), the Poisson intensity for deaths in the APC model is then written for $a = 1, 2, \dots$ as

$$\begin{aligned} \lambda_{a,c} &= \kappa_c \cdot f_{a|c} \\ &= \kappa_c \cdot S_{a-1|c} \cdot h_{a|c} \\ &= \kappa_c \prod_{u \leq a-1} (1 - h_{u,0})^{e^{\psi u}} \left[1 - (1 - h_{a,0})^{e^{\psi a}} \right]. \end{aligned} \quad (4.10)$$

For $a = 0$, the Poisson intensity is

$$\lambda_{0,c} = \kappa_c \left[1 - (1 - h_{0,0})^{e^{\psi 0}} \right].$$

Any valid APC model must ensure that $\kappa_c > 0$ for all c .

Let $a_j \in \mathbb{N}_0$ and $c_j \in \mathbb{Z}$ be the age-at-death and cohort respectively for individuals $j = 1, 2, \dots, n$. The data can be written in vectors as $\mathbf{a}^T = (a_1, a_2, \dots, a_n)$ and $\mathbf{c}^T = (c_1, c_2, \dots, c_n)$. While the linear predictor ψ_a consists of the parameter vector $\boldsymbol{\beta}$, the base hazard function $h_{a,0}$ consists of the parameter vector $\boldsymbol{\alpha}$. The likelihood function for the data can be written as

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \mathbf{a}, \mathbf{c}) = \prod_{j=1}^n f_{a_j|c_j}.$$

The logarithm of the likelihood function for the data is then

$$\sum_{j=1}^n \log f_{a_j|c_j}. \quad (4.11)$$

Similar to Section 4.1, the purpose of this section is to establish a method to fit $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to data by maximising the log likelihood function. To avoid

repetition, we do not discuss the method in as much detail. It is helpful to consider a simple dataset and a simple parametric form for $f_{a_j|c_j}$. We draw $n = 1000$ observations of the age-at-death from a geometric distribution with parameter $\pi_1 = 0.4$ and draw 1000 observations of cohort from a binomial distribution with trial parameter 5 and success probability 0.5:

```
n <- 1000; a <- rgeom(n,0.4); c <- rbinom(n,5,0.5).
```

Suppose that the baseline hazard function follows a discrete Weibull distribution so that

$$h_{a,0} = 1 - \frac{(\alpha_1)^{(a+1)\alpha_2}}{(\alpha_1)^{a\alpha_2}},$$

where $0 < \alpha_1 < 1$ and $\alpha_2 > 0$. The geometric and discrete Weibull distributions were defined in Table 3.3. We write the discrete Weibull hazard function in R as a function of data vector \mathbf{a} and parameter vector $\boldsymbol{\alpha}^T = (\alpha_1, \alpha_2)$:

```
h0 <- function(a,alpha) {
1- ( ( alpha[1])^((a+1)^alpha[2]) ) /
( alpha[1]^(a^alpha[2]) ) )
}
Nalpha <- 2.
```

The function `h0` creates a vector of baseline hazard functions evaluated at the observed ages at death,

$$\mathbf{h}_0^T = (h_{a_1,0}, h_{a_2,0}, \dots, h_{a_n,0}). \quad (4.12)$$

Suppose that we specify the covariate vector at age a_j as $\mathbf{x}_{a_j}^T = (c_j, p_j, p_j^2)$ with a corresponding parameter vector $\boldsymbol{\beta}^T = (\beta_1, \beta_2, \beta_3)$, where β_1 , β_2 and β_3 can take values on \mathbb{R} . The matrix of covariates at death is

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_{a_1}^T \\ \mathbf{x}_{a_2}^T \\ \vdots \\ \mathbf{x}_{a_n}^T \end{pmatrix} = \begin{pmatrix} c_1 & p_1 & p_1^2 \\ c_2 & p_2 & p_2^2 \\ \vdots & \vdots & \vdots \\ c_n & p_n & p_n^2 \end{pmatrix}.$$

The following R function, `Xmake`, takes the data vectors `a` and `c` as inputs and produces the matrix `X`:

```
Xmake <- function(a,c) { cbind( c, a+c, (a+c)^2 ) }
Nbeta <- 3.
```

In this example, the linear predictor at age a_j is $\psi_{a_j} = \mathbf{x}_{a_j}^T \boldsymbol{\beta} = \beta_1 c_j + \beta_2 p_j + \beta_3 p_j^2$ and the hazard function at age a_j is

$$\begin{aligned} h_{a_j|c_j} &= (1 - h_{a_j,0})^{\exp(\psi_{a_j})} \\ &= \left(\frac{(\alpha_1)^{(a_j+1)^{\alpha_2}}}{(\alpha_1)^{(a_j)^{\alpha_2}}} \right)^{\exp(\beta_1 c_j + \beta_2 p_j + \beta_3 p_j^2)}. \end{aligned} \quad (4.13)$$

The probability mass function implied by our choice of $h_{a_j,0}$ and ψ_{a_j} is obtained for $a = 1, 2, \dots$ by the definition

$$f_{a_j|c_j} = \prod_{u \leq a_j-1} (1 - h_{u|c_j}) \cdot h_{a_j|c_j}. \quad (4.14)$$

For $a = 0$, the probability mass function is obtained as

$$f_{0|c_j} = h_{0|c_j} = (1 - h_{0,0})^{\exp(\psi_0)} \quad (4.15)$$

$$= (\alpha_1)^{\exp(c_j(\beta_1 + \beta_2) + \beta_3 c_j^2)}. \quad (4.16)$$

We establish a function `apc.disc` in the statistical package R to calculate the log likelihood function for a particular value of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Details on this function are given in Appendix 9.2. For our particular specification of `n`, `a`, `c`, `h0`, `Nalpha`, `Xmake`, `Nbeta` and `scalePar`, we find the set of parameters that maximise the log likelihood by running the following optimisation procedure:

```
init <- rep(0.1, Nalpha+Nbeta)
fit <- optim( init, apc.disc,
control=list(fnscale=-1,maxit=5000) )
fit$par <- fit$par / scalePar.
```

After typing `fit` into R, we find that the maximum likelihood estimates of the parameters are $\hat{\alpha}_1 = 0.607$, $\hat{\alpha}_2 = 0.927$, $\hat{\beta}_1 = -0.019$, $\hat{\beta}_2 = 0.037$

and $\hat{\beta}_3 = 0.000$ to three decimal places. The corresponding value of the log likelihood function is -1699.540. Recalling (4.13) and (4.14), the fitted hazard function is

$$h_{a|c} = \left(\frac{(0.607)^{(a+1)^{0.927}}}{(0.607)^{a^{0.927}}} \right)^{\exp(-0.019c+0.037p)}. \quad (4.17)$$

The probability mass function can be obtained for $a = 1, 2, \dots$ by definition (4.14) as the product of $h_{a|c}$ and

$$\prod_{u=0}^{a-1} \left(1 - \left(\frac{(0.607)^{(u+1)^{0.927}}}{(0.607)^{u^{0.927}}} \right)^{\exp(-0.019c+0.037(u+c))} \right).$$

For $a = 0$, the probability mass function is obtained using (4.16) as $f_{0|c} = (0.607)^{\exp(0.017c)}$.

Since the discrete Weibull distribution is a special case of the geometric distribution for $\alpha_1 = 1 - \pi_1 = 0.6$ and $\alpha_2 = 1$, our estimates $\hat{\alpha}_1$ and $\hat{\alpha}_2$ provide a very accurate description of the data. The observed age at deaths were simulated from a geometric distribution with no period or cohort dependence, so the parameters $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ should be close to zero. Indeed, the β parameters are close to zero. The loss of accuracy experienced when fitting a continuous model in Section 4.1 did not occur in this discrete example.

In Figure 4.2, a bar plot is produced for the $n = 1000$ observed age at deaths drawn from a geometric distribution. Two fitted probability mass functions $\hat{f}_{a|1}$ and $\hat{f}_{a|4}$ are also plotted. The two curves provide a reasonable description of the data since the bars and the fitted density both peak in the first year of age and are both decreasing on average thereafter. The two curves do not change very much with cohort which reflects the fact that the data are generated with no cohort dependence. So the fitted mass function $\hat{f}_{a|c}$ seems to provide a good description of the geometric data.

The method described in this section can be applied to many other examples. The parameterisation of the APC model is determined by `h0` and `Xmake`. The parametric form for `h0` could be specified according to any of the models

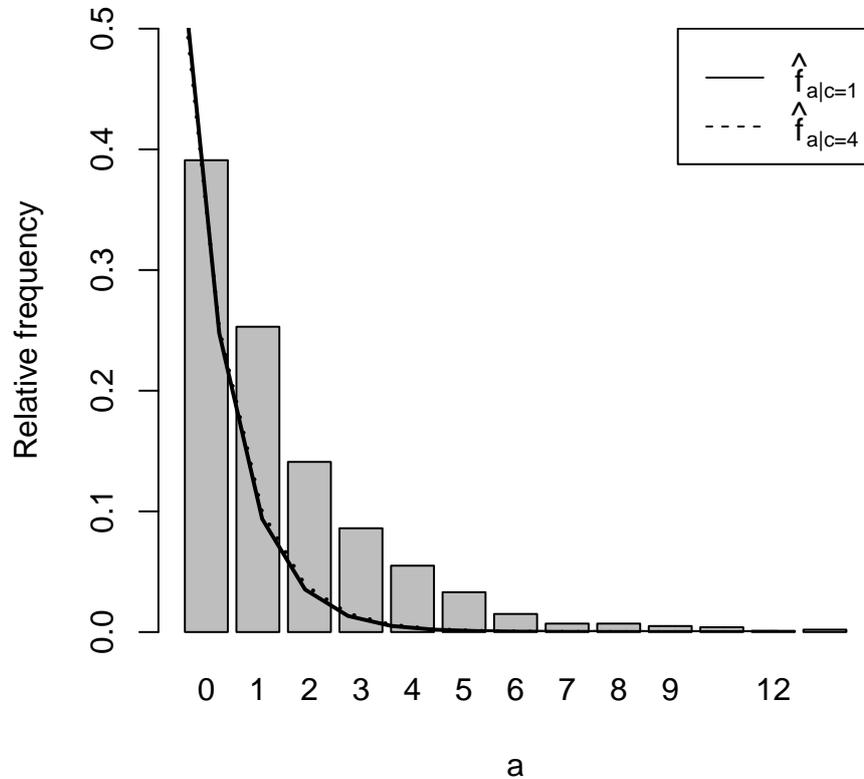


Figure 4.2: A bar plot for the observed age at deaths for individuals $j = 1, 2, \dots, 1000$, which are drawn from a geometric distribution with parameter 0.4. The two curved lines are the fitted probability mass functions for cohorts $c = 1$ and $c = 4$ for the model described in equation (4.17).

from Table 3.3. For example, the Geometric, Xekalaki and Salvia–Bollinger models would be coded respectively as

```
h0 <- function(a,alpha) { alpha[1] }
h0 <- function(a,alpha) { 1 / ( alpha[1] + (alpha[2]*a) ) }
h0 <- function(a,alpha) { alpha[1] / (a+1) }.
```

The function `Xmake` is specified in the same way as the case for continuous time described in Section 4.1.

4.3 Survival setting in the literature

We used the terms “birth” and “death” in Chapter 2 as unifying terms for an entry event and an exit event, respectively. This is helpful because, in many cases, entry and exit are at the birth and death of an individual (Kermack et al. 1934; Osmond and Gardner 1982; Choi et al. 2016). The entry event is very often at birth, but there are various exit events that have been studied in the APC literature. Death from any cause has been studied by Kermack et al. (1934) and Lee and Hsieh (1996). The most common exit event seems to be death due to cancer. There are examples of APC survival models with exit at death due to lung cancer (Stevens and Moolgavkar 1984; Peto et al. 1995), cervix cancer (Barrett 1973), bladder cancer (Barrett 1978; Osmond and Gardner 1982), breast cancer (Clayton and Schifflers 1987b; Choi et al. 2016) and prostate cancer (James and Segal 1982; Holford 1983). Frost (1940) studied deaths due to tuberculosis, and more recently Kramer et al. (2015) studied deaths due to heart disease.

Rather than exits at death, many researchers have studied exits at the diagnosis of a disease, especially cancer. There are examples of APC survival models with exit at the diagnosis of bladder cancer (Clayton and Schifflers 1987a), cervix cancer (Sasieni and Adams 2001), thyroid cancer (Liu et al. 2001), breast cancer (Moolgavkar et al. 1979; Leung et al. 2002), lung cancer (Murayama et al. 2006) and testis cancer (Carstensen 2007). Greenberg et al. (1950) is a landmark paper on APC survival modelling because they provide perhaps the first consideration of the effects of period and cohort on the hazard function. Greenberg et al. carried out a study for the diagnosis of syphilis. An “individual” in our framework is not necessarily a human. Dealler and Kent (1995) and Anderson et al. (1996) studied the diagnosis of Bovine Spongiform Encephalopathy (BSE) for cattle. These two papers on BSE are very important

in this thesis and are discussed in detail in Chapter 7. There are also examples of exit events that are not related to death or disease diagnosis such as homicide arrest (Fu 2008) and hip fracture (Jean et al. 2013).

A natural starting point for an APC analysis is to consider the fit of an independence model. An independence model means that the survival distribution is assumed not to vary with covariates such as period or cohort. In discrete time, we can use (4.10) to write the independence model as

$$\lambda_{a,c} = \kappa_c \cdot f_a.$$

The independence model may be deemed not to fit the data well, but it is still the natural starting point. If necessary, modifications are then made to the independence model to account for departures from independence due to period and cohort effects. In Section 4.2, we showed that modifications to allow for covariate effects of period and cohort should be made according to either a CLL or DL model. The probability function f_a was written in equation (3.5) as a product of S_{a-1} and h_a . When allowing for a dependence on period and cohort, the probability function $f_{a|c} \equiv f_{a,\psi_a}$ is a product of $S_{a-1|c}$ and $h_{a|c}$.

Letting $w_{a,c} = \kappa_c \cdot S_{a-1|c}$ be the expected number of individuals born in cohort c who are at risk of death at age a , the Poisson intensity for deaths which is a product of κ_c and $f_{a|c}$ can be rewritten in terms of the hazard function as

$$\lambda_{a,c} = w_{a,c} \cdot h_{a|c}.$$

In real studies, time is truly continuous, but data on the observed ages at death and periods at death are rounded down to the nearest and published in a contingency table categorised in terms of age and period. This form of data is discussed further in Section 6.1. Data are usually provided on both the number of deaths and the number at risk (Barrett 1973; Moolgavkar et al. 1979; Lee and Hsieh 1996). The value of cohort in the tables is assumed to be determined by the equation $c = p - a$. The number at risk is often converted to person-years to derive the total time exposed to risk of death at age a for those born in cohort c . Carstensen (2007, pages 3022–3023) shows how to convert

the population data to person-years. The number at risk or person-years are treated as known and the discrete Poisson intensity $\lambda_{a,c}$ is fitted to the rounded data by only estimating the hazard function $h_{a|c}$.

An independence model is often deemed to be unsuitable as a description of the Poisson intensity apparent in the age-by-period data. Recalling Table 3.5, the PH model can be written in discrete time as the following additive function:

$$\log h_{a|c} = \eta_a + \psi_a,$$

where $\eta_a = h_{a,0}$. Modifications to the discrete independence model $\lambda_{a,c}$ are usually made according to a PH assumption (Holford 1983; Liu et al. 2001; Kramer et al. 2015). In Table 3.5, we stated that the PH model is invalid in discrete time because, in discrete time, the hazard function is a probability and so it has a range on the interval $[0, 1]$. It is inappropriate to multiply the discrete hazard function by an arbitrary constant. The independence model should be parameterised in continuous time and modifications should be made to $h(a | c)$ according to equation (4.1).

The Kronecker delta is a function of two variables which takes value one if the variables are equal and zero otherwise. For variables u and v , the Kronecker delta can be written as

$$z_{u,v} = \begin{cases} 0 & \text{if } u \neq v \\ 1 & \text{if } u = v. \end{cases} \quad (4.18)$$

Suppose that a^- and a^+ represent the smallest and largest ages at death in the two contingency tables. Leung et al. (2002) and Jean et al. (2013) initially fitted an independence model parameterised with dummy variables so that

$$\log h_a = \sum_{u=a^-}^{a^+} z_{u,a} \cdot \alpha_u = \alpha_a.$$

The values of α_a are unconstrained. They dismissed the fit of the independence model and assumed that the effect of cohort or period on the logarithm of the

hazard function is linear. That is, they fitted the following models:

$$\log h_{a|c} = \alpha_a + \pi p,$$

$$\log h_{a|c} = \alpha_a + \gamma c.$$

The linear trend parameters π and γ are unconstrained real numbers.

Suppose that p^- and c^+ represent the smallest and largest periods at death in the two contingency tables, and let c^- and c^+ denote the smallest and largest observed cohorts. The fit of the two models with linear effects for period and cohort were also deemed to be unsuitable as a description of the Poisson intensity underlying the data. Both research groups made further modifications to their models and fitted the following three models to the contingency table data:

$$\log h_{a|c} = \alpha_a + \sum_{u=p^-}^{p^+} z_{u,p} \cdot \pi_u = \alpha_a + \pi_p, \quad (4.19)$$

$$\log h_{a|c} = \alpha_a + \sum_{u=c^-}^{c^+} z_{u,c} \cdot \gamma_u = \alpha_a + \gamma_c, \quad (4.20)$$

$$\log h_{a|c} = \alpha_a + \pi_p + \gamma_c. \quad (4.21)$$

The three models were also fitted in APC studies conducted by Holford (1983), Liu et al. (2001) and Kramer et al. (2015) to describe departures from the independence model. In a classic study, Kermack et al. (1934) fitted the model described in equation (4.20) to data on all-cause mortality. The final model described in equation (4.21) has received much attention in the literature because of a fundamental confounding problem. The problem is that it is not possible to identify the linear trends of α_a , π_p and γ_c . Numerous attempts have been made to produce meaningful estimates of the linear trend parameters such as the minimisation of a penalty function by Osmond and Gardner (1982) and the use of an intrinsic estimator by Yang and Land (2013, Chapter 5). In Section 5.3, we demonstrate the issue of confounding for the dummy variable parameterisation of the APC model.

Clayton and Schifflers (1987b) assigned dummy variables to describe the

base hazard function, but assumed that the effects of period and cohort on the base hazard are linear as opposed to nonlinear. Their model can be obtained by replacing π_p and γ_c with some linear trends πp and γc . This model has the same confounding issue as the model described in (4.21). To avoid confounding issues, Moolgavkar et al. (1979) replaced the period parameters in equation (4.21) with an interaction term in age and cohort. Their fitted model can be written as

$$\begin{aligned}\log h_{a|c} &= \alpha_a + \gamma_c + \sum_{u=a^-}^{a^+} z_{u,a} \cdot \xi_u \cdot \sum_{u=c^-}^{c^+} z_{u,c} \cdot \delta_u \\ &= \alpha_a + \gamma_c + \xi_a \delta_c.\end{aligned}$$

Lee and Hsieh (1996) modified the independence model by assigning a cubic polynomial to describe the effect of cohort:

$$\log h_{a|c} = \alpha_a + \gamma_1 c + \gamma_2 c^2 + \gamma_3 c^3.$$

In this thesis, we assume that deaths follow a Poisson model. While it is common for APC studies to adopt the Poisson assumption, there are some cases in which a different distribution is assumed for deaths. Barrett (1973) assumes that deaths follow a Neyman Type A model and fit the dummy variable parameterisation of the PH model from equation (4.21) to data on cervix cancer. Similarly, deaths are assumed to follow a Negative Binomial model by Jean et al. (2013) in a study of data on hip fractures.

4.4 Regression setting

An important distinction we make in this chapter is between two settings for APC modelling: survival and regression. This distinction has not been made explicit in the literature. The distinction between survival and regression is important because the issues discussed in Chapters 6 and 7 are only relevant to

the survival setting. The issue of misrounding does not apply to the regression setting. The fundamental problem of confounding is explained in Chapter 5 and is compared for the survival and regression settings. In this chapter, we have so far discussed APC models in the survival setting. The purpose of this section is to define the APC regression model and to present some examples of regression models from the literature. We use the term “regression model” to refer to the broad class of statistical models known as generalised linear models (GLMs) (Nelder and Wedderburn 1972).

Consider a response variable Y conditional on a general set of covariates $\mathbf{x}^T = (x_1, \dots, x_q)$ with the corresponding parameter vector $\boldsymbol{\beta}^T = (\beta_1, \dots, \beta_q)$. An introductory text on statistical modelling will usually start with a basic linear regression model (Dobson and Barnett 2008, page 45):

$$Y \mid \mathbf{x}^T \sim \text{Normal}(\mathbf{x}^T \boldsymbol{\beta}, \sigma_Y^2). \quad (4.22)$$

A GLM is a generalisation of the regression model in (4.22) to allow for various response distributions including the Normal distribution. A suitable link function, denoted by g , should be chosen to constrain a parameter such as the mean $\mathbb{E}[Y]$ to an appropriate interval (Dobson and Barnett 2008, pages 51–52):

$$g(\mathbb{E}[Y]) = \mathbf{x}^T \boldsymbol{\beta}.$$

While an identity link $g(\mathbb{E}[Y]) = \mathbb{E}[Y]$ is suitable for a normally distributed response, it is not so appropriate for a Poisson response and a logarithmic link $g(\mathbb{E}[Y]) = \log(\mathbb{E}[Y])$ should be used to constrain the response mean to be non-negative.

An APC regression model has a response variable that is not related to a time-gap between birth and death, and this response can be measured multiple times while the individual is alive. In an APC survival model, the response variable is the age-at-death, A , for an individual and the response distribution is a probability distribution for age that is considered to be conditional on period and cohort. The age-at-death can only be measured once. The response distribution for regression modelling is considered to be conditional on age,

period and cohort. So age is a part of the response distribution in survival modelling, whereas age is a covariate in the regression model. Period and cohort are covariates in any APC model.

There are significantly fewer examples for APC regression models in the literature than for APC survival models. APC regression models have been considered in a range of subject areas such as in the study of obesity (Diouf et al. 2010), in the study of religious attendance (Hayward and Krause 2015), and in the study of deer hunting licenses (Winkler and Warnke 2013). In the survival setting, the response character is determined by the time concept since in continuous time, for example, the age-at-death is continuous and defined on the interval \mathbb{R}^+ .

The time concept is still very important in APC regression modelling because time impacts on the covariates age, period and cohort. Time does not influence the response distribution or link function of the regression model. In a survival context, the time concept influences both the response distribution and link function due to the response variable being a statistic based on age. In a survival context, the link function refers to the functional and we concluded in Chapter 3 that only the PH functional was suitable to describe the survival distribution in continuous time, while only the CLL or DL functionals were suitable in discrete time.

Recalling the linear predictor from (3.10), the function `glm` can be used in the statistical package R to fit a GLM to data:

$$\text{glm}(\text{formula}, \text{family}). \quad (4.23)$$

The argument `formula` is a formula object with the response variable Y on the left and the covariates on the right. The argument `family` is an argument for both the distribution of Y and the link function g . The `family` argument can be specified for example as `poisson`, `gaussian` and `binomial`.

For example, Xu et al. (1995) studied an individual's expiratory volume in one second (FEV1) measurable in litres. The response variable has a range

on the positive real numbers \mathbb{R}^+ . They assumed that the response variable was normally distributed and chose an identity link function. They assigned a quadratic polynomial to describe the effect of age, assigned dummy variables to describe the effect of period, and proposed the use of an interaction term between age and cohort, so that $\mathbb{E}[Y] = \mu + \alpha_1 a + \alpha_2 a^2 + \pi_p + \gamma_c$. They fitted the following APC model:

$$\text{glm}(Y \sim \text{poly}(a,2)+\text{factor}(p)+a:\text{factor}(c), \text{gaussian}).$$

Schomerus et al. (2015) studied the desire for social distance from a person with depression and schizophrenia. The desire for social distance was measured as a sum-score out of 28. This means that the response variable has a range of $\{0, 1, 2, \dots, 28\}$. They assumed that the response variable is normally distributed and that the effects of age and cohort were linear on the desire for social distance, but that period had a nonlinear effect. That is, the response variable and covariates were related by the equation $\mathbb{E}[Y] = \mu + \alpha a + \pi_p + c$. They fitted the following APC model:

$$\text{glm}(Y \sim a+\text{factor}(p)+c, \text{gaussian}).$$

Suppose that an experiment is conducted where the outcome is binary such that there is a success event and a failure event. The experiment is repeated for N trials. Let Y_ℓ be the outcome of the ℓ th trial which takes a value of one for success and a value of zero for failure. Let $Z = \sum_{\ell=1}^N Y_\ell$ be a count for the total number of success experiments. Diouf et al. (2010) conducted a study on obesity. The response variable in their study is binary where the success event is the observation of obesity. Similarly, Schwadel (2011) conducted a study on religious attendance and the response variable is binary. In this second case, the success event is the observation of religious attendance. Diouf et al. (2010) and Schwadel (2011) assumed that Z followed a Binomial distribution with trial parameter N and a probability parameter θ . Dummy variables were assigned to age, period and cohort and a logit link function was chosen to

describe the relationship between θ and covariates, so that

$$Z \mid a, p, c \sim \text{Binomial} \left(N, \theta = \frac{1}{1 + e^{-\mathbf{x}^T \boldsymbol{\beta}}} \right).$$

They both fitted the following APC model:

```
glm( cbind(Z, N-Z) ~ factor(a)+factor(p)+factor(c), binomial ).
```

When `family` is set to `binomial`, the response variable on the left hand side of the `formula` object must be specified as a data frame with the first column corresponding to the number of success events and the second column corresponding to the number of failure events.

4.5 Conclusion

In Chapter 3, we explained that the survival distribution of a continuous APC model, $f(a \mid c)$, should be specified according to a PH functional. We also explained that the survival distribution of a discrete APC model, $f_{a|c}$, should be specified according to either a CLL or DL functional. In this chapter, we presented two new functions coded in the statistical package R to fit the survival distribution of an APC model to data by maximising a likelihood function. The function `apc.cont` specifies a PH formula for $f(a \mid c)$, while `apc.disc` specifies a CLL formula for $f_{a|c}$.

Both functions estimate the baseline survival distribution and linear predictor simultaneously. The data consist of $j = 1, 2, \dots, 1000$ random simulations of the age-at-death and cohort, which are real numbers in continuous time and integers in discrete time. The likelihood function is a product of the 1000 individual probability functions, $f(a_j \mid c_j)$ and $f_{a_j|c_j}$. A benefit of using `apc.cont` and `apc.disc` is that, to calculate the probability function for each individual, they both calculate the hazard function for each individual at all intermediate times between birth and death. So our estimation method

accounts for the entire survival experience of individuals and considers the exact time-gaps between deaths. Both `apc.cont` and `apc.disc` seem to be valid methods for the estimation of $f(a | c)$ and $f_{a|c}$ since our fitted models $\hat{f}(a | c)$ and $\hat{f}_{a|c}$ both gave a reasonable fit to the simulated data.

The existing R function `coxph` fits survival models specified according to a PH functional by maximising a partial likelihood function. The partial likelihood function for age-dependent covariates is presented by Collett (2015, page 297). We can use `coxph` to fit APC models by specifying period and/or cohort as covariates. A drawback with using `coxph` to fit APC models is that the parameter estimates only depend on the rank order of the ages at death, so the exact time-gaps between deaths are not considered.

We also explained that the way of handling dependence in the APC literature is not consistent with our findings. Data are typically rounded and published in a contingency table categorised in terms of age and period. Statisticians usually fit a discrete independence model $\lambda_{a,c} = \kappa_c \cdot f_a$ to describe the age-by-period data. Modifications are then made to the independence model using a PH functional as opposed to a CLL or DL functional. In Chapter 6, we explain that, since the discrete independence model is equivalent to a continuous independence model discretised over an age-by-cohort region in the Lexis diagram, the age-by-period data are used for model fitting as if they are rounded age-by-cohort. We then investigate whether the need for modifications to the independence model is only apparent in the literature due to the misrounded treatment of the data.

We also distinguished between survival and regression as two settings for APC modelling. In a survival context, an APC model considers how the age-at-death distribution varies with period and cohort. In a regression context, an APC model considers how some response variable, which is unrelated to the death of an individual, varies with age, period and cohort. The confounding issue discussed in Chapter 5 is relevant to both regression and survival models. However, the issues discussed in Chapters 6 and 7 are only relevant to the

survival models.

Chapter 5

Confounding in age-period-cohort models

It is important to consider issues of confounding when modifications are made to the independence models $\lambda(a, c)$ and $\lambda_{a,c}$. Johnston et al. (2018, page 1958) define confounding as the situation in which the relationship between two variables is distorted because of a strong relationship between one or two of the variables and a third variable included in the analysis. Since age, period and cohort are linearly determined by the equation $\text{cohort} = \text{period} - \text{age}$, the relationship between the age-at-death distribution and period is distorted by the inclusion of cohort in a survival model. The confounding concept can be extended to a situation with four variables to say that the relationship between age, period and a response variable is distorted by the inclusion of cohort in a regression model. Thus, cohort is confounded with age and period (Rodgers 1982, page 775).

APC models suffer from the most severe case of confounding, the issue of identifiability, in which the relationship between variables cannot be identified uniquely (Mason et al. 1973; Holford 1983; Smith and Wakefield 2016). In APC literature, confounding is synonymous with a lack of identifiability. The confounding issue in APC models is that it is impossible to separate parameters

for linear terms in age, period and cohort (Mason et al. 1973; Rodgers 1982). It is also impossible to separate parameters for quadratic terms and two-way interaction terms in age, period and cohort (Rodgers 1982; Fienberg and Mason 1985).

In continuous and discrete time, it is only necessary to consider two variables in the formulation of an APC model. Identifiability issues are not exclusive to APC modelling and occur in other literature such as errors-in-variables modelling (Draper and Smith 1998, page 90) and analysis of variance (Draper and Smith 1998, pages 474–478) as a result of having too many parameters to estimate given the available data. An estimate of the model parameters can still be obtained by removing parameters until the number of parameters is equal to the number of data observations. Parameters of APC models cannot be identified even after dealing with overparameterisation because there are still linear dependencies between variables in the model formulation. One approach to dealing with APC confounding is to study models parameterised with only two of age, period and cohort (Clayton and Schifflers 1987a; Lee and Hsieh 1996). Glenn (2005) and Smith and Wakefield (2016) summarise the approaches used to identify the parameters of APC models.

In Section 5.1, we explain the APC confounding issue in more detail. Confounding is compared for models with interaction terms and without interaction terms. In Sections 5.2 and 5.3, we introduce data and look at confounding issues through the design matrix. A distinction is made between confounding arising due to overparameterisation (Draper and Smith 1998, page 474) and confounding arising due to the APC linear identity. In Section 5.2, we explore confounding for an APC regression model parameterised with orthogonal polynomials. In Section 5.3, we demonstrate confounding for the APC survival model parameterised with dummy variables in (4.21). We show that confounding issues arise between parameters for first-order orthogonal polynomials, but it is not so clear how issues arise for factor variables. We highlight a special case in which the factor model has no confounding issue in relation to the APC linear identity.

5.1 Confounding

In general, the purpose of APC modelling is to separate the effects of age, period and cohort on some phenomenon. However, it is impossible to separate the three effects into distinct contributions because of a fundamental confounding problem. Since age, period and cohort are linearly determined by the equation $\text{cohort} = \text{period} - \text{age}$, the relationship between the age-at-death distribution and period is distorted by the inclusion of cohort in a survival model. The relationship between age, period and a response variable is distorted by the inclusion of cohort in a regression model.

A model formulated in terms of age, period and cohort is equivalent to a model formulated in terms of age and period or age and cohort. Let $\phi(a, p, c)$ denote a function in age, period and cohort. Letting u_1 , u_2 and u_3 be real numbers, a linear function of age, period and cohort can be written as

$$\phi(a, p, c) = u_1a + u_2p + u_3c. \quad (5.1)$$

By substituting $c = p - a$ and $p = a + c$ into (5.1), the linear function can be written equivalently as $\phi(a, p, c) = b_1a + b_2p$ or $\phi(a, p, c) = e_1a + e_2c$, where

$$b_1 = u_1 - u_3, \quad e_1 = u_1 + u_2, \quad b_2 = e_2 = u_2 + u_3. \quad (5.2)$$

Equation (5.2) shows that each parameter for a linear term is a compound parameter that accounts for contributions from two variables rather than one variable. In particular, e_1 is not the true coefficient for age and is instead a summation of the coefficients for age and period. The period coefficient is absorbed by the coefficients for age and cohort to form the compound parameters e_1 and e_2 . The compound parameter e_2 was discussed by Clayton and Schifflers (1987b, page 474) and named as net drift.

The APC models presented in equations (4.4) and (4.13) do not suffer from confounding, because $h(a_j | c_j)$ and $h_{a_j|c_j}$ cannot be written in terms of

a linear combination of a_j , p_j and c_j . In the first case, the logarithm of the hazard function is

$$\log h(a_j | c_j) = \gamma(\alpha_1, \alpha_2) + (\alpha_1 - 1) \log a_j + \beta_1 c_j + \beta_2 p_j, \quad (5.3)$$

where $\gamma(\alpha_1, \alpha_2)$ is a constant determined by the values of α_1 and α_2 .

There can also be confounding in APC models due to the inclusion of quadratic terms and second-order interaction terms for age, period and cohort (Rodgers 1982, page 783). Letting u_1, \dots, u_6 be real numbers, consider the nonlinear function

$$\phi(a, p, c) = u_1 a^2 + u_2 p^2 + u_3 c^2 + u_4 ap + u_5 ac + u_6 pc. \quad (5.4)$$

By substituting $c = p - a$ and $p = a + c$ into (5.4), the nonlinear function can be written as $\phi(a, p, c) = b_1 a^2 + b_2 p^2 + b_3 ap$ or $\phi(a, p, c) = e_1 a^2 + e_2 c^2 + e_3 ac$, where

$$b_1 = u_1 + u_3 - u_5, \quad e_1 = u_1 + u_2 + u_4, \quad b_2 = e_2 = u_2 + u_3 + u_6 \quad (5.5)$$

$$b_3 = u_4 + u_5 - u_6 - 2u_3, \quad e_3 = u_4 + u_5 + u_6 + 2u_2.$$

Parameter e_2 is not the true coefficient for a quadratic cohort term, but is instead a summation of coefficients for the quadratic period term, the quadratic cohort term and the interaction term between period and cohort. The net drift parameter in (5.2), written as $u_2 + u_3$, which accounts for the first-order effects of period and cohort, is comparable to the parameter written as $u_2 + u_3 + u_6$ in (5.5) which accounts for the second-order effects of period and cohort.

If the three two-way interaction terms are not included in the model, then the parameters for quadratic terms can be identified uniquely. The APC confounding problem can be defined more generally for an N th order case in which it is impossible to identify uniquely the coefficients for N th order terms in age, period and cohort when the model contains all N th order interaction terms. The confounding problem was discussed for cases $N = 1$ and $N = 2$, but can also be applied to cases $N = 3, 4, 5, \dots$

5.2 Confounding for polynomial functions

In this section, we introduce data and look at confounding issues through the design matrix. A design matrix is a matrix of coefficients for parameters with each column corresponding to a different parameter and each row corresponding to a different observation index. The design matrix for a generalised linear model is presented by Dobson and Barnett (2008, page 37). The observation index used in this chapter is displayed in Figure 5.1.

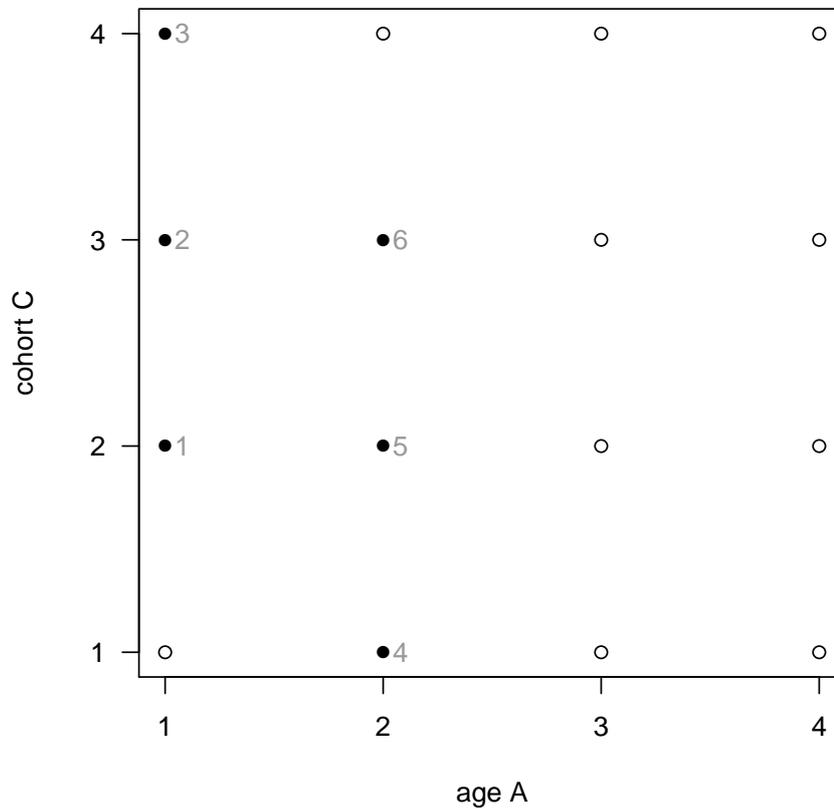


Figure 5.1: A Lexis diagram in age-cohort space for discrete time. Age, period and cohort are only defined at the vertices. Period is calculated as $p = a + c - 2$. The six filled vertices represent cells of a hypothetical age-by-period contingency table for deaths. For example, cell two corresponds to $a = 1$, $p = 2$ and $c = 3$. This display of age-by-period cells in age-cohort space is similar to the data in Table 7.1.

Suppose that age a and period p are defined on the sets $\{1, 2, \dots, L^{(a)}\}$ and $\{1, 2, \dots, L^{(p)}\}$ respectively, and that data on death counts are published in the form of a two-way contingency table categorised in terms of age and period. If the number of levels for age and period is $L^{(a)} = 2$ and $L^{(p)} = 3$, then there are $L^{(a)}L^{(p)} = 6$ cells of the age-by-period table. The number of cells is chosen to be small for illustrative purposes, but the number of cells could be chosen to be very large. Cohort is redefined as $c = p - a + L^{(a)}$ so that cohort is defined on the set $\{1, 2, \dots, L^{(c)}\}$ (Barrett 1973; Clayton and Schifflers 1987b). The number of levels for cohort is equal to

$$L^{(c)} = L^{(a)} + L^{(p)} - 1. \quad (5.6)$$

An age-by-period table with two age levels and three period levels has $L^{(c)} = 4$ levels for cohort. In Figure 5.1, we assign a numeric label to each cell of this hypothetical age-by-period table and depict the corresponding value of age, period and cohort for each cell.

Consider a response variable Y and a covariate x that is categorical with $L^{(x)}$ levels. Letting $\psi_q(x)$ denote a q th order polynomial in $x \in \{1, 2, \dots, L^{(x)}\}$, we assume in this section that polynomials are orthogonal such that

$$\sum_{x=1}^{L^{(x)}} \psi_{q_1}(x) \cdot \psi_{q_2}(x) = 0 \quad (5.7)$$

for $q_1 \neq q_2$ and for $q_1, q_2 \leq L^{(x)} - 1$ (Draper and Smith 1998, page 462). The levels of age, period and cohort are $L^{(a)} = 2$, $L^{(p)} = 3$ and $L^{(c)} = 4$ respectively, so we only need to consider polynomials up to order three. Letting $\bar{x} = \frac{L^{(x)}+1}{2}$ be the mean of x , Draper and Smith (1998, page 466) present the following formulae for orthogonal polynomials of orders one to three:

$$\begin{aligned} \psi_1(x) &= \lambda_1^{(x)}(x - \bar{x}) \\ \psi_2(x) &= \lambda_2^{(x)} \left([x - \bar{x}]^2 - \frac{[L^{(x)}]^2 - 1}{12} \right) \\ \psi_3(x) &= \lambda_3^{(x)} \left([x - \bar{x}]^3 - \frac{[3[L^{(x)}]^2 - 7][x - \bar{x}]}{20} \right). \end{aligned} \quad (5.8)$$

A polynomial of order zero is always equal to one and the variables $\lambda_1^{(x)}$, $\lambda_2^{(x)}$ and $\lambda_3^{(x)}$ are constants chosen to ensure $\psi_q(x)$ takes an integer value.

A regression model for Y parameterised with orthogonal polynomials in a has the vector form

$$\mathbf{a}^T \boldsymbol{\alpha} = \sum_{u=1}^{L(a)} \alpha_{u-1} \psi_{u-1}(a)$$

where $\boldsymbol{\alpha}^T = (\alpha_0 \dots \alpha_{L(a)-1})$ and $\mathbf{a}^T = (\psi_0(a) \dots \psi_{L(a)-1}(a))$. Letting $a(l)$ denote the level of a for the l th cell of the age-by-period table, the design matrix of this regression model can be written as

$$\mathbf{A} = (a_{u,v}) \quad \text{with} \quad a_{u,v} = \psi_{v-1}(a(u)), \quad (5.9)$$

where $u \in \{1, 2, \dots, L^{(a)}L^{(p)}\}$ and $v \in \{1, 2, \dots, L^{(a)}\}$. For example, we use definition (5.8) to calculate the the second row of \mathbf{A} as

$$(\psi_0(a(2)) \quad \psi_1(a(2))) = (\psi_0(1) \quad \psi_1(1)) = (1 \quad -1).$$

A regression model for Y parameterised with orthogonal polynomials in a , p and c has the vector form

$$\mathbf{x}^T \boldsymbol{\beta} = \sum_{u=1}^{L(a)} \alpha_{u-1} \psi_{u-1}(a) + \sum_{v=1}^{L(p)} \pi_{v-1} \psi_{v-1}(p) + \sum_{w=1}^{L(c)} \gamma_{w-1} \psi_{w-1}(c) \quad (5.10)$$

where $\boldsymbol{\beta}^T = (\alpha_0 \dots \alpha_{L(a)-1} \quad \pi_0 \dots \pi_{L(p)-1} \quad \gamma_0 \dots \gamma_{L(c)-1})$ and \mathbf{x}^T is equal to

$$(\psi_0(a) \quad \dots \quad \psi_{L(a)-1}(a) \quad \psi_0(p) \quad \dots \quad \psi_{L(p)-1}(p) \quad \psi_0(c) \quad \dots \quad \psi_{L(c)-1}(c)).$$

By defining the design matrix for each covariate according to (5.9) with the range of u staying the same, and merging the three resulting matrices \mathbf{A} , \mathbf{P} and \mathbf{C} to form a full design matrix, the design matrix of the APC regression

model parameterised with orthogonal polynomials is derived as

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \mathbf{x}_3^T \\ \mathbf{x}_4^T \\ \mathbf{x}_5^T \\ \mathbf{x}_6^T \end{pmatrix} = \begin{pmatrix} \mathbf{A} & \mathbf{P} & \mathbf{C} \end{pmatrix} = \begin{pmatrix} 1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & 3 \\ 1 & -1 & 1 & 0 & -2 & 1 & 1 & -1 & -3 \\ 1 & -1 & 1 & 1 & 1 & 1 & 3 & 1 & 1 \\ 1 & 1 & 1 & -1 & 1 & 1 & -3 & 1 & -1 \\ 1 & 1 & 1 & 0 & -2 & 1 & -1 & -1 & 3 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -3 \end{pmatrix}.$$

The cross-product of a design matrix, \mathbf{X} , is written as $\mathbf{X}^T\mathbf{X}$. A model has an identifiability issue if the cross-product of its design matrix has a zero determinant (Draper and Smith 1998, Chapter 16). A zero determinant indicates that there are linear dependencies between columns of the design matrix. The cross-product for the design matrix of the orthogonal APC regression model has a zero determinant because of overparameterisation, that is, there are more parameters than observations (Draper and Smith 1998, page 474). We choose to set $\pi_0 = \gamma_0 = \gamma_3 = 0$ to remove three columns from the full design matrix and derive the following reduced matrix:

$$\mathbf{X}_r = \begin{pmatrix} \psi_1 & \psi_1 & \psi_1 \\ 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & -1 & 0 & -2 & 1 & -1 \\ 1 & -1 & 1 & 1 & 3 & 1 \\ 1 & 1 & -1 & 1 & -3 & 1 \\ 1 & 1 & 0 & -2 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & -1 \end{pmatrix}.$$

The number of observations is equal to the number of remaining model parameters, and, in theory, the cross-product $\mathbf{X}_r^T\mathbf{X}_r$ should not have a zero determinant. However, the determinant is still zero because there is a linear dependency between columns of \mathbf{X}_r relating to the first-order polynomials in age, period and cohort. Specifically, the following equation holds true for all

$l \in \{1, 2, \dots, L^{(a)}L^{(p)}\}$:

$$\psi_1(a(l)) + \psi_1(c(l)) - 2\psi_1(p(l)) = 0. \quad (5.11)$$

For example, the quantity in (5.11) can be calculated for the second row of the design matrix as

$$\begin{aligned} \psi_1(a(2)) + \psi_1(c(2)) - 2\psi_1(p(2)) &= \psi_1(1) + \psi_1(3) - 2\psi_1(2) \\ &= -1 + 1 - 0 \\ &= 0. \end{aligned}$$

There are no further linear dependencies between columns after setting one of α_1 , π_1 and γ_1 equal to zero to produce a final design matrix, \mathbf{X}_{r2} . Notice that, it is not necessary to set $\gamma_3 = 0$ once removing two intercept parameters and one parameter for a first-order polynomial.

The number of linearly independent parameters for model (5.10), which we label as ρ_{orth} , is equal to the number of parameters minus the number of parameter constraints. Since it is necessary to remove three columns from the original design matrix to derive a final matrix without linear dependencies and the number of levels for cohort is defined according to (5.6), we can write ρ_{orth} in terms of the number of levels for age and period as

$$\rho_{\text{orth}} = \left(L^{(a)} + L^{(p)} + L^{(c)} \right) - 3 = 2 \left(L^{(a)} + L^{(p)} \right) - 4. \quad (5.12)$$

Two constraints were imposed to remove the linear dependencies caused by overparameterisation. The third constraint is non-standard for general regression modelling and occurs in the context of APC modelling due to the linear identity.

The number of degrees of freedom for the orthogonal APC regression model, which we label as δ_{orth} and which is equal to the number of observations minus the number of linearly independent parameters, can be written as

$$\delta_{\text{orth}} = L^{(a)}L^{(p)} - \left[2 \cdot \left(L^{(a)} + L^{(p)} \right) - 4 \right] = \left(L^{(a)} - 2 \right) \left(L^{(p)} - 2 \right). \quad (5.13)$$

The equation (5.13) only holds for levels of age and period greater than one. The orthogonal APC regression model has six linearly independent parameters to be estimated given data for six cells of an age-by-period table, so there are zero degrees of freedom. Even if the number of initial parameters is substantially lower than the number of cells in an age-by-period table, such that the number of degrees of freedom is large, there will still be three linear dependencies in the model design matrix.

5.3 Confounding for factor variables

The Kronecker delta function was defined in equation (4.18). In this section, we will use $z_{u,a}$, $x_{v,p}^{(p)}$ and $x_{w,c}^{(c)}$ to denote Kronecker delta functions for age, period and cohort. Supposing that we incorporate an intercept to the APC survival model parameterised with dummy variables in (4.21), the base model can be written in vector form as

$$\log h_{a|c} = \mu + \alpha_a = \mathbf{z}^T \boldsymbol{\alpha} \quad (5.14)$$

where

$$\mathbf{z}^T = \left(1 \quad z_{1,a} \quad z_{2,a} \quad \cdots \quad z_{L^{(a)},a} \right)$$

and $\boldsymbol{\alpha}^T = \left(\mu \quad \alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_{L^{(a)}} \right)$. Let \mathbf{Z} denote the design matrix for the base model and suppose that $a(l)$ is the level of age for the l th row of \mathbf{Z} . If $\mathbf{e}_{a(l)}$ is a vector of length $L^{(a)}$ with a one in the $a(l)$ th place and zeros elsewhere, then the l th row of \mathbf{Z} is

$$\mathbf{z}_l^T = \left(1 \quad \mathbf{e}_{a(l)}^T \right). \quad (5.15)$$

For example, we can write $\mathbf{e}_{a(2)}^T = \left(1 \quad 0 \right)$ so that the second row of the design

matrix is $\mathbf{z}_2^T = (1 \ 1 \ 0)$. The full design matrix for the base model is

$$\mathbf{Z} = \begin{pmatrix} \mathbf{z}_1^T \\ \mathbf{z}_2^T \\ \mathbf{z}_3^T \\ \mathbf{z}_4^T \\ \mathbf{z}_5^T \\ \mathbf{z}_6^T \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}.$$

The cross-product $\mathbf{Z}^T \mathbf{Z}$ has a zero determinant because, for all l , there is a linear dependency between columns relating to the intercept parameter:

$$\sum_{u=1}^{L^{(a)}} z_{u,a(l)} = 1. \quad (5.16)$$

For example, the second row of the design matrix has a dependency $z_{1,1} + z_{2,1} = 1 + 0 = 1$. This dependency is standard for analysis of variance and is demonstrated by Draper and Smith (1998, page 478). We choose to set $\alpha_1 = 0$ to remove a column from \mathbf{Z} and derive the following reduced matrix:

$$\mathbf{Z}_r = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

The determinant of $\mathbf{Z}_r^T \mathbf{Z}_r$ is non-zero. The number of degrees of freedom for the base model, which we denote as δ_{base} , is calculated similarly to (5.13) as

$$\delta_{\text{base}} = L^{(a)} (L^{(p)} - 1).$$

The base model has four degrees of freedom so the design matrix for the systematic part of the APC model can consist of four columns at most if the base model and systematic part are to be estimated simultaneously.

Letting $x_{v,p}^{(p)}$ and $x_{w,c}^{(c)}$ also be Kronecker delta functions, the factor APC survival model has a systematic part that can be written in vector form as

$$\mathbf{x}^T \boldsymbol{\beta} = \pi_p + \gamma_c \quad (5.17)$$

where

$$\mathbf{x}^T = \left(x_{1,p}^{(p)} \quad x_{2,p}^{(p)} \quad \cdots \quad x_{L^{(p)},p}^{(p)} \quad x_{1,c}^{(c)} \quad x_{2,c}^{(c)} \quad \cdots \quad x_{L^{(c)},c}^{(c)} \right)$$

and $\boldsymbol{\beta}^T = \left(\pi_1 \quad \pi_2 \quad \cdots \quad \pi_{L^{(p)}} \quad \gamma_1 \quad \gamma_2 \quad \cdots \quad \gamma_{L^{(c)}} \right)$. Letting \mathbf{X} denote the design matrix for the systematic part of the model, the l th row of \mathbf{X} can be written similarly to (5.15) as

$$\mathbf{x}_l^T = \left(\mathbf{e}_{p(l)}^T \quad \mathbf{e}_{c(l)}^T \right). \quad (5.18)$$

For example, the second row of \mathbf{X} corresponds to $p = 2$ and $c = 3$ so that $\mathbf{e}_{p(2)}^T = (0 \ 1 \ 0)$ and $\mathbf{e}_{c(2)}^T = (0 \ 0 \ 1 \ 0)$. The full design matrix for the systematic part of the model is

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \mathbf{x}_3^T \\ \mathbf{x}_4^T \\ \mathbf{x}_5^T \\ \mathbf{x}_6^T \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

The cross-product $\mathbf{X}^T \mathbf{X}$ has a zero determinant because the systematic part has more parameters than there are observations. There is also a linear dependency similar to (5.16) for each covariate because, for all l ,

$$\sum_{v=1}^{L^{(p)}} x_{v,p(l)}^{(p)} = \sum_{w=1}^{L^{(c)}} x_{w,c(l)}^{(c)} = 1.$$

We stated that, if the base model and systematic part are both to be estimated, then \mathbf{X} should have at most four columns because $\delta_{\text{base}} = 4$. We choose to set $\pi_1 = \gamma_1 = \gamma_2 = 0$ to remove three columns from the design matrix for the

systematic part and derive the following reduced matrix:

$$\mathbf{X}_r = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}.$$

We merge the matrices \mathbf{Z}_r and \mathbf{X}_r to form a design matrix for the full factor APC survival model, so that

$$\mathbf{M}_r = (\mathbf{Z}_r, \mathbf{X}_r) = \left(\begin{array}{cc|cccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \end{array} \right).$$

There is no confounding issue because $\mathbf{M}_r^T \mathbf{M}_r$ has a non-zero determinant.

In Section 5.2, we explained that confounding arises in an orthogonal APC regression model due a linear dependency between first-order polynomials in age, period and cohort. This confounding is a consequence of the linear identity between age, period and cohort. However, the factor APC survival model described in this section has no confounding issue in relation to the linear identity. Here, confounding in relation to the APC linear identity depends on the number of levels for age and period in the age-by-period table because confounding only arises if $L^{(a)}$ and $L^{(p)}$ are both greater than two.

The factor APC survival model has an initial design matrix $\mathbf{M} = (\mathbf{Z}, \mathbf{X})$. Each row of the design matrix corresponds to a cell in an age-by-period contingency table with a total of $L^{(a)}L^{(p)}$ cells. Each column of the design matrix

corresponds to a parameter so the total number of model parameters is

$$L^{(a)} + L^{(p)} + L^{(c)} + 1 = 2 \left(L^{(a)} + L^{(p)} \right).$$

We must set three parameters equal to zero to remove the three sum-to-one linear dependencies that are standard for analysis of variance (Draper and Smith 1998, page 478). The number of remaining parameters is then $2 \left(L^{(a)} + L^{(p)} \right) - 3$. The parameters should be identifiable if the number of cells is greater than or equal to the number of remaining parameters such that

$$L^{(a)}L^{(p)} \geq 2 \left(L^{(a)} + L^{(p)} \right) - 3.$$

This inequality implies that

$$\left(L^{(a)} - 2 \right) \left(L^{(p)} - 2 \right) \geq 1. \quad (5.19)$$

However, if $L^{(a)}$ and $L^{(p)}$ are both greater than two so that (5.19) is satisfied, then $\mathbf{M}^T\mathbf{M}$ still has a zero determinant because there is an additional linear dependency that is attributable to the APC linear identity. A fourth constraint is then necessary to identify the model parameters. If at least one of $L^{(a)}$ and $L^{(p)}$ is equal to two, then there is one parameter too many to identify, and, hence, a fourth constraint not attributable to the APC linear identity is necessary to identify the model parameters. The number of degrees of freedom for the factor model was stated by Fienberg and Mason (1985, page 72) as equal to

$$\delta_{\text{factor}} = L^{(a)}L^{(p)} - \left[2 \cdot \left(L^{(a)} + L^{(p)} \right) - 4 \right] = \left(L^{(a)} - 2 \right) \left(L^{(p)} - 2 \right) \quad (5.20)$$

for $L^{(a)} \geq 2$ and $L^{(p)} \geq 2$.

5.4 Conclusion

Confounding issues can arise in models that are parameterised in terms of age, period and cohort. There is confounding if a model consists of an additive

combination of linear terms in age, period and cohort, as in equation (5.1). There is also confounding if a model consists of an additive combination of quadratic terms and two-way interaction terms in age, period and cohort, as in equation (5.4). However, not all models parameterised simultaneously in terms of the three variables suffer from confounding. In particular, the two models presented in equations (4.4) and (4.13) do not suffer from a lack of identifiability.

To overcome issues of confounding, one option is to parameterise models with only two of the three variables. However, if a model consists of an additive combination of two linear terms, then caution is needed when interpreting the parameters of the linear terms. For example, coefficients for the linear age and cohort terms, a and c , would absorb the coefficient of the linear period term, p . It is necessary to interpret the two model parameters in terms of the third ignored variable.

The existence of confounding in a model can be checked by assessing whether there are linear dependencies in the columns of the model design matrix. Linear dependencies between columns can arise due to the APC linear identity as well as due to overparameterisation. The appearance of confounding in relation to the linear identity is clear for polynomial functions, but is not so clear for factor variables. In a model parameterised with orthogonal polynomials, there is a clear linear dependency between the columns for first-order polynomials in age, period and cohort. It is not so clear where the linear dependency arises in the columns of a design matrix for a model parameterised with factor variables.

Chapter 6

Misrounding effects for age-period-cohort modelling

The purpose of this chapter is to show the potentially misleading effects that can be apparent when studying an APC model under a misrounded treatment of data. Data are usually made available for model fitting in the form of a two-way contingency table categorised in terms of age and period (Frost 1940; Greenberg et al. 1950; Clayton and Schifflers 1987b). To estimate the Poisson intensity, researchers recategorise the age-by-period data in terms of an age-by-cohort rounding by adopting a particular surrogate convention. One very popular surrogate convention is to assume that the linear identity holds true so that cohort = period – age (Clayton and Schifflers 1987a; Murayama et al. 2006). An age-by-period rounding corresponds to a parallelogram region in a Lexis diagram, whereas an age-by-cohort rounding corresponds to a Lexis square.

The applied models from Section 4.3 are typically formulated in discrete time. Under independence, a discrete model is equivalent to a model discretised age-by-cohort. The fitting of a discrete model to age-by-period data means that the data are used for model fitting as if the data are rounded age-by-cohort. This misrounded treatment of data has been carried out in many

applied APC studies such as Clayton and Schifflers (1987a), Dealler and Kent (1995) and Jean et al. (2013). In Sections 6.3, 6.4 and 6.5, we show that three potentially misleading effects can appear for a misrounded treatment of data. The three misrounding effects can be summarised as: (i) an apparent under-reporting of deaths; (ii) an apparent longer life expectancy; and (iii) an apparent lengthening of life expectancy. Misrounding effects can only be significant if the cohort intensity is changing rapidly.

Each effect is a qualitative interpretation of a mathematical result. It is helpful to remove statistical fluctuations in the observations so we work with the Poisson intensities, that is, the expected counts. The first effect relates to a difference between expected counts, the second effect relates to a difference between expected count ratios and the third effect relates to a difference between expected count cross-ratios. It is possible for three, two, one or even none of the effects to hold true as each effect depends on the particular circumstances. Many statisticians could have been affected by these three effects to some extent due to the widespread adoption of misrounding. In particular, the third misrounding effect could have important implications for studies of human mortality.

In Section 6.1, we formulate Poisson models for age-by-period and age-by-cohort regions in the Lexis diagram. In Section 6.2, we define a misrounded treatment and an exact treatment for an age-by-period rounding of data. We also distinguish between three different surrogate conventions. Chapter 7 provides a detailed illustration of the ideas in this chapter for the case of Bovine Spongiform Encephalopathy (BSE) incidence. It is often sufficient to study an APC model under a misrounded treatment because misrounding effects are not significant if the cohort intensity is constant or changing very slowly. Note that, only the rounding down convention from Section 2.2 is used in Chapters 6 and 7.

6.1 Two-way rounding

A rounding of age, period and cohort can be presented in different ways. Ideally age, period and cohort would all be rounded simultaneously and the death counts would be displayed in triangles of the Lexis diagram as in Figure 2.1. A three-way rounding is often not used by data collectors and a two-way rounding in terms of age and period is used instead (Clayton and Schifflers 1987a, pg. 451). A two-way rounding is a grouping of a pair of adjacent lower and upper Lexis triangles that forms either an age-by-cohort square, an age-by-period parallelogram or a period-by-cohort parallelogram. Note that, an age-by-cohort rounding has a unique integer for age and cohort, but has two possible integers for period. Carstensen (2007, page 3024) refers to age-by-period regions as A-sets and age-by-cohort regions as C-sets.

Recalling Section 2.2, we assume in this chapter that the age-at-death, period-at-death and cohort are rounded down to the nearest integer so that

$$I = \lfloor A \rfloor, J = \lfloor P \rfloor \text{ and } K = \lfloor C \rfloor.$$

Letting $R_{i,k}^{(AC)}$ denote an age-by-cohort Lexis region and letting $R_{i,j}^{(AP)}$ denote an age-by-period Lexis region, we can write

$$R_{i,k}^{(AC)} = \{(a, p, c) : I = i, K = k\}$$

and

$$R_{i,j}^{(AP)} = \{(a, p, c) : I = i, J = j\}.$$

Recalling the definition of lower and upper Lexis triangles in (2.15) and (2.16), we pair the lower triangle $T_{i,k}^{(L)}$ and the upper triangle $T_{i,k}^{(U)}$ to write

$$R_{i,k}^{(AC)} = T_{i,k}^{(L)} \cup T_{i,k}^{(U)}. \quad (6.1)$$

If i and k are known, then j can either take a value of $i + k$ or $i + k + 1$. The Lexis square (6.1) is displayed in Figure 6.1 for a rounding down convention.

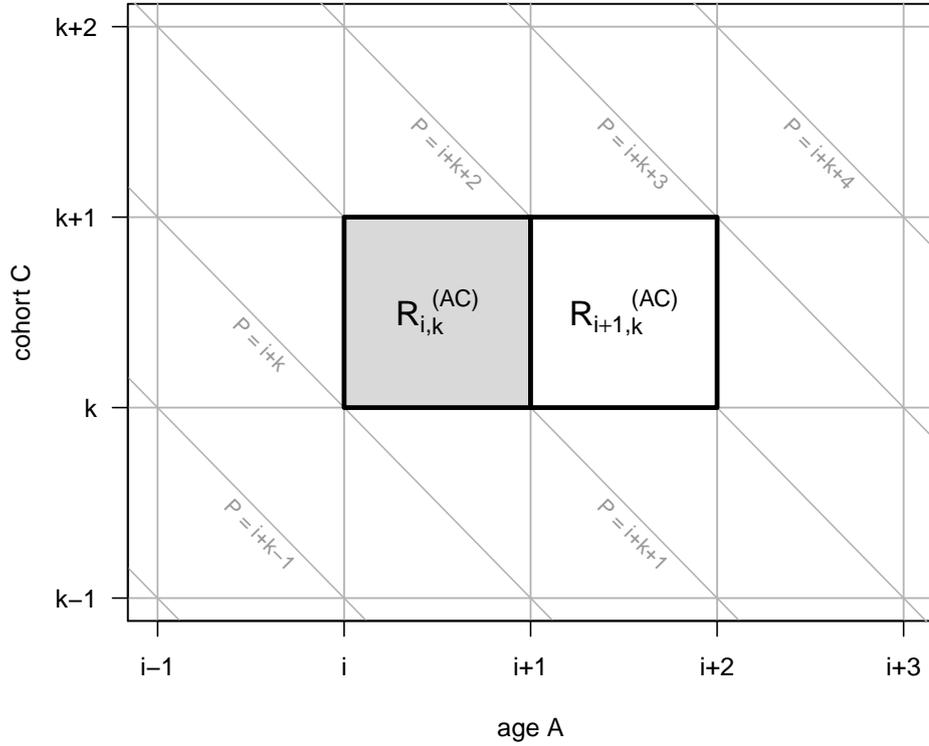


Figure 6.1: A Lexis diagram in age-cohort space with some highlighted age-by-cohort regions.

An age-by-period region can be defined similarly as a pairing of triangles $T_{i,j-i}^{(L)}$ and $T_{i,j-i-1}^{(U)}$. however, we must consider two scenarios because $R_{i,j}^{(AP)}$ can be one of two parallelogram regions in Figure 2.1. The first scenario is that if $j = i + k$ then

$$R_{i,i+k}^{(AP)} = T_{i,k}^{(L)} \cup T_{i,k-1}^{(U)}. \quad (6.2)$$

The second scenario is that if $j = i + k + 1$ then

$$R_{i,i+k+1}^{(AP)} = T_{i,k+1}^{(L)} \cup T_{i,k}^{(U)}. \quad (6.3)$$

The Lexis parallelograms (6.2) and (6.3) are displayed in Figure 6.2 for a rounding down convention.

The Poisson intensity $\lambda(a, c)$ from equation (2.5) should be viewed as the intensity of deaths at one particular instant in the age-cohort space of Figure

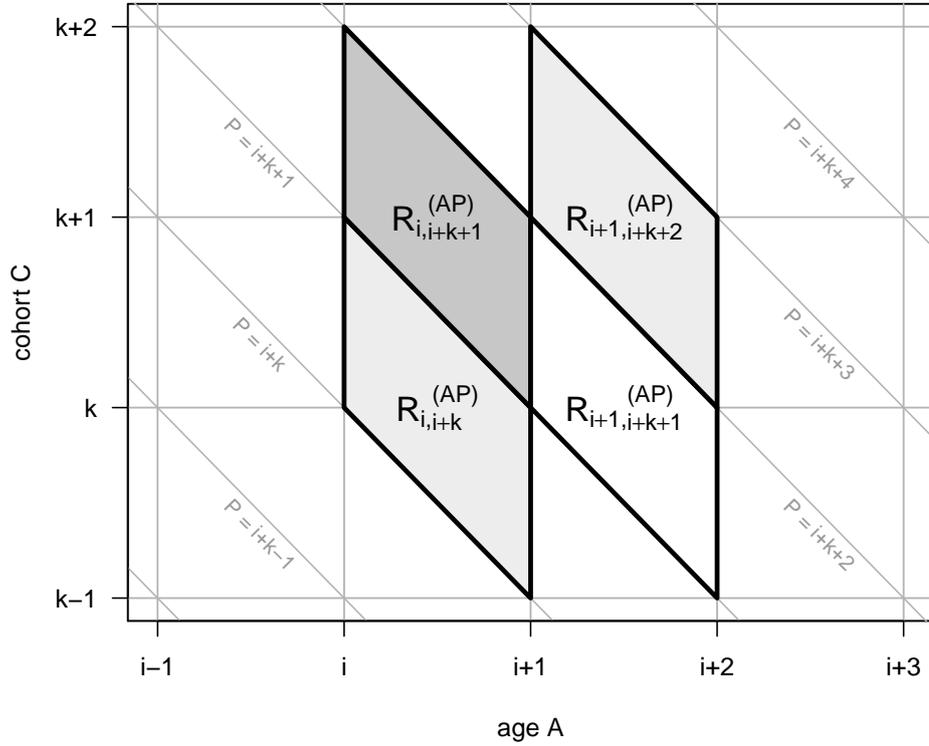


Figure 6.2: A Lexis diagram in age-cohort space with some highlighted age-by-period regions. In Section 6.3, we compare the number of deaths occurring in regions $R_{i,k}^{(AC)}$ and $R_{i,i+k}^{(AP)}$. In Section 6.4, we compare the ratio of deaths in regions $R_{i+1,i+k+1}^{(AP)}$ and $R_{i,i+k}^{(AP)}$ with the ratio of deaths in regions $R_{i+1,k}^{(AC)}$ and $R_{i,k}^{(AC)}$. In Section 6.5, we investigate how the ratio of deaths in regions $R_{i+1,i+k+1}^{(AP)}$ and $R_{i,i+k}^{(AP)}$ changes for unit increases in k .

2.1. The number of deaths for cohort k at age i , denoted as $N_{i,k}$, follows a Poisson distribution:

$$N_{i,k} \sim \text{Poisson}(\nu_{i,k}), \quad (6.4)$$

where

$$\begin{aligned} \nu_{i,k} &= \iint_{R_{i,k}^{(AC)}} \lambda(a, c) \, dc \, da \\ &= \int_i^{i+1} \int_k^{k+1} f(a | c) \cdot \kappa(c) \, dc \, da. \end{aligned} \quad (6.5)$$

The number of deaths in period j at age i , denoted as $M_{i,j}$, follows a Poisson distribution:

$$M_{i,j} \sim \text{Poisson}(\mu_{i,j}), \quad (6.6)$$

where

$$\begin{aligned} \mu_{i,j} &= \iint_{R_{i,j}^{(\text{AP})}} \lambda(a, c) \, dc \, da \\ &= \int_i^{i+1} \int_{j-a}^{j+1-a} f(a \mid c) \cdot \kappa(c) \, dc \, da. \end{aligned} \quad (6.7)$$

We have integrated $\lambda(a, c)$ over the region $R_{i,k}^{(\text{AC})}$ to define the Poisson intensity for age-by-cohort counts, $\nu_{i,k}$, and we have integrated $\lambda(a, c)$ over the region $R_{i,j}^{(\text{AP})}$ to define the Poisson intensity for age-by-period counts, $\mu_{i,j}$. The data typically made available for an APC analysis are observations of $M_{i,j}$ (Greenberg et al. 1950; Clayton and Schifflers 1987a; Liu et al. 2001).

In Section 3.1, we described a survival distribution that is independent of covariates, and, in Section 3.3, we described a survival distribution that is conditional on covariates. The expressions for $\nu_{i,k}$ and $\mu_{i,j}$ can be written more simply under the following independence assumption for the survival distribution:

$$\lambda(a, c) = \kappa(c) \cdot f(a). \quad (6.8)$$

Recall from Section 3.1 that $F(a) = \int_0^a f(u) \, du$ and let the integrated intensity at cohort c be written as

$$Q(c) = \int_0^c \kappa(v) \, dv. \quad (6.9)$$

Under the independence assumption from (6.8), the Poisson intensities described in (6.5) and (6.7) can be simplified to

$$\nu_{i,k} = [F(i+1) - F(i)] \cdot [Q(k+1) - Q(k)] = f_i^* \cdot \kappa_k^* \quad (6.10)$$

and

$$\mu_{i,j} = \int_0^1 f(a+i) [Q(j-i-a+1) - Q(j-i-a)] \, da. \quad (6.11)$$

The new expression for $\nu_{i,k}$ is much simpler since it is a function of i multiplied by a function of k , but the new expression for $\mu_{i,j}$ is still quite complex. The probability function in rounded time, f_i^* , was defined in (3.8), and the cohort intensity in rounded time, κ_k^* , is equal to $Q(k+1) - Q(k)$. The two Poisson intensities defined in (6.10) and (6.11) are key quantities in this chapter.

6.2 Misrounding

Data must be categorised in terms of age and cohort in order to estimate $\lambda(a, c)$. Since data are typically made available as observations of $M_{i,j}$, we recategorise $M_{i,j}$ in terms of age and cohort to derive a surrogate for $N_{i,k}$ which we denote as $N_{i,k}^{(s)}$. We outline three surrogate conventions in Table 6.1. The first convention labelled by $s = 1$ assumes that $k = j - i$ so that $M_{i,j}$ can be indexed by age and cohort as $M_{i,i+k} = N_{i,k}^{(1)}$. This first surrogate convention is analogous to the basic continuous identity in (2.2) and has been used by many authors including Moolgavkar et al. (1979, page 494), Clayton and Schifflers (1987a, page 459) and Murayama et al. (2006, page 4). We later illustrate this first surrogate convention in Table 7.1. We show in Section 6.3 that there is only equality between the expectations of $N_{i,k}^{(s)}$ and $N_{i,k}$, denoted as $\nu_{i,k}^{(s)}$ and $\nu_{i,k}$, for a constant cohort intensity.

In terms of the Lexis regions in Figures 6.1 and 6.2, we can say that:

1. the first surrogate convention $s = 1$ treats a count from region $R_{i,i+k}^{(AP)}$ as the count for region $R_{i,k}^{(AC)}$,
2. the second surrogate convention $s = 2$ treats a count from region $R_{i,i+k+1}^{(AP)}$ as the count for region $R_{i,k}^{(AC)}$,
3. the third surrogate convention $s = 3$ treats an average of the counts from regions $R_{i,i+k}^{(AP)}$ and $R_{i,i+k+1}^{(AP)}$ as the count for region $R_{i,k}^{(AC)}$.

The distinction between the three surrogate conventions has not been made before. Applying a surrogate convention is necessary, but misrounding is not necessary and can be problematic. Note that, rather than a basic surrogate convention, the Human Mortality Database smooths age-by-period counts via a regression equation to derive age-by-cohort counts (Wilmoth et al. 2017 (accessed August 22, 2018, Section 4.2)).

Table 6.1: There are three ways, labelled by s for “surrogate”, in which we can use $M_{i,j}$ to derive a surrogate for $N_{i,k}$. The surrogate $N_{i,k}^{(s)}$ has an expectation denoted as $\nu_{i,k}^{(s)}$. We defined $N_{i,k}$ and $M_{i,j}$ in (6.4) and (6.6), respectively.

s	$N_{i,k}^{(s)}$	$\nu_{i,k}^{(s)}$
1	$M_{i,i+k}$	$\mu_{i,i+k}$
2	$M_{i,i+k+1}$	$\mu_{i,i+k+1}$
3	$\frac{1}{2}(M_{i,i+k} + M_{i,i+k+1})$	$\frac{1}{2}(\mu_{i,i+k} + \mu_{i,i+k+1})$

Recalling (3.9) and (3.11), the probability density function $f(a | c)$ is a function of the parameter vectors $\boldsymbol{\alpha}^T = (\alpha_1, \dots, \alpha_m)$ and $\boldsymbol{\beta}^T = (\beta_1, \dots, \beta_q)$. Supposing that the cohort intensity is also parametric, we can write $\kappa(c)$ as a function of a parameter vector $\boldsymbol{\xi}^T = (\xi_1, \dots, \xi_r)$. The bivariate function $\lambda(a, c) = f(a | c) \cdot \kappa(c)$ is then a function of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$. Let

$$\mathbf{N}^{(s)} \equiv \left(N_{i,k}^{(s)} \right)_{i,k}$$

be a matrix of random surrogate death counts for $i \in \{i^- + 1, \dots, i^+\}$ and $k \in \{k^- + 1, \dots, k^+\}$. Suppose that we replace a with i and replace c with k , so that the Poisson intensity indexed in discrete time in (2.7) is written as

$$\lambda_{i,k} = \kappa_k \cdot f_{i|k} \cong \nu_{i,k}. \quad (6.12)$$

Recalling (6.10), the Poisson intensities $\lambda_{i,k}$ and $\nu_{i,k}$ are equivalent under the independence assumption because we can write $\nu_{i,k}$ as a product of a cohort

intensity and a survival distribution:

$$\lambda_{i,k} = f_i \kappa_k = f_i^* \kappa_k^* = \nu_{i,k}. \quad (6.13)$$

We stated in Section 4.3 that a natural starting point for an APC analysis is to consider the fit of an independence model, and, if necessary, modifications are made to the independence model to account for departures from independence due to period and cohort effects. Data made available for an APC analysis are usually observations of $M_{i,j}$ (Greenberg et al. 1950; Clayton and Schifflers 1987a; Liu et al. 2001). It is important to discretise the independence model $\lambda(a, c)$ correctly to reflect that the data are rounded age-by-period. We can summarise our discussion in Section 4.3 to say that researchers typically estimate the discrete independence model $\lambda_{i,k} = \nu_{i,k}$, given observations $\mathbf{N}^{(1)} = \mathbf{n}^{(1)}$, by maximising the following Poisson likelihood function for $s = 1$ (Holford 1983; Dealler and Kent 1995; Kramer et al. 2015):

$$L_{A1}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\xi} \mid \mathbf{N}^{(s)} = \mathbf{n}^{(s)}) = \prod_{i=i^-+1}^{i^+} \prod_{k=k^-+1}^{k^+} \frac{(\lambda_{i,k})^{n_{i,k}^{(s)}} \exp(-\lambda_{i,k})}{n_{i,k}^{(s)}!}. \quad (6.14)$$

However, this means that the independence model is discretised incorrectly, or is “misrounded”, say, because the fitting of $\lambda_{i,k}$ to age-by-period data means that the age-by-period data are used for model fitting as if the data are rounded age-by-cohort. That is, the observations in Lexis parallelograms are used for model fitting as if they are observations in Lexis squares.

An exact treatment of the age-by-period data would be to estimate the Poisson intensity $\nu_{i,k}^{(s)}$ by maximising the following likelihood function:

$$L_{E1}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\xi} \mid \mathbf{N}^{(s)} = \mathbf{n}^{(s)}) = \prod_{i=i^-+1}^{i^+} \prod_{k=k^-+1}^{k^+} \frac{(\nu_{i,k}^{(s)})^{n_{i,k}^{(s)}} \exp(-\nu_{i,k}^{(s)})}{n_{i,k}^{(s)}!}. \quad (6.15)$$

In Appendix 9.3, we demonstrate how to implement the exact approach to fitting the independence model. The two likelihood functions are labelled with subscripts “A” or “E” to indicate an approximate or exact treatment of data in the fitting of an independence model. The distinction between exact and approximate treatments of data has not been made before.

In Section 4.3, we showed that the discrete independence model is usually modified by a PH assumption. We will investigate whether the need for modifications to the independence model is only apparent due to the misrounded treatment of the age-by-period data. In the remainder of this chapter, we show that misleading conclusions can arise when fitting the independence model for a misrounded treatment of data. The potential for misleading conclusions can be overcome by adopting an exact treatment of the age-by-period data. We note that the potential for misleading conclusions is most severe when the cohort intensity is changing rapidly. If the cohort intensity is changing slowly, then it would be sufficient to carry out a misrounded treatment of the age-by-period data. In Chapter 7, the case of Bovine Spongiform Encephalopathy (BSE) is used as an illustrative example to investigate the implications of misrounding on the outcome of an APC study. BSE is a prime example since the cohort intensity for BSE incidence is rapidly changing.

6.3 Under-reporting of deaths

The first misrounding effect is that there is a strict inequality between the Poisson intensities for $N_{i,k}^{(s)}$ and $N_{i,k}$ which holds for $s = 1$ and $s = 2$ under certain conditions for the cohort intensity and survival distribution. Some important conditions for the cohort intensity are that $\kappa(c)$ is strictly increasing so that $\frac{d}{dc}\kappa(c) > 0$ for all c , strictly decreasing so that $\frac{d}{dc}\kappa(c) < 0$ for all c , or constant so that $\frac{d}{dc}\kappa(c) = 0$ for all c . The power function, $\kappa(c) = \xi_1 c^{\xi_2}$ with $\xi_1 > 0$, has a first derivative equal to $\xi_1 \xi_2 c^{\xi_2 - 1}$. Hence, the power function is strictly increasing for $\xi_2 > 0$, strictly decreasing for $\xi_2 < 0$ and constant for $\xi_2 = 0$. An important condition for the survival distribution is the independence assumption defined in (6.8).

If the cohort intensity is strictly increasing and the survival distribution

is independent of cohort, then

$$\nu_{i,k}^{(1)} < \nu_{i,k} \quad \forall i, k, \quad (6.16)$$

$$\nu_{i,k}^{(2)} > \nu_{i,k} \quad \forall i, k. \quad (6.17)$$

However, if the cohort intensity is strictly decreasing and the survival distribution is independent of cohort, then the inequalities reverse so that

$$\nu_{i,k}^{(1)} > \nu_{i,k} \quad \forall i, k, \quad (6.18)$$

$$\nu_{i,k}^{(2)} < \nu_{i,k} \quad \forall i, k. \quad (6.19)$$

If the cohort intensity is the same for all cohorts, so that $\kappa(c) = \xi$ for all c , then for $s = 1$ and $s = 2$,

$$\nu_{i,k}^{(s)} = \nu_{i,k} \quad \forall i, k. \quad (6.20)$$

The five results (6.16)-(6.20) are proved in Theorem 6.1. Note that, strict inequalities exist for the first and second surrogate conventions, but there may not be strict inequalities for the third convention.

The percentage loss in accuracy incurred by carrying out a misrounded treatment of data as opposed to an exact treatment can be calculated for each (i, k) cell as $100 \cdot \left| \frac{\nu_{i,k}^{(s)}}{\nu_{i,k}} - 1 \right|$. This is because, under a misrounded treatment of data, it is assumed that the observation in cell (i, k) of the surrogate data has an expected value of $\nu_{i,k}$. In reality, the observation in cell (i, k) has an expected value of $\nu_{i,k}^{(s)}$. For example, if $\nu_{i,k}^{(s)} > \nu_{i,k}$ for all i and k , then a statistician's estimate of $\nu_{i,k}$ under a misrounded treatment of data would be overstated for each (i, k) cell.

In Figures 6.3 and 6.4, we illustrate inequalities (6.16) and (6.18) respectively for a particular choice of $f(a)$ and $\kappa(c)$. In Appendix 9.3, we outline a method to calculate $\frac{\nu_{i,k}^{(1)}}{\nu_{i,k}}$ for a particular parametric form for $\lambda(a, c)$, which consists of a Gamma survival distribution and an exponential cohort intensity. This method can be applied to any parametric form. In Figure 6.3, the percentage loss in accuracy is most severe for earlier cohorts since the percentage loss is approximately 45% for cohort 2001 and decreases to approximately 15%

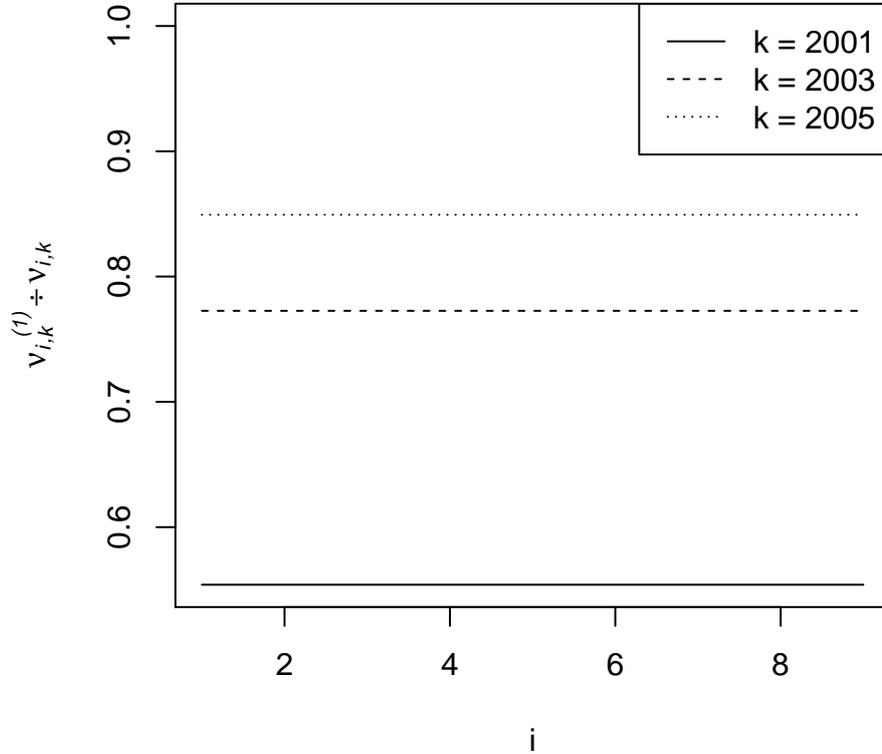


Figure 6.3: The calculation of $\frac{\nu_{i,k}^{(1)}}{\nu_{i,k}}$ for an exponential survival distribution $A \sim \text{Exponential}(\kappa_1 = 0.5)$ with $a \geq 0$ and a cohort intensity $\kappa(c) = [c - 2000]^2$ with $c \geq 2000$. The exponential family was described in Table 3.2. The survival distribution is independent of cohort and the cohort intensity is strictly increasing.

for cohort 2005. In Figure 6.4, the percentage loss in accuracy is severe for cohort 2001 at approximately 295-1140%. For cohort 2003 and 2005, the loss in accuracy is not so severe and has a range of approximately 10-75%.

The first misrounding effect tells us that $\frac{\nu_{i,k}^{(1)}}{\nu_{i,k}}$ is strictly greater than, strictly less than or strictly equal to one for certain conditions, but this effect does not tell us about the pattern of change in $\frac{\nu_{i,k}^{(1)}}{\nu_{i,k}}$ with i and k . In Figure 6.3, the independence of the ratio $\frac{\nu_{i,k}^{(1)}}{\nu_{i,k}}$ with i is explained in Section 6.4 by the second misrounding effect. The convergence of $\frac{\nu_{i,k}^{(1)}}{\nu_{i,k}}$ towards one for

unit increases in k is not explained by any of the three misrounding effects discussed in this chapter. In this thesis, we use a Riemann Sum approximation to estimate $\nu_{i,k}^{(1)}$ and $\nu_{i,k}$. The Riemann approximations, written in code as the functions `parallelogram` and `square`, are discussed in detail in Appendix 9.3.

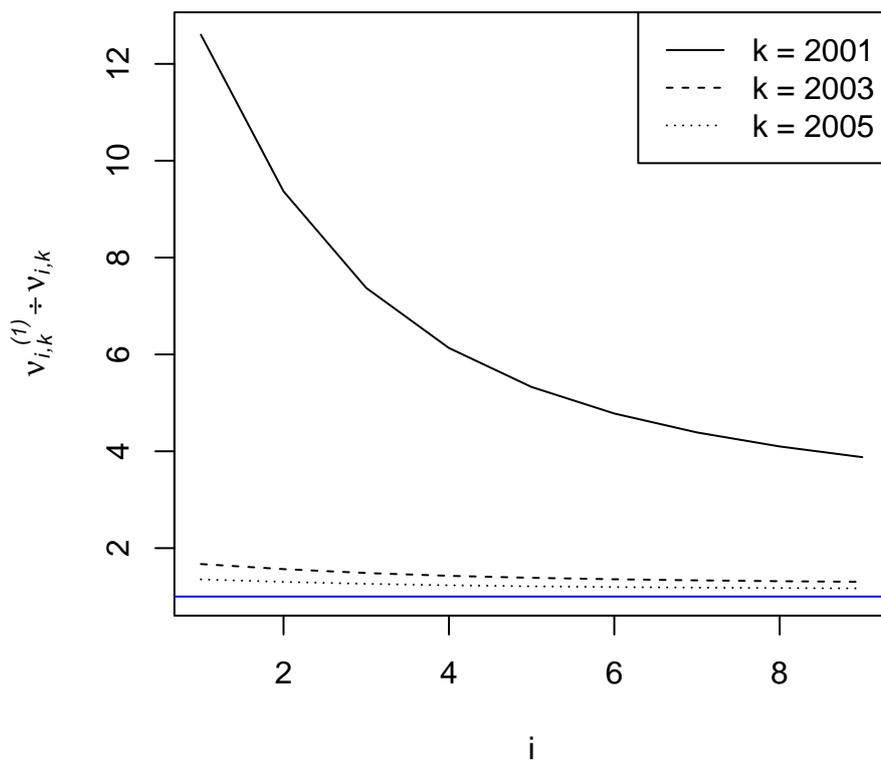


Figure 6.4: The calculation of the ratio $\frac{\nu_{i,k}^{(1)}}{\nu_{i,k}}$ for a Gamma survival distribution $A \sim \text{Gamma}(\pi_7 = 3, \kappa_7 = 20)$ with $a \geq 0$ and a cohort intensity $\kappa(c) = [c - 2000]^{-2}$ with $c \geq 2000$. The gamma family was described in Table 3.2. The survival distribution is independent of cohort and the cohort intensity is strictly decreasing. A horizontal line is drawn at the point the y-axis equals one.

The first result (6.16) means that if the survival distribution is the same for all individuals and the number of births is increasing on average over time, then we should expect to encounter a strict understatement of deaths under misrounding for the first surrogate convention. The third result (6.18) means

that if instead the number of births is expected to decrease over time, then we expect to encounter a strict overstatement of deaths under misrounding for the first surrogate convention. A very similar interpretation can be made for the second and fourth results in (6.17) and (6.19). The fifth result (6.20) means that if the survival distribution is the same for all individuals and the number of births is not changing over time, then an approximate modelling approach is equivalent to an exact modelling approach.

The first four results are apparent because of a misrounded treatment of data. An exact treatment of data outlined in (6.15) must be adopted to ensure these four results are not apparent. In Chapter 7, we estimate the cohort intensity for BSE incidence which we deem to be strictly increasing with cohort until a particular time point and strictly decreasing thereafter. This leads to an apparent understating of case numbers followed by an apparent overstating of case numbers.

Theorem 6.1: Under the independence model, $\lambda(a, c) = \kappa(c) \cdot f(a)$, the following results hold true under particular surrogate conventions:

- (a) If $\kappa(c)$ is strictly increasing, then $\nu_{i,k}^{(1)} < \nu_{i,k}$ for all i and k . But if $\kappa(c)$ is strictly decreasing, then $\nu_{i,k}^{(1)} > \nu_{i,k}$ for all i and k . Also, if $\kappa(c)$ is constant, then $\nu_{i,k}^{(1)} = \nu_{i,k}$ for all i and k .
- (b) If $\kappa(c)$ is strictly increasing, then $\nu_{i,k}^{(2)} > \nu_{i,k}$ for all i and k . But if $\kappa(c)$ is strictly decreasing, then $\nu_{i,k}^{(2)} < \nu_{i,k}$ for all i and k . Also, if $\kappa(c)$ is constant, then $\nu_{i,k}^{(2)} = \nu_{i,k}$ for all i and k .

Proof: (a) Recall the expressions for $\nu_{i,k}$ and $\mu_{i,j}$ in (6.10) and (6.11). Also recalling the definition of $\nu_{i,k}^{(1)}$ from Table 6.1, we can write $\nu_{i,k}^{(1)} - \nu_{i,k}$ as

$$\int_0^1 f(a+i) [Q(k+1-a) - Q(k-a)] da - \int_0^1 f(a+i) [Q(k+1) - Q(k)] da.$$

Simplifying the subtraction of the two integrals to get one integral, we get the following:

$$\int_0^1 f(a+i) \{ [Q(k+1-a) - Q(k-a)] - [Q(k+1) - Q(k)] \} da. \quad (6.21)$$

Let us define a new function for the cumulative change in the cohort intensity over a unit interval:

$$\phi(a, k) = Q(k+1-a) - Q(k-a). \quad (6.22)$$

If $\kappa(c)$ is strictly increasing, then $\phi(a, k)$ is strictly decreasing with a because, for all a and all k ,

$$\frac{\partial}{\partial a} \phi(a, k) = \kappa(k-a) - \kappa(k+1-a) < 0.$$

Further, a function $\phi(a, k)$ which is strictly decreasing with a implies that, for all $a \in (0, 1)$ and all k ,

$$Q(k+1-a) - Q(k-a) < Q(k+1) - Q(k).$$

Hence, if $\kappa(c)$ is strictly increasing then expression (6.21) is negative for all i and k and we conclude that $\nu_{i,k}^{(1)} - \nu_{i,k} < 0$ for all i and k .

If $\kappa(c)$ is strictly decreasing, then $\phi(a, k)$ is strictly increasing with a because, for all a and all k , $\frac{\partial}{\partial a} \phi(a, k) = \kappa(k-a) - \kappa(k+1-a) > 0$. A function $\phi(a, k)$ which is strictly increasing with a implies that, for all $a \in (0, 1)$ and all k ,

$$Q(k+1-a) - Q(k-a) > Q(k+1) - Q(k).$$

Hence, if $\kappa(c)$ is strictly decreasing, then expression (6.21) is positive for all i and k and we conclude that $\nu_{i,k}^{(1)} - \nu_{i,k} > 0$ for all i and k .

Also, if $\kappa(c)$ is constant with c , then $\phi(a, k)$ is constant with a since $\frac{\partial}{\partial a} \phi(a, k) = \kappa(k-a) - \kappa(k+1-a) = 0$. Further, a function $\phi(a, k)$ which is constant with a implies that, for all $a \in (0, 1)$ and all k ,

$$Q(k+1-a) - Q(k-a) = Q(k+1) - Q(k).$$

Hence, if $\kappa(c)$ is constant, then expression (6.21) is equal to zero for all i and k and we conclude that $\nu_{i,k}^{(1)} - \nu_{i,k} = 0$ for all i and k .

(b) Recalling the definition of $\nu_{i,k}^{(2)}$ from Table 6.1, we can write $\nu_{i,k}^{(2)} - \nu_{i,k}$ as

$$\int_0^1 f(a+i) \{ [Q(k+2-a) - Q(k+1-a)] - [Q(k+1) - Q(k)] \} da. \quad (6.23)$$

We showed in part (a) of the proof that, if $\kappa(c)$ is strictly increasing, then $\phi(a, k)$ is strictly decreasing with a , so that, for all $a \in (0, 1)$ and all k ,

$$Q(k+2-a) - Q(k+1-a) > Q(k+1) - Q(k).$$

Hence, if $\kappa(c)$ is strictly increasing then (6.23) is positive for all i and k and we conclude that $\nu_{i,k}^{(2)} - \nu_{i,k} > 0$ for all i and k .

We also showed in part (a) that, if $\kappa(c)$ is strictly decreasing, then $\phi(a, k)$ is strictly increasing with a , so that, for all $a \in (0, 1)$ and all k ,

$$Q(k+2-a) - Q(k+1-a) < Q(k+1) - Q(k).$$

Hence, if $\kappa(c)$ is strictly decreasing, then (6.23) is negative for all i and k and we conclude that $\nu_{i,k}^{(2)} - \nu_{i,k} < 0$ for all i and k .

Finally, we showed in part (a) that if $\kappa(c)$ is constant with c , then $\phi(a, k)$ is constant with a , so that, for all $a \in (0, 1)$ and all k ,

$$Q(k+2-a) - Q(k+1-a) = Q(k+1) - Q(k).$$

Hence, if $\kappa(c)$ is constant, then (6.23) is equal to zero for all i and k and we conclude that $\nu_{i,k}^{(2)} - \nu_{i,k} = 0$ for all i and k . This completes the proof.

6.4 Longer life expectancy

The second misrounding effect is that there is a strict inequality between the ratio of Poisson intensities for $N_{i,k}^{(s)}$ and $N_{i,k}$, which holds for $s = 1$ and $s = 2$,

under certain conditions for the cohort intensity and survival distribution. Relative to the first misrounding effect, there are no additional assumptions for the cohort intensity, but there is one new condition for the survival distribution in addition to the independence assumption.

Suppose that $\log f(a)$ is twice differentiable. Some important conditions for the survival distribution are that $f(a)$ is log concave so that $\frac{d^2}{da^2} \log f(a) < 0$ for all a , log convex so that $\frac{d^2}{da^2} \log f(a) > 0$ for all a , or log linear so that $\frac{d^2}{da^2} \log f(a) = 0$ for all a . Recall the Gamma family from Table 3.2. The probability function for the Gamma family is log concave for $\kappa_7 > 1$, log convex for $\kappa_7 < 1$ and log linear for $\kappa_7 = 1$ because

$$\frac{d^2}{da^2} \log f^{(\text{Gamma})}(a, \kappa_7, \pi_7) = \frac{1 - \kappa_7}{a^2}. \quad (6.24)$$

The independence assumption implies that the ratio of Poisson intensities for $N_{i,k}$ is independent of cohort k so that

$$\frac{\nu_{i+1,k}}{\nu_{i,k}} = \frac{\int_0^1 f(a+i+1) da}{\int_0^1 f(a+i) da} = \frac{f_{i+1}^*}{f_i^*} \equiv \zeta_i. \quad (6.25)$$

However, the ratio of Poisson intensities for $N_{i,k}^{(s)}$ in most cases is dependent on cohort because, for $s = 1$ and $s = 2$,

$$\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}} = \frac{\int_0^1 f(a+i+1) \cdot [Q(k-a+s) - Q(k-a+s-1)] da}{\int_0^1 f(a+i) \cdot [Q(k-a+s) - Q(k-a+s-1)] da}. \quad (6.26)$$

The ratio $\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}}$ is independent of cohort when the cohort intensity changes at an exponential rate so that $\kappa(c) = \xi_1 \exp(\xi_2 c)$ with $\xi_1 \in \mathbb{R}^+$ and $\xi_2 \in \mathbb{R}$, because

$$\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}} = \frac{\int_0^1 f(a+i+1)e^{-\xi_2 a} da}{\int_0^1 f(a+i)e^{-\xi_2 a} da} \equiv \omega_i \neq \zeta_i. \quad (6.27)$$

In Chapter 7, we assume that the cohort intensity for BSE incidence increases at an exponential rate and then decreases exponentially. Even though $\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}}$ is independent of cohort for a cohort intensity that changes at an exponential rate, it is still not equal to the ratio of Poisson intensities for $N_{i,k}$. If either

the cohort intensity is constant so that $\kappa(c) = \xi$ or the survival distribution is strictly log linear, then the ratios of Poisson intensities are equal for $s = 1$ and $s = 2$:

$$\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}} = \frac{\nu_{i+1,k}}{\nu_{i,k}} \equiv \zeta_i \quad \forall i, k. \quad (6.28)$$

If the cohort intensity is strictly increasing and the independent survival distribution is strictly log concave, or if the cohort intensity is strictly decreasing and the independent survival distribution is strictly log convex, then, for $s = 1$ and $s = 2$,

$$\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}} > \frac{\nu_{i+1,k}}{\nu_{i,k}} \equiv \zeta_i \quad \forall i, k. \quad (6.29)$$

However, if the cohort intensity is strictly increasing and the independent survival distribution is strictly log convex, or if the cohort intensity is strictly decreasing and the independent survival distribution is strictly log concave, then for $s = 1$ and $s = 2$, the inequality reverses so that

$$\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}} < \frac{\nu_{i+1,k}}{\nu_{i,k}} \equiv \zeta_i \quad \forall i, k. \quad (6.30)$$

The results in (6.28), (6.29) and (6.30) are proved in Theorem 6.2. Similar to the first misrounding effect, there may not be strict inequalities for the third surrogate convention.

The percentage loss in accuracy incurred by carrying out a misrounded treatment of data as opposed to an exact treatment can also be calculated for each (i, k) cell in terms of the relative ratio as $100 \cdot \left| \left(\frac{\nu_{i+1,k}}{\nu_{i,k}} \div \frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}} \right) - 1 \right|$. This is because it is assumed under a misrounded treatment of data that the ratio of counts taken from cells (i, k) and $(i + 1, k)$ has an expected value of $\frac{\nu_{i+1,k}}{\nu_{i,k}} = \zeta_i$. In reality, the ratio of counts taken from cells (i, k) and $(i + 1, k)$ has an expected value of $\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}}$, which can not be written as independent of k except in the case of an exponential cohort intensity or a log linear survival distribution. For example, if $\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}} > \frac{\nu_{i+1,k}}{\nu_{i,k}}$, then, in general, a statistician's

estimate of $\nu_{i,k}$ under a misrounded treatment of data would have ratios that are overstated and varying with cohort for each pair of (i, k) cells.

We illustrate the results (6.28) and (6.30) in Figures 6.5 and 6.6, respectively. An exponential survival distribution is strictly log linear so leads to an equality $\frac{\nu_{i+1,k}}{\nu_{i,k}} \div \frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}} = 1$ for all i and k . This equality explains why the curves in Figure 6.3 are independent of i for each cohort year. In this case, there is no loss in accuracy in terms of the count ratios when carrying out a misrounded treatment of data. A log concave Gamma survival distribution coupled with a decreasing power function leads to an inequality $\frac{\nu_{i+1,k}}{\nu_{i,k}} \div \frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}} > 1$ for all i and k . The convergence of the curves in Figure 6.6 towards one is later explained by the third misrounding effect. In this second case, a statistician incurs a 0.1-34.8% loss in accuracy in terms of the count ratios.

The results presented in this section can be interpreted in terms of the expected conditional age-at-death, $\mathbb{E}[A | c]$. Under an independence assumption, the logarithm of the ratio of Poisson intensities for $N_{i,k}$, written as $\log\left(\frac{\nu_{i+1,k}}{\nu_{i,k}}\right)$, is a discretised analogue of the relative rate of change in the survival distribution, $\frac{d}{da} \log f(a)$. If $\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}}$ is strictly less than $\frac{\nu_{i+1,k}}{\nu_{i,k}}$, then under misrounding, the relative rate of change in the survival distribution appears too small at each age and may appear to be changing with cohort. We illustrate in Figure 6.7 that if $\frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}}$ is strictly less than $\frac{\nu_{i+1,k}}{\nu_{i,k}}$, then the survival distribution will appear too far to the left so that the expected age-at-death will appear too small. The expected age-at-death also appears to be changing with cohort so that $\frac{d}{dc} \mathbb{E}[A | c] \neq 0$.

In general, we can say that if $\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}}$ is strictly less than $\frac{\nu_{i+1,k}}{\nu_{i,k}}$, then $\mathbb{E}[A | c]$ will appear too small and may appear to change with cohort when in fact the expected age-at-death is not changing with cohort due to the independence assumption. Similarly, if $\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}}$ is strictly greater than $\frac{\nu_{i+1,k}}{\nu_{i,k}}$, then $\mathbb{E}[A | c]$ will appear too large and may appear to change with cohort. Also, if $\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}}$ is strictly equal to $\frac{\nu_{i+1,k}}{\nu_{i,k}}$, then the expected age-at-death will appear correctly and appear unchanging with cohort under misrounding.

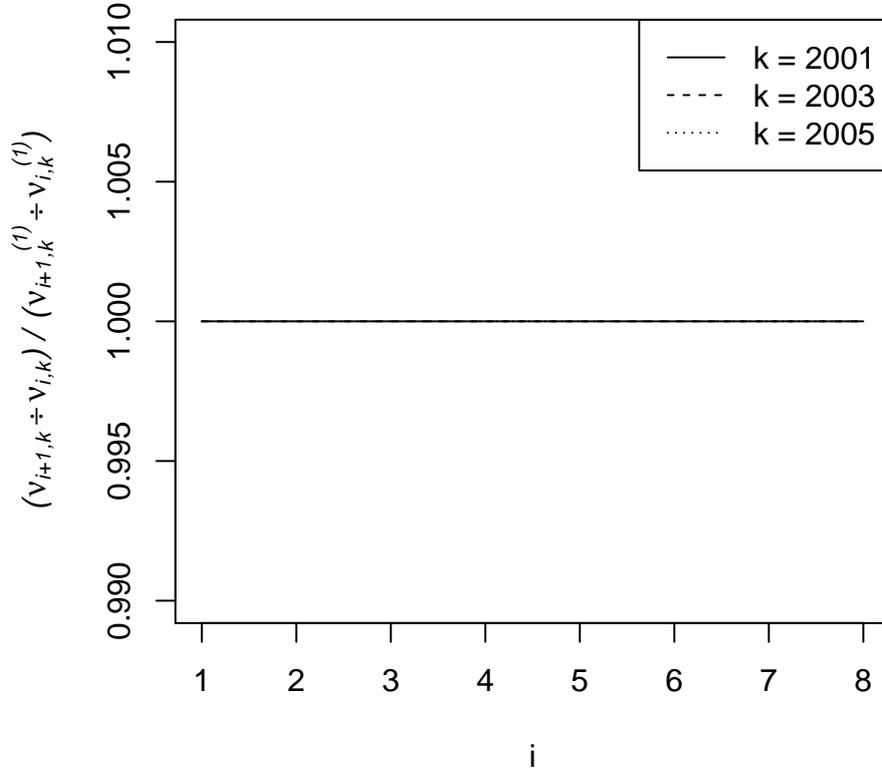


Figure 6.5: The calculation of $\frac{\nu_{i+1,k}}{\nu_{i,k}} \div \frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}}$ for an exponential survival distribution $A \sim \text{Exponential}(\kappa_1 = 0.5)$ with $a \geq 0$ and a cohort intensity $\kappa(c) = [c - 2000]^2$ with $c \geq 2000$. The exponential family was described in Table 3.2 and the calculation of $\frac{\nu_{i,k}^{(1)}}{\nu_{i,k}}$ was illustrated in Figure 6.3. The survival distribution is independent and log linear, while the cohort intensity is strictly increasing.

One interpretation of the second result (6.29) is that if the survival distribution is strictly log concave and the same for all individuals, and the number of births is increasing on average over time, then the expected age-at-death will appear too large and possibly changing with cohort under misrounding for the first and second surrogate conventions. If instead the survival distribution is strictly log convex, the third result (6.30) tells us that the expected age-at-death will appear too small and possibly changing with cohort under misrounding. We show in Chapter 7 that certain conditions led the expected

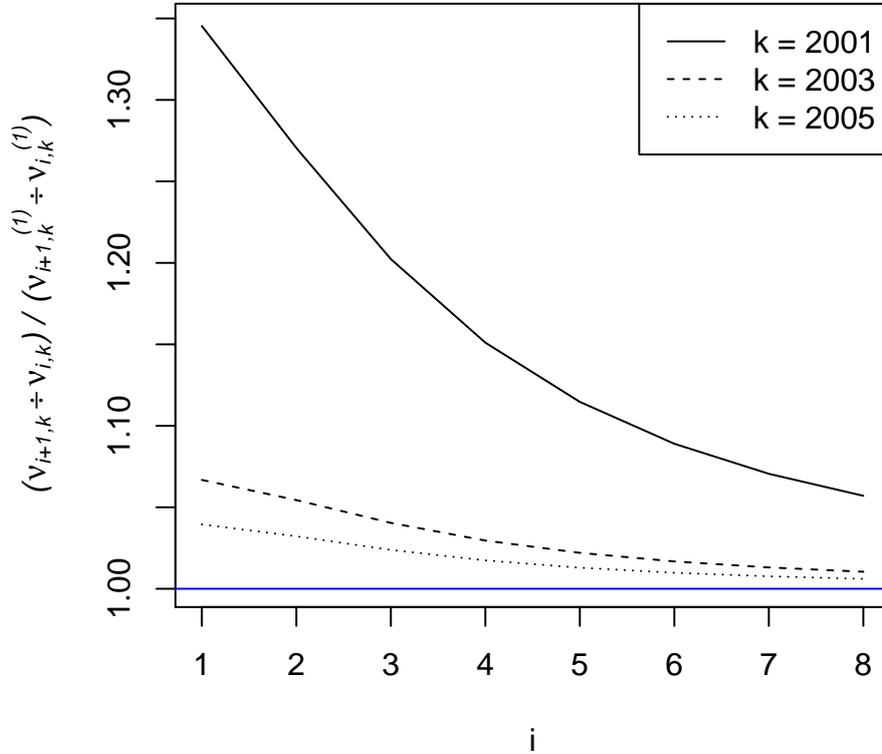


Figure 6.6: The calculation of $\frac{\nu_{i+1,k}}{\nu_{i,k}} \div \frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}}$ for a Gamma survival distribution $A \sim \text{Gamma}(\pi_7 = 3, \kappa_7 = 20)$ with $a \geq 0$ and a cohort intensity $\kappa(c) = [c - 2000]^{-2}$ with $c \geq 2000$. The gamma family was described in Table 3.2 and the calculation of $\frac{\nu_{i,k}^{(1)}}{\nu_{i,k}}$ was illustrated in Figure 6.4. The survival distribution is independent and strictly log concave, while the cohort intensity is strictly decreasing.

age-at-death for BSE incidence to appear too large for cattle born before a certain time instant and then appear too small for cattle born after that time instant.

Theorem 6.2: Under the independence model, $\lambda(a, c) = \kappa(c) \cdot f(a)$, the following results hold true under surrogate conventions $s = 1$ and $s = 2$:

- If $\kappa(c)$ is strictly increasing and $f(a)$ is strictly log concave, or if $\kappa(c)$ is strictly decreasing and $f(a)$ is strictly log convex, then $\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}} > \frac{\nu_{i+1,k}}{\nu_{i,k}} = \zeta_i$ for all i and k .
- If $\kappa(c)$ is strictly increasing and $f(a)$ is strictly log convex, or if $\kappa(c)$ is strictly decreasing and $f(a)$ is strictly log concave, then $\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}} < \frac{\nu_{i+1,k}}{\nu_{i,k}} = \zeta_i$ for all i and k .
- If $\kappa(c)$ is constant or $f(a)$ is strictly log linear, then $\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}} = \frac{\nu_{i+1,k}}{\nu_{i,k}} = \zeta_i$ for all i and k .

Proof: Recall the definition of $\phi(a, k)$ from (6.22). Suppose

$$\psi(a, i) = f(a + i + 1)/f(a + i). \quad (6.31)$$

We use equations (6.25) and (6.26) to express the ratios $\frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}}$ and $\frac{\nu_{i+1,k}}{\nu_{i,k}}$ in terms of the functions $\psi(a, i)$ and $\phi(a, k)$:

$$\begin{aligned} \frac{\nu_{i+1,k}}{\nu_{i,k}} &= \frac{\int_0^1 f(a + i + 1) da}{\int_0^1 f(a + i) da} \\ &= \frac{\int_0^1 f(a + i)\psi(a, i) da}{\int_0^1 f(a + i) da} \end{aligned}$$

and

$$\begin{aligned} \frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}} &= \frac{\int_0^1 f(a + i + 1) [Q(k + 1 - a) - Q(k - a)] da}{\int_0^1 f(a + i) [Q(k + 1 - a) - Q(k - a)] da} \\ &= \frac{\int_0^1 f(a + i + 1)\phi(a, k) da}{\int_0^1 f(a + i)\phi(a, k) da} \\ &= \frac{\int_0^1 f(a + i)\psi(a, i)\phi(a, k) da}{\int_0^1 f(a + i)\phi(a, k) da}. \end{aligned} \quad (6.32)$$

The ratio difference $\frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}} - \frac{\nu_{i+1,k}}{\nu_{i,k}}$ has a denominator equal to

$$\int_0^1 f(a + i) da \int_0^1 f(a + i)\phi(a, k) da$$

and a numerator equal to

$$\begin{aligned} & \int_0^1 f(a+i) da \int_0^1 f(a+i)\psi(a,i)\phi(a,k) da \\ & - \int_0^1 f(a+i)\phi(a,k) da \int_0^1 f(a+i)\psi(a,i) da. \end{aligned} \quad (6.33)$$

The denominator is always positive since $f(a) > 0$ for all a and $\kappa(c) > 0$ for all c . So only the numerator needs to be evaluated. It is helpful to introduce a dummy variable, w , to clarify the product of integrals. By introducing w for the second integrand in each integral product, the numerator of (6.33) is equal to

$$\int_0^1 \int_0^1 f(a+i)f(w+i)[\psi(w,i)\phi(w,k) - \psi(w,i)\phi(a,k)] da dw. \quad (6.34)$$

If instead w is introduced for the first integrand in each integral product, the numerator of (6.33) is equal to

$$\int_0^1 \int_0^1 f(a+i)f(w+i)[\psi(a,i)\phi(a,k) - \psi(a,i)\phi(w,k)] da dw. \quad (6.35)$$

Expressions (6.34) and (6.35) are mathematically equivalent so we can take an average of (6.34) and (6.35) to derive the following double integral:

$$\frac{1}{2} \int_0^1 \int_0^1 f(a+i)f(w+i)[\psi(w,i) - \psi(a,i)][\phi(w,k) - \phi(a,k)] da dw.$$

For $\frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}} - \frac{\nu_{i+1,k}}{\nu_{i,k}}$ to be strictly negative or non-negative, it is sufficient to show that

$$[\psi(w,i) - \psi(a,i)][\phi(w,k) - \phi(a,k)] \quad (6.36)$$

is negative or non-negative respectively for all w, a, i and k . This can be determined by assessing the monotonicity properties of $\psi(\cdot)$ and $\phi(\cdot)$. By taking the logarithm of $\psi(a,i)$ from (6.31) and differentiating with respect to a , we derive the equation

$$\frac{d}{da}[\log(\psi(a,i))] = \frac{d}{da}\log(f(a+i+1)) - \frac{d}{da}\log(f(a+i)). \quad (6.37)$$

If $f(a)$ is strictly log concave, then $\log(f(a))$ is strictly concave and the quantity in (6.37) is strictly negative. The slope of $\log(\psi(a,i))$ is then always negative,

and $\psi(a, i)$ is strictly decreasing with a . So if $f(a)$ is strictly log convex then $\psi(a, i)$ is strictly increasing with a , and if $f(a)$ is strictly log linear then $\psi(a, i)$ is constant with a . We showed in the proof of Theorem 6.1 that: (i) if $\kappa(c)$ is strictly increasing, then $\phi(a, k)$ is strictly decreasing with a ; (ii) if $\kappa(c)$ is strictly decreasing, then $\phi(a, k)$ is strictly increasing with a ; and (iii) if $\kappa(c)$ is constant, then $\phi(a, k)$ is constant with a .

If $\kappa(c)$ is strictly increasing and $f(a)$ is strictly log concave, then the inequality $\frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}} - \frac{\nu_{i+1,k}}{\nu_{i,k}} > 0$ is true for all i and k since:

- if $w < a$, then $\psi(w, i) > \psi(a, i)$ and $\phi(w, k) > \phi(a, k)$
- if $w > a$, then $\psi(w, i) < \psi(a, i)$ and $\phi(w, k) < \phi(a, k)$.

If $\kappa(c)$ is strictly decreasing and $f(a)$ is strictly log convex, then the inequality $\frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}} - \frac{\nu_{i+1,k}}{\nu_{i,k}} > 0$ is true for all i and k since:

- if $w < a$, then $\psi(w, i) < \psi(a, i)$ and $\phi(w, k) < \phi(a, k)$
- if $w > a$, then $\psi(w, i) > \psi(a, i)$ and $\phi(w, k) > \phi(a, k)$.

If $\kappa(c)$ is strictly increasing and $f(a)$ is strictly log concave, then the inequality $\frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}} - \frac{\nu_{i+1,k}}{\nu_{i,k}} < 0$ is true for all i and k since:

- if $w < a$, then $\psi(w, i) < \psi(a, i)$ and $\phi(w, k) > \phi(a, k)$
- if $w > a$, then $\psi(w, i) > \psi(a, i)$ and $\phi(w, k) < \phi(a, k)$.

If $\kappa(c)$ is strictly decreasing and $f(a)$ is strictly log concave, then the inequality $\frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}} - \frac{\nu_{i+1,k}}{\nu_{i,k}} < 0$ is true for all i and k since:

- if $w < a$, then $\psi(w, i) > \psi(a, i)$ and $\phi(w, k) < \phi(a, k)$
- if $w > a$, then $\psi(w, i) < \psi(a, i)$ and $\phi(w, k) > \phi(a, k)$.

Hence, if $\kappa(c)$ is constant and/or $f(a)$ is strictly log linear, then the equality $\frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}} - \frac{\nu_{i+1,k}}{\nu_{i,k}} = 0$ holds true for all i and k because, at least one of the equalities $\psi(w, i) = \psi(a, i)$ and $\phi(w, k) = \phi(a, k)$ will hold true for all values of a and w .

Recalling equation (6.26), the ratio $\frac{\nu_{i+1,k}^{(2)}}{\nu_{i,k}^{(2)}}$ can be expressed in terms of $\psi(a, i)$ and $\phi(a, k)$ as the following:

$$\begin{aligned} \frac{\nu_{i+1,k}^{(2)}}{\nu_{i,k}^{(2)}} &= \frac{\int_0^1 f(a+i+1) [Q(k+2-a) - Q(k+1-a)] da}{\int_0^1 f(a+i) [Q(k+2-a) - Q(k+1-a)] da} \\ &= \frac{\int_0^1 f(a+i+1)\phi(a, k+1) da}{\int_0^1 f(a+i)\phi(a, k+1) da} \\ &= \frac{\int_0^1 f(a+i)\psi(a, i)\phi(a, k+1) da}{\int_0^1 f(a+i)\phi(a, k+1) da}. \end{aligned}$$

By taking the ratio difference and introducing a dummy variables to clarify the product of integrals as in (6.34) and (6.35), we derive a numerator for $\frac{\nu_{i+1,k}^{(2)}}{\nu_{i,k}^{(2)}} - \frac{\nu_{i+1,k}}{\nu_{i,k}}$ that is mathematically equivalent to

$$\frac{1}{2} \int_0^1 \int_0^1 f(a+i)f(w+i)[\psi(w, i) - \psi(a, i)][\phi(w, k+1) - \phi(a, k+1)] da dw.$$

The inequalities derived in the previous paragraph remain the same for the second surrogate convention as we only need to replace $\phi(\cdot, k)$ with $\phi(\cdot, k+1)$. This completes the proof.

6.5 Lengthening of life expectancy

The third misrounding effect is that there is a strict inequality for the cross-ratio of Poisson intensities for $N_{i,k}^{(s)}$, which holds for $s = 1$ and $s = 2$, under certain conditions for the cohort intensity and survival distribution. The third misrounding effect is described in Theorem 6.3. Relative to the second misrounding effect, there are no new assumptions for the survival distribution, but there is one new condition for the cohort intensity in addition to the strict monotonicity condition. Suppose that $\log \kappa(c)$ is twice differentiable. An im-

portant new condition is that $\kappa(c)$ is log concave, log convex or log linear. For example, a power function $\kappa(c) = \xi_1 c^{\xi_2}$ is log concave for $\xi_2 > 0$, log convex for $\xi_2 < 0$ and log linear for $\xi_2 = 0$ because $\frac{d^2}{dc^2} \log \kappa(c) = -\frac{\xi_2}{c^2}$.

Recalling (6.25), the independence assumption implies that the cross-ratio of Poisson intensities for $N_{i,k}$ is equal to one so that, for all i and k ,

$$\frac{\nu_{i+1,k+1}}{\nu_{i,k+1}} \div \frac{\nu_{i+1,k}}{\nu_{i,k}} = 1. \quad (6.38)$$

However, the cross-ratio of Poisson intensities for $N_{i,k}^{(s)}$ may not be equal to one for $s = 1$ and $s = 2$. If the survival distribution is independent of cohort and either the cohort intensity or survival distribution is strictly log linear, then, for $s = 1$ and $s = 2$, the cross-ratio of Poisson intensities for $N_{i,k}^{(s)}$ is equal to one for all i and k :

$$\frac{\nu_{i+1,k+1}^{(s)}}{\nu_{i,k+1}^{(s)}} \div \frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}} = 1. \quad (6.39)$$

If the cohort intensity and the independent survival distribution are both strictly log concave or strictly log convex, then, for $s = 1$ and $s = 2$, and for all i and k ,

$$\frac{\nu_{i+1,k+1}^{(s)}}{\nu_{i,k+1}^{(s)}} \div \frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}} < 1. \quad (6.40)$$

However, if the cohort intensity is strictly log convex and the independent survival distribution is strictly log concave, or if the cohort intensity is strictly log concave and the independent survival distribution is strictly log convex, then for $s = 1$ and $s = 2$, the inequality reverses so that, for all i and k ,

$$\frac{\nu_{i+1,k+1}^{(s)}}{\nu_{i,k+1}^{(s)}} \div \frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}} > 1. \quad (6.41)$$

For other conditions, there is not a strict equality or inequality for the cross-ratio of Poisson intensities for $N_{i,k}^{(s)}$. There may not be strict inequalities or equalities for the third surrogate convention. Theorem 6.3 proves that the three results (6.39), (6.40) and (6.41) hold true. In Figures 6.5 and 6.6, and later on in Figure 7.4, the value of $\frac{\nu_{i+1,k+1}^{(1)}}{\nu_{i,k+1}^{(1)}} \div \frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}}$ can be determined by holding i

constant and increasing k by one. For each figure, this theorem can be seen to hold.

The results in this section can be interpreted in terms of the change in the expected age-at-death with cohort, $\frac{d}{dc} \mathbb{E}[A | c]$. Under an independence assumption, the logarithm of the cross-ratio of Poisson intensities for $N_{i,k}$, written as $\log \left(\frac{\nu_{i+1,k+1}}{\nu_{i,k+1}} \div \frac{\nu_{i+1,k}}{\nu_{i,k}} \right)$, is a discretised analogue of the mixed derivative, $\frac{d^2}{da dc} \log f(a | c)$. If $\frac{\nu_{i+1,k+1}^{(s)}}{\nu_{i,k+1}^{(s)}}$ is strictly greater than $\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}}$, then the mixed derivative will appear to be positive so that, under misrounding, the relative rate of change in the survival distribution, $\frac{d}{da} \log f(a | c)$, appears to be increasing with cohort. In Figure 6.7, we show that, if $\frac{\nu_{i+1,k+1}^{(1)}}{\nu_{i,k+1}^{(1)}}$ is strictly greater than $\frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}}$, then the survival distribution will appear to be shifting to the right for newer cohorts. Hence, the expected age-at-death will appear to be getting larger over time so that $\frac{d}{dc} \mathbb{E}[A | c] > 0$ for all c .

In general, we can say that if $\frac{\nu_{i+1,k+1}^{(s)}}{\nu_{i,k+1}^{(s)}}$ is strictly greater than $\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}}$, then under misrounding, it will incorrectly appear that the expected age-at-death is increasing across generations, so that $\frac{d}{dc} \mathbb{E}[A | c] > 0$ for all c , when in fact the expected age-at-death is not changing with cohort. Similarly, if $\frac{\nu_{i+1,k+1}^{(s)}}{\nu_{i,k+1}^{(s)}}$ is strictly less than $\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}}$, then under misrounding, it will incorrectly appear that the expected age-at-death is decreasing across generations so that $\frac{d}{dc} \mathbb{E}[A | c] < 0$ for all c . Also, if $\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}}$ is strictly equal to $\frac{\nu_{i+1,k+1}^{(s)}}{\nu_{i,k+1}^{(s)}}$, then under misrounding, the expected age-at-death will correctly appear to stay the same over time so that $\frac{d}{dc} \mathbb{E}[A | c] = 0$.

One interpretation of the second result (6.40) is that if the relative rate of change in the survival distribution and cohort intensity are strictly increasing on average, then newer generations of individuals will appear to live shorter lives on average under misrounding for the first and second surrogate conventions. In relation to (6.41), if the relative rate of change in the survival distribution is strictly decreasing and the relative rate of change in the cohort intensity is strictly increasing, then newer generations of individuals will

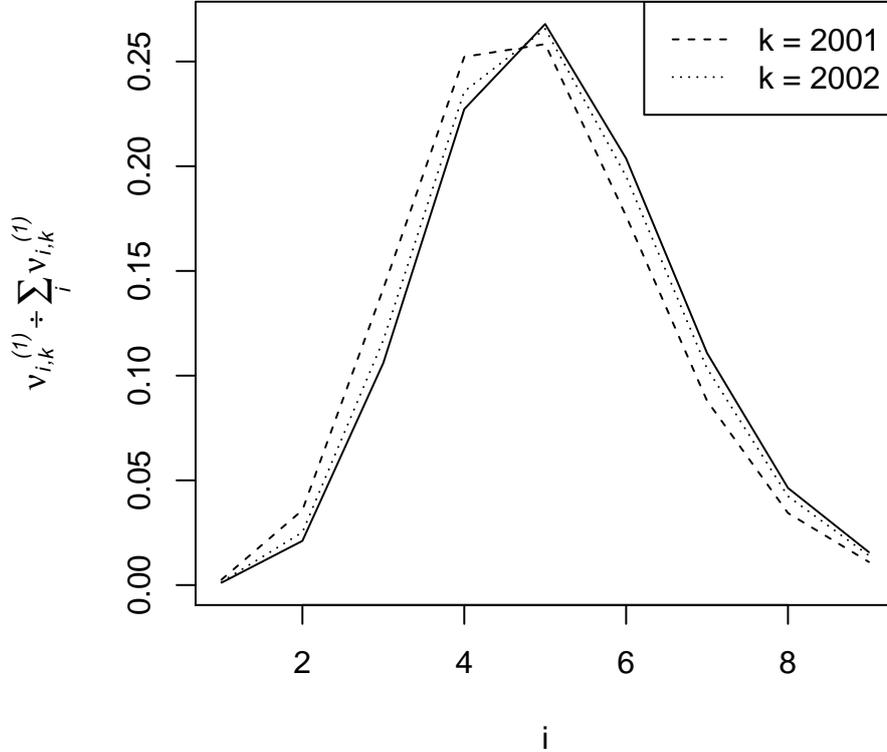


Figure 6.7: The calculation of $\frac{\nu_{i,k}^{(1)}}{\sum_i \nu_{i,k}^{(1)}}$ for a Gamma survival distribution $A \sim \text{Gamma}(\pi_7 = 3, \kappa_7 = 20)$ with $a \geq 0$ and a cohort intensity $\kappa(c) = [c - 2000]^{-2}$ with $c \geq 2000$. The quantity $\frac{\nu_{i,k}^{(1)}}{\sum_i \nu_{i,k}^{(1)}}$ is a discrete representation of the survival distribution that is apparent under misrounding for $s = 1$. The solid line is equal to f_i^* . The survival distribution is log concave and the cohort intensity is log convex.

appear to live longer lives on average under misrounding. For this second scenario, we plot a discrete representation of the survival distribution, $\frac{\nu_{i,k}^{(1)}}{\sum_i \nu_{i,k}^{(1)}}$, in Figure 6.7, which is apparent under misrounding for $s = 1$. By holding i constant and increasing k by one, we can determine from Figure 6.6 that the plotted values of $\frac{\nu_{i,k}^{(1)}}{\sum_i \nu_{i,k}^{(1)}}$ correspond to the inequality $\frac{\nu_{i+1,k+1}^{(1)}}{\nu_{i,k+1}^{(1)}} \div \frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}} > 1$.

In Chapter 7, we fit a log linear cohort intensity to data on BSE incidence. So, under our model, the expected age-at-death for BSE incidence appears to stay the same over time. The conclusion of an increasing expected age-at-death

is often reached in APC studies for human mortality (Clayton and Schifflers 1987a; Lee and Hsieh 1996; Kramer et al. 2015) and seems to be an important topic of debate for society. It would be interesting for further research to assess the extent to which the apparent lengthening of life for humans is attributable to the third misrounding effect.

We can combine the results of the second and third misrounding effects in Theorems 6.2 and 6.3 respectively, to fully describe the appearance of the survival distribution. For example, in Figure 6.7, we illustrate that if the cohort intensity is strictly log convex and strictly decreasing, and if the independent survival distribution is strictly log concave, then the expected age-at-death appears too small but is increasing for newer cohorts towards the true expected age-at-death which would be apparent for a misrounded treatment of data. This true expected age-at-death is indicated by the true survival distribution, $f_i^* = \int_i^{i+1} f(a) da$, which is the solid line depicted in Figure 6.7.

Theorem 6.3: Under the independence model, $\lambda(a, c) = \kappa(c) \cdot f(a)$, the following results hold true under surrogate conventions $s = 1$ and $s = 2$:

- If both $\kappa(c)$ and $f(a)$ are strictly log concave or strictly log convex, $\frac{\nu_{i+1, k+1}^{(s)}}{\nu_{i, k+1}^{(s)}} < \frac{\nu_{i+1, k}^{(s)}}{\nu_{i, k}^{(s)}}$ for all i and k .
- If one of $\kappa(c)$ and $f(a)$ is strictly log concave and the other is strictly log convex, then $\frac{\nu_{i+1, k+1}^{(s)}}{\nu_{i, k+1}^{(s)}} > \frac{\nu_{i+1, k}^{(s)}}{\nu_{i, k}^{(s)}}$ for all i and k .
- If at least one of $\kappa(c)$ and $f(a)$ is strictly log linear, then $\frac{\nu_{i+1, k+1}^{(s)}}{\nu_{i, k+1}^{(s)}} = \frac{\nu_{i+1, k}^{(s)}}{\nu_{i, k}^{(s)}}$ for all i and k .

Proof: Recalling (6.22), (6.31) and (6.32), the absolute difference between

$\frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}}$ and $\frac{\nu_{i+1,k+1}^{(1)}}{\nu_{i,k+1}^{(1)}}$ can be written as

$$\frac{\nu_{i+1,k+1}^{(1)}}{\nu_{i,k+1}^{(1)}} - \frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}} = \frac{\nu_{i+1,k+1}^{(1)} \cdot \nu_{i,k}^{(1)} - \nu_{i+1,k}^{(1)} \cdot \nu_{i,k+1}^{(1)}}{\nu_{i,k+1}^{(1)} \cdot \nu_{i,k}^{(1)}}, \quad (6.42)$$

where the numerator is equal to

$$\begin{aligned} & \int_0^1 f(a+i)\psi(a,i)\phi(a,k+1) da \cdot \int_0^1 f(a+i)\phi(a,k) da \\ & - \int_0^1 f(a+i)\psi(a,i)\phi(a,k) da \cdot \int_0^1 f(a+i)\phi(a,k+1) da. \end{aligned} \quad (6.43)$$

The denominator of (6.42) is positive because $\kappa(c) > 0$ for all c and so $\nu_{i,k}^{(1)} > 0$ for all i and k . It is sufficient to work with the numerator to find a strict inequality or equality for (6.42).

We introduce a dummy variable w for the second integrand in each integral product so that the numerator (6.43) is equal to

$$\int_0^1 \int_0^1 f(a+i)f(w+i)\psi(a,i)g(a,w,k) da dw; \quad (6.44)$$

where

$$g(a,w,k) \equiv \phi(a,k+1)\phi(w,k) - \phi(a,k)\phi(w,k+1).$$

Notice that the function g is symmetric because $g(a,w,k) = -g(w,a,k)$. We can instead introduce w for the first integrand in each integral product so that the numerator is equal to

$$- \int_0^1 \int_0^1 f(a+i)f(w+i)\psi(w,i)g(a,w,k) da dw. \quad (6.45)$$

Expressions (6.44) and (6.46) are mathematically equivalent so that we can take an average of (6.44) and (6.46) to derive the following double integral:

$$\frac{1}{2} \int_0^1 \int_0^1 f(a+i)f(w+i) [\psi(a,i) - \psi(w,i)] g(a,w,k) da dw. \quad (6.46)$$

Consider a continuous function

$$e(x) = \frac{Q(x+2) - Q(x+1)}{Q(x+1) - Q(x)}$$

with a domain and range on the interval \mathbb{R}^+ . The function $e(x)$ can be written in terms of $\kappa(c)$ as

$$e(x) = \frac{\int_1^2 \kappa(x+u) \, du}{\int_0^1 \kappa(x+v) \, dv}.$$

The quantity $\frac{d}{dx} \log e(x)$ is equal to the ratio

$$\frac{\int_1^2 \kappa'(x+u) \, du \cdot \int_0^1 \kappa(x+v) \, dv - \int_0^1 \kappa'(x+v) \, dv \cdot \int_1^2 \kappa(x+u) \, du}{\int_1^2 \kappa(x+u) \, du \cdot \int_0^1 \kappa(x+v) \, dv}.$$

The denominator of $\frac{d}{dx} \log e(x)$ is positive since $\kappa(c) > 0$ for all c . So only the numerator needs to be evaluated. The numerator of $\frac{d}{dx} \log e(x)$ can be written as the following double integral:

$$\int_1^2 \int_0^1 [\kappa'(x+u)\kappa(x+v) - \kappa(x+u)\kappa'(x+v)] \, dv \, du.$$

The inequality $\kappa'(x+u)\kappa(x+v) - \kappa(x+u)\kappa'(x+v) > 0$ implies that

$$\frac{\kappa'(x+u)}{\kappa(x+u)} - \frac{\kappa'(x+v)}{\kappa(x+v)} = \frac{d}{du} \log \kappa(x+u) - \frac{d}{dv} \log \kappa(x+v) > 0.$$

Further, if $\kappa(c)$ is log convex so that $\frac{d}{dc} \log \kappa(c)$ is strictly increasing with c , then $\frac{d}{dx} \log e(x) > 0$, and, hence, $\frac{d}{dx} e(x) > 0$ for all x and all $u > v$.

Similarly, if $\kappa(c)$ is log concave so that $\frac{d}{dc} \log \kappa(c)$ is strictly decreasing with c , then $\frac{d}{dx} \log e(x) < 0$, and, hence, $\frac{d}{dx} e(x) < 0$ for all x and all $u > v$. Also, if $\kappa(c)$ is log linear so that $\frac{d}{dc} \log \kappa(c)$ does not change with c , then $\frac{d}{dx} \log e(x) = 0$, and, hence, $\frac{d}{dx} e(x) = 0$ for all x and all $u > v$. In summary,

1. if $\kappa(c)$ is log convex, then $e(x)$ is strictly increasing
2. if $\kappa(c)$ is log concave, then $e(x)$ is strictly decreasing
3. if $\kappa(c)$ is log linear, then $e(x)$ is constant.

The difference function $e(k-a) - e(k-w)$ is equivalent to a ratio with a positive denominator, $\phi(a, k) \cdot \phi(w, k)$, and a numerator equal to $g(a, w, k)$.

Hence,

$$e(k-a) - e(k-w) > 0 \implies g(a, w, k) > 0,$$

$$e(k-a) - e(k-w) < 0 \implies g(a, w, k) < 0,$$

$$e(k-a) - e(k-w) = 0 \implies g(a, w, k) = 0.$$

If $e(x)$ is strictly increasing, then $e(k-a) - e(k-w) > 0$ for $a < w$ and $e(k-a) - e(k-w) < 0$ for $a > w$. It is helpful to note that $k-a > k-w$ for $a < w$. Therefore, if $\kappa(c)$ is strictly log convex, then

1. if $a < w$, then $g(a, w, k) > 0$;
2. if $a > w$, then $g(a, w, k) < 0$.

Similarly, if $e(x)$ is strictly decreasing, then the inequalities reverse so that $e(k-a) - e(k-w) < 0$ for $a < w$ and $e(k-a) - e(k-w) > 0$ for $a > w$. Therefore, if $\kappa(c)$ is strictly log concave, then

1. if $a < w$, then $g(a, w, k) < 0$;
2. if $a > w$, then $g(a, w, k) > 0$.

If $e(x)$ is constant, then $e(k-a) - e(k-w) = 0$ for all a and w . Therefore, if $\kappa(c)$ is strictly log linear, then $g(a, w, k) = 0$ for all a and w .

In the proof of Theorem 6.2, we showed that:

- if $f(a)$ is strictly log concave, then $\psi(a, i) - \psi(w, i) < 0$ for $a > w$;
- if $f(a)$ is strictly log convex, then $\psi(a, i) - \psi(w, i) > 0$ for $a > w$;
- if $f(a)$ is strictly log linear, then $\psi(a, i) - \psi(w, i) = 0$ for all a and w .

Overall, if $\kappa(c)$ and $f(a)$ are both strictly log concave, then the inequality $\frac{\nu_{i+1, k+1}^{(1)}}{\nu_{i, k+1}^{(1)}} - \frac{\nu_{i+1, k}^{(1)}}{\nu_{i, k}^{(1)}} < 0$ is true for all i and k since:

- if $a < w$, then $\psi(a, i) > \psi(w, i)$ and $g(a, w, k) < 0$
- if $a > w$, then $\psi(a, i) < \psi(w, i)$ and $g(a, w, k) > 0$.

If $\kappa(c)$ and $f(a)$ are both strictly log convex, then the inequality $\frac{\nu_{i+1, k+1}^{(1)}}{\nu_{i, k+1}^{(1)}} - \frac{\nu_{i+1, k}^{(1)}}{\nu_{i, k}^{(1)}} < 0$ is true for all i and k since:

- if $a < w$, then $\psi(a, i) < \psi(w, i)$ and $g(a, w, k) > 0$
- if $a > w$, then $\psi(a, i) > \psi(w, i)$ and $g(a, w, k) < 0$.

If $\kappa(c)$ is strictly log convex and $f(a)$ is strictly log concave, then the inequality $\frac{\nu_{i+1, k+1}^{(1)}}{\nu_{i, k+1}^{(1)}} - \frac{\nu_{i+1, k}^{(1)}}{\nu_{i, k}^{(1)}} > 0$ is true for all i and k since:

- if $a < w$, then $\psi(a, i) > \psi(w, i)$ and $g(a, w, k) > 0$
- if $a > w$, then $\psi(a, i) < \psi(w, i)$ and $g(a, w, k) < 0$.

If $\kappa(c)$ is strictly log concave and $f(a)$ is strictly log convex, then the inequality $\frac{\nu_{i+1, k+1}^{(1)}}{\nu_{i, k+1}^{(1)}} - \frac{\nu_{i+1, k}^{(1)}}{\nu_{i, k}^{(1)}} > 0$ is true for all i and k since:

- if $a < w$, then $\psi(a, i) < \psi(w, i)$ and $g(a, w, k) < 0$
- if $a > w$, then $\psi(a, i) > \psi(w, i)$ and $g(a, w, k) > 0$.

If at least one of $\kappa(c)$ and $f(a)$ is strictly log linear, then the equality $\frac{\nu_{i+1, k+1}^{(1)}}{\nu_{i, k+1}^{(1)}} - \frac{\nu_{i+1, k}^{(1)}}{\nu_{i, k}^{(1)}} = 0$ is true for all i and k since at least one of the equalities, $\psi(a, i) = \psi(w, i)$ and $g(a, w, k) = 0$, holds true for all a and w .

For surrogate convention $s = 2$, we would replace $g(a, w, k)$ with $g(a, w, k+1)$. All of the inequalities and equalities from the previous paragraph would still hold true. This completes the proof.

6.6 Conclusion

Data on deaths are usually rounded and published in a contingency table categorised in terms of age and period. An age-by-cohort contingency table is rarely made available for analysis. Statisticians take the age-by-period rounded data as surrogates for age-by-cohort rounded data and fit the discrete independence model $\lambda_{i,k} = \kappa_k \cdot f_i$ to the surrogate data. We describe three possible surrogate conventions labelled by s , but it is most common to adopt the first convention $s = 1$ in which the value of cohort for each (i, j) cell of the age-by-period table is determined as $k = j - i$.

The discrete independence model $\lambda_{i,k}$ is equivalent to a continuous independence model that has been discretised over an age-by-cohort region in the Lexis diagram, written as $\nu_{i,k}$. So the age-by-period data are used for model fitting as if the data are rounded age-by-cohort. We compared this misrounded treatment of the data with an exact treatment. An exact treatment of data is to fit the continuous independence model that has been discretised over an age-by-period region in the Lexis diagram, written as $\nu_{i,k}^{(s)}$, to the surrogate data. The exact treatment means that the age-by-period data are used for model fitting as if the data are rounded age-by-period.

Under a misrounded treatment of data, it is assumed that the observation in cell (i, k) of the surrogate data has an expected value of $\nu_{i,k}$. In reality, the observation in cell (i, k) has an expected value of $\nu_{i,k}^{(s)}$. In certain circumstances, there are strict inequalities between $\nu_{i,k}$ and $\nu_{i,k}^{(s)}$ for all (i, k) cells which means that on average the surrogate data either overstate or understate the number of deaths. A statistician's estimate of $\nu_{i,k}$ would be strictly overstated or understated under a misrounded treatment of data. The percentage loss in accuracy incurred by carrying out a misrounded treatment of data as opposed to an exact treatment can be calculated for each (i, k) cell as $100 \cdot \left| \frac{\nu_{i,k}^{(s)}}{\nu_{i,k}} - 1 \right|$. A statistician might encounter an apparent over-reporting or under-reporting

in the surrogate data relative to the discrete independence model, which is not apparent in reality and is not apparent when carrying out an exact treatment of the surrogate data.

It is also assumed under a misrounded treatment of data that the ratio of counts taken from cells (i, k) and $(i + 1, k)$ has an expected value of $\frac{\nu_{i+1,k}}{\nu_{i,k}} = \zeta_i$. That is, the relative change in the count data with age is the same for all cohorts. In reality, the ratio of counts taken from cells (i, k) and $(i + 1, k)$ has an expected value of $\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}}$, which in general is not the same for all cohorts such that $\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}} \neq \zeta_i$. The percentage loss in accuracy incurred by carrying out a misrounded treatment of data as opposed to an exact treatment can be calculated for each (i, k) cell in terms of the relative ratio as $100 \cdot \left| \left(\frac{\nu_{i+1,k}}{\nu_{i,k}} \div \frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}} \right) - 1 \right|$.

Certain conditions lead to strict inequalities between $\frac{\nu_{i+1,k}}{\nu_{i,k}}$ and $\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}}$ and these inequalities have an interpretation in terms of the location of the apparent survival distribution. For example, if $\frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}} > \frac{\nu_{i+1,k}}{\nu_{i,k}}$, then the survival distribution apparent under the misrounded treatment of data is further to the right than the true distribution, $f(a)$. Further, if $\frac{\nu_{i+1,k+1}^{(s)}}{\nu_{i,k+1}^{(s)}} > \frac{\nu_{i+1,k}^{(s)}}{\nu_{i,k}^{(s)}}$, then the survival distribution apparent under misrounding is shifting to the right for newer cohorts, thereby suggesting an increasing longevity when in fact longevity is not changing under the independence model.

In summary, a misrounded treatment of data can be problematic when assessing the fit of an independence model due to the potential for misleading conclusions in relation to the understating or overstating of cell counts, the understating or overstating of longevity, and false changes in longevity over time. The potential for misleading conclusions can be overcome by carrying out an exact treatment of the data. In theory, the concept of misrounding has significant implications in the APC literature due to the widespread fitting of discrete independence models to age-by-period data. However, we caution that the effect of misrounding on a statistical analysis can only be significant

when the cohort intensity is changing rapidly. In Chapter 4, we explained that modifications are made to the discrete independence model using a PH functional. In Chapter 7, we investigate for the illustrative example of BSE whether the need for modifications to the independence model is only apparent due to the misrounded treatment of data.

Chapter 7

Misrounding effects in the case of BSE incidence

Bovine Spongiform Encephalopathy (BSE) is a fatal neurodegenerative disease in cattle which is perhaps better known as “mad cow disease”. BSE was first observed in cattle in the United Kingdom (UK) in 1986 and its incidence grew rapidly over the following years to cause considerable havoc in the cattle industry (Donnelly and Ferguson 2000, pages 9–10). The Ministry of Agriculture, Fisheries and Farming (MAFF) collected information for each BSE case and provided public summaries of this database in a contingency table rounded age-by-period. In contrast, the Central Veterinary Laboratory (CVL) compiled a more extensive database on BSE cases and were able to construct a contingency table rounded age-by-cohort (Donnelly and Ferguson 2000, pages 25–29).

Dealler and Kent (1995) and Anderson et al. (1996) analysed the BSE incidence data to assess the suitability of the independence model $\lambda(a, c) = f(a) \cdot \kappa(c)$ and to investigate various modifications. Dealler and Kent studied the MAFF data for a misrounded treatment of data and found large departures from the independence model. Anderson et al. studied the CVL data for an exact treatment of data and found much smaller departures from the

independence model. In this chapter, we argue that the two research groups came to dramatically different conclusions because the exact treatment of data outlined in Section 6.2 is necessary for the study of BSE incidence due to a cohort intensity that was changing rapidly and a survival distribution that was sharply peaked.

In Section 7.1, a parametric form is specified for the independence model based on a naive inspection of BSE incidence data rounded as age-by-period. We use a simple parametric form to capture the main features about $\lambda(a, c)$. This then allows us to capture the main features about the expected counts in age-by-cohort and age-by-period regions of the Lexis diagram, $\nu_{i,k}$ and $\nu_{i,k}^{(1)}$. We choose a Gamma model to describe $f(a)$ and an exponential model to describe $\kappa(c)$. In Section 7.2, we fit the model to the age-by-period data for an exact treatment and integrate the model over Lexis parallelograms to calculate our estimate of the age-by-period Poisson intensity $\nu_{i,k}^{(1)}$. We compare the estimate of $\nu_{i,k}^{(1)}$ to the MAFF data to assess how well the independence model describes the BSE incidence data.

In Section 7.3, the fitted independence model is integrated over Lexis squares to calculate our estimate of $\nu_{i,k}$. We compare our estimates of $\nu_{i,k}^{(1)}$ and $\nu_{i,k}$ to illustrate the three misrounding effects described in Theorems 6.1, 6.2 and 6.3 for the case of BSE incidence. Under our model, the expected age-at-onset appears too large for cattle born before the ban, appears to decrease for cattle born at around the time of the ban, and appears too small for cattle born after the ban. These misrounding effects are mathematical consequences of the model assumptions such as a log linear cohort intensity and a log concave survival distribution. In Section 7.4, we simulate observations of the number of BSE cases under our fitted independence model and explain whether the misrounding effects are apparent for noisy data. We also explain how modifications could be made to the independence model to allow the expected age at onset to vary with cohort.

7.1 Model for BSE incidence

In order to follow the cause of the disease, the Ministry of Agriculture, Fisheries and Farming (MAFF) collected information from mid-1989 onwards about each BSE case (Dealler and Kent 1995) including:

- the date of birth, $C \in \mathbb{R}$,
- the date of disease onset, $P \in \mathbb{R}$,
- the age at onset, $A = P - C \in \mathbb{R}^+$.

A date can be converted to a decimal. For example, the cohort of a cow born on the 18th January 1985 can be written as the decimal $1985 + \frac{18-1}{365} = 1985.047$. The minus one deduction in the fractional part of cohort means that the 1st January rather than the 31st December is converted to an integer. For some cases, the date of birth and date of onset were either unknown or estimated. When only the month and year were recorded, the dates were entered into the database as the first day of the month.

MAFF adopted the rounding down approach from Section 2.2 to present a cattle's date of birth, date of onset and age at onset as years:

$$I = \lfloor A \rfloor, J = \lfloor P \rfloor, K = \lfloor C \rfloor.$$

The continuous identity $A = P - C$ transformed under discretisation into the two near-identities

$$I = J - K \quad \text{and} \quad I = J - K - 1.$$

A cow born in year $k = 1980$ that is diagnosed with BSE in year $j = 1985$ could have an age at onset of either $i = 4$ or $i = 5$.

MAFF provided public summaries of this database in the form of a two-way contingency table categorised age-by-period and discretised into years.

Each cell of the MAFF table is an observation of the Poisson model, $M_{i,j}$, from equation (6.6), which counts the number of BSE cases in the Lexis parallelogram $R_{i,j}^{(AP)}$. Dealler and Kent (1995) applied the first surrogate convention from Table 6.1 to recategorise the MAFF table in terms of age and cohort so that $M_{i,i+k} \equiv N_{i,k}^{(1)}$. Recalling Figures 6.1 and 6.2, the observed number of BSE cases in the Lexis parallelogram $R_{i,i+k}^{(AP)}$ is treated by Dealler and Kent as an observation for the number of BSE cases in a Lexis square $R_{i,k}^{(AC)}$.

In Table 7.1, we present the observations of $N_{i,k}^{(1)}$ for the incidence of BSE. The first surrogate convention is adopted very often in the APC literature and some examples can be found for other case studies such as in Clayton and Schifflers (1987a, page 459) and Murayama et al. (2006, page 4). The number of BSE cases was published for ages two to ten and periods 1989 to 1993. Since there was no observation for cattle contracting BSE aged ten in period 1989, the year of birth is defined from 1980 to 1991. It is possible that cattle can contract BSE after age ten and after period 1993, but the MAFF data do not cover ages above ten years or periods above 1993. Each positive diagonal of Table 7.1 represents a certain period, for example, the diagonal from 49 to 17 displays the number of cattle contracting BSE in period 1989.

The simplest model for BSE incidence is an independence model in which the number of BSE cases, indexed in continuous time by cohort c and age a , follows a Poisson process with an intensity described in (6.8) as

$$\lambda(a, c) = f(a) \cdot \kappa(c).$$

Here $f(a)$ is the probability density function of the age at onset given that a cow is eventually deemed to have BSE, and $\kappa(c)$ is the intensity for cattle born at time c which are eventually deemed to have BSE. The Poisson model assumes that BSE cases occur independently. Infected cattle feed is believed to be the major source of BSE transmission (Dealler and Kent 1995, page 6). A Poisson model would not be so suitable if an infected cow can pass on the disease to another cow due to close proximity and a spatial dependence would then need to be considered. In this chapter, we choose a Poisson model to

Table 7.1: Each cell is an observation of $N_{i,k}^{(1)}$ for the incidence of BSE, where $N_{i,k}^{(1)}$ is a random variable for a count in the region $R_{i,i+k}^{(AP)}$ from Figure 6.2.

$n_{i,k}^{(1)}$	Age i								
	2	3	4	5	6	7	8	9	10
1980								17	18
1981							62	43	40
1982						198	123	83	50
1983					879	521	225	172	83
1984				2275	1918	950	440	244	
Cohort k 1985			2557	4065	2561	1268	632		
1986		781	4399	5741	4073	1983			
1987	49	1744	8847	10907	7865				
1988	73	4227	16039	17637					
1989	85	2015	7497						
1990	40	1208							
1991	23								

describe BSE incidence.

Dealler and Kent (1995) and Anderson et al. (1996) analysed BSE incidence data to assess the validity of the independence model. Various modifications to the independence model would then be investigated if the model did not provide a sufficient description of the data. Even though the independence model may be deemed as unsuitable for a given dataset, the independence model is still the natural starting point of an APC analysis. For example, the independence model has been deemed as unsuitable for the study of human mortality (Clayton and Schifflers 1987a; Kramer et al. 2015). An age-cohort perspective was deemed to be more appropriate than an age-period perspective because long-term exposures rather than current exposures were deemed to be the primary driver of BSE incidence.

We can specify a suitable parametric model for the Poisson intensity,

$\lambda(a, c)$, based on a naive inspection of Table 7.1. Our parameterisation of $\lambda(a, c)$ is chosen based on inspection because our intention is for $\lambda(a, c)$ to capture the main features of the MAFF data. We can then capture the main features of the three quantities $\frac{\nu_{i,k}^{(1)}}{\nu_{i,k}}$, $\frac{\nu_{i+1,k}}{\nu_{i,k}} \div \frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}}$ and $\frac{\nu_{i+1,k+1}^{(1)}}{\nu_{i,k+1}^{(1)}} \div \frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}}$ outlined in Theorems 6.1, 6.2 and 6.3, in order to investigate how a misrounded treatment of data affected the study of BSE incidence. Akaike information criterion (AIC) is a measure of how well a statistical model fits some data and it penalises for the number of model parameters (Dobson and Barnett 2008, page 137). The use of AIC to discriminate between various parametric models would be helpful if we were to build a more sophisticated model of BSE incidence.

In this section, we explain that for the case of BSE incidence, the main features of the MAFF data are that the age at onset density is unimodal and sharply peaked at around age five, and that the cohort intensity is rapidly increasing until mid-1988 and rapidly decreasing thereafter. Our simplistic model consists of four parameters: there are two parameters for $\kappa(c)$ and two parameters for $f(a)$. To fully describe the data and achieve small model residuals, it would be necessary to formulate a sophisticated model that consists of a large number of parameters. We expect that our simplistic model for BSE incidence will produce large residuals.

It appears that BSE incidence is approximately doubling with each unit increase in cohort until cohort 1988, and that the number of cases is approximately halving thereafter. For instance, the number of cattle contracting BSE at age eight increases from 62 to 123, and then increases from 123 to 225. The number of cases for age three decreases from 4227 to 2015, and then decreases from 2015 to 1208. A feed ban was introduced in mid-1988 to control the epidemic outbreak of BSE (Dealler and Kent 1995, page 3) and the halving in BSE incidence for cattle born after mid-1988 is compatible with an effective feed ban.

We choose to represent the cohort intensity for BSE incidence as the fol-

lowing function:

$$\kappa(c) = \xi_1 \exp(-\xi_2 |c - 1988.5|) \quad \text{for } c \geq 1979, \quad (7.1)$$

where $\xi_1 > 0$ and $\xi_2 > 0$. The parameter ξ_1 is the peak of the cohort intensity, whereas ξ_2 is the rate of change in the cohort intensity. The cohort intensity is assumed to be increasing exponentially before the feed ban and decreasing exponentially after the ban. The rate of increase and decrease is assumed to be the same so that the cohort intensity is symmetric at around the introduction of the feed ban. The MAFF table has a cohort range of [1979, 1992) because the Lexis parallelogram $R_{i,i+k}^{(AP)}$ extends over cohorts $[k - 1, k + 1)$. We assume that no cattle born at times $c < 1979$ contracted BSE, but that cattle born at times $c > 1992$ can potentially contract BSE. This cohort intensity is strictly log linear.

Table 7.2: Each cell is an observation of $\frac{N_{i+1,k}^{(1)}}{N_{i,k}^{(1)}}$ for the incidence of BSE, where $N_{i,k}^{(1)}$ and $N_{i+1,k}^{(1)}$ are random variables for a count in the regions $R_{i,i+k}^{(AP)}$ and $R_{i+1,i+1+k}^{(AP)}$ respectively. These Lexis regions are depicted in Figure 6.2.

	$\frac{n_{i+1,k}^{(1)}}{n_{i,k}^{(1)}}$	i							
		2	3	4	5	6	7	8	9
	1980								1.06
	1981							0.69	0.93
	1982						0.62	0.67	0.60
	1983					0.59	0.43	0.76	0.48
	1984				0.84	0.50	0.46	0.55	
k	1985			1.59	0.63	0.50	0.50		
	1986		5.63	1.31	0.71	0.49			
	1987	35.59	5.07	1.23	0.72				
	1988	57.90	3.79	1.10					
	1989	23.71	3.72						
	1990	30.20							

Most cattle appear to contract BSE at five years old. The survival dis-

tribution seems to be right skewed because there is a greater spread of case numbers to the right of the peak at age five. A mode at age five coupled with right skew means that we should expect cattle to contract BSE at around age six. The suitability of the independence model to describe BSE incidence can be judged based on an inspection of the count ratios in Table 7.2. We stated in Section 6.5 that if the independence model is true, then for certain conditions for $\lambda(a, c)$, we can expect to see count ratios that are decreasing with k . The count ratio for cattle aged four decreases from 1.59 to 1.31, and then decreases from 1.31 to 1.23. Since the count ratios in each column of Table 7.2 tend to decrease with k , the data could be compatible with the third misrounding effect outlined in Theorem 6.3 and the independence model may be suitable. Before modifications can be made to the independence model, we must first conclude whether the decreasing pattern in the count ratios is compatible with the third misrounding effect.

We choose to represent the survival distribution for BSE incidence as the Gamma model from Table 3.2:

$$f(a) = \frac{\alpha_1^{\alpha_2} a^{\alpha_2-1} \exp(-\alpha_1 a)}{\Gamma(\alpha_2)} \quad \text{for } a > 0, \quad (7.2)$$

where $\alpha_1 > 0$ and $\alpha_2 > 0$. A benefit of the Gamma model relative to the Weibull model is that it can be easily parameterised in terms of the mean and variance. It is more intuitive to interpret the age at onset density in terms of the mean and variance than in terms of the rate and shape parameters, α_1 and α_2 . The expectation and variance of the age-at-onset are determined respectively as

$$\mu_A \equiv \mathbb{E}[A] = \frac{\alpha_2}{\alpha_1} \quad \text{and} \quad \sigma_A^2 \equiv \text{Var}[A] = \frac{\alpha_2}{\alpha_1^2}. \quad (7.3)$$

Recall from (6.24), that the Gamma model is strictly log concave for $\alpha_2 > 1$, log convex for $\alpha_2 < 1$ and log linear for $\alpha_2 = 1$. The survival distribution would only have right skew if it has the property of strict log concavity.

7.2 Exact treatment of data

In Section 6.2, we defined exact and misrounded treatments of age-by-period data for the estimation of the continuous Poisson intensity, $\lambda(a, c)$. An exact treatment refers to correctly using the age-by-period data as if the data are rounded age-by-period, while a misrounded treatment refers to incorrectly using the age-by-period data as if the data are rounded age-by-cohort. Each cell of the MAFF data from Table 7.1 should be assumed to count the number of BSE cases in a Lexis parallelogram rather than a Lexis square. In Sections 6.3, 6.4 and 6.5, we explained that misleading conclusions can arise under certain circumstances when carrying out a misrounded treatment of the data. Misleading conclusions can be avoided by carrying out an exact treatment of the data.

Our model for BSE incidence consists of four parameters to be estimated. There are two parameters for the cohort intensity, denoted by ξ_1 and ξ_2 , and there are two parameters for the survival distribution, denoted by μ_A and σ_A^2 . In this section, we carry out an exact treatment of the data by choosing estimates of the parameters ξ_1 , ξ_2 , μ_A and σ_A^2 to maximise the likelihood function described in equation (6.15). The exact likelihood function can be written more specifically for the case study of BSE as

$$L_{\text{BSE}} \left(\mu_A, \sigma_A^2, \xi_1, \xi_2 \mid \mathbf{N}^{(1)} = \mathbf{n}^{(1)} \right) = \prod_{i=2}^{10} \prod_{k=1980}^{1991} \frac{\left(\nu_{i,k}^{(1)} \right)^{n_{i,k}^{(1)}} \exp \left(-\nu_{i,k}^{(1)} \right)}{n_{i,k}^{(1)}!},$$

where

$$\nu_{i,k}^{(1)} = \iint_{R_{i,i+k}^{(\text{AP})}} \lambda(a, c) \, dc \, da = \int_i^{i+1} \int_{i+k-a}^{i+k+1-a} \lambda(a, c) \, dc \, da.$$

There are no parameters for the effects of period and cohort on the survival distribution due to the independence assumption. In Appendix 9.3, we give a detailed explanation on how to estimate the model parameters for an exact treatment in the statistical package R.

Anderson et al. (1996, page 782) used an exact treatment of the BSE incidence data. The Central Veterinary Laboratory (CVL) collected information on each BSE case including the date of birth, date of onset and age at onset (Donnelly and Ferguson 2000, pages 25–29). Anderson et al. were permitted full access to the CVL database (Donnelly and Ferguson 2000, Preface) and were able to construct a contingency table cross-categorised in terms of age and cohort. The data in each cell could correctly be assumed to count the number of cases in a Lexis square. Dealler and Kent (1995) used a misrounded treatment of the MAFF data because they assumed that each cell of the MAFF table counted the number of BSE cases in a Lexis square. The parameters of an independence model were chosen by Dealler and Kent to maximise the approximate likelihood function described in equation (6.14).

The Poisson intensity for an age-by-cohort region in the Lexis diagram, $\nu_{i,k}$, which was first introduced in equation (6.5), can be written as

$$\int_{a=0}^{a=1} \int_{c=0}^{c=1} \lambda(i+a, k+c) dc da.$$

A Riemann Sum approximation to $\nu_{i,k}$ is

$$\frac{1}{n^2} \sum_{u=1}^n \sum_{v=1}^n \lambda\left(i + \frac{u}{n}, k + \frac{v}{n}\right). \quad (7.4)$$

A Riemann Sum for a square region outside of the APC modelling context has previously been presented by Adams (2006, page 755). This Riemann Sum can be simplified to a product of two cumulative distribution functions:

$$\left[\frac{1}{n} \sum_{u=1}^n f\left(i + \frac{u}{n}\right) \right] \cdot \left[\frac{1}{n} \sum_{v=1}^n \kappa\left(k + \frac{v}{n}\right) \right]. \quad (7.5)$$

Similarly, the Poisson intensity for an age-by-period region, $\nu_{i,k}^{(1)}$, which was introduced in Table 6.1, can be rewritten as

$$\int_0^1 \int_0^1 \lambda(i+a, k+c-a) dc da.$$

We express the Poisson intensity $\nu_{i,k}^{(1)}$ as the following Riemann Sum:

$$\frac{1}{n^2} \sum_{u=1}^n \sum_{v=1}^n \lambda\left(i + \frac{u}{n}, k + \frac{v}{n} - \frac{u}{n}\right). \quad (7.6)$$

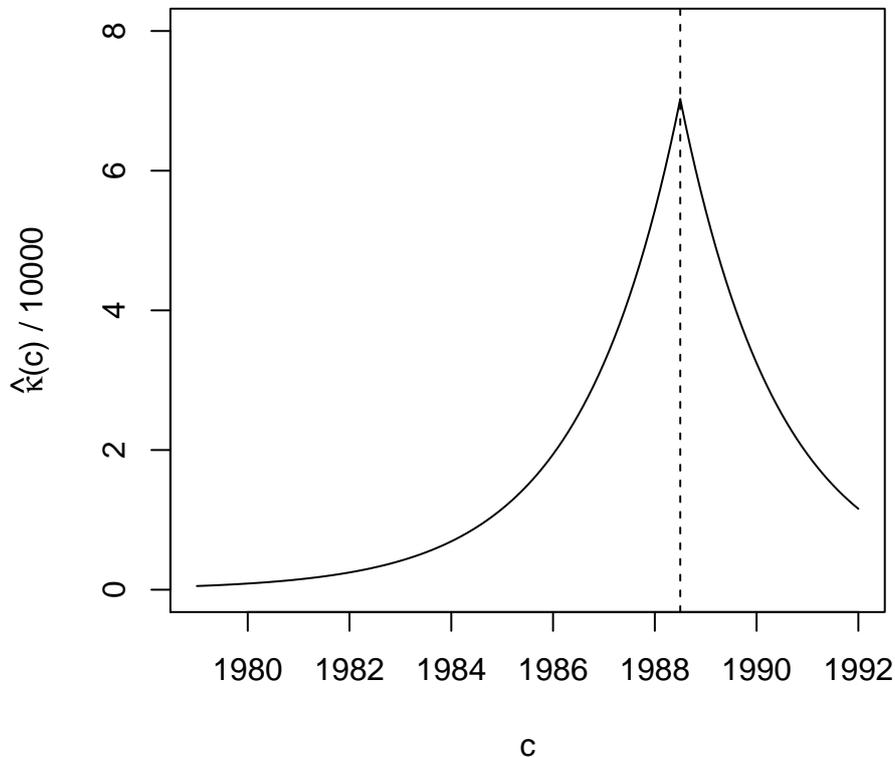


Figure 7.1: An estimate of the intensity for cattle born at time c which are eventually deemed to have BSE. The estimate is denoted as $\hat{\kappa}(c)$ and the estimate was obtained by maximising an exact likelihood function. The specific parametric formula for $\hat{\kappa}(c)$ is stated in (7.8). The dashed vertical line indicates the point in time for the introduction of the feed ban.

The quantity n breaks a year into n intervals. The value of the continuous Poisson intensity is assumed not to change within each interval. A choice of $n = 12$ will split a year into months and $n = 365$ will split a year into days. The larger n is chosen to be, the more accurate our calculation of $\nu_{i,k}^{(1)}$ will be and also the more computing power will be required for the calculation. We consider n values of 1 and 365 to be too small and too large respectively, and choose n to be equal to 30 so that the year is approximately divided into fortnights. Values of n larger than 30 made very little difference to our

calculation of $\nu_{i,k}^{(1)}$.

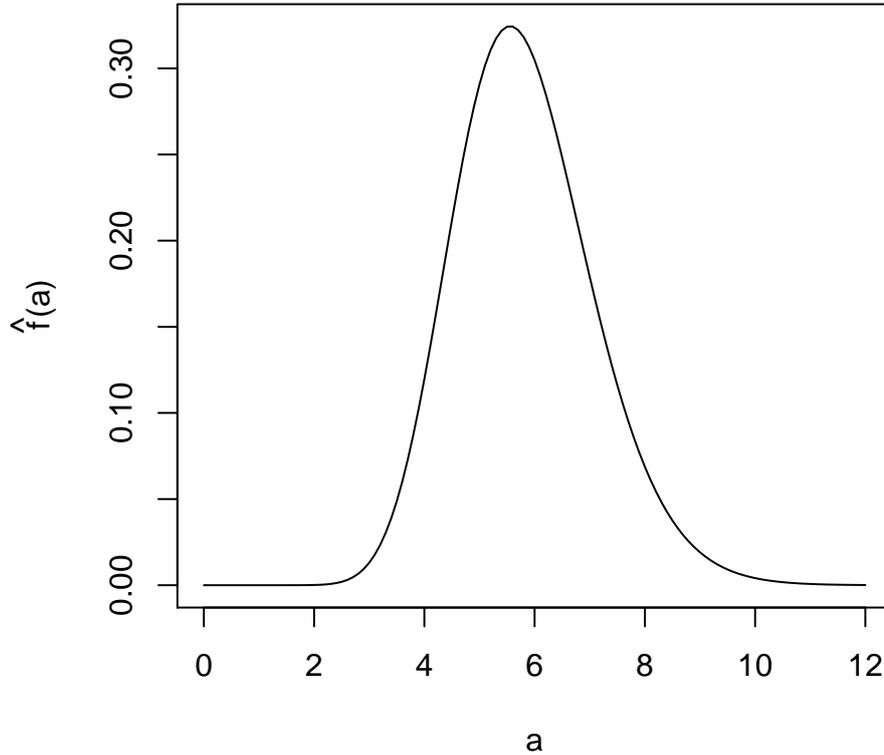


Figure 7.2: An estimate of the probability density function for the age at onset of BSE, given that a cow is eventually deemed to have BSE. The estimate $\hat{f}(a)$ is plotted for $a \in [0, 12)$ and the estimate was obtained by maximising a likelihood function. The specific parametric formula for $\hat{f}(a)$ is stated in (7.9).

By running the R code presented in Appendix 9.3, we obtained the following parameter estimates to three decimal places:

$$\hat{\xi}_1 = 70,248.730, \hat{\xi}_2 = 0.515, \hat{\mu}_A = 5.825, \hat{\sigma}_A^2 = 1.572. \quad (7.7)$$

Hence, the cohort intensity estimated for BSE incidence is

$$\hat{\kappa}(c) = 70,248.730 \cdot \exp(-0.515 |c - 1988.5|) \quad \text{for } c \geq 1979. \quad (7.8)$$

The parameters $\hat{\mu}_A$ and $\hat{\sigma}_A$ imply that the survival distribution estimated for

BSE incidence can be written in terms of the shape and rate parameters as

$$\hat{f}(a) = 1.324 \cdot 10^{-7} a^{20.584} \exp(-3.705a) \quad \text{for } a > 0; \quad (7.9)$$

where $\hat{\alpha}_1 = \frac{\hat{\mu}_A}{\hat{\sigma}_A^2} = 3.705$ and $\hat{\alpha}_2 = \frac{\hat{\mu}_A^2}{\hat{\sigma}_A^2} = 21.584$. The estimated survival distribution is strictly log concave because $\hat{\alpha}_2$ is greater than one. Our estimate of the continuous Poisson intensity can be deduced as the product of $\hat{\kappa}(c)$ and $\hat{f}(a)$.

The estimates of the cohort intensity and survival distribution for BSE incidence are displayed in Figures 7.1 and 7.2, respectively. The estimated cohort intensity is rapidly increasing before the feed ban and rapidly decreasing after the ban. The cohort intensity is symmetric about the dashed vertical line and peaks at 70,248.730. The estimated survival distribution has a mean of 5.825 and has a slight right skew. The survival distribution peaks very close to the mean and is very sharply peaked so that there is almost no cases of BSE below age three and above age ten. Hence, the estimated Poisson intensity seems to be a suitable model for the incidence of BSE based on our inspection of the MAFF table in Section 7.1.

In Table 7.3, we present the estimate of the age-by-period discretised Poisson intensity, $\hat{\nu}_{i,k}^{(1)}$, under our fitted model. We can assess the suitability of the independence model for BSE incidence, $\hat{\lambda}(a, c)$, by comparing the fit of the age-by-period discretised Poisson intensities to the MAFF data. To account for both the magnitude and sign of the residual, we consider a Pearson residual for each (i, k) cell (Dobson and Barnett 2008, pg. 167):

$$E_{i,k} = \frac{N_{i,k}^{(1)} - \nu_{i,k}^{(1)}}{\sqrt{\nu_{i,k}^{(1)}}}.$$

In Section 7.1, we explained that our intention was for $\hat{\lambda}(a, c)$ to capture the main features of the data. The parametric form specified for our model is simplistic and we expect the Pearson residuals to be large.

Let X_1, \dots, X_m be independent and identically distributed random variables with expectation μ_X and variance σ_X^2 . The sample mean is denoted by

Table 7.3: Each cell is an estimate of the age-by-period discretised Poisson intensity, $\nu_{i,k}^{(1)}$, for the independence model described in (7.8) and (7.9). The calculations are rounded to the nearest integer.

$\hat{\nu}_{i,k}^{(1)}$	i								
	2	3	4	5	6	7	8	9	10
1980	3	49	191	292	230	112	37	9	2
1981	5	82	320	488	385	187	63	16	3
1982	8	137	535	817	644	313	105	26	5
1983	14	229	895	1367	1078	523	176	44	9
1984	24	384	1498	2287	1804	876	294	74	15
k 1985	40	642	2508	3828	3020	1466	492	124	25
1986	66	1075	4197	6407	5054	2454	824	207	41
1987	111	1799	7025	10724	8459	4107	1378	346	69
1988	185	2968	11459	17336	13580	6559	2192	549	109
1989	222	3253	11669	16746	12638	5942	1947	480	94
1990	139	2012	7153	10208	7675	3600	1177	290	57
1991	83	1202	4274	6099	4586	2151	703	173	34

$\bar{X} = \frac{1}{m} \sum_{u=1}^m X_u$. The central limit theorem (CLT) states that, as $m \rightarrow \infty$, the sample mean tends in distribution to a normal distribution with mean μ_X and variance $\frac{\sigma_X^2}{m}$ (Lipschutz and Schiller 1998, page 188). Given $\bar{X} \sim N\left(\mu_X, \frac{\sigma_X^2}{m}\right)$, there is a 0.95 probability that

$$Z = \frac{\bar{X} - \mu_X}{\sqrt{\sigma_X^2/m}} \sim N(0, 1)$$

will take a value in the interval $(-1.96, 1.96)$. The formula for the Pearson residual can be derived by substituting $m = 1$ and $X_1 = N_{i,k}^{(1)}$ into the expression for Z . If it is reasonable to assume that $E_{i,k} \sim N(0, 1)$ for all i and k , then our model can be considered to fully describe the MAFF data when the absolute value of each Pearson residual is less than 1.96.

A measure of the total fit of our model to the 44 cells in the MAFF data can be assessed by carrying out a chi-squared test. Let $\chi_q^2(5\%)$ denote the

Table 7.4: Each cell is an estimate of the Pearson residual $e_{i,k} = \frac{n_{i,k}^{(1)} - \nu_{i,k}^{(1)}}{\sqrt{\nu_{i,k}^{(1)}}}$ to one decimal place. The quantities $n_{i,k}^{(1)}$ and $\hat{\nu}_{i,k}^{(1)}$ are displayed in Tables 7.1 and 7.3, respectively.

$\hat{e}_{i,k}$	i								
	2	3	4	5	6	7	8	9	10
1980								2.5	11.8
1981							-0.1	6.9	20.8
1982						-6.5	1.8	11.0	19.5
1983					-6.1	-0.1	3.7	19.3	25.1
1984				-0.3	2.7	2.5	8.5	19.8	
k 1985			1.0	3.8	-8.4	-5.2	6.3		
1986		-9.0	-3.1	-8.3	-13.8	-9.5			
1987	-5.9	-1.3	21.7	1.8	-6.5				
1988	-8.2	23.1	42.8	2.3					
1989	-9.2	-21.7	-38.6						
1990	-8.4	-17.9							
1991	-6.6								

critical value at a 5% significance level for a chi-squared distribution with q degrees of freedom. If $E_{i,k} \sim N(0, 1)$ for all i and k , then the sum of squares,

$$S^* = \sum_{i=2}^{10} \sum_{k=1980}^{1991} E_{i,k}^2,$$

follows a chi-squared distribution with $44 - 4 = 40$ degrees of freedom (Dobson and Barnett 2008, pages 167–168). The number of degrees of freedom is calculated as the number of non-empty cells in Table 7.4 minus the number of estimated parameters. The value of $\chi_{40}^2(5\%)$ can be obtained from Lipschutz and Schiller (1998, page 361) as 55.8. Therefore, if it is reasonable to assume that $E_{i,k} \sim N(0, 1)$ for all i and k , then our model can be considered to fully describe the MAFF data at a 5% significance level when S^* is less than 55.8.

The values of the Pearson residuals for our model are displayed in Table 7.4. An inspection of this table suggests that the independence model does not

give a good description of the MAFF data because the majority of the residuals are outside of the interval $(-1.96, 1.96)$ and the sum of squares is clearly much greater than 55.8. In particular, the model fit is quite poor in the upper-right and lower-left parts of Table 7.4. It is clear that some modifications would help to improve the fit of $\hat{\lambda}(a, c)$. In this case, the modifications relate to building a more sophisticated independence model rather than introducing a dependence on period or cohort. One recommendation is to assign less probability mass to younger ages to overcome the overstating of BSE incidence, and to assign more mass to older ages to overcome the understating of incidence.

Our simplistic model only has four parameters and modifications to achieve a good fit will require the incorporation of a large number of parameters. The purpose of $\hat{\lambda}(a, c)$ is to capture the main features of the data such as a sharply peaked survival density with a mean of approximately six years as well as a cohort intensity that increases rapidly and then decreases rapidly. Figures 7.1 and 7.2 show that $\hat{\lambda}(a, c)$ does capture the main features of the data. Our fitted model allows us to capture the main features about $\nu_{i,k}$ and $\nu_{i,k}^{(1)}$, so that we can then capture the main features about the three quantities $\frac{\nu_{i,k}^{(1)}}{\nu_{i,k}}$, $\frac{\nu_{i+1,k}}{\nu_{i,k}} \div \frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}}$ and $\frac{\nu_{i+1,k+1}^{(1)}}{\nu_{i,k+1}^{(1)}} \div \frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}}$. Further, in Section 7.3, our fitted model allows us to investigate approximately how the three misrounding effects outlined in Theorems 6.1, 6.2 and 6.3 impacted on the study of BSE incidence by Dealler and Kent. We conclude that our simplistic model provides a sufficient fit for us to demonstrate the flaws of an analysis for BSE incidence under a misrounded treatment of data.

7.3 Misrounding effects for BSE incidence

In this section, we show that for our estimate of the Poisson intensity, there are three misleading effects that appear when using a misrounded treatment of

the MAFF data to fit an independence model. Under a misrounded treatment, the MAFF data are assumed to be observations in Lexis squares as opposed to Lexis parallelograms. This means that the number of BSE cases in cell (i, k) of the MAFF table is assumed to have an expected value of $\nu_{i,k}$ rather than its true expectation $\nu_{i,k}^{(1)}$. All three effects are mathematical consequences of the model assumptions such as a cohort intensity that is strictly log linear, strictly increasing before the feed ban and strictly decreasing after the feed ban. Also, the survival distribution is assumed to be independent of cohort and strictly log concave.

The first misrounding effect is about an apparent under-reporting of case numbers before the feed ban and an apparent over-reporting of case numbers after the feed ban. The age-by-cohort discretised Poisson intensity, $\nu_{i,k}$, is the expected number of cases in the Lexis square, $R_{i,k}^{(AC)}$, as depicted in Figure 6.1. The estimate of this Poisson intensity is presented in Table 7.5 and can be calculated from our independence model using the product formula, $[\hat{F}(i+1) - \hat{F}(i)] \cdot [\hat{Q}(k+1) - \hat{Q}(k)]$. A comparison of each pair of (i, k) cells in Tables 7.3 and 7.5 shows that the following strict inequalities hold true for any i :

$$\hat{\nu}_{i,k}^{(1)} < \hat{\nu}_{i,k} \text{ for } k = 1980, 1981, \dots, 1988; \quad (7.10)$$

$$\hat{\nu}_{i,k}^{(1)} > \hat{\nu}_{i,k} \text{ for } k = 1989, 1990, 1991. \quad (7.11)$$

For example, the number of cattle born in year 1982 that contract BSE aged three is expected to be 137 for an age-by-period rounding, but is expected to be 188 for an age-by-cohort rounding.

In Theorem 6.1, we proved that certain inequalities hold between Poisson intensities for the number of deaths in Lexis parallelograms and in Lexis squares. The two inequalities (7.10) and (7.11) hold true because, in our model for BSE incidence, the cohort intensity is strictly increasing before the feed ban and is strictly decreasing after the ban. However, it is surprising that a strict inequality holds for cohort $k = 1988$ since the cohort intensity is increasing and then decreasing during the time interval (1988, 1989). The equality of obser-

Table 7.5: Each cell is an estimate of the age-by-cohort discretised Poisson intensity, $\nu_{i,k}$, for the independence model described in (7.8) and (7.9). The calculations are rounded to the nearest integer.

$\hat{\nu}_{i,k}$	i								
	2	3	4	5	6	7	8	9	10
1980	4	67	250	369	284	136	45	11	2
1981	7	112	418	617	475	227	75	19	4
1982	12	188	700	1033	796	380	126	31	6
1983	21	315	1172	1729	1332	636	211	52	10
1984	35	527	1961	2893	2228	1064	353	88	17
k 1985	58	883	3282	4842	3730	1780	590	147	29
1986	97	1477	5494	8104	6242	2980	987	246	49
1987	162	2472	9194	13564	10447	4987	1653	411	81
1988	233	3562	13247	19543	15052	7185	2381	592	117
1989	158	2409	8961	13219	10181	4860	1611	401	79
1990	94	1440	5354	7898	6083	2904	962	239	47
1991	56	860	3199	4719	3635	1735	575	143	28

variations for some of the paired (i, k) cells is only apparent due to our rounding of $\hat{\nu}_{i,k}^{(1)}$ and $\hat{\nu}_{i,k}$ to the nearest integer.

In Figure 7.3, we plot the ratio of discretised Poisson intensities, $\frac{\hat{\nu}_{i,k}^{(1)}}{\hat{\nu}_{i,k}}$, as a function of i for two cohorts before the ban and for two cohorts after the ban. It can be seen that the two inequalities hold true since $\frac{\hat{\nu}_{i,k}^{(1)}}{\hat{\nu}_{i,k}} < 1$ for the pre-ban cohorts and $\frac{\hat{\nu}_{i,k}^{(1)}}{\hat{\nu}_{i,k}} > 1$ for post-ban cohorts. Before the feed ban, the ratio $\frac{\hat{\nu}_{i,k}^{(1)}}{\hat{\nu}_{i,k}}$ is independent of cohort and takes values between 0.682 and 0.845 to three decimal places. The estimated ratio of Poisson intensities after the ban is independent and takes values between 1.206 and 1.482 to three decimal places. For $\frac{\hat{\nu}_{i,k}^{(1)}}{\hat{\nu}_{i,k}}$, the patterns of change with age i before and after the ban are later explained by the second misrounding effect. The independence of $\frac{\hat{\nu}_{i,k}^{(1)}}{\hat{\nu}_{i,k}}$ with k is not explained by any of the three misrounding effects.

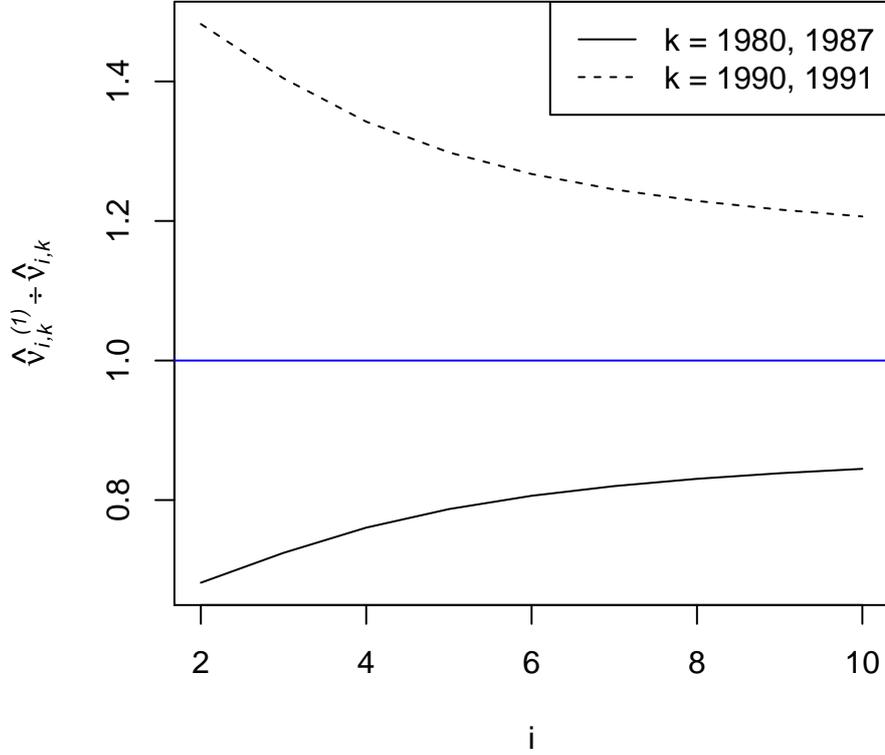


Figure 7.3: The curved lines are calculations for the estimated ratio of Poisson intensities $\frac{\hat{\nu}_{i,k}^{(1)}}{\hat{\nu}_{i,k}}$ for particular cohorts. The values of $\hat{\nu}_{i,k}^{(1)}$ and $\hat{\nu}_{i,k}$ are presented in Tables 7.3 and 7.5. Each curved line consists of nine points which have been joined together with straight lines. The horizontal line drawn at the point the y-axis equals one indicates an idealistic value for the ratio of intensities.

The level of apparent under-reporting for any cohort before the feed ban, which can be measured as a percentage as $\left(1 - \frac{\hat{\nu}_{i,k}^{(1)}}{\hat{\nu}_{i,k}}\right) \cdot 100$, takes values between 15.50 and 31.85 to two decimal places. The percentage of apparent over-reporting after the feed ban, $\left(\frac{\hat{\nu}_{i,k}^{(1)}}{\hat{\nu}_{i,k}} - 1\right) \cdot 100$, takes values between 20.65 and 48.23 to two decimal places. The levels of under-reporting and over-reporting are also decreasing before and after the ban for unit increases in cohort. The levels of under-reporting and over-reporting are quite large due to the fast rate of change in the cohort intensity over time. Hence, if the independence model

is suitable and a misrounded treatment of data is adopted for model fitting, then we expect the number of cases to appear too small before the feed ban and too large after the ban.

The percentage loss in accuracy incurred by carrying out a misrounded treatment of data as opposed to an exact treatment can be calculated for each (i, k) cell as $100 \cdot \left| \frac{\hat{\nu}_{i,k}^{(s)}}{\hat{\nu}_{i,k}} - 1 \right|$. This is because, under a misrounded treatment of data, it is assumed that the observation in cell (i, k) of the MAFF data has an expected value of $\nu_{i,k}$. In reality, the observation in cell (i, k) has an expected value of $\nu_{i,k}^{(1)}$. A statistician's estimate of $\nu_{i,k}$ under a misrounded treatment of data would be strictly overstated after the ban by around 21-48% and understated before the ban by around 16-32%. Overall, a statistician incurs a loss in accuracy of approximately 16-48% per cell of the MAFF table.

The second misrounding effect is about an expected age-at-onset for BSE that appears too large before the ban and too small after the ban. The estimated ratios of age-by-cohort and age-by-period discretised Poisson intensities, $\frac{\hat{\nu}_{i+1,k}}{\hat{\nu}_{i,k}}$ and $\frac{\hat{\nu}_{i+1,k}^{(1)}}{\hat{\nu}_{i,k}^{(1)}}$, are presented in Tables 7.6 and 7.7, respectively. A comparison of the estimates for each pair of (i, k) cells shows that the following strict inequalities hold true for any i :

$$\frac{\hat{\nu}_{i+1,k}^{(1)}}{\hat{\nu}_{i,k}^{(1)}} > \frac{\hat{\nu}_{i+1,k}}{\hat{\nu}_{i,k}} \text{ for } k = 1980, 1981, \dots, 1988; \quad (7.12)$$

$$\frac{\hat{\nu}_{i+1,k}^{(1)}}{\hat{\nu}_{i,k}^{(1)}} < \frac{\hat{\nu}_{i+1,k}}{\hat{\nu}_{i,k}} \text{ for } k = 1989, 1990, 1991. \quad (7.13)$$

For example, the number of cases for cattle born in year 1982 increases by a factor of 3.941 between ages three and four under an age-by-period rounding, but only increases by a factor of 3.754 under an age-by-cohort rounding.

In Theorem 6.2, we proved that certain inequalities hold between ratios of Poisson intensities for the number of deaths in Lexis parallelograms and in Lexis squares. The two inequalities (7.12) and (7.13) hold true because, in our model for BSE incidence, the survival distribution is log concave and the cohort intensity is strictly increasing before the ban and is strictly decreasing

Table 7.6: Each cell is a calculation for the ratio of age-by-cohort discretised Poisson intensities, $\frac{\hat{\nu}_{i+1,k}}{\hat{\nu}_{i,k}}$, for the independence model described in (7.8) and (7.9). Our calculations of $\hat{\nu}_{i,k}$ are displayed in Table 7.5. The calculations are presented to three decimal places.

$\frac{\hat{\nu}_{i+1,k}}{\hat{\nu}_{i,k}}$	i							
	2	3	4	5	6	7	8	9
$\forall k$	15.480	3.754	1.485	0.774	0.479	0.332	0.249	0.198

Table 7.7: Each cell is an estimate of the ratio of age-by-period rounded Poisson intensities, $\frac{\hat{\nu}_{i+1,k}^{(1)}}{\hat{\nu}_{i,k}^{(1)}}$, for the independence model described in (7.8) and (7.9). Our calculations of $\hat{\nu}_{i,k}^{(1)}$ are displayed in Table 7.3. The calculations are presented to three decimal places.

$\frac{\hat{\nu}_{i+1,k}^{(1)}}{\hat{\nu}_{i,k}^{(1)}}$	i							
	2	3	4	5	6	7	8	9
≤ 1987	16.451	3.941	1.537	0.793	0.487	0.337	0.252	0.199
k 1988	16.299	3.901	1.524	0.787	0.485	0.335	0.251	0.199
1989	14.888	3.623	1.445	0.758	0.472	0.329	0.247	0.197
≥ 1990	14.666	3.588	1.436	0.755	0.471	0.328	0.247	0.196

after the ban. Similar to our interpretation of the first misrounding effect, it is surprising that there is a strict inequality for cohort year 1988.

In Figure 7.4, we plot the relative difference between ratios for age-by-cohort and age-by-period intensities, $\frac{\hat{\nu}_{i+1,k}}{\hat{\nu}_{i,k}} \div \frac{\hat{\nu}_{i+1,k}^{(1)}}{\hat{\nu}_{i,k}^{(1)}}$, as a function of i for two cohorts before the ban and two cohorts after the ban. It can be seen that the two inequalities (7.12) and (7.13) hold true. The estimated ratio for age-by-cohort intensities is a function of i and is independent of k :

$$\frac{\hat{\nu}_{i+1,k}}{\hat{\nu}_{i,k}} \equiv \hat{\zeta}_i = \frac{\hat{F}(i+2) - \hat{F}(i+1)}{\hat{F}(i+1) - \hat{F}(i)} \quad \forall i, k.$$

The estimated ratio for age-by-period intensities is largely independent of cohort because the value of $\frac{\hat{\nu}_{i+1,k}^{(1)}}{\hat{\nu}_{i,k}^{(1)}}$ is independent of k for cohorts 1980 to 1987. The ratio then decreases with cohort for all i until cohort 1989 and thereafter the ratio is independent of cohort. We explained in (6.27) that these indepen-

dent ratios are a mathematical consequence of a log linear cohort intensity.

The relative ratio $\frac{\hat{\nu}_{i+1,k}}{\hat{\nu}_{i,k}} \div \frac{\hat{\nu}_{i+1,k}^{(1)}}{\hat{\nu}_{i,k}^{(1)}}$ takes values between 0.941 and 0.993 before the ban and values between 1.001 and 1.055 after the ban. In Section 6.4, we explained that strict inequalities such as the inequalities described in (7.12) and (7.13) can be interpreted in terms of the expected age-at-onset. Further, we can say that if the independence model is suitable and a misrounded treatment of data is adopted for model fitting, then the expected age-at-onset for BSE appears too large before the feed ban and too small after the ban. However, the discrepancy from the true expected age-at-onset will not be significant since the estimate of the relative ratio is close to one for all i and k .

The percentage loss in accuracy incurred by carrying out a misrounded treatment of data as opposed to an exact treatment can also be calculated for each (i, k) cell in terms of the relative ratio as $100 \cdot \left| \left(\frac{\hat{\nu}_{i+1,k}}{\hat{\nu}_{i,k}} \div \frac{\hat{\nu}_{i+1,k}^{(s)}}{\hat{\nu}_{i,k}^{(s)}} \right) - 1 \right|$. This is because it is assumed under a misrounded treatment of data that the ratio of counts taken from cells (i, k) and $(i+1, k)$ has an expected value of $\frac{\hat{\nu}_{i+1,k}}{\hat{\nu}_{i,k}} = \hat{\zeta}_i$. In reality, the ratio of counts taken from cells (i, k) and $(i+1, k)$ has an expected value of $\frac{\hat{\nu}_{i+1,k}^{(s)}}{\hat{\nu}_{i,k}^{(s)}}$. A statistician's estimate of $\nu_{i,k}$ under a misrounded treatment of data would have ratios that are overstated before the ban by around 0.7-5.9% and that are understated after the ban by around 0.1-5.5%. Overall, a statistician incurs a loss in accuracy in terms of count ratios of approximately 0.1-5.9% for each pair of cells in the MAFF table. The 0.1-5.9% loss in accuracy in the count ratios is much less than the 16-48% loss for the counts.

The final misrounding effect is about an apparent decrease in the expected-at-onset at around the time of the feed ban. In Theorem 6.3, we proved that certain inequalities hold between count cross-ratios of Poisson intensities for the number of deaths in Lexis parallelograms. Our estimates in Table 7.7 show

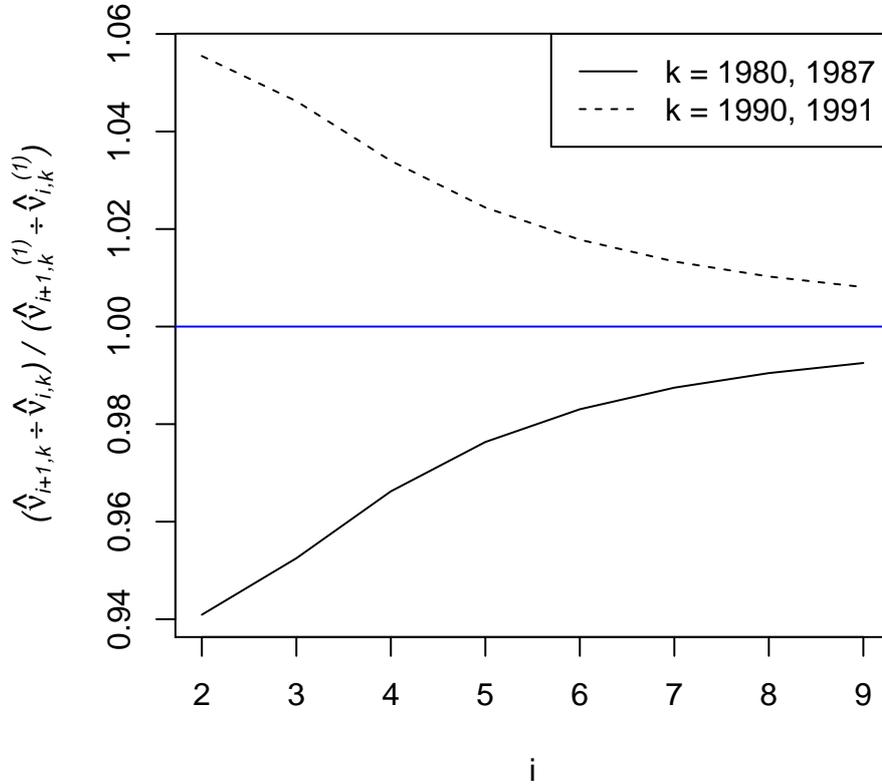


Figure 7.4: The curved lines are calculations for the relative difference between ratios for age-by-cohort and age-by-period Poisson intensities, $\frac{\hat{\nu}_{i+1,k}}{\hat{\nu}_{i,k}} \div \frac{\hat{\nu}_{i+1,k}^{(1)}}{\hat{\nu}_{i,k}^{(1)}}$, for particular cohorts. The values of $\frac{\hat{\nu}_{i+1,k}}{\hat{\nu}_{i,k}}$ and $\frac{\hat{\nu}_{i+1,k}^{(1)}}{\hat{\nu}_{i,k}^{(1)}}$ are presented in Tables 7.6 and 7.7. Each curved line consists of eight points which have been joined together with straight lines. The horizontal line drawn at the point the y-axis equals 1.0 indicates an idealistic value of the relative difference between ratios. The cohort intensity is strictly log linear and the survival distribution is strictly log concave.

that the following equalities and inequalities hold true for any i :

$$\frac{\hat{\nu}_{i+1,k+1}^{(1)}}{\hat{\nu}_{i,k+1}^{(1)}} = \frac{\hat{\nu}_{i+1,k}^{(1)}}{\hat{\nu}_{i,k}^{(1)}} \quad \text{for } k = 1980, 1981, \dots, 1986, 1990. \quad (7.14)$$

$$\frac{\hat{\nu}_{i+1,k+1}^{(1)}}{\hat{\nu}_{i,k+1}^{(1)}} < \frac{\hat{\nu}_{i+1,k}^{(1)}}{\hat{\nu}_{i,k}^{(1)}} \quad \text{for } k = 1987, 1988, 1989. \quad (7.15)$$

The equality (7.14) implies that the survival distribution for BSE incidence

appears to be unchanging before the feed ban and unchanging after the ban. The equalities are a consequence of a strictly log linear cohort intensity. However, the inequality (7.15) implies that the survival distribution is shifting to the left at around the time of the feed ban.

Building on our discussion of the second misrounding effect for BSE incidence, we can say that the expected age-at-onset appears too large before the feed ban, appears to decrease below the true expected age-at-onset at around the time of the ban, and then appears too small after the ban. This apparent shift in the survival distribution is indicated by the shift between the two curves in Figure 7.4 for cohorts 1987 and 1990. Since our independence model assumes that the survival distribution is independent of cohort, the change in the expected age-at-onset is only apparent due to misrounding. Hence, if the independence model is true and a misrounded treatment of data is adopted for model fitting, then the expected age-at-onset for BSE will appear to decrease at around the time of the feed ban.

7.4 Implications for BSE incidence

In this chapter, we have shown that if an independence model $\lambda(a, c) = f(a) \cdot \kappa(c)$ is true for BSE incidence and a misrounded treatment is carried out for the MAFF data rounded age-by-period, then there are misleading effects that arise when we assess the suitability of the independence model. One effect is that the expected number of BSE cases appears to be too small for cohorts $c < 1988.5$ and too large for cohorts $c > 1988.5$. Another effect is that the expected age-at-onset for BSE appears too large for cattle born before the ban, appears to decrease for cattle born at around the time of the ban, and appears too small for cattle born after the ban.

The MAFF data presented in Table 7.1 consists of signal and noise. By

signal, we refer to the underlying true intensity of deaths indexed in continuous time by age and cohort, $\lambda(a, c)$. A substantial amount of noise in the data can make it difficult to identify the signal. For instance, Dealler and Kent (1995, page 5) stated that powerful commercial and political forces could have led to a substantial under-reporting of cases after the feed ban. This is because there was a large fall after the ban in the level of compensation offered to farmers for BSE cases. Also, veterinary officers were sometimes reluctant to record cases of BSE as it would be expensive for MAFF.

The true number of deaths in age-by-period regions and age-by-cohort regions of the Lexis diagram, denoted by $\nu_{i,k}^{(s)}$ and $\nu_{i,k}$, can be derived as a consequence of the signal. Our estimate of $\lambda(a, c)$ derived for an exact treatment of the MAFF data was presented in equations (7.8) and (7.9). The estimate $\hat{\lambda}(a, c)$ was discretised to obtain our corresponding estimates of $\nu_{i,k}^{(s)}$ and $\nu_{i,k}$. Estimates of the ratios $\frac{\nu_{i,k}^{(1)}}{\nu_{i,k}}$ and $\frac{\nu_{i+1,k}}{\nu_{i,k}} \div \frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}}$ were illustrated in Figures 7.3 and 7.4 for each some (i, k) cells. The misleading effects that resulted from a misrounded treatment of the MAFF data were only discussed in terms of the two ratios, that is, they were only discussed in terms of the signal. The misleading effects may not be apparent when accounting for noise in the data.

Suppose that we simulate values of $n_{i,k}^{(1)}$ from a Poisson model with parameter $\hat{\nu}_{i,k}^{(1)}$ and simulate values of $n_{i,k}$ from a Poisson model with parameter $\hat{\nu}_{i,k}$. In Figure 7.5, the simulated count ratio $\frac{n_{i,k}^{(1)}}{n_{i,k}}$ is plotted for cohorts 1980, 1987 and 1991. This plot incorporates noise into Figure 7.3. For cohorts 1980 and 1987, the ratio $\frac{n_{i,k}^{(1)}}{n_{i,k}}$ is less than one for all ages with the exception of the cell corresponding to cohort 1980 and age ten. If we exclude this outlier, we can say that $\frac{n_{i,k}^{(1)}}{n_{i,k}}$ takes values between 0.300 and 0.800 for cohort 1980 and takes values between 0.667 and 0.954 for cohort 1987 to three decimal places. In comparison, the underlying ratio $\frac{\nu_{i,k}^{(1)}}{\nu_{i,k}}$ was estimated to take values between 0.682 and 0.845 to three decimal places.

For cohort 1991, the ratio $\frac{n_{i,k}^{(1)}}{n_{i,k}}$ is strictly greater than one and it only takes values between 1.196 and 1.464. In comparison, the underlying ratio

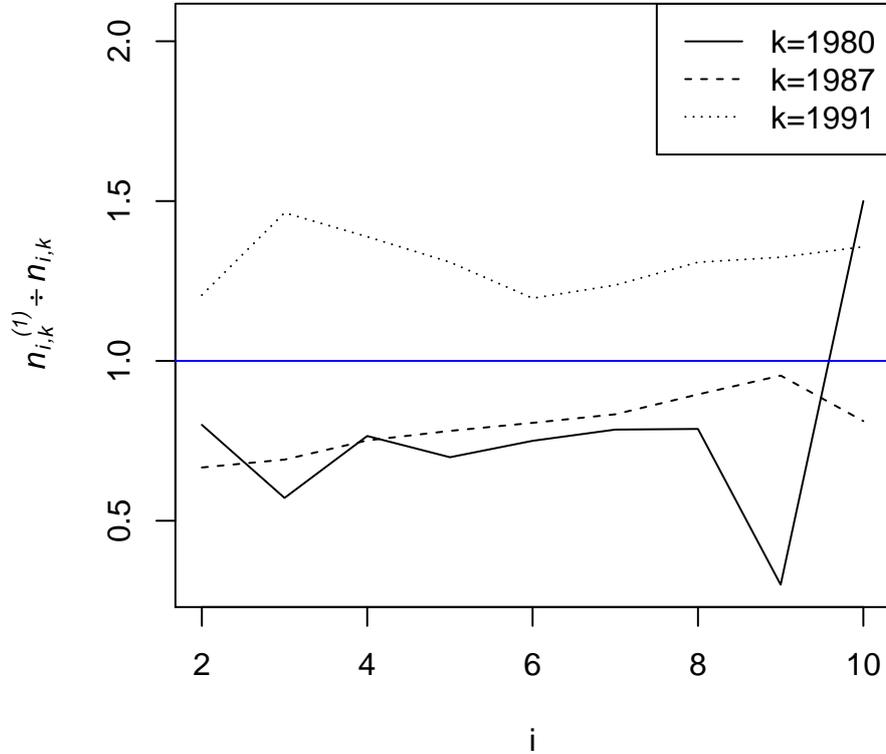


Figure 7.5: A simulation of the ratio of counts, $\frac{n_{i,k}^{(1)}}{n_{i,k}}$, for particular cohorts. The expected value of the simulated count ratio is $\frac{\hat{\nu}_{i,k}^{(1)}}{\hat{\nu}_{i,k}}$ and its values are presented in Figure 7.3.

$\frac{\nu_{i,k}^{(1)}}{\nu_{i,k}}$ was estimated to take values between 1.206 and 1.482 to three decimal places. Since the simulated count ratio is not very close to one, and the count ratio largely satisfies the strict inequalities outlined in Theorem 6.1 for certain conditions about the cohort intensity, we can conclude that the level of under-reporting before the feed ban and the level of over-reporting after the ban that are apparent in the signal are also apparent in the noisy data. Hence, if the independence model is true and a misrounded treatment is carried out for the MAFF data, there are misleading effects that are apparent in relation to the under-reporting and over-reporting of BSE cases. The first misleading effect was likely to have affected the analysis of BSE incidence conducted by Dealler

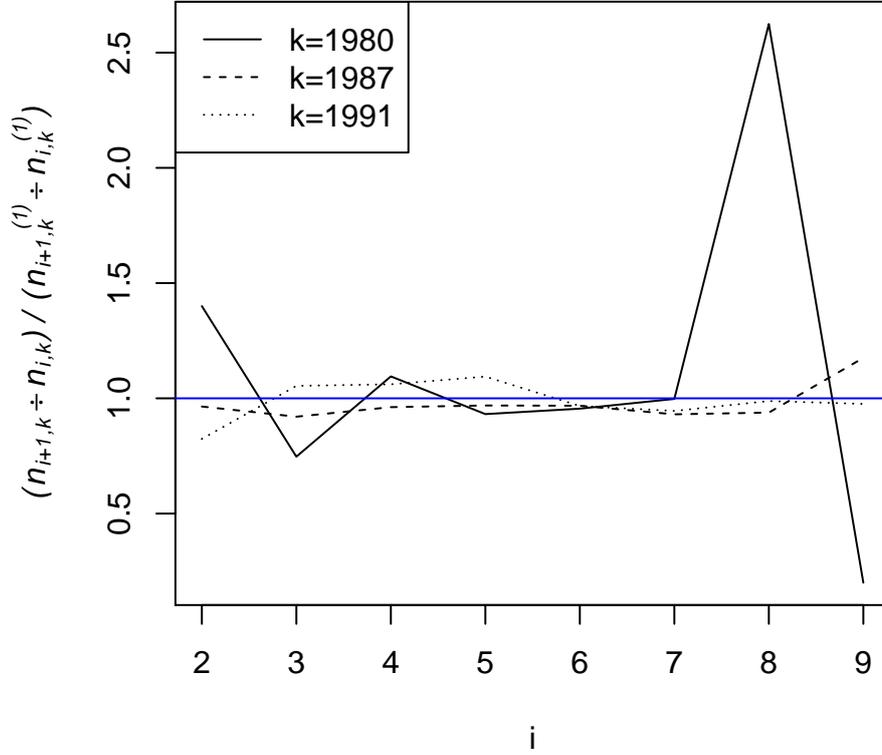


Figure 7.6: A simulation of the relative ratio, $\frac{n_{i+1,k}}{n_{i,k}} \div \frac{n_{i+1,k}^{(1)}}{n_{i,k}^{(1)}}$, for cohorts 1980, 1987 and 1991. The expected value of the simulated relative ratio is $\frac{\hat{\nu}_{i+1,k}}{\hat{\nu}_{i,k}} \div \frac{\hat{\nu}_{i+1,k}^{(1)}}{\hat{\nu}_{i,k}^{(1)}}$. So this plot can be viewed as introducing noise to Figure 7.4.

and Kent (1995).

In Figure 7.6, the simulated relative ratio $\frac{n_{i+1,k}}{n_{i,k}} \div \frac{n_{i+1,k}^{(1)}}{n_{i,k}^{(1)}}$ is plotted for cohorts 1980, 1987 and 1991. This plot incorporates noise into Figure 7.4. The simulations of the relative ratio fluctuate with age above and below one for all cohorts. In comparison, the underlying relative ratio $\frac{\nu_{i+1,k}}{\nu_{i,k}} \div \frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}}$ is strictly less than one for cohorts 1980 and 1987 and is strictly greater than one for cohort 1991. Since the simulated relative ratio is close to one, and it does not satisfy the strict inequalities outlined in Theorems 6.2 and 6.3, we can conclude that the decrease in the expected age-at-onset for BSE incidence over

time that is apparent in the signal is not apparent for the noisy data. Hence, if an independence model is true and a misrounded treatment is carried out for the MAFF data, the second misleading effect in relation to the changing age-at-onset distribution is not apparent and is unlikely to have affected the analysis of BSE incidence conducted by Dealler and Kent (1995).

In Table 7.2, we took the counts for cells (i, k) and $(i + 1, k)$ in the MAFF table and calculated their ratio. It was apparent that the count ratio was decreasing on average with cohort for each year of age. We stated towards the end of Section 7.1 that before modifications are made to the independence model, we must first conclude whether the decreasing pattern in Table 7.2 is attributable to Theorem 6.3. Since the decrease in the expected age-at-onset for BSE incidence over time is apparent in the signal but not apparent for the noisy data, the decreasing pattern in the observed ratios cannot be attributed to Theorem 6.3. A model for BSE incidence should allow for a slight dependence on cohort.

The conclusions reached in Chapter 3 inform us how dependence should be handled in an APC model. If we were to allow the survival distribution to vary with period and cohort, then a prime example which also does not suffer from the confounding issues discussed in Chapter 5 would be to specify the age at onset density according to equations (4.4) and (4.5). If we were to allow the survival distribution to vary only with cohort, then we could extend our model for BSE to specify that the age at onset follows a Gamma distribution where the mean μ_A varies with cohort. Recalling the gamma density from (9.3), we can write, for $a > 0$ and $c \geq 1979$,

$$f(a | c) = \frac{\alpha_1(c)^{\alpha_2(c)} a^{\alpha_2(c)-1} \exp(-\alpha_1(c)a)}{\Gamma(\alpha_2(c))}, \quad (7.16)$$

where

$$\alpha_1(c) = \frac{\mu_A \cdot \exp(-\beta(c - 1980))}{\sigma_A^2} \equiv \frac{\mu_A(c)}{\sigma_A^2} \quad (7.17)$$

and

$$\alpha_2(c) = \frac{(\mu_A \cdot \exp(-\beta(c - 1980)))^2}{\sigma_A^2} \equiv \frac{\mu_A^2(c)}{\sigma_A^2}. \quad (7.18)$$

We assume that the expected age at onset is a function of c , which means that the shape and rate parameters are also functions of c . The variance does not change with c . If $\beta > 0$, then the expected age at onset density is decreasing with c .

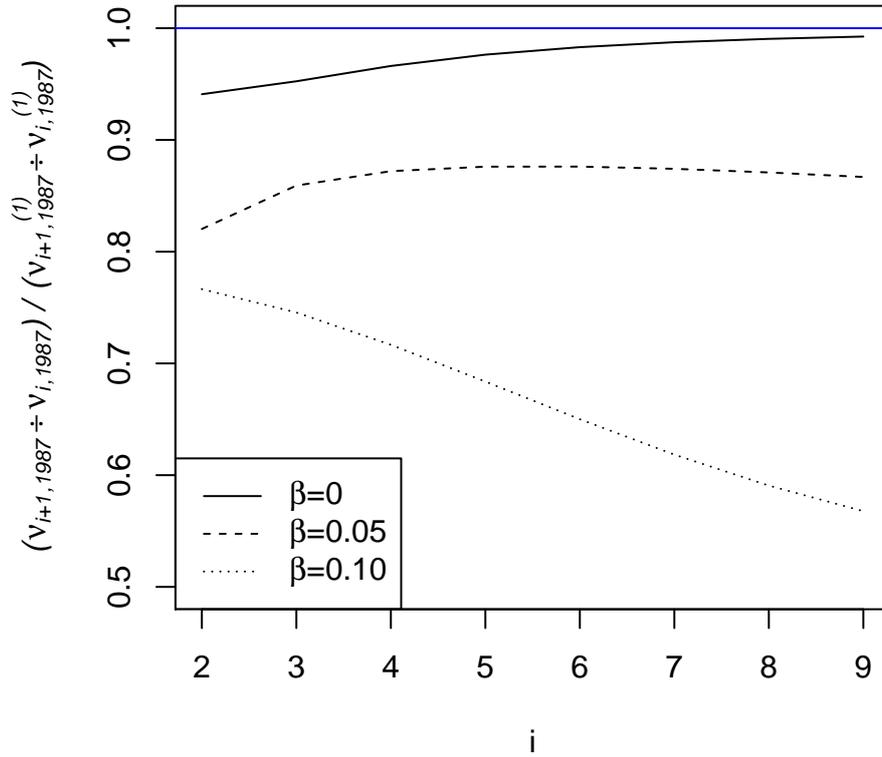


Figure 7.7: A calculation of $\frac{\nu_{i+1,1987}}{\nu_{i,1987}} \div \frac{\nu_{i+1,1987}^{(1)}}{\nu_{i,1987}^{(1)}}$ under a dependence model with $\kappa(c)$ and $f(a | c)$ specified according to equations (7.1) and (7.16), respectively. The values of the cross ratio are evaluated at the parameters $\xi_1 = 70, 248.730$, $\xi_2 = 0.515$, $\mu_A = 5.825$, $\sigma_A^2 = 1.572$. The three curves differ in their value of β .

Suppose that the Poisson intensity is written in its dependence form, $\lambda(a, c) = \kappa(c) \cdot f(a | c)$. Let cohort intensity $\kappa(c)$ take the form of the exponential model in equation (7.1) and let $f(a | c)$ take the form of the gamma density in (7.16). Under dependence, the discretised Poisson intensities $\nu_{i,k}$ and $\nu_{i,k}^{(1)}$ are defined according to equations (6.5) and (6.7), where $\mu_{i,i+k} = \nu_{i,k}^{(1)}$. The

value of β has a significant effect on the expected count ratio, $\frac{\nu_{i,k}^{(1)}}{\nu_{i,k}}$, and the expected count cross ratio, $\frac{\nu_{i+1,k}}{\nu_{i,k}} \div \frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}}$. In Figure 7.7, the expected count cross ratio for cohort 1987 is plotted for various β values, where the other parameters are set equal to the maximum likelihood estimates presented in (7.7) that were calculated for an independence model. That is, $\xi_1 = 70,248.730$, $\xi_2 = 0.515$, $\mu_A = 5.825$ and $\sigma_A^2 = 1.572$. In Appendix 9.4, we explain how the three curves were obtained. Cohort 1987 is used as an illustrative example and any cohort could have been chosen for this illustration. Note that, for cohort $c = 1987$, the mean $\mu_A(c)$ is equal to 5.825 for $\beta = 0$, 4.105 for $\beta = 0.05$ and 2.893 for $\beta = 0.10$.

For $\beta = 0$, we obtain the curve from Figure 7.4. As β increases, the mean of the age at onset changes with cohort at a faster rate. The cross ratio is very close to one for $\beta = 0$, but the values of the cross ratio diverge away from one for each increase in β . The gap between the curves for $\beta = 0$ and $\beta = 0.10$ is quite substantial. If an independence model is true, so that $\beta = 0$, a statistician would incur a small loss in accuracy of between 0.7-5.9% when carrying out a misrounded treatment of data as opposed to an exact treatment. However, if a dependence model is true with β equal to 0.5, then a statistician would incur a loss of between 23.4-43.2%. Thus, it appears that an exact treatment of data is necessary in case studies where the survival distribution is changing rapidly over time. A misrounded treatment could be sufficient in cases where the survival distribution does not change over time.

7.5 Conclusion

We assessed the suitability of an independence model $\lambda(a, c) = \kappa(c) \cdot f(a)$ for the incidence of BSE. We argued that a simplistic parameterisation of the independence model with a gamma age at onset density and an exponential

cohort intensity captured the main features of the MAFF data. The model was estimated for an exact treatment of the MAFF data and the model was discretised to obtain estimates of the quantities $\nu_{i,k}$ and $\nu_{i,k}^{(1)}$. We then investigated whether the misleading effects outlined in Theorems 6.1, 6.2 and 6.3 impacted significantly on the study of BSE incidence conducted by Dealler and Kent (1995). Theorem 6.1 led to a substantial loss in accuracy in their statistical analysis. The decreasing count ratios in the MAFF data were found not to be compatible with Theorem 6.3 and a slight dependence should be incorporated into the model.

A misrounded treatment of data was an issue for the study of BSE incidence due to a cohort intensity that was rapidly changing and a survival distribution that was sharply peaked. A misrounded treatment of the data led Dealler and Kent (1995) to find dramatic departures from the independence model. The independence model was deemed to be unsuitable for the study of BSE incidence. Anderson et al. (1996, pages 782–784) adopted an exact treatment of data for the study of BSE incidence and as a result found much smaller departures from the independence model. The smaller departures are consistent with our findings. We argue that the BSE incidence data are compatible with a survival distribution that has a slight dependence on cohort and that Dealler and Kent (1995) were misled by the effects of misrounding. Some small modifications to the independence model were necessary in the study of BSE incidence.

Note that, an age-by-cohort rounding of data is not necessarily superior to age-by-period rounded data for a statistical analysis. If the data are rounded in terms of age and period, as in the MAFF data, then the Poisson intensity should be discretised over a Lexis parallelogram. The estimate of the Poisson intensity $\hat{\lambda}(a, c)$ would be the same for the exact treatment of age-by-period data in which $n_{i,k}^{(1)}$ is assumed to have an expected value of $\nu_{i,k}^{(1)}$, and for the exact treatment of age-by-cohort data in which $n_{i,k}$ is assumed to have an expected value of $\nu_{i,k}$. The knowledge that there are two methods of deriving the same estimate is helpful because data are almost always rounded in terms

of age and period and are rarely made available as rounded age-by-cohort. Data presented as rounded age-by-cohort by the Human Mortality Database (HMD) are actually obtained as a smoothing of age-by-period data (Wilmoth et al. 2017 (accessed August 22, 2018, Section 4.2)).

A misrounded treatment of data has been adopted by Dealler and Kent and by other researchers because age-by-period counts and age-by-cohort counts are a close approximation to each other in most circumstances, and it is easier to discretise $\lambda(a, c)$ over a Lexis square than it is to discretise $\lambda(a, c)$ over a Lexis parallelogram. A misrounded treatment is sufficient for studies in which the cohort intensity is not changing or is changing very slowly. Care is needed when the cohort intensity is changing rapidly as in the case of BSE.

Chapter 8

Conclusion

In this thesis, we investigated the potential for misleading conclusions when assessing the fit of an independence model for a misrounded treatment of data, and assessed whether the need for modifications to the independence model is only apparent due to the misleading conclusions. This research is highly relevant due to the common publication of data as rounded age-by-period coupled with the common mistreatment of the data as rounded age-by-cohort. We found that the potential for misleading conclusions is significant when the cohort intensity or survival distribution is changing significantly over time.

Findings

In Chapter 3, we found that modifications to the independence model are contingent on whether we introduce effects of only cohort, only period, or both cohort and period. If the survival distribution is assumed to vary only with cohort, then there are many possible ways to handle dependence in the APC model. For instance, in continuous time, the dependence can be introduced by functionals such as PO, LLS and PH. If the survival distribution is assumed

to vary only with period or with both cohort and period, then there are few ways to handle dependence because only functionals for the hazard function are appropriate. The density function $f(a)$ should be modified according to a PH functional, while the mass function f_a should be modified according to either a CLL functional or a DL functional.

In Chapter 4, we presented two new functions coded in the statistical package R to fit the survival distribution of an APC model to data. The function `apc.cont` specifies a PH formula for $f(a | c)$, while `apc.disc` specifies a CLL formula for $f_{a|c}$. A benefit of using `apc.cont` and `apc.disc` is that they account for the entire survival experience of individuals and consider the exact time-gaps between deaths. We also explained that the way of handling dependence in the APC literature is not consistent with our findings. Data are typically rounded and published in a contingency table categorised in terms of age and period. Statisticians usually fit a discrete independence model $\lambda_{a,c} = \kappa_c \cdot f_a$ to describe the age-by-period data. Modifications are then made to the independence model using a PH functional as opposed to a CLL or DL functional. We also distinguished between survival and regression as two settings for APC modelling. The concept of misrounding is only relevant to survival models.

In Chapter 5, we provided an overview for the concept of confounding in APC models. We explained that confounding does not only relate to linear terms in age, period and cohort as there can also be a lack of identifiability for quadratic terms and cubic terms when a model contains certain interaction terms. Not all models parameterised simultaneously in terms of the three variables suffer from confounding. Caution is needed when interpreting the parameters of linear terms in models parameterised with only two variables because the coefficients of the two linear terms absorb the coefficient of the third ignored variable. Linear dependencies between columns of the model design matrix can arise due to the APC linear identity as well as due to overparameterisation. The appearance of confounding in relation to the linear identity is clear for polynomial functions, but is not so clear for factor variables.

In Chapter 6, we explained that statisticians usually take age-by-period rounded data as surrogates for age-by-cohort rounded data and fit the discrete independence model $\lambda_{i,k} = \kappa_k \cdot f_i$ to the surrogate data. The age-by-period data are used for model fitting as if the data are rounded age-by-cohort. We found that this misrounded treatment of the data can be problematic when assessing the fit of an independence model due to the potential for misleading conclusions. In certain circumstances, we found that a statistician encounters an apparent over-reporting or under-reporting in the surrogate data relative to the discrete independence model, which is not apparent in reality and is not apparent when carrying out an exact treatment of the surrogate data. Also, in certain circumstances, the expected age at death may appear to be increasing or decreasing over time under a misrounded treatment of data, when in fact longevity is not changing under the independence model. The potential for misleading conclusions can be overcome by carrying out an exact treatment of the data. However, we caution that the effect of misrounding on a statistical analysis can only be significant when the cohort intensity is changing rapidly.

In Chapter 7, we assessed the suitability of an independence model $\lambda(a, c) = \kappa(c) \cdot f(a)$ for the incidence of BSE. We found that a misrounded treatment of the data was an issue for the study of BSE incidence due to a cohort intensity that was rapidly changing and a survival distribution that was sharply peaked. A misrounded treatment of the data led Dealler and Kent (1995) to find dramatic departures from the independence model. The independence model was deemed to be unsuitable for the study of BSE incidence. Anderson et al. (1996, pages 782–784) adopted an exact treatment of data for the study of BSE incidence and as a result found much smaller departures from the independence model. A substantial loss in accuracy relating to the apparent under-reporting and over-reporting of BSE cases affected the statistical analysis of Dealler and Kent (1995). We found that the decreasing count ratios in the MAFF data were not to be compatible with an independence model and concluded that a slight dependence should be incorporated into the model.

An age-by-cohort rounding of data is not necessarily superior to age-by-

period rounded data for a statistical analysis. The estimate of the Poisson intensity $\hat{\lambda}(a, c)$ would be the same for the exact treatment of age-by-period data in which $n_{i,k}^{(1)}$ is assumed to have an expected value of $\nu_{i,k}^{(1)}$, and for the exact treatment of age-by-cohort data in which $n_{i,k}$ is assumed to have an expected value of $\nu_{i,k}$. A misrounded treatment of data has been adopted by Dealler and Kent (1995) and by other researchers because age-by-period counts and age-by-cohort counts are a close approximation to each other in most circumstances, and it is easier to discretise $\lambda(a, c)$ over a Lexis square than it is to discretise $\lambda(a, c)$ over a Lexis parallelogram. A misrounded treatment is sufficient for studies in which the cohort intensity is not changing or is changing very slowly.

The contents of this thesis can be used by other statisticians. This thesis should allow a statistician to better appreciate different time concepts, in particular, the difference between models specified in discrete time and rounded time. Statisticians could implement an exact treatment of age-by-period data to avoid the misleading effects that can be apparent for a misrounded treatment of data. The extent to which a statistician should be worried about the effects of misrounding depends on the circumstances. If the number of births or the expected age-at-death for an individual is slowly changing over time, then it is sufficient to adopt a misrounded treatment. However, if the number of births or the expected age-at-death is rapidly changing over time, then a statistician should adopt an exact treatment. For any circumstance, a good practice for statisticians would be to discretise APC models correctly in order to describe exactly the rounding of the data.

The contents of this thesis could eventually form a textbook on APC modelling. The existing textbooks on APC modelling aim to demonstrate and critique the innovative approaches used to overcome the confounding problem (Fienberg and Mason 1985; Glenn 2005; Yang and Land 2013; O'Brien 2015). These approaches involve manipulating the model specification, the method of parameter estimation or even the rounding of the data in order to tease out the simultaneous effects of age, period and cohort on a response variable.

The textbooks are not so suitable as an introduction to APC modelling. Our exploration of the misrounding concept provides a fresh perspective to the existing literature.

Future research

In this thesis, we explained that statisticians usually fit a discrete model $\lambda_{i,k} = \kappa_k \cdot f_{i|k}$ to data rounded age-by-period. When the survival distribution is independent of period and cohort so that $f_{i|k} = f_i$, the discrete model is equivalent to a continuous model discretised over an age-by-cohort region in the Lexis diagram so that $\lambda_{i,k} = \nu_{i,k}$. This meant that the age-by-period data were used for model fitting as if they were rounded age-by-cohort, and we called this a misrounded treatment of the data.

However, when the survival distribution varies with period and/or cohort, the discrete model is not equivalent to a continuous model discretised in terms of age and cohort so that $\lambda_{i,k} \neq \nu_{i,k}$. The discrete model can perhaps be written in terms of the continuous model $\lambda(a, c) = \kappa(c) \cdot f(a | c)$ as a product of $\kappa_k = \int_k^{k+1} \kappa(c) dc$ and $f_{i|k} = \int_i^{i+1} \int_k^{k+1} f(a | c) dc da$. Under dependence, the discrete model does not make sense as a discretisation of the continuous model over a region in the Lexis diagram. The concept of a misrounded treatment becomes a concept of a misdiscretised treatment of the age-by-period data. Hence, we can investigate the potential for misleading conclusions when fitting APC models for a misdiscretised treatment of the data. This would involve considering the expected count ratio $\frac{\nu_{i,k}^{(s)}}{\lambda_{i,k}}$ as opposed to $\frac{\nu_{i,k}^{(s)}}{\nu_{i,k}}$ in order to assess the loss in accuracy of carrying out a misdiscretised treatment rather than an exact treatment of the age-by-period data.

Under independence, the relative rate of change in the cohort intensity is a primary driver of disparities between expected counts, count ratios and count

cross-ratios. There may not be strict inequalities between counts, count ratios and count cross-ratios for a dependence assumption. However, if the speed of change in the survival distribution with cohort is a primary driver of disparities, then these disparities could potentially be larger relative to the disparities for an independence assumption. Figure 7.7 suggested that a rapidly changing survival distribution would lead to large disparities from one. Further, given a misdiscretised treatment of the data, the change in the apparent survival distribution over time could be noticeably different to the change in the true survival distribution over time.

One particularly interesting case study is human mortality. The lengthening of human lives over time has been reported by academic papers (Kermack et al. 1934; Lee and Hsieh 1996) and by various media outlets. We can assess the extent to which the apparent lengthening of human lives is attributable to misrounding or misdiscretisation. I suspect that misrounding or misdiscretisation effects would not be significant for human mortality because the number of births changes slowly over time. It may then be unnecessary to study human mortality for an exact treatment of the data. It would be interesting to consider how a slowly changing cohort intensity coupled with a changing survival distribution would impact on disparities for expected counts, count ratios and count cross-ratios.

The case study of mortality due to Pleural Malignant Mesothelioma (PMM) may be interesting because PMM has a cohort intensity that is rapidly increasing and then rapidly decreasing. The shape of the cohort intensity for BSE incidence is closely related to the volume of infected cattle feed, whereas the shape of the cohort intensity for PMM mortality is closely related to the volume of asbestos (Murayama et al. 2006). PMM has been studied for European countries by Peto et al. (1999) and for Japan by Murayama et al. (2006). Peto et al. and Murayama et al. both report that the risk of PMM mortality is increasing for newer cohorts. We can explore whether this apparent decrease in life expectancy over time is compatible with an independence model.

We fitted a simplistic model for BSE incidence that consisted of four parameters. The Pearson residuals of this model were quite large, but the model allowed us to illustrate the fundamental flaws of studying an independence model for a misrounded treatment of data. It would be interesting to fit more sophisticated models with the necessary number of parameters to produce small Pearson residuals, and, hence, to replicate the effects of misrounding or misdiscretisation encountered in various case studies including BSE, PMM and human mortality. These models would allow for a survival distribution dependent on cohort if an independence assumption is not suitable.

Chapter 9

Appendix

This chapter contains a glossary of key terms used throughout the thesis and contains code used for fitting APC models and independence models.

9.1 Glossary

It is helpful to think of cohort, period and age as the time-of-birth, current time and the time elapsed since birth for an individual, respectively. We provide a more general definition below. Note that, by time, we refer to calendar time.

Cohort - The random time of birth for an individual, $C = c$.

Period - The time at which an individual is observed after birth, p . This is the time scale on which the period-at-death is defined.

Age - The time-gap between cohort and period, $a = p - c$, that is, period minus cohort. This is the time scale on the age-at-death is defined.

Linear identity - The equation $a = p - c$ is the linear identity.

Cohort intensity - It is helpful to think of the cohort intensity as the number or rate of individuals born at time c .

Survival distribution - The probability distribution for the age-at-exit for an individual, possibly conditional on covariates. This can be expressed in continuous time, for example, as a probability density function or a hazard function.

Poisson intensity - The expected number or intensity of deaths at a particular point in the age-cohort space.

Independence model - A special case of the Poisson intensity in which the survival distribution does not vary with covariates, particularly period and cohort.

APC model - In a survival context, the model considers how the age-at-death distribution varies with period and/or cohort. In this thesis, APC model is synonymous with the Poisson intensity. In a regression context, the model considers how the distribution of a non-survival response variable varies with age, period and cohort.

Base model - The survival distribution for a reference individual or reference group of individuals that share a particular set of covariate values. The base model can be viewed as the independence model before modifications are made to incorporate the effects of period and cohort.

Covariate model - A covariate model takes the base model and describes the survival distribution for any individual by a departure from the base model through $\psi = \mathbf{x}(a)^T \boldsymbol{\beta}$, where \mathbf{x} is a vector of covariates at age a . The base model corresponds to $\psi(a) = 0$ for all a . If the covariates are period and cohort, then the covariate model is an APC model.

Functional - A function of a function. In this thesis, functionals are used

to introduce dependence to the independence model in order to formulate a covariate model.

Confounding - In this thesis, confounding is synonymous with a lack of identifiability.

Surrogate - Taking age-by-period data and recategorising the data in terms in age and cohort means that the age-by-period data are used as surrogates for age-by-cohort data.

Age-by-period data - A two-way contingency table categorised in terms of rounded age and rounded period. Each cell of the table counts the number of deaths for a particular age and period. An age-by-period region in the Lexis diagram is a parallelogram.

Age-by-cohort data - A two-way contingency table categorised in terms of rounded age and rounded cohort. Each cell of the table counts the number of deaths for a particular age and cohort. An age-by-cohort region in the Lexis diagram is a square.

Misrounded treatment / Misrounding - This concept requires some data, a model and a treatment of data in the model. Misrounding means that data rounded age-by-period are used for model fitting as if the data are rounded age-by-cohort.

Exact treatment - Data rounded age-by-period are used for model fitting as if the data are rounded age-by-period.

9.2 Estimation of the age-period-cohort model

In Sections 4.1 and 4.2, we established a method to fit APC models to data by maximising a likelihood function. The functions `apc.cont` and `apc.disc`

written in the statistical package R were key components of the method since they calculate the log likelihood function for a particular choice of α and β . In this section, we give full details of the two functions.

Continuous time

The following function calculates the value of the log likelihood function (4.2) for a particular choice of (α^T, β^T) , where $f(a_j | c_j)$ is defined according to the Riemann Sum in (4.6):

```
apc.cont <- function(param) {

##part one
alpha <- param[1:Nalpha] / scalePar
beta <- param[(Nalpha+1):(Nalpha+Nbeta)] / scalePar

##part two
Psi <- matrix(0,n,m)
for (ell in (1:m)) {
X <- Xmake( a*(ell/m), c )
Psi[,ell] <- X %%% beta
}

##part three
H <- matrix(0,n,m)
for (ell in (1:m)) {
H[,ell] <- exp(Psi[,ell])* h0(a*(ell/m),alpha)
}

##part four
S <- exp( - H %%% rep(1/m,m) )
f <- S*H[,m]
```

```
sum( log(f) )
}
```

Before running `apc.cont`, it is necessary for the user to specify `n`, `a`, `c`, `h0`, `Nalpha`, `Xmake`, `Nbeta`, `m` and `scalePar`. The argument `param` is the full parameter vector

$$(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T) = (\alpha_1, \alpha_2, \dots, \alpha_{n_\alpha}, \beta_1, \beta_2, \dots, \beta_{n_\beta}), \quad (9.1)$$

where n_α and n_β denote the size of vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Notice that, the last line of code in `apc.cont` is `sum(log(f))` which is the log likelihood function.

We have split the R function `apc.cont` into four parts. Part one defines the argument `param` as the vector (9.1) multiplied by a constant `scalePar`. For this example, if `scalePar <- 100`, then the vector `param <- c(1,1,2,2)` implies that `alpha <- c(0.01,0.01)` and `beta <- c(0.02,0.02)`. The maximum likelihood estimates of the parameters that are returned when applying `optim` to `apc.cont` are divided by `scalePar` to produce the unscaled parameters. We do not use scaling in this example, so we set `scalePar <- 1`.

Part two produces a matrix of linear predictors, `Psi`. First, a covariate matrix for the ℓ th intermediate time is constructed through `X`. Specifically, the matrix `X` is code for

$$\mathbf{X}_\ell = \begin{pmatrix} \mathbf{x} (a_1 \cdot \frac{\ell}{m})^T \\ \mathbf{x} (a_2 \cdot \frac{\ell}{m})^T \\ \vdots \\ \mathbf{x} (a_n \cdot \frac{\ell}{m})^T \end{pmatrix} = \begin{pmatrix} c_1 & (a_1 \cdot \frac{\ell}{m}) + c_1 \\ c_2 & (a_2 \cdot \frac{\ell}{m}) + c_2 \\ \vdots & \vdots \\ c_n & (a_n \cdot \frac{\ell}{m}) + c_n \end{pmatrix},$$

where $\mathbf{X}_m = \mathbf{X}$. The rows of `Psi` correspond to different individuals labelled by $j = 1, 2, \dots, n$ and the columns consider the value of the linear predictor at some intermediate times before and at death for each individual. The intermediate times are labelled by $\ell = 1, 2, \dots, m$ and the final column of `Psi` is

the observed linear predictor at death. The matrix `Psi` is code for

$$\mathbf{\Psi} = \begin{pmatrix} \psi(a_1 \cdot \frac{1}{m}) & \psi(a_1 \cdot \frac{2}{m}) & \cdots & \psi(a_1) \\ \psi(a_2 \cdot \frac{1}{m}) & \psi(a_2 \cdot \frac{2}{m}) & \cdots & \psi(a_2) \\ \vdots & \vdots & \vdots & \vdots \\ \psi(a_n \cdot \frac{1}{m}) & \psi(a_n \cdot \frac{2}{m}) & \cdots & \psi(a_n) \end{pmatrix}.$$

The ℓ th column of $\mathbf{\Psi}$ is equal to $\mathbf{X}_\ell \boldsymbol{\beta}$.

Part three produces a matrix of hazard functions, \mathbf{H} . Recalling (4.12), we can consider a vector of baseline hazard functions at an ℓ th intermediate time between birth and death, so that

$$\mathbf{h}_{0,\ell}^T = \left(h_0 \left(a_1 \cdot \frac{\ell}{m} \right), h_0 \left(a_2 \cdot \frac{\ell}{m} \right), \dots, h_0 \left(a_n \cdot \frac{\ell}{m} \right) \right). \quad (9.2)$$

The ℓ th column of \mathbf{H} is a product of the ℓ th column of $\mathbf{\Psi}$ and of $\mathbf{h}_{0,\ell}$. The element in the j th row and ℓ th column of \mathbf{H} is $h(a_j \cdot \frac{\ell}{m} | c_j) = h_0(a_j \cdot \frac{\ell}{m}) \cdot \psi(a_j \cdot \frac{\ell}{m})$. In full, \mathbf{H} is code for the hazard function for each individual considered at a series of times before and at death:

$$\mathbf{H} = \begin{pmatrix} h(a_1 \cdot \frac{1}{m} | c_1) & h(a_1 \cdot \frac{2}{m} | c_1) & \cdots & h(a_1 | c_1) \\ h(a_2 \cdot \frac{1}{m} | c_2) & h(a_2 \cdot \frac{2}{m} | c_2) & \cdots & h(a_2 | c_2) \\ \vdots & \vdots & \vdots & \vdots \\ h(a_n \cdot \frac{1}{m} | c_n) & h(a_n \cdot \frac{2}{m} | c_n) & \cdots & h(a_n | c_n) \end{pmatrix}.$$

Part four calculates the Riemann Sum approximation to the log likelihood function, `sum(log(f))`. First, a vector of survivor functions written in approximate form as Riemann Sums is calculated for each individual at their age of death. The vector denoted in code by `S` can be written as

$$\mathbf{S} = \begin{pmatrix} \exp \left(-\frac{1}{m} \cdot \sum_{\ell=1}^m h \left(a_1 \cdot \frac{\ell}{m} \mid c_1 \right) \right) \\ \exp \left(-\frac{1}{m} \cdot \sum_{\ell=1}^m h \left(a_2 \cdot \frac{\ell}{m} \mid c_2 \right) \right) \\ \vdots \\ \exp \left(-\frac{1}{m} \cdot \sum_{\ell=1}^m h \left(a_n \cdot \frac{\ell}{m} \mid c_n \right) \right) \end{pmatrix}.$$

The vector \mathbf{S} is calculated by taking the product of \mathbf{H} and a vector of length m with each element being $\frac{1}{m}$, and then by taking the exponential of minus the calculated product. A vector of probability density functions at death, denoted by \mathbf{f} , is then calculated as a product of \mathbf{S} and the final column of matrix \mathbf{H} . The vector of probability densities is

$$\mathbf{f} = \begin{pmatrix} f(a_1 | c_1) \\ f(a_2 | c_2) \\ \vdots \\ f(a_n | c_n) \end{pmatrix},$$

where $f(a_j | c_j)$ is defined in (4.6). The log likelihood function defined in (4.2) is calculated as `sum(log(f))`.

Discrete time

The following R function calculates the value of the log likelihood function (4.11) for a particular choice of $(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)$, where $f_{a_j|c_j}$ is defined according to equations (4.13) and (4.14):

```
apc.disc <- function(param) {
  ##Part one
  alpha <- param[1:Nalpha] / scalePar
  beta <- param[(Nalpha+1):(Nalpha+Nbeta)] / scalePar

  for (j in 1:n) {
    ##Part two
    X <- Xmake( 0:a[j], c[j] )
    Psi <- X %*% beta
    H <- 1 - ( (1-h0(0:a[j],alpha))^(exp(Psi)) )
  }
}
```

```

##Part three
if (a[j]==0) { f[j] <- H }
else { S <- prod( (1-H)[1:a[j]] );
f[j] <- S*H[a[j]+1] }

}

##Part four
sum ( log(f) )

}.

```

The argument `param` has the same meaning as in equation (9.1). Before running `apc.disc`, it is necessary for the user to specify `n`, `a`, `c`, `h0`, `Nalpha`, `Xmake`, `Nbeta` and `scalePar`. The variable `scalePar` was defined in Section 4.1 as a scaling variable for α and β and we choose to set `scalePar` to one in this example.

We have split the R function `apc.disc` into four parts. Part one defines the argument `param`. Part two first defines a covariate matrix for the j th individual through `X`. This j th covariate matrix, which we will denote as \mathbf{X}_j , can be written as

$$\mathbf{X}_j = \begin{pmatrix} \mathbf{x}_0^T \\ \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_{a_j}^T \end{pmatrix} = \begin{pmatrix} c_j & c_j & (c_j)^2 \\ c_j & c_j + 1 & (c_j + 1)^2 \\ \vdots & \vdots & \vdots \\ c_j & c_j + a_j & (c_j + a_j)^2 \end{pmatrix}.$$

Each row of \mathbf{X}_j corresponds to the covariates for the j th individual at a particular age and the rows are ordered by ascending age. The first row corresponds to age zero and the final row corresponds to the age at death. So the number of rows in \mathbf{X}_j will vary between individuals. The following vector of linear

predictors for the j th individual is constructed through `Psi`:

$$\boldsymbol{\psi}_j = \mathbf{X}_j \boldsymbol{\beta} = \begin{pmatrix} \mathbf{x}_0^T \boldsymbol{\beta} \\ \mathbf{x}_1^T \boldsymbol{\beta} \\ \vdots \\ \mathbf{x}_{a_j}^T \boldsymbol{\beta} \end{pmatrix}.$$

Also, a vector of hazard functions for the j th individual is constructed through `H`, which can be written as

$$\mathbf{h}_j = \begin{pmatrix} h_{0|c_j} \\ h_{1|c_j} \\ \vdots \\ h_{a_j|c_j} \end{pmatrix} = \begin{pmatrix} 1 - (1 - h_{0,0})^{\exp(\mathbf{x}_0^T \boldsymbol{\beta})} \\ 1 - (1 - h_{1,0})^{\exp(\mathbf{x}_1^T \boldsymbol{\beta})} \\ \vdots \\ 1 - (1 - h_{a_j,0})^{\exp(\mathbf{x}_{a_j}^T \boldsymbol{\beta})} \end{pmatrix}.$$

In part three of `apc.disc`, the probability mass function $f_{a_j|c_j}$ is defined for the j th individual. In Section 3.1, we stated that, with the exception of age zero, the probability mass function is defined according to the product formula (3.6). At age zero, the probability mass function is defined as equal to the hazard function. We use an `if` command to distinguish between these two definitions, so that if $a_j = 0$, the probability mass function is defined in correspondence with equation (4.15) as equal to $h_{0|c_j} = 1 - (1 - h_{0,0})^{\exp(\mathbf{x}_0^T \boldsymbol{\beta})}$.

For ages $a = 1, 2, \dots$, the code `S` calculates the survivor function for the j th individual, which is defined as $S_{a_j-1|c_j} = \prod_{u \leq a_j-1} (1 - h_{u|c_j})$ in the product formula of equation (4.14). Notice that, the length of the vector `H` is equal to `a[j]+1`. The function $f_{a_j|c_j}$, written in code as `f[j]`, is then obtained by multiplying $S_{a_j-1|c_j}$ with the final element of the vector `h_j`.

The `for` command carries out the calculation of $f_{a_j|c_j}$ in parts two and three for all n individuals to produce a vector $\mathbf{f}^T = (f_{a_1|c_1}, f_{a_2|c_2}, \dots, f_{a_n|c_n})$. Part four takes the logarithm and sum of `f` to calculate the log likelihood function described in (4.11).

9.3 Estimation of the independence model

In Section 7.2, we estimate the independence model for the exact treatment of data in the case of BSE incidence. The purpose of this section is to provide details on how the exact treatment was carried out to produce the parameters estimates presented in equation (7.7). We also show how the parameter estimates were used to calculate expected count ratios and count cross-ratios. The independence model consists of the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\xi}$ since $f(a)$ consists of parameters $\boldsymbol{\alpha}^T = (\alpha_1, \dots, \alpha_m)$ and $\kappa(c)$ consists of parameters $\boldsymbol{\xi}^T = (\xi_1, \dots, \xi_r)$. Under a Poisson model, the likelihood function for the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\xi}$, given data on deaths provided in a contingency table categorised in terms of age and period, is

$$L_{\text{EI}}(\boldsymbol{\alpha}, \boldsymbol{\xi} \mid \mathbf{N}^{(1)} = \mathbf{n}^{(1)}) = \prod_{i=i^-+1}^{i^+} \prod_{k=k^-+1}^{k^+} \frac{\left(\nu_{i,k}^{(1)}\right)^{n_{i,k}^{(1)}} \exp\left(-\nu_{i,k}^{(1)}\right)}{n_{i,k}^{(1)}!},$$

where

$$\begin{aligned} \nu_{i,k}^{(1)} &= \mu_{i,i+k} = \iint_{R_{i,i+k}^{(\text{AP})}} \lambda(a, c) \, dc \, da \\ &= \int_i^{i+1} \int_{i+k-a}^{i+k+1-a} \lambda(a, c) \, dc \, da. \end{aligned}$$

Model estimation

The random number of deaths in period j at age i is denoted as $M_{i,j}$. An age-by-period contingency table of observed counts can be written as $\mathbf{m} = (m_{i,j})$. By applying the first surrogate convention, so that $m_{i,i+k} = n_{i,k}^{(1)}$, the age-by-period table is presented in terms of age and cohort. The resulting table can be written in matrix form as $\mathbf{n}^{(1)} = \left(n_{i,k}^{(1)}\right)$. An example is presented in Table 7.1 for the case of BSE incidence. The code `Data` is a data frame for $\mathbf{n}^{(1)}$.

```

Data <- matrix(c(
rep(NA,7),17,18,
rep(NA,6),62,43,40,
rep(NA,5),198,123,83,50,
rep(NA,4),879,521,225,172,83,
rep(NA,3),2275,1918,950,440,244,rep(NA,1),
rep(NA,2),2557,4065,2561,1268,632,rep(NA,2),
rep(NA,1),781,4399,5741,4073,1983,rep(NA,3),
49,1744,8847,10907,7865,rep(NA,4),
73,4227,16039,17637,rep(NA,5),
85,2015,7497,rep(NA,6),
40,1208,rep(NA,7),
23,rep(NA,8)
),
nrow=12, ncol=9, byrow=TRUE)

Data <- as.data.frame(Data).

```

It is helpful but not essential to label the rows and columns of the data. The columns correspond to ages $i = 2, 3, \dots, 10$, while the rows correspond to cohorts $k = 1980, 1981, \dots, 1991$.

```

colnames(Data) <- 2:10
rownames(Data) <- 1980:1991.

```

In Chapter 7, the following Gamma model is chosen to describe $f(a)$ for the case of BSE incidence:

$$f(a) = \frac{\alpha_1^{\alpha_2} a^{\alpha_2-1} \exp(-\alpha_1 a)}{\Gamma(\alpha_2)} \quad \text{for } a > 0, \quad (9.3)$$

where $\alpha_1 > 0$ and $\alpha_2 > 0$. We reparameterise the Gamma model in terms of the mean and variance. Letting α_1^* and α_2^* denote the mean and variance of the age-at-onset, the rate parameter α_1 (`rateGam`) and the shape parameter

α_2 (`shapeGam`) are determined as

$$\alpha_1 = \frac{\alpha_1^*}{\alpha_2^*} \quad \text{and} \quad \alpha_2 = \frac{(\alpha_1^*)^2}{\alpha_2^*}. \quad (9.4)$$

Also, `alpha` is a vector of length two in which `alpha[1]` is α_1^* and `alpha[2]` is α_2^* . In R, we write a function `f` which takes age a and a parameter vector $\boldsymbol{\alpha}^T = (\alpha_1^*, \alpha_2^*)$ as inputs and calculates the probability density function described in (9.3):

```
shapeGam <- function(alpha) { (alpha[1]^2)/ alpha[2] }
rateGam <- function(alpha) { alpha[1]/alpha[2] }
f <- function(a,alpha) { dgamma(a, shape=shapeGam(alpha),
rate=rateGam(alpha)) }.
```

In Chapter 7, we choose the following symmetric exponential model to describe $\kappa(c)$ for the case of BSE incidence:

$$\kappa(c) = \xi_1 \exp(-\xi_2 |c - 1988.5|) \quad \text{for } c > 1979, \quad (9.5)$$

where $\xi_1 > 0$ and $\xi_2 > 0$. In R, we write a function `q` which takes cohort c and a parameter vector $\boldsymbol{\xi}^T = (\xi_1, \xi_2)$ as inputs and calculates the cohort intensity described in (9.5):

```
q <- function(c,xi) {100000*xi[1]* exp(-xi[2]* abs(c-1988.5))} .
```

The parameter ξ_2 corresponds to `xi[2]` and parameter ξ_1 corresponds to `100000*xi[1]`. The parameter `xi[1]` is multiplied by 100,000 because the optimisation procedure could be time consuming if the optimal value of `xi[1]` is very large.

The independence model can then be defined as a product of $f(a)$ and $\kappa(c)$. In R, we write a function `lambda` which takes a , c , $\boldsymbol{\alpha}$ and $\boldsymbol{\xi}$ as inputs and produces the independence model:

```
lambda <- function(a,c,alpha,xi) { f(a,alpha) * q(c,xi) }.
```

The Poisson intensity for a parallelogram region in the Lexis diagram,

$\nu_{i,k}^{(1)}$, can be written as

$$\int_{a=0}^{a=1} \int_{c=0}^{c=1} \lambda(i+a, k+c-a) dc da.$$

Since $\nu_{i,k}^{(1)}$ is difficult to calculate by numerical integration, we express the Poisson intensity as the following Riemann Sum from (7.6):

$$\frac{1}{n^2} \sum_{u=1}^n \sum_{v=1}^n \lambda\left(i + \frac{u}{n}, k + \frac{v}{n} - \frac{u}{n}\right).$$

Before we calculate the Riemann Sum, it is necessary to specify the value of Riemann Sum divisor, n , as well as the number of rows and columns in the data. It also helpful to state the lower limits for age and cohort in the age-by-period table described by `Data`.

```
c0 <- as.numeric( rownames(Data)[1] ) -1 ##lower limit for cohort
a0 <- as.numeric( colnames(Data)[1] ) ##lower limit for age
nAge <- ncol(Data) ##number of columns
nCoh <- nrow(Data) ##number of rows
n <- 30.
```

The larger n is chosen to be, the more accurate will be our approximation to $\nu_{i,k}^{(1)}$. Note that, $\nu_{i,k}^{(1)}$ is the expected count in a Lexis region $R_{i,i+k}^{(AP)}$ and this region depicted in Figure 6.2 extends over ages $[i, i+1)$ and cohorts $[k-1, k+1)$. In R, we write a function `parallelogram` which takes parameter vectors α and ξ as inputs and returns a matrix with elements corresponding to the Riemann Sum approximation to $\nu_{i,k}^{(1)}$ for ages $i = 2, 3, \dots, 10$ and cohorts $k = 1980, 1981, \dots, 1991$:

```
u <- c(); v <- c()
parallelogram <- function(alpha,xi) {
pA <- matrix(nrow=nCoh, ncol=1)
A <- matrix(nrow=nCoh, ncol=nAge)
for (i in 1:nAge) {
for (k in 1:nCoh) {
for (delta2 in 1:n) {
```

```

for (delta1 in 1:n) {
u[delta1] <- lambda( a0 +i-1 +(delta2/n), c0+ k + (delta1-delta2)/n,
alpha,xi ) *(1/(n^2))
}
v[delta2] <- sum(u)
}
pA[k,] <- sum(v)
}
A[,i] <- pA
}
A
}.

```

The code `delta2` is used for variable u and `delta1` is used for variable v . Also, i is replaced with $a0+i$ and k is replaced with $c0+k$, where $a0$ is the lower bound for age in the dataset and $c0$ is the lower bound for cohort.

We next carry out an optimisation procedure to find the values of the parameter vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\xi}$ which maximise the likelihood function described at the start of this section $L_{E1}(\boldsymbol{\alpha}, \boldsymbol{\xi} \mid \mathbf{N}^{(1)} = \mathbf{n}^{(1)})$. The logarithm of the likelihood function is

$$\sum_{i=i^-+1}^{i^+} \sum_{k=k^-+1}^{k^+} \left[n_{i,k}^{(1)} \left(\log \nu_{i,k}^{(1)} \right) - \nu_{i,k}^{(1)} \right] + \text{constant}.$$

For the case of BSE incidence, $i^- = 1$, $i^+ = 10$, $k^- = 1979$ and $k^+ = 1991$. In R, we write a function `MLE.stat` that takes `parallelogram` and `Data` as inputs and returns a matrix with each element corresponding to the contribution to the log likelihood function for cell (i, k) , $\left[n_{i,k}^{(1)} \left(\log \nu_{i,k}^{(1)} \right) - \nu_{i,k}^{(1)} \right]$. We also write a function `MLEvalueOP` that takes the full parameter vector $(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T) = (\alpha_1, \alpha_2, \xi_1, \xi_2)$ as an input and returns the summation of `MLE.stat` over all cells, that is, it returns the value of the log likelihood function. Specifically, we write

```
nAlpha <- 2; nXi <- 2
```

```

MLEvalueOP <- function(param) {
alpha <- param[1:nAlpha]
xi <- param[(nAlpha+1):(nAlpha+nXi)]
MLE.stat <- (Data*log(parallelogram(alpha,xi)) -
parallelogram(alpha,xi)
ifelse( any(is.na( log(parallelogram(alpha,xi)) ) ),
NA, sum(MLE.stat, na.rm=TRUE) )
}.

```

It is necessary to specify the length of $\boldsymbol{\alpha}$ and $\boldsymbol{\xi}$, denoted in code by `nAlpha` and `nXi`, in order to define the full parameter vector `param`.

We then use the existing R function `optim` to find the vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\xi}$ that maximise the likelihood function $L_{E1}(\boldsymbol{\alpha}, \boldsymbol{\xi} \mid \mathbf{N}^{(1)} = \mathbf{n}^{(1)})$. In R, this means that we find the value of `param` that maximises `MLEvalueOP`:

```

optim( c(1,1,1,1), MLEvalueOP, control=list(fnscale=-1) ).

```

We choose to start the optimisation procedure at an initial vector $(\boldsymbol{\alpha}^T, \boldsymbol{\xi}^T) = (1, 1, 1, 1)$. Note that, the command `fnscale=-1` states that we want to maximise `MLEvalueOP` as opposed to minimise. The output of `optim` was the parameter estimates $\hat{\alpha}_1^* = 5.825$, $\hat{\alpha}_2^* = 1.572$, $\hat{\xi}_1 = 0.702$ and $\hat{\xi}_2 = 0.515$. These estimates are presented in equation (7.7).

Using the fitted model

The maximum likelihood estimates of the model parameters, which are presented in equation (7.7), can be written in R as

```

alphaM <- c(5.825,1.572); xiM <- c(0.702,0.515).

```

The Poisson intensity for a square region in the Lexis diagram, $\nu_{i,k}$, which was

first introduced in equation (6.5), can be written as

$$\int_{a=0}^{a=1} \int_{c=0}^{c=1} \lambda(i+a, k+c) dc da.$$

We express the Poisson intensity as the following Riemann Sum:

$$\frac{1}{n^2} \sum_{u=1}^n \sum_{v=1}^n \lambda\left(i + \frac{u}{n}, k + \frac{v}{n}\right). \quad (9.6)$$

In R, we write a function `square` which takes parameter vectors α and β as inputs and returns a matrix with each element corresponding to the Riemann Sum approximation of $\nu_{i,k}$ for ages $i = 2, 3, \dots, 10$ and cohorts $k = 1980, 1981, \dots, 1991$:

```
u2 <- c(); v2 <- c()
square <- function(alpha,xi) {
pB <- matrix(nrow=nCoh, ncol=1); B <- matrix(nrow=nCoh, ncol=nAge)
for (i in 1:nAge) {
for (k in 1:nCoh) {
for (delta2 in 1:n) {
for (delta1 in 1:n) {
u2[delta1] <- lambda( a0 +i-1 +(delta2/n), c0+ k + (delta1/n),
alpha,xi ) *(1/(n^2))
}
v2[delta2] <- sum(u2)
}
pB[k,] <- sum(v2)
}
B[,i] <- pB
}
B
}.

```

In R, we obtain the maximum likelihood estimates of $\nu_{i,k}^{(1)}$ and $\nu_{i,k}$ by substituting `alphaM` and `xiM` into the functions `parallelogram` and `square`:

```
parallM <- parallelogram(alphaM, xiM)
```

```

squareM <- square(alphaM, xiM)
round(parallM,0) ##presented in Table 7.3
round(squareM,0) ##presented in Table 7.5.

```

The ratio $\nu_{i,k}^{(1)} \div \nu_{i,k}$ is used to calculate the percentage loss in accuracy in carrying out a misrounded treatment of data as opposed to an exact treatment. In R, we calculate $\nu_{i,k}^{(1)} \div \nu_{i,k}$ by taking the ratio of `parallM` and `squareM`:

```

round(parallM/squareM,2) ##plotted in Figure 7.3.

```

We simulate values of $N_{i,k}^{(1)}$ and $N_{i,k}$ from a Poisson model with parameters $\hat{\nu}_{i,k}^{(1)}$ and $\hat{\nu}_{i,k}$. The quantities $\hat{\nu}_{i,k}^{(1)}$ and $\hat{\nu}_{i,k}$ are our maximum likelihood estimates for the expected number of deaths in age-by-period and age-by-cohort regions of the Lexis diagram, denoted by $\nu_{i,k}^{(1)}$ and $\nu_{i,k}$. The simulations $n_{i,k}^{(1)}$ and $n_{i,k}$ can be viewed as incorporating noise into $\hat{\nu}_{i,k}^{(1)}$ and $\hat{\nu}_{i,k}$. Our aim is to assess whether the patterns in $\hat{\nu}_{i,k}^{(1)}$ and $\hat{\nu}_{i,k}$ are apparent for noisy data. To derive the simulations, we type the following code into R:

```

p.simA <- matrix(nrow=nCoh, ncol=1)
simA <- matrix(nrow=nCoh, ncol=nAge)
for (i in 1:nAge) {
  for (k in 1:nCoh) {
    p.simA[k,] <- rpois(1,parallM[k,i])
  }
  simA[,i] <- p.simA
}
simA

```

```

p.simB <- matrix(nrow=nCoh, ncol=1)
simB <- matrix(nrow=nCoh, ncol=nAge)
for (i in 1:nAge) {
  for (k in 1:nCoh) {
    p.simB[k,] <- rpois(1,squareM[k,i])
  }
}
simB

```

```

}
simB[,i] <- p.simB
}
simB.

```

The function `rpois(1, parallM[k,i])` simulates one value from a Poisson distribution with parameter $\hat{\nu}_{i,k}^{(1)}$, while `rpois(1, squareM[k,i])` simulates one value from a Poisson distribution with parameter $\hat{\nu}_{i,k}$. The code `simA` and `simB` are matrices with cell (i, k) corresponding to the observations $n_{i,k}^{(1)}$ and $n_{i,k}$, respectively.

The ratio of simulated counts, $n_{i,k}^{(1)} \div n_{i,k}$, can be calculated in R for all (i, k) cells as

```
round(simA/simB,2) ##plotted in Figure 7.5.
```

This ratio can be compared with `round(parallM/squareM,2)` to assess whether the expected pattern in $N_{i,k}^{(1)} \div N_{i,k}$ is also apparent for the simulated data.

9.4 Modifying the independence model

In Section 7.4, we modified our independence model for BSE incidence to allow the mean of the age at onset to vary with cohort. The Poisson intensity of the APC model is $\lambda(a, c) = \kappa(c) \cdot f(a | c)$, where the age at onset density $f(a | c)$ is defined according to the gamma density in (7.16) and the cohort intensity is defined in (7.1). The cohort intensity is the same as in the independence model from Appendix 9.3 and is written in R as

```

q <- function(c,xi) {
100000*xi[1]* exp(-xi[2]* abs(c-1988.5) )
}.

```

In our code, `alpha[1]` is the mean and `alpha[2]` is the variance of the age at onset. The rate and shape parameters $\alpha_1(c)$ and $\alpha_2(c)$ defined in equations (7.17) and (7.18) can be written as

```
rateGam <- function(alpha,beta,c) {
(alpha[1]* exp(-beta[1]*(c-1980))) / alpha[2]
}

shapeGam <- function(alpha,beta,c) {
( (alpha[1]* exp(-beta[1]*(c-1980)))^2 ) / alpha[2]
}.
```

The functions `rateGam` and `shapeGam` take `alpha`, `beta` and `c` as inputs and produce values of $\alpha_1(c)$ and $\alpha_2(c)$. To introduce dependence, we multiply the mean `alpha[1]` by a function of cohort, `exp(-beta[1]*(c-1980))`, where `beta[1]` is a parameter for the cohort effect on μ_A . Since `exp(-beta[1]*(c-1980))` is equal to one for cohort 1980, the cohort effect β is relative to cohort 1980.

The age at onset density $f(a | c)$ and the Poisson intensity $\lambda(a, c) = f(a | c) \cdot \kappa(c)$ can be written in R as

```
f <- function(a,c,alpha,beta) {
dgamma(a, shape=shapeGam(alpha,beta,c),
rate=rateGam(alpha,beta,c))
}

lambda <- function(a,c,alpha,beta,xi) {
f(a,c,alpha,beta)*q(c,xi)
}.
```

Our aim in this section is to obtain the curves illustrated in Figure 7.7. This figure shows the effect of β on the expected count ratio, $\frac{\nu_{i,k}^{(1)}}{\nu_{i,k}}$, and the expected count cross-ratio, $\frac{\nu_{i+1,k}}{\nu_{i,k}} \div \frac{\nu_{i+1,k}^{(1)}}{\nu_{i,k}^{(1)}}$. The function `parallelogram` introduced in Appendix 9.3 takes `alpha` and `xi` as inputs and produces a matrix with each element corresponding to a Riemann Sum approximation of $\nu_{i,k}^{(1)}$ in the case of

independence. The following function `parallelogramR` takes cohort `k` and the parameter vectors `alpha`, `beta` and `xi` as inputs and produces a vector,

$$\mathbf{v}_k^{(1)T} = \left(\frac{\nu_{3,k}^{(1)}}{\nu_{2,k}^{(1)}}, \frac{\nu_{4,k}^{(1)}}{\nu_{3,k}^{(1)}}, \dots, \frac{\nu_{10,k}^{(1)}}{\nu_{9,k}^{(1)}} \right) :$$

```
a0 <- 2; u <- c(); v <- c(); aRatio <- c()
parallelogramR <- function(k,alpha,beta,xi) {
  A <- c()
  for (i in 1:nAge ) {
    for (delta2 in 1:n) {
      for (delta1 in 1:n) {
        u[delta1] <- lambda( a0 +i-1 +(delta2/n), k +(delta1-delta2)/n,
          alpha, beta, xi ) *(1/(n^2))
      }
      v[delta2] <- sum(u)
    }
    pA <- sum(v)
    A[i] <- pA
  }
  for (j in 1:8) {aRatio[j] <- A[j+1]/A[j]}
  aRatio
}
```

The function `square` introduced in Appendix 9.3 takes `alpha` and `xi` as inputs and produces a matrix with each element corresponding to a Riemann Sum approximation of $\nu_{i,k}$ in the case of independence. The following function `squareR` takes cohort `k` and the parameter vectors `alpha`, `beta` and `xi` as inputs and produces a vector,

$$\mathbf{v}_k^T = \left(\frac{\nu_{3,k}}{\nu_{2,k}}, \frac{\nu_{4,k}}{\nu_{3,k}}, \dots, \frac{\nu_{10,k}}{\nu_{9,k}} \right) :$$

```
u2 <- c(); v2 <- c(); bRatio <- c()
squareR <- function(k,alpha,beta,xi) {
```

```

B <- c()
for (i in 1:nAge) {
  for (delta2 in 1:n) {
    for (delta1 in 1:n) {
      u2[delta1] <- lambda( a0+ i-1 +(delta2/n), k +(delta1/n),
alpha,beta,xi ) *(1/(n^2))
    }
    v2[delta2] <- sum(u2)
  }
  pB <- sum(v2)
  B[i] <- pB
}
for (j in 1:8) {bRatio[j] <- B[j+1]/B[j]}
bRatio
}.

```

The maximum likelihood estimates of the model parameters, which are presented in equation (7.7), can be written in R as

```
alphaM <- c(5.825,1.572); xiM <- c(0.702,0.515).
```

Suppose the parameters in $\lambda(a, c) = \kappa(c) \cdot f(a | c)$ are set to $\mu_A = 5.825$, $\sigma_A^2 = 1.572$, $\xi_1 = 0.702$ and $\xi_2 = 0.515$, with β yet to be specified. The following code, `R.ratio1`, takes vector \mathbf{v}_{1987} and divides it by $\mathbf{v}_{1987(1)}$ to produce a vector of cross ratios,

$$\left(\frac{\nu_{3,1987}}{\nu_{2,1987}} \div \frac{\nu_{3,1987}^{(1)}}{\nu_{2,1987}^{(1)}}, \frac{\nu_{4,1987}}{\nu_{3,1987}} \div \frac{\nu_{4,1987}^{(1)}}{\nu_{3,1987}^{(1)}}, \dots, \frac{\nu_{10,1987}}{\nu_{9,1987}} \div \frac{\nu_{10,1987}^{(1)}}{\nu_{9,1987}^{(1)}} \right) :$$

```

R.ratio1 <- squareR( 1987, alphaM, 0, xiM ) /
parallelogramR( 1987, alphaM, 0, xiM ).

```

In `R.ratio1`, the parameter β is set equal to zero. The following code, `R.ratio2` and `R.ratio3`, also produce a vector of cross ratios for cohort 1987, but with $\beta = 0.05$ and $\beta = 0.10$:

```
R.ratio2 <- squareR( 1987, alphaM, 0.05, xiM ) /  
parallelogramR( 1987, alphaM, 0.05, xiM )
```

```
R.ratio3 <- squareR( 1987, alphaM, 0.10, xiM ) /  
parallelogramR( 1987, alphaM, 0.10, xiM ).
```

This demonstration can be carried out for other cohorts.

Bibliography

- G. Adams and R. Watson. A Discrete Time Parametric Model for the Analysis of Failure Time Data. *Australian Journal of Statistics*, 31(3):365–384, 1989.
- R. A. Adams. *Calculus: A Complete Course*. Prentice Hall Canada, 6th edition, 2006.
- R. M. Anderson, C. A. Donnelly, N. M. Ferguson, M. E. J. Woolhouse, C. J. Watt, H. J. Udy, S. MaWhinney, S. P. Dunstan, T. R. E. Southwood, and J. W. Wilesmith. Transmission Dynamics and Epidemiology of BSE in British Cattle. *Nature*, 382(6594):779–788, 1996.
- J. C. Barrett. Age, Time and Cohort Factors in Mortality from Cancer of the Cervix. *Journal of Hygiene*, 71(2):253–259, 1973.
- J. C. Barrett. The Redundant Factor Method and Bladder Cancer Mortality. *Journal of Epidemiology and Community Health*, 32(4):314–316, 1978.
- H. M. Blalock. The Identification Problem and Theory Building: The Case of Status Inconsistency. *American Sociological Review*, 31(1):52–61, 1966.
- B. Carstensen. Age-Period-Cohort Models for the Lexis Diagram. *Statistics in Medicine*, 26(15):3018–3045, 2007.
- Y. Choi, Y. Kim, S. K. Park, H.-R. Shin, and K.-Y. Yoo. Age-Period-Cohort Analysis of Female Breast Cancer Mortality in Korea. *Cancer Research and Treatment*, 48(1):11–19, 2016.

- D. Clayton and E. Schifflers. Models for Temporal Variation in Cancer Rates. I: Age-Period and Age-Cohort Models. *Statistics in Medicine*, 6(4):449–467, 1987a.
- D. Clayton and E. Schifflers. Models for Temporal Variation in Cancer Rates. II: Age-Period-Cohort models. *Statistics in Medicine*, 6(4):469–481, 1987b.
- D. Collett. *Modelling Survival Data in Medical Research*. Chapman and Hall/CRC, London, 3rd edition, 2015.
- D. R. Cox. Regression Models and Life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(202):187–202, 1972.
- D. R. Cox and D. Oakes. *Analysis of Survival Data*. Monographs on Statistics and Applied Probability. Chapman and Hall/CRC, 1984.
- CT4. *Models, Core Technical: Core Reading for the 2013 Examinations*. The Faculty of Actuaries and Institute of Actuaries, London, 2012.
- H. T. Davis and M. L. Feldstein. The Generalised Pareto Law as a Model for Progressively Censored Survival Data. *Biometrika*, 66(2):299–306, 1979.
- S. F. Dealler and J. T. Kent. BSE: An Update on the Statistical Evidence. *British Food Journal*, 97(8):3–18, 1995.
- D. C. M. Dickson, M. R. Hardy, and H. R. Waters. *Actuarial Mathematics for Life Contingent Risks*. Cambridge University Press, 1st edition, 2009.
- I. Diouf, M. A. Charles, P. Ducimetère, A. Basdevant, E. Eschwege, and B. Heude. Evolution of Obesity Prevalence in France: An Age-Period-Cohort Analysis. *Epidemiology*, 21(3):360–365, 2010.
- A. J. Dobson and A. Barnett. *An Introduction to Generalized Linear Models*. London: Chapman and Hall/CRC Press, 3rd edition, 2008.
- C. A. Donnelly and N. M. Ferguson. *Statistical Aspects of BSE and vCJD: Models for Epidemics*. Chapman and Hall/CRC Press, 2000.

- N. R. Draper and H. Smith. *Applied Regression Analysis*. Wiley, New York, 3rd edition, 1998.
- I. Elbatal and N. S. Butt. A New Generalisation of Quadratic Hazard Rate Distribution. *Pakistan Journal of Statistics and Operation Research*, 9(4): 343–361, 2014.
- S. E. Fienberg and W. M. Mason. Specification and Implementation of Age, Period and Cohort Models. In W. M. Mason and S. E. Fienberg, editors, *Cohort Analysis in Social Research: Beyond the Identification Problem*, pages 45–88. Springer, New York, 1985.
- W. H. Frost. The Age Selection of Mortality from Tuberculosis in Successive Decades. *The Milbank Memorial Fund Quarterly*, 18(1):61–66, 1940.
- W. J. Fu. A Smoothing Cohort Model in Age–Period–Cohort Analysis With Applications to Homicide Arrest Rates and Lung Cancer Mortality Rates. *Sociological Methods and Research*, 36(3):327–361, 2008.
- N. D. Glenn. Cohort Analysts’ Futile Quest: Statistical Attempts to Separate Age, Period and Cohort Effects. *American Sociological Review*, 41(5):900–904, 1976.
- N. D. Glenn. *Cohort Analysis*. Quantitative Applications in the Social Sciences. California: Sage Publications, Inc, 2nd edition, 2005.
- B. G. Greenberg, J. J. Wright, and C. G. Sheps. A Technique For Analysing Some Factors Affecting the Incidence of Syphilis. *Journal of the American Statistical Association*, 45(251):373–399, 1950.
- P. L. Gupta, R. C. Gupta, and R. C. Tripathi. On the Monotonic Properties of Discrete Failure Ratesonotonic Properties of Discrete Failure Rates. *Journal of Statistical Planning and Inference*, 65(2):255–268, 1997.
- D. R. Hayward and N. Krause. Aging, Social Development, and Cultural Factors in Changing Patterns of Religious Involvement Over a 32-Year Period:

- An Age–Period–Cohort Analysis of 80 Countries. *Journal of Cross-Cultural Psychology*, 46(8):979–995, 2015.
- U. Hjorth. A Reliability Distribution With Increasing, Decreasing, Constant and Bathtub-Shaped Failure Rates. *Technometrics*, 22(1):99–107, 1980.
- T. R. Holford. The Estimation of Age, Period and Cohort Effects for Vital Rates. *Biometrics*, 39(2):311–324, 1983.
- I. R. James and M. R. Segal. On a Method of Mortality Analysis Incorporating Age-Year Interaction, with Application to Prostate Cancer Mortality. *Biometrics*, 38(2):433–443, 1982.
- S. Jean, S. O’Donnell, C. Lagacé, P. Walsh, C. Bancej, J. P. Brown, S. Morin, A. Papaioannou, S. B. Jaglal, and W. D. Leslie. Trends in Hip Fracture Rates in Canada: An Age-Period-Cohort Analysis. *Journal of Bone and Mineral Research*, 28(6):1283–1289, 2013.
- N. L. Johnson, A. W. Kemp, and S. Kotz. *Univariate Discrete Distributions*. Wiley, 3rd edition, 2005.
- R. Johnston, K. Jones, and D. Manley. Confounding and Collinearity in Regression Analysis: A Cautionary Tale and an Alternative Procedure, Illustrated by Studies of British Voting Behaviour. *Quality and Quantity*, 52(4):1957–1976, 2018.
- J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. Wiley, New Jersey, 2nd edition, 2002.
- W. O. Kermack, A. G. McKendrick, and P. L. McKinlay. Death-Rates in Great Britain and Sweden: Expression of Specific Mortality Rates as Products of Two Factors, and some Consequences thereof. *Journal of Hygiene*, 34(4):433–547, 1934.
- M. R. Kramer, A. L. Valderrama, and M. L. Casper. Decomposing Black-White Disparities in Heart Disease Mortality in the United States, 1973–2010: An

- Age-Period-Cohort Analysis. *American Journal of Epidemiology*, 182(4): 302–312, 2015.
- L. L. Kupper, J. M. Janis, A. Karmous, and B. G. Greenberg. Statistical Age-Period-Cohort Analysis: A Review and Critique. *Journal of Chronic Diseases*, 38(10):811–830, 1985.
- C. D. Lai and D. Q. Wang. A Finite Range Discrete Life Distribution. *International Journal of Reliability, Quality and Safety Engineering*, 2(2):147–160, 1995.
- J. F. Lawless. *Statistical Models and Methods for Lifetime Data*. Wiley, New York, 1982.
- W. C. Lee and R. L. Hsieh. Estimating Life Expectancy Using an Age-Cohort Model in Taiwan. *Journal of Epidemiology and Community Health*, 50(2): 214–217, 1996.
- G. M. Leung, T. Q. Thach, T. H. Lam, A. J. Hedley, W. Foo, R. Fielding, P. S. F. Yip, E. M. C. Lau, and C. M. Wong. Trends in Breast Cancer Incidence in Hong Kong between 1973 and 1999: An Age-Period-Cohort Analysis. *British Journal of Cancer*, 87(9):982–988, 2002.
- W. H. R. A. Lexis. *Einleitung in die Theorie der Bevölkerungsstatistik (Introduction to the Theory of Population Statistics)*. Karl J Trübner, Strassburg (Strasbourg), 1875.
- S. Lipschutz and J. J. Schiller. *Schaum's Outline of Introduction to Probability and Statistics*. McGraw Hill Professional, 1998.
- S. Liu, R. Semenciw, A. M. Ugnat, and Y. Mao. Increasing Thyroid Cancer Incidence in Canada 1970-1996: Time Trends and Age-Period-Cohort Effects. *British Journal of Cancer*, 85(9):1335–1339, 2001.
- K. O. Mason, W. M. Mason, H. H. Winsborough, and K. W. Poole. Some Methodological Issues in Cohort Analysis of Archival Data. *American Sociological Review*, 38(2):242–258, 1973.

- S. H. Moolgavkar, R. G. Stevens, and J. A. Lee. Effect of Age on Incidence of Breast Cancer in Females. *Journal of the National Cancer Institute*, 62(3): 493–501, 1979.
- T. Murayama, K. Takahashi, Y. Natori, and N. Kurumatani. Estimation of Future Mortality from Pleural Malignant Mesothelioma in Japan Based on an Age-Cohort Model. *American Journal of Industrial Medicine*, 49(1):1–7, 2006.
- T. Nakagawa and S. Osaki. The Discrete Weibull Distribution. *IEEE Transactions on Reliability*, 5(300-301), 1975.
- J. A. Nelder and R. W. M. Wedderburn. Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*., 135(3):370–384, 1972.
- R. M. O’Brien. *Age-Period-Cohort Models: Approaches and Analyses with Aggregate Data*. Statistics in the Social and Behaviourial Sciences. Chapman and Hall/CRC, Boca Raton, Florida, 2015.
- C. Osmond and M. J. Gardner. Age, Period and Cohort Models Applied to Cancer Mortality Rates. *Statistics in Medicine*, 1(3):245–259, 1982.
- J. Peto, F. Matthews, J. Hodgson, and J. R. Jones. Continuing Increase in Mesothelioma Mortality in Britain. *The Lancet*, 345(8949):535–539, 1995.
- J. Peto, A. Decarli, C. L. Vecchia, F. Levi, and E. Negri. The European Mesothelioma Epidemic. *British Journal of Cancer*, 79(3-4):666–672, 1999.
- W. L. Rodgers. Estimable Functions of Age, Period and Cohort Effects. *American Sociological Review*, 47(6):774–787, 1982.
- N. B. Ryder. The Cohort as a Concept in the Study of Social Change. In M. A. Hardy, editor, *Studying Aging and Social Change: Conceptual and Methodological Issues*, chapter 3, pages 66–92. Sage Publications Inc, California, 1997.
- A. A. Salvia and R. C. Bollinger. On Discrete Hazard Functions. *IEEE Transactions on Reliability*, 31(5):458–459, 1982.

- P. Sasieni and J. Adams. Changing Rates of Adenocarcinoma and Adenosquamous Carcinoma of the Cervix in England. *The Lancet*, 357(9267):1490–1493, 2001.
- G. Schomerus, S. V. der Auwera, H. Matschinger, S. E. Baumeister, and M. C. Angermeyer. Do Attitudes Towards Persons with Mental Illness Worsen During the Course of Life? An Age-Period-Cohort Analysis. *Acta Psychiatrica Scandinavica*, 132(5):357–364, 2015.
- P. Schwadel. Age, Period and Cohort Effects on Religious Activities and Beliefs. *Social Science Research*, 40(1):181–192, 2011.
- V. Shkolnikov, M. Barbieri, and J. Wilmoth. *The Human Mortality Database*. The Human Mortality Database, 2018 (accessed January 27, 2018). URL www.mortality.org.
- T. R. Smith and J. Wakefield. A Review and Comparison of Age-Period-Cohort Models for Cancer Incidence. *Statistical Science*, 31(4):591–610, 2016.
- W. E. Stein and R. Dattero. A New Discrete Weibull Distribution. *IEEE Transactions on Reliability*, 33(2):196–197, 1984.
- R. G. Stevens and S. H. Moolgavkar. A Cohort Analysis of Lung Cancer and Smoking in British Males. *American Journal of Epidemiology*, 119(4):624–641, 1984.
- C. Vandeschrick. The Lexis Diagram, a Misnomer. *Demographic Research*, 4(3):97–124, 2001.
- G. N. Wilkinson and C. E. Rogers. Symbolic Description of Factorial Models for Analysis of Variance. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 22(3):392–399, 1973.
- J. R. Wilmoth, K. Andreev, D. Jdanov, D. A. Gleij, and T. Riffe. *Methods Protocol for the Human Mortality Database, Version 6*, 2017 (accessed August 22, 2018). URL <https://www.mortality.org/Public/Docs/MethodsProtocol.pdf>.

- R. Winkler and K. Warnke. The Future of Hunting: An Age-Period-Cohort Analysis of Deer Hunter Decline. *Population and Environment*, 34(4):460–480, 2013.
- E. Xekalaki. Hazard Functions and Life Distributions in Discrete Time. *Communications in Statistics - Theory and Methods*, 12(21):2503–2509, 1983.
- X. Xu, N. Laird, D. W. Dockery, J. P. Schouten, B. Rijcken, and S. T. Weiss. Age, Period, and Cohort Effects on Pulmonary Function in a 24-year Longitudinal Study. *American Journal of Epidemiology*, 141(6):554–566, 1995.
- Y. Yang and K. C. Land. *Age-Period-Cohort Analysis: New Models, Methods, and Empirical Applications*. Chapman and Hall/CRC Interdisciplinary Statistics, Boca Raton, Florida. CRC Press, 2013.